Optimal network design and inventory planning for clinical trial supply chains

Zhao, Hui

2019

Zhao, H. (2019). Optimal network design and inventory planning for clinical trial supply chains. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/82988>

<https://doi.org/10.32657/10220/47563>

OPTIMAL NETWORK DESIGN AND INVENTORY

PLANNING FOR CLINICAL TRIAL SUPPLY

CHAINS

ZHAO HUI

SCHOOL OF MECHANICAL AND AEROSPACE

ENGINEERING

2019

**OPTIMAL NETWORK DESIGN AND INVENTORY
PLANNING FOR CLINICAL TRIAL SUPPLY
CHAINS**

ZHAO HUI

School of Mechanical and Aerospace Engineering

**A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy**

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

28-Jan-2019

Date

Zhao Hui

ZHAO HUI

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

28-Jan-2019

Date

Kan Wu

WU KAN

Authorship Attribution Statement

This thesis contains material from 2 papers published in the following peer-reviewed journals where I was the first and/or corresponding author.

Chapter 4 is published as Hui Zhao, Kan Wu, and Edward Huang. Clinical trial supply chain design based on the Pareto-optimal trade-off between time and cost. *IIE Transactions* 50.6 (2018): 512-524. DOI: 10.1080/24725854.2017.1395978.

The contributions of the co-authors are as follows:

1. Prof. Kan Wu provided the initial project direction.
2. I prepared the manuscript drafts. The manuscript was revised by Prof. Kan Wu and Prof. Edward Huang.
3. I co-designed the study with Prof. Kan Wu and Prof. Edward Huang, and performed all the model formulation, proposition and algorithm proposal, numerical experiments and data analysis for this work.

Chapter 5 is published as Hui Zhao, Edward Huang, Runliang Dou, and Kan Wu. A multi-objective production planning problem with the consideration of time and cost in clinical trials. *Expert Systems with Applications* (2019). In press. Available online 11 Jan 2019. DOI: 10.1016/j.eswa.2019.01.038

The contributions of the co-authors are as follows:

1. Prof. Kan Wu provided the initial project direction.
2. I prepared the manuscript drafts. The manuscript was revised by Prof. Kan Wu, Prof. Edward Huang and Prof. Runliang Dou.

-
3. I co-designed the study with Prof. Kan Wu, Prof. Edward Huang and Prof. Runliang Dou, and performed all the model formulation, numerical experiments and data analysis for this work.

28-Jan-2019

Date

Zhao Hui

ZHAO HUI

Abstract

The clinical trial plays a key role in the competitiveness of a pharmaceutical company. It is a critical step in new drug development and characterized by long duration, high cost, and high potential return. Due to the increasing complexity of new drug development, the profit is decreasing and the pharmaceutical companies strive to reduce the cost. In this thesis, we strive to improve the efficiency in three problems of clinical trial supply chains, including the inventory allocation, network design, and production planning problems. For each of these problems, we make a discussion of the relevant literature, modeling approaches, and algorithms. Due to the patient recruitment uncertainty in clinical trial supply chains, these problems are formulated as stochastic programming models.

The inventory allocation problem considers the efficient coordination of clinical drugs in clinical trial supply chains. Due to the specific characteristics of clinical trial supply chains, the inventory policy is analytically intractable. In this study, we develop an inventory allocation model to determine the distribution plan and inventory levels of clinical drugs at each location of the supply chain in each period. The total cost of transportation, holding, and stock-out is minimized in the model. Unlike classical supply chains, the demand for clinical drugs is hard to predict and stock-outs may occur due to the patient recruitment uncertainty. When the number of recruited patients at a clinic is lower than expected, the clinical drugs at the clinic can be shipped back to the distribution center for later redistribution or even shipped directly to other clinics. To solve the model efficiently, we apply Benders decomposition as the solution approach and propose two acceleration techniques. Numerical results show that our model

gives a lower cost than the (s, S) policy in most cases.

In addition, efficient clinical trial supply chain configurations are identified in the network design problem. Since the clinical trial duration is included in the 20 years of patent life, its reduction may increase the drug's commercialization time under patent protection and the total sales revenue for the specific drug significantly. However, when the duration is reduced by increasing the number of clinical sites in the supply chain, the total cost is increased simultaneously. Hence, the duration and cost become a tradeoff and Pareto-optimal supply chain configurations are used to improve the clinical trial efficiency. In this study, we propose a multi-objective clinical site selection model which considers the tradeoff between time and cost of clinical trials. An efficiency curve representing the Pareto-optimal tradeoff is provided for decision makers to design the supply chain. To identify all Pareto-optimal supply chain configurations efficiently, we propose some propositions and develop an algorithm accordingly. Some optimality cuts are also derived to improve the solving efficiency of this problem.

The overproduction reduction of clinical drugs is considered in the production planning problem. Overproduction is commonly seen in clinical trial supply chains and pharmaceutical companies strive to reduce it due to the significant drug production cost nowadays. When the overproduction is reduced in clinical trial supply chains, stock-outs may occur when the allocated inventory cannot meet the stochastic demand at some clinical sites. As a result, the clinical trial duration may increase. To reduce the overproduction without the duration increase in clinical trials, the inventory in the clinical trial supply chain should be coordinated efficiently. In this study, we propose a multi-objective multi-stage stochastic model to optimize the production quantity, where the clinical trial

duration and the total production and operational costs are minimized. The progressive hedging algorithm is applied as the solution approach. In the numerical experiments, we study this algorithm's performance and find the optimal production quantity of clinical drugs.

Acknowledgement

First of all, I would like to extend my sincere gratitude to my supervisor, Prof. Kan Wu, for this interesting topic, his instructive advices, useful suggestions, and strict requirements. His incredibly thorough attitude and diligence to work also influence me deeply, which I will always learn from him.

In addition, I am sincerely grateful to Prof. Edward Huang, whose profound professional knowledge, approachability and patience really help me a lot in my research.

I also would like to thank my family and friends. Their encouragement and support give me confidence and strength.

Contents

Chapter 1	Introduction.....	1
1.1	Background of Clinical Trials	1
1.2	Clinical Trial Supply Chains	6
1.3	Research Scope.....	8
1.4	Organization of the Thesis.....	11
Chapter 2	Literature Review.....	13
2.1	Pharmaceutical Supply Chains.....	13
2.2	Clinical Trial Supply Chain Management	14
2.2.1	Inventory Allocation	14
2.2.2	Network Design	19
2.2.3	Production Planning.....	21
2.3	Relevant Modeling Approaches and Algorithms.....	26
2.3.1	Optimization Approaches for Problems Under Uncertainty ..	26
2.3.1.1	Stochastic Programming Models	26
2.3.1.2	Chance-Constrained Models	28
2.3.1.3	Simulation-Based Optimization.....	28
2.3.2	Modeling Approaches in Stochastic Programming Models...	29
2.3.2.1	Sample Average Approximation	29
2.3.2.2	Comparison between Two-stage and Multi-stage Stochastic Models	33
2.3.2.3	The Rolling Horizon Scheme.....	34
2.3.3	Algorithms for Stochastic Programming Models	35
2.3.3.1	Benders Decomposition	36

2.3.3.2	Progressive Hedging Algorithm.....	39
2.3.3.3	Comparison.....	43
2.3.4	Algorithms for Multi-objective Models.....	44
2.4	Summary of Research Gaps	45
Chapter 3	Inventory Allocation in Clinical Trial Supply Chains.....	49
3.1	Introduction	49
3.2	Model Development.....	51
3.2.1	Assumptions.....	51
3.2.2	Mathematical Model	53
3.2.3	Model Analysis	61
3.3	Proposed Methodologies	63
3.3.1	Reformulation Method.....	63
3.3.2	Row Generation Strategy.....	66
3.4	Numerical Experiments.....	71
3.4.1	Numerical Study	71
3.4.2	Performance Analysis	73
3.4.3	Cost Comparison with the (s, S) Policy	74
3.5	Conclusion.....	80
	Appendix.....	82
Chapter 4	Network Design in Clinical Trial Supply Chains	85
4.1	Introduction	85
4.2	Model Development.....	87
4.2.1	Assumptions.....	87
4.2.2	Mathematical Model	90
4.3	Propositions and Optimality Cuts.....	93

4.3.1	Pareto-optimal Propositions.....	93
4.3.2	Optimality Cut Properties	98
4.4	Algorithm	100
4.5	Numerical Experiments	102
4.5.1	Numerical Study	102
4.5.2	The Efficiency Curve	103
4.5.3	Performance Analysis	108
4.6	Conclusion.....	110
Chapter 5	Production Planning in Clinical Trial Supply Chains	113
5.1	Introduction	113
5.2	Model Development	116
5.2.1	Assumptions.....	116
5.2.2	Production Planning Model	118
5.2.2.1	A Deterministic Model.....	119
5.2.2.2	Stochastic Programming Extensions.....	122
5.3	Numerical Experiments	126
5.3.1	Numerical Study	127
5.3.2	Effects of ρ on Convergence and Solution Quality	128
5.3.3	The Solving Efficiency Comparison.....	130
5.3.4	Production and Distribution Cost.....	132
5.4	Conclusion.....	138
	Appendix.....	141
Chapter 6	Conclusion and Future Work	143
6.1	Conclusion.....	143
6.2	Future Work.....	147

Publications..... 149

References 151

List of Figures

Figure 1-1. A clinical trial supply chain	7
Figure 2-1. The rolling horizon scheme	35
Figure 2-2. Benders decomposition	38
Figure 3-1. The delivery routes in a clinical trial supply chain	50
Figure 3-2. The procedure of the row generation strategy	70
Figure 4-1. The optimality cuts in the properties	100
Figure 4-2. The comparison between the lower and upper bounds in the SAA method.....	105
Figure 4-3. The efficiency curve	106
Figure 4-4. Total cost's increase with the delay cost	107
Figure 4-5. Total time's decrease with the delay cost	107
Figure 4-6. The optimality cuts on the efficiency curve	110
Figure 5-1. A scenario tree	123
Figure 5-2. The errors of the solutions in all scenarios obtained in PHA	130
Figure 5-3. The expected total cost of the solution obtained in PHA	130
Figure 5-4. The total inventory distribution cost	133
Figure 5-5. The production and distribution cost	134
Figure 5-6. The distribution cost for each delay value	135
Figure 5-7. The distribution cost for each γ value	136
Figure 5-8. The total cost for each delay value	136

Figure 5-9. The total cost for each γ value.....	137
--	-----

Figure 5-10. The tradeoff between the total cost and clinical trial delay.....	138
---	-----

List of Tables

Table 2-1. The progressive hedging algorithm	41
Table 2-2. An example of the cycle behavior in progressive hedging algorithm	42
Table 3-1. Summary of the notations	54
Table 3-2. Statistics in the SAA method	72
Table 3-3. Comparison of all accelerating techniques	73
Table 3-4. Comparison of two acceleration strategies	74
Table 3-5. Parameter settings	75
Table 3-6. Non-stationary patient arrival rates and their corresponding optimal (s, S) policy	75
Table 3-7. The average total cost comparison between our approach and the (s, S) policy with transshipment.....	76
Table 3-8. Cost comparison between our approach and the (s, S) policy with transshipment	77
Table 4-1. Summary of the notations	89
Table 4-2. The function of finding the Pareto-optimal set.....	102
Table 4-3. The upper bound in the SAA method	104
Table 4-4. The Pareto-optimal tradeoffs	105
Table 4-5. The performance of the optimality cuts	109
Table 5-1. Summary of the notations in the deterministic model	119

Table 5-2. Summary of the notations in the stochastic model	123
Table 5-3. The solving efficiency comparison.....	132
Table 5-4. The production and inventory distribution cost.....	135

Chapter 1 Introduction

1.1 Background of Clinical Trials

New drug development is characterized by high investment and low success rate (Laínez et al. 2012, Shah 2004). It follows a long sequence of steps, including the discovery, animal trials, Food and Drug Administration (FDA) application, product and process development, three phases of clinical trials, FDA filing and approval, and launch to the drug market of the United States (Colvin and Maravelias 2008). It is estimated that only around five of 5000 to 10000 discovered compounds perform well enough and can be selected to take the clinical trials (Profile 2010), and only one of the five can get the approval of FDA and enter the U.S. drug marketplace (DiMasi et al. 2003). If a candidate fails, all investment is wasted. About 67.4 billion dollars has been invested in Research and Development (R&D) in major pharmaceutical companies in 2010. It is about 17% of the total annual sales of these companies (Khanna 2012).

Despite the high expense and low success rate, a successful drug is rewarding. It can not only help patients alleviate their pains or extend their lives, but also can generate large profits for the responsible pharmaceutical company. Under the FDA's regulations, each new drug is granted 20 years of patent protection. Once a new drug successfully enters the drug marketplace, a patent-protected drug can be highly profitable. A blockbuster can generate \$1 billion or more in annual sales revenue, which not only compensates all previous R&D expenses but also generates a high profit for the company (Mullin 2011).

Clinical trials play a key role in the process of developing new drugs. A new drug has to be tested on human beings in order to determine its safety, efficacy,

and dose levels before it can be approved by the FDA. A clinical trial usually includes three phases. A drug that successfully passes all three phases may receive the approval from the FDA and enter the U.S. drug marketplace. The trial is terminated immediately when the drug fails in any phase. Besides the low success rate, clinical trials are also characterized by high cost and long duration (Laínez, et al. 2012, Roberts et al. 2003, Shah 2004). On average, 37% of the total R&D expense in the new drug development are spent on clinical trials (Fleischhacker 2009). A clinical trial usually lasts for 6 to 8 years (Kaitin and DiMasi 2011, Shah 2004) and takes more than 70% of the total time (8 to 10 years (Sundaramoorthy et al. 2012)) in the new drug development.

Since the clinical trial duration is included in the patent life (Grabowski and Vernon 2000), its reduction has the potential to bring more profit to pharmaceutical companies. While a drug can make a significant profit in its patent life, the profit will be largely reduced after patent expiration due to the rise of generic substitutes at much lower prices (Martinez and Goldstein 2007). During the patent life, the gross margin for a typical drug can be 90% to 95% (Martinez and Goldstein 2007). The price may drop to a value close to the production cost after the patent expires (Martinez and Goldstein 2007). On average, a day's loss in patent life can cause a loss of more than \$1 million in revenue for a typical drug (Clemento 1999). Even if a trial fails, it still consumes the limited resources (e.g. time, cost, and labor) and may delay the schedule of other trials. Hence, probably the most important driver in the pharmaceutical industry is the time-to-market (Shah 2004). All pharmaceutical companies strive to commercialize their new drugs as soon as possible. The clinical trial duration should be minimized to gain the most profit.

The clinical trial duration is the time spent on collecting enough valid trial results. Prior to the commencement of each clinical trial, the clinical team determines the patient accrual goal and implementation regulations, such as the dosage levels, treatment interval (the length of the period between two treatments, e.g. one week), and treatment period (the number of periods required for a patient to participate in the trial, e.g. several weeks). The patient accrual goal is the number of valid trial results required in this clinical trial. If a patient takes the trial according to the implementation regulations, the trial data collected on him/her is deemed valid. Otherwise, the trial data will be invalid if the patient drops out the trial due to the stock-outs at the clinical site, the dissatisfaction of the trial because of no observed improvement or loss of interest, etc.

Patient recruitment is often considered to be a bottleneck for the timely completion of a clinical trial (Drennan 2002, Getz and De Bruin 2000, Thiers et al. 2008). The patients taking a clinical trial should satisfy certain criteria, such as the diagnostic profile and age (Buzby et al. 1988). Due to the increasing complexity of drug development, clinical trials require more patients to participate than ever before. The patient accrual goal can be a large number in Phase III of clinical trials. But unfortunately, both the enrollment and retention rates of the participants are decreasing in existing regions (Profile 2010, Thiers, et al. 2008). As a result, enrolling enough qualified patients can be time-consuming at only one or a few locations. In addition, since clinical trials are usually conducted in developed countries, such as the U.S. and European countries, the cost is quite high. Clinical trials are obliged to go global (Meixell and Gargeya 2005, Thiers, et al. 2008). The globalization has been attributed as the main factor of the rapid speed in new drug development of some

pharmaceutical companies (Getz and De Bruin 2000).

While the globalization of clinical trials accelerates the patient recruitment process by increasing the number of participating clinical sites (Roberts, et al. 2003, Rowland and McMahon 2004, Van Arnum 2011), it may increase the cost in the supply chain. The cost includes clinical sites' enrollment costs and clinical drugs' production, delivery, and holding costs. The enrollment cost is the cost of recruiting a clinical site into a clinical trial. It is between \$20,000 and \$30,000 for a typical clinical site (Miseta 2013).

A larger clinical trial supply chain can be more complicated and hard to coordinate (Shah 2004). Since clinical drugs are in the development stage, they can only be obtained from the responsible pharmaceutical company. In addition, clinical trials are conducted in a double-blind manner, i.e., when a treatment is given, neither the patient nor the doctor knows if it is a placebo or a real drug. Hence, the inventories at both the distribution centers (DCs) and clinical sites are managed and controlled by the pharmaceutical company. Since the globalization of clinical trials makes the supply chain involve multiple continents, languages, cultures, regulations and so on, it is much more complicated to provide the right drug supply at the right quantity to the right place at the right time in a cost-efficient way (Abdelkafi et al. 2009, Pedroso and Nakano 2009, Van Arnum 2009, Van Arnum 2011). Due to the demand uncertainty at clinical sites, more inventory is required to maintain the same service level in a larger supply chain (Croxtton and Zinn 2005, Fleischhacker et al. 2015).

Clinical drugs are commonly overproduced to hedge the stock-outs at clinical sites and then avoid the delay in clinical trials (Abdelkafi, et al. 2009, Fleischhacker, et al. 2015). When a stock-out occurs at a commercial supply

chain, a patient can have the prescription replaced by a generic substitution or fulfilled at another pharmacy. However, it does not apply to stock-outs in a clinical trial supply chain. Since a clinical trial may fail at each of the three phases and the required drug amount can be considerable in later phases (especially in Phase III), the production of clinical drugs is usually conducted more than once to reduce waste. Due to the long supply lead time and notable changeover cost, drugs are commonly produced in large batches and a substantial replenishment cycle exists between the arrivals of two consecutive production batches from a plant to a DC. Hence, the initial inventory at the beginning of a replenishment cycle is the sum of the volume of the production batch and the unused drugs from the previous cycle (if any). The total inventory in the supply chain decreases in a replenishment cycle and will only increase after the next production batch arrives. Due to the patient recruitment uncertainty, the demands for clinical drugs at clinical sites are hard to predict (Chen et al. 2012, Shah 2004). Stock-outs may occur when the allocated inventory cannot meet the demand at some clinical sites. To avoid the trial delay caused by stock-outs at clinical sites, overproduction is commonly seen in clinical trial supply chains.

The pharmaceutical companies are facing increasingly challenging circumstances and striving to reduce the cost. The new drug development becomes more and more diverse and complicated (Abdelkafi, et al. 2009). The investment in R&D is growing, but the number of new drugs approved by FDA has declined (Martinez and Goldstein 2007). In 2007, the FDA approved 17 new molecular entities and 2 biological applications. This was the least number of new drugs approved by FDA since 1983. A new drug's effective patent life has also been shortened due to the increase of its development time (Shah 2004).

Hence, the profit margin has started to decline and the pharmaceutical companies strive to reduce cost (Fleischhacker 2009, Martinez and Goldstein 2007).

The overproduction of clinical drugs is to be reduced in pharmaceutical companies (Abdelkafi, et al. 2009). By reducing the overage from an average of over 100% to under 50%, GlaxoSmithKline saved \$120 million in 2009 (Fleischhacker, et al. 2015). Bristol-Myers Squibb also estimated \$40 million could be further saved by overage reduction (Powell 2010). Since the traditional chemistry-based way of developing drugs becomes less efficient (produces fewer and fewer drugs), today's drug development tends to be based on biotechnology (Martinez and Goldstein 2007). Biotech drugs are usually complicated and expensive to produce (Abdelkafi, et al. 2009). Unlike commercial drugs, clinical drugs will be invalid and have to be disposed if the trial is unsuccessful. Hence, pharmaceutical companies try to reduce the cost through reducing the overproduction of clinical drugs.

1.2 Clinical Trial Supply Chains

The clinical trial is a research study which tests how well a new medical approach works in people. It consists of therapeutic drugs going through a series of research experiments, or phases, to gain scientifically supported insight (Cheng 2008), and aims to test if the specific drugs can be utilized by the patients to alleviate their pains or extend their lives.

The clinical trial typically proceeds through three successive phases (DiMasi, et al. 2003). In phase I, it tests the metabolic actions, safe dose levels, and the best way to give a new treatment, such as infusion or injection. Since this is the first trial on human beings, little is known about the possible risks and benefits of the treatment under the clinical trial. Only a small number, usually 20-

100 healthy volunteers are enrolled in this phase (Cheng 2008, DiMasi, et al. 2003). The duration can be up to one month. The trial in phase II is conducted to get the evidence on the treatment's side effects and efficacy (i.e., whether the treatment works against a specific disease). The number of patients required in this phase is typically 100-500 (Colvin and Maravelias 2008). This phase usually lasts for several months. In phase III, at least 1000-1500 patients are recruited to obtain the dosage regimens and statistically significant data about effectiveness, by making comparisons between the new and standard treatments, in preparation for FDA's approval. Several years are often spent here. Moreover, many drugs that work successfully in animal trials fail miserably in human beings, and most drugs cannot pass all three phases. If the test fails in some phases, the clinical trial is terminated immediately and all leftover drugs should be disposed of.

Similar to commercial drugs, clinical drugs have a complete supply chain from suppliers to retailers. According to the description in the literature (Chen, et al. 2012, Shah 2004, Sousa et al. 2011), a clinical trial supply chain can be depicted in Figure 1-1. The supply chain consists of the manufacture and distribution parts.

Figure 1-1. A clinical trial supply chain

A clinical trial supply chain starts with the manufacture of active pharmaceutical ingredient (API) at the primary manufacturing sites, and ends with the patients receiving clinical trials at the clinical sites. At the very beginning,

raw materials are produced into API in the form of powder at primary sites. The manufacture usually consists of either several chemical syntheses and separation processes, or fermentation and purification processes (Chen, et al. 2012, Shah 2004). By adding “excipient” inert materials along with further processing, the manufacture at secondary manufacturing sites converts the API to drugs of different forms, such as capsules, tablets or liquid, etc. After packaging and labeling, the final products will be dispatched to all DCs around the world. The supply lead time from the primary manufacturing sites to DCs can be up to one year or longer (Shah 2004).

Besides the new drug, two other products, including a placebo (a product without the API) and a comparator (a commercial drug with the same targeting disease), are also produced and used in the clinical trials to test the new drug’s safety and efficacy (Chen, et al. 2012). When a treatment is given, neither the patient nor the doctor knows what exactly it is. This double-blind study is designed to guarantee the impartiality and integrity of the test. In order to assist the double-blind study and avoid psychological biases, the placebo and the comparator have the same form, packaging, and labeling as the new drug.

When the production is finished, all the products will be dispatched to DCs all over the world and distributed to the local clinical sites from the corresponding DC. Patients will be recruited at the clinical sites and take the clinical trials there.

1.3 Research Scope

Under increasingly challenging circumstances, an efficient clinical trial supply chain is required in pharmaceutical companies to reduce the cost without the delay in the completion of clinical trials. In this research, we make a study of the

clinical trial supply chains to improve their efficiency in three topics, including the network design, production planning, and inventory allocation.

As discussed, while the globalization of clinical trials accelerates the patient recruitment process and reduces the clinical trial duration, it increases the cost simultaneously. Although enrolling the most clinical sites will lead to the shortest clinical trial duration, the total costs of enrolling clinical sites and distributing clinical drugs can be significant for the pharmaceutical companies to afford. Hence, the time and cost required by a clinical trial become a tradeoff. Due to the diverse cost structures and stochastic patient recruitment processes among clinical sites, the time and cost may differ greatly with respect to supply chain configurations. To have an efficient clinical trial supply chain, clinical sites should be selected elaborately.

The production quantity of clinical drugs will also increase due to the scale increase of clinical trial supply chains. Due to the patient recruitment uncertainty at clinical sites, more inventory is required to maintain the same service level in a larger supply chain (Croxtan and Zinn 2005, Fleischhacker, et al. 2015). A simple example in the literature (Fleischhacker 2009) shows that to maintain a 99% service level, the inventory required will increase by 9.5% if the number of identical clinical sites increases from one to two, and by 69.5% if it increases to 45. Since clinical drugs are expensive nowadays and should be disposed of after an unsuccessful trial, the pharmaceutical companies strive to reduce the overproduction.

If the production quantity is simply reduced without transshipments to coordinate the supply chain, the stock-outs at clinical sites and clinical trial delay can hardly be avoided. When the production quantity is closer to the actual

demand, stock-outs may occur at clinics even if there is still inventory in the supply chain. This would happen when the decision makers overestimate the demand and ship large quantities to some clinics, and thus the DC has insufficient inventory to replenish other clinics at the later stage of a replenishment cycle. In this situation, if the inventory at a clinic with excess inventory can be shipped to clinics facing or going to face stock-outs, the clinical trial delay caused by stock-outs may be avoided. According to the discussion by Rudi et al. (2001), clinical drugs, which are characterized by long supply lead time, short selling seasons and high demand uncertainty, are eligible for and benefited from transshipments. While the delivery and holding costs were considered to be negligible in the past, they have skyrocketed in recent years and can take 10% to 20% of the final value of a drug (Papageorgiou et al. 2001). Hence, transshipments can be applied to coordinate the inventories in clinical trial supply chains and improve the system performance.

We consider three problems in this study, including the network design (site selection), production planning, and inventory allocation in clinical trial supply chains. Since the overproduction reduction can hardly be achieved without the efficient inventory coordination between DCs and clinical sites, the inventory allocation decisions should be considered in the production planning problem. The inventory allocation decisions and costs can also differ greatly with respect to the supply chain structures. Hence, both the production planning and network design problems are dependent on the inventory allocation decisions. In this research, we study the inventory allocation problem first. The site selection problem is discussed subsequently to provide an appropriate supply chain structure for the clinical trials. The production planning problem is studied at last.

Due to the patient recruitment uncertainty, all problems are formulated as stochastic programming models and solved by certain algorithms.

1.4 Organization of the Thesis

In this research, we will make an investigation of the problems in clinical trial supply chains. The literature review is given in Chapter 2, including the studies about the problems in clinical trial supply chains, the relevant modeling approaches, and algorithms. The research gaps are also discussed. In addition, the inventory allocation, network design, production planning problems are studied in detail respectively in a subsequent chapter. Each problem is formulated as a mathematical model and solved by a specific algorithm. Finally, we summarize this research and provide some directions for future research in Chapter 6.

Chapter 2 Literature Review

In this chapter, we present a literature review of the problems in clinical trial supply chains. An overview of the pharmaceutical supply chains is given in Section 2.1. In Section 2.2, we present the studies about the problems in clinical trial supply chain management respectively, including the inventory allocation, network design, and production planning problems. In addition, the relevant modeling approaches and algorithms of these problems are given in Section 2.3. Finally, we discuss the research gaps between our problems and the studies in the literature.

2.1 Pharmaceutical Supply Chains

Since the end of the last millennium, considerable attention has been drawn to pharmaceutical supply chains due to their specific characteristics. Problems and corresponding related works have been discussed and summarized in the literature (Laínez, et al. 2012, Shah 2004). Shah (2004) also gave a comprehensive review of this industry, including the changes happening in recent years and the components of an entire pharmaceutical supply chain. The characteristics of the pharmaceutical industry were summarized by Laínez, et al. (2012) as well.

Due to its high risk but high potential return, the new drug development, which is critical to the development of the pharmaceutical industry, has gained special attention from researchers. The researchers mainly studied the selection of promising new drugs from the candidate list in the early phase testing for further development, the priority determination for drugs in development, the clinical trial supply chain management, the capacity planning before clinical trial

completion, etc. (Gatica et al. 2003, Laínez, et al. 2012, Papageorgiou, et al. 2001, Shah 2004).

Simulation methods and the interactive voice response system (IVRS) have been applied in clinical trial supply chain management. The simulation was introduced to design the clinical trial (Kimko and Duffull 2002). The importance of IT technology was emphasized in clinical trials (Grossman and Bates 2008). IVRS has been applied in clinical trials to monitor the real-time demand at clinical sites, make responses, and make inventory allocations automatically based on predefined policies (Byrom 2002, Waters et al. 2010). Alshawhi et al. (2003) studied the information sharing and coordination in the pharmaceutical R&D supply chain through data warehousing. When IVRS is applied in simulation models, the cost reduction and performance improvement can be predicted in the clinical trial supply chains (Dowlman et al. 2004, Peterson et al. 2004).

2.2 Clinical Trial Supply Chain Management

Due to its critical role in the pharmaceutical industry, the clinical trial supply chain has drawn special attention from researchers. In this section, we discuss the studies relevant to the inventory allocation, network design, and production planning problems in clinical trial supply chains respectively.

2.2.1 Inventory Allocation

The studies about the inventory allocation problem are discussed in this section. At first, we present a literature review of this problem in clinical trial supply chains. We also conduct a review of this problem's studies in other applications, which have features similar to the clinical trial supply chains, such the limited

inventory, non-stationary stochastic demand, and transshipment options. In addition, since the lot-sizing problem optimizes the ordering quantities and is similar to the inventory allocation problem, it is discussed under the multi-echelon setting in this section.

Few inventory allocation models for clinical trial supply chains have been proposed in the literature. Abdelkafi, et al. (2009) attempted to find a stationary inventory policy by combining the simulation and optimization. The best policy was selected by balancing the various costs against the risk of stock-outs. Since it was taken from a few simulation scenarios, the global optimum was not guaranteed. Chen, et al. (2012) optimized the production and distribution of the entire clinical trial supply chain through an integrated simulation-optimization model, which combined the mixed-integer linear programming with the discrete event simulation. Chen et al. (2012) also extended the model and studied the safety stocks by implementing the risk pooling strategy at distribution centers and the (Q, r) policy at clinical sites. Fleischhacker, et al. (2015) proposed an optimization model to find the optimal stationary inventory policy at each clinical site in the supply chain. All the researchers above assumed stationary inventory policies and/or developed simulation-optimization models. None of them has optimized the inventory allocation in clinical trial supply chains with the consideration of the limited inventory, non-stationary stochastic demand, and transshipments.

Similar to clinical drugs, fast-fashion goods (e.g. fashion apparel) also have a long replenishment cycle and highly volatile demand (Caro and Gallien 2010, Mantrala and Rao 2001). However, their supply chain is managed in a decentralized manner, which is different from clinical trial supply chains. Most

researchers discussed a single retailer's ordering policy (Choi et al. 2011, Gurnani and Tang 1999, Xiao and Jin 2011). Although Caro and Gallien (2010), Correa (2007), and Li et al. (2014) addressed the inventory allocation among multi-retailers, the inventory coordination via transshipments was not considered.

Most inventory allocation problems in multi-echelon supply chains consider stationary inventory policies. There are three basic inventory policies obtained at a single location (Wang 2011), including the continuous-review batch-ordering (Q, r) policy (Axsäter 2000), the fixed-interval order-up-to (S, T) policy (Graves 1996), and the periodic-review base-stock (s, S) policy (Porteus 1990). For each of these policies, Wang (2011) conducted a comprehensive literature review on its extension in centralized multi-echelon supply chains. Li and Wang (2007) reviewed the studies on coordinated inventory policies in centralized and decentralized multi-echelon supply chains with deterministic or stochastic demand respectively.

It can be very expensive to approximate non-stationary inventory policies by stationary ones in problems with non-stationary stochastic demand (Tunc et al. 2011). Since problems with non-stationary stochastic demand are ill-structured from a theoretical point of view and inherently complicated from a computational point of view, the number of researchers studying the non-stationary stochastic demand is much less than that of stationary demand (Graves and Willems 2008). The time-varying (s, S) policy and base-stock policy are proved to be optimal respectively when the ordering cost includes or not include a positive constant setup cost at a single stage or a pure serial system (Clark and Scarf 1960, Scarf 1959). They are further studied at a single stage (Bollapragada and Morton 1999, Morton and Pentico 1995) and in a multi-stage serial system

(Graves 1999, Iida 2001). Federgruen and Zipkin (1984) considered the periodic inventory allocation in a multi-echelon supply chain with a depot and several demand locations. They assumed the depot acted as a transshipment center and no inventory was carried at the central depot. To find the optimal ordering decisions, they formulated this problem as a dynamic program and proposed an approximation to solve it due to its large dimension of the state space. Graves and Willems (2008) studied a periodic review base-stock policy in a multi-echelon supply chain with non-stationary stochastic demand.

In supply chains with limited inventory, the optimal inventory distribution cannot be captured by stationary policies in general. Since the inventory ordering decisions at a retailer are fulfilled depending on the stock availability in a distribution system with limited inventory, they cannot be made independently of those at other retailers and the stationary inventory policies obtained at a single retailer are usually not optimal in multi-echelon supply chains (Wang 2011). Federgruen and Zipkin (1984) minimized the expected cost through a combined inventory allocation and vehicle routing model. Chien et al. (1989) formulated a mixed integer program to allocate the limited amount of inventory among customers with a fleet of vehicles in order to obtain the maximal profit. To maximize the effectiveness of the post-disaster humanitarian relief, Pérez-Rodríguez and Holguín-Veras (2015) developed models to distribute the scarce available supplies to populations in need. In addition, Fumero and Vercellis (1999) optimized the production and distribution planning in an integrated model, where the inventory allocation and vehicle routing decisions depended on the limited production quantity. These researchers consider the allocation of limited inventory in supply chains with either single period or deterministic demand.

Although the inventory pooling has been studied by many researchers in a multi-echelon supply chain, most of them studied the inventory distribution with transshipments under a specific inventory policy, such as (Q, r) , (s, S) or $(S-I, S)$ policy (Paterson et al. 2011). The optimality of these inventory policies is proved under single retailer settings (Axsäter 2007). On the other hand, the optimal policies in a multi-echelon supply chain with transshipments can only be obtained within a single period (Dong and Rudi 2004, Zhang 2005), deterministic demand (Herer and Tzur 2003), few retailers (e.g. two or three, (Herer and Rashit 1999, Herer and Tzur 2001, Robinson 1990, Yang and Qin 2007)), or multiple retailers with identical cost parameters at every retailer (Robinson 1990). In addition, due to the limited inventory in clinical trial supply chains, the inventory requests of clinical sites according to these inventory policies can hardly be fulfilled after the inventory at the DC is depleted.

Dynamic programming has been used to develop inventory policies with transshipments in multi-retailer supply chains. However, due to the high dimension of the state space, it is infeasible to solve the model for systems of practical size (Archibald et al. 2009, Archibald et al. 2010, Paterson et al. 2012). The problem can only be solved optimally with two retailers (Olsson 2009, Paterson, et al. 2012, Yang and Qin 2007) or few transshipment routes (Kukreja and Schmidt 2005). The multi-retailer cases are often solved through heuristics (Agrawal et al. 2004, Archibald, et al. 2009, Archibald, et al. 2010).

Similar to our problem, the lot-sizing problem optimizes the (production and/or) replenishment quantity at each site (Karimi et al. 2003). The joint ordering decisions are usually considered in multi-echelon lot-sizing problems, where the DC and retailers make inventory orders jointly to minimize the total

cost of the supply chain. Inventory policies (e.g. the integer-ratio and power-of-two policies (Roundy 1985)) and heuristic solution methods (Federgruen and Tzur 1999, Wee and Yang 2004) are developed to find the optimal decisions. On the other hand, the clinical trial supply chain only has a limited initial inventory for a specific clinical drug and can hardly be further replenished from the drug plants due to the long replenishment cycle. Hence, the studies in the literature of lot sizing problem are not applicable to our setting.

None of the above studies has considered the inventory allocation problem in a multi-echelon supply chain with non-stationary stochastic demand, limited inventory, and transshipment options. The inventory policy is also analytically intractable if it considers the transshipment and direct and reverse replenishment jointly in multi-period multi-clinic clinical trial supply chains. We propose a numerical approach through stochastic programming to find the optimal inventory allocation decisions in this study.

2.2.2 Network Design

The network design problem in clinical trial supply chains has been discussed by some researchers. Getz and De Bruin (2000) stated that better site selection (i.e., network design) could improve the efficiency of patient recruitment in clinical trial supply chains. The patient accrual can be accelerated by increasing the number of enrolled clinical sites (Roberts, et al. 2003).

While problems considering the cost minimization are widely addressed in the literature (Beamon 1998, Farahani et al. 2014, Melo et al. 2009), few researchers have studied the duration minimization in clinical trials or the cost in clinical trial supply chain design. Miseta (2013) discussed clinical trials' high enrollment costs and low patient accrual rates at some clinical sites, and claimed

that the clinical trial cost should be reduced with improved site selection. Fleischhacker (2009) considered the duration minimization in clinical trial network design problems. They identified the optimal number of clinical sites to minimize the total time spent on new drug development. However, they did not consider the cost in clinical trials and assumed the patient arrival rates were identical at all clinical sites, which may not be true in practice. The clinical trial duration was also minimized in terms of revenue maximization in the literature (Colvin and Maravelias 2008, Colvin and Maravelias 2009, Colvin and Maravelias 2010). The authors considered the uncertainty of the outcome in each phase of each clinical trial and optimized the resource-constrained scheduling of all present clinical trials. In addition, Grossmann and co-authors considered the duration minimization in new product development, using an approach that is also applicable to clinical trials (Jain and Grossmann 1999, Maravelias and Grossmann 2002, Maravelias and Grossmann 2004, Schmidt and Grossmann 1996).

Network design problems have been widely discussed in other applications in the literature. Several review papers have provided a comprehensive summary of this topic (Beamon 1998, Erengüç et al. 1999, Garcia and You 2015, Meixell and Gargeya 2005, Pontrandolfo 1999, Vidal and Goetschalckx 1997). Ramezani et al. (2013) also made a systematic review of this topic in terms of the network structures and modeling approaches, such as models' features (the number of products and periods, deterministic or stochastic parameters), decisions (location/allocation, facility capacity), and objectives.

Since the multi-criteria decision making was introduced to management sciences, more than one performance measures have been considered in network

design problems to evaluate the effectiveness and efficiency of the system (Altıparmak et al. 2006, Farahani et al. 2010). Beamon (1998) classified these measures as either qualitative or quantitative. Qualitative ones include the customer satisfaction, flexibility, information and material flow integration, effective risk management, and supplier performance in the supply chain. Quantitative performance measures can be categorized as either the cost scope (e.g. cost minimization or profit maximization) or the customer responsiveness (e.g. fill rate maximization or activity time minimization) (Beamon 1998).

A few network design problems in other applications are similar to the problem in clinical trial supply chains and consider the tradeoff between time and cost. The tradeoff has been modeled in multi-objective problems of facility location (Arntzen et al. 1995, Badri et al. 1998, da Graça Costa et al. 2008, Pasandideh et al. 2013), project schedule (Hadjiconstantinou and Klerides 2010, Klerides and Hadjiconstantinou 2015), JIT inventory allocation (Farahani and Elahipanah 2008), order allocation (Chan and Chung 2004), and transportation network design (Chen et al. 2010). However, since a clinical trial is stopped when its patient accrual goal is reached, the clinical trial duration is different from those in above studies, where the duration to be minimized is the total time required to complete a predetermined number of jobs with limited resources and fixed processing times.

2.2.3 Production Planning

The production planning problem has been studied jointly with inventory distribution in clinical trial supply chains. Chen, et al. (2012) proposed a decentralized model to optimize the production and distribution phases respectively and combined the solutions in a discrete event simulation to evaluate

the supply chain's performance. The model was also extended to a centralized simulation-optimization one, where the (Q, r) policy and a risk pooling strategy were applied at the clinical sites and distribution centers respectively in the clinical trial supply chain (Chen, et al. 2012). In addition, by balanced the inventory overage cost against the total shipping cost, Fleischhacker, et al. (2015) obtained the optimal production quantity and inventory ordering policy at each location in the supply chain (including the regional distribution centers and clinical sites). Due to the lack of the knowledge of the entire system, decentralized models may not find optimal solutions. The solutions can also be far from the optimal ones if stationary inventory policies are applied for supply chains with non-stationary stochastic demand (Tunc, et al. 2011). However, due to the lack of reliable forecasts, the demand at each clinical site is highly uncertain in clinical trial supply chains and usually follows a non-stationary Poisson distribution (Chen, et al. 2012). In addition, only the cost but not the clinical trial duration minimization is considered in the production planning of clinical trial supply chains in these researches.

Production planning in clinical trial supply chains has also been studied with the consideration of the drugs' failure rates and demand realization in the literature. A clinical trial may fail in each of its three phases. By considering each drug's failure rate, Gatica, et al. (2003) optimized the production schedule for a certain number of clinical drugs with the finite available manufacturing resources. Fleischhacker and Zhao (2011) demonstrated that for a clinical drug with deterministic demand, the Wagner-Whitin model could be generalized by incorporating the drug's failure risk and the salvage cost of leftover drugs to find the optimal production decisions. They also considered the demand uncertainty

and determined the production time and quantities in a two-period production model based on the first period's demand realization (Fleischhacker and Zhao 2013). In the production planning of clinical trial supply chains, none of these researches considers the duration minimization in clinical trials.

The production and distribution planning problem has been widely discussed in other applications. Erengüç, et al. (1999) categorized the decisions in such problems by supplier, plant and distribution stages, and reviewed the relevant studies in each stage. Beamon (1998) divided the models of the integrated supply chain design and analysis in four categories: deterministic analytic models, stochastic analytic models, economic models, and simulation models. In addition, Vidal and Goetschalckx (1997) and Mula et al. (2010) reviewed the studies with strategic and tactical decisions respectively in production and distribution planning problems. Vidal and Goetschalckx (1997) discussed the mixed-integer programming strategic design models and identified the formulations' relevant factors, solution methods' characteristics, and computational times. Mula, et al. (2010) selected the mathematical programming models with tactical and/or operational decisions and reviewed them from eight aspects, including the supply chain structure, decision level, modeling approach, purpose, shared information, limitations, novelty, and practical applications.

The tactical production and distribution planning models in the literature have similar (production and distribution) decision variables to our problem. In multi-period multi-retailer supply chains, most studies consider multiple products and minimize the total production, holding and transportation costs subject to certain production capacities and/or resource constraints (e.g. operation time, manpower, and machines). The stock-out was disallowed in deterministic

models (Boudia et al. 2006, Chandra and Fisher 1994, Jang et al. 2002) or minimized in stochastic cases (Gupta and Maranas 2003, Lee and Kim 2002). In addition, a single product model was considered by Boudia et al. (2008), where the inventory was distributed via a fleet of a limited number of vehicles to completely fulfill the deterministic demand at retailers with minimal delivery cost. However, since the products in the studies above and clinical drugs are in different phases of new product development, i.e., in the commercialization and R&D phases respectively, the problem settings in those studies are different from ours.

Since most studies above are about commercial products, they do not consider the overage cost but aim to minimize or eliminate the stock-out cost. For commercial products, the total demand in the entire planning horizon (selling season) is usually greater than the production capacity (Boudia, et al. 2006, Boudia, et al. 2008, Chandra and Fisher 1994, Gupta and Maranas 2003, Jang, et al. 2002). In addition, the storage capacity at each retailer may not accommodate all products required in the entire planning horizon (Boudia, et al. 2006, Boudia, et al. 2008, Lee and Kim 2002). As a result, such products might be produced very often. The inventory holding and overage costs are reduced or even eliminated consequently (Chandra and Fisher 1994). Since a salvage cost may be obtained from the leftovers, the total overage cost is negligible for commercial products. On the other hand, failing to fulfill the customer demand may result in loss of immediate revenue or even market share (Gupta and Maranas 2003, Lee and Kim 2002). Hence, the stock-outs are minimized in these studies.

On the contrary, clinical drugs are still in R&D phases, their overage should be reduced and the stock-out is not considered explicitly in our problem. Due to

the significant potential revenue of clinical drugs, the clinical trial duration minimization is a key concern in new drug development. The duration will be minimal if no patient faces stock-outs at clinical sites before the completion of the clinical trial. In other words, stock-outs after the clinical trial's completion will not increase the clinical trial duration. Thus, we should only minimize the stock-outs within the clinical trial duration. Since the clinical trial duration is uncertain due to the patient recruitment uncertainty at clinical sites, the stock-outs to be minimized cannot be expressed explicitly. Rather than the stock-out cost, the overage cost is an important concern in clinical trial supply chains. Compared with the commercial products, the total amount of drugs required in a clinical trial is quite trivial. It can hardly exceed the production or storage capacity. In addition to the notable changeover cost and long supply lead time in drug production, as well as the negligible holding cost due to the small volumes, all drugs required in a clinical trial are usually produced in a large batch. Due to the increasing complexity of drug production, the production cost of clinical drugs can be considerable, especially for the emerging biotech drugs (Abdelkafi, et al. 2009, Martinez and Goldstein 2007). Furthermore, since the clinical drugs are in the R&D phase, the leftovers cannot be used elsewhere and should only be disposed of after an unsuccessful trial. Hence, overproduction should be reduced in clinical trials.

In addition, while inventory policies are widely used in literature, they are analytically intractable in the distribution of clinical drugs due to the specific characteristics of clinical trial supply chains. While the stock-out or backorder cost is minimized by conventional inventory policies, the duration minimization is the main concern in clinical trial supply chains. Although no stock-out means

the minimal duration, the stock-outs out of the duration will not change the duration. Even if the objective is to minimize the stock-out or backorder cost, stationary inventory policies may result in a much higher cost in case of non-stationary stochastic demand (Tunc, et al. 2011). The time-varying (s, S) policy and base-stock policy are proved to be the optimal inventory policies for non-stationary stochastic demand at a single stage or a pure serial system but not in distribution networks (Clark and Scarf 1960, Scarf 1959). In addition, optimal inventory policies with transshipments can only be obtained under very restricted conditions, such as single period (Dong and Rudi 2004, Zhang 2005), deterministic demand (Herer and Tzur 2003), few retailers (e.g. two or three, (Herer and Rashit 1999, Herer and Tzur 2001, Robinson 1990, Yang and Qin 2007)), or multiple identical retailers (Robinson 1990).

2.3 Relevant Modeling Approaches and Algorithms

2.3.1 Optimization Approaches for Problems Under Uncertainty

2.3.1.1 Stochastic Programming Models

Stochastic programming is an approach for modeling optimization problems with uncertain parameters (Shapiro and Philpott 2007). The models can be two-stage or multi-stage.

Two-stage stochastic programming models can be formulated as follows (Shapiro et al. 2009),

$$\min f_1(x) + E[Q(x, \xi)], s. t. x \in X \subseteq R^n, \quad (2.1)$$

where $Q(x, \xi)$ is the optimal value of the second-stage problem, i.e.,

$$Q(x, \xi) = \min f_2(y, \xi), s. t. T(x, \xi) + W(y, \xi) = h(\xi), y \in Y \subseteq R^m. \quad (2.2)$$

Here x and y are the decision variables in the first and second stages

respectively. Since ξ is a random variable, the coefficients of x and y are uncertain parameters in the second-stage problem. While x should be determined before any realization of ξ becomes available, y is determined after a specific value of ξ is known. The model is linear if its objective functions and constraints are all linear (Shapiro, et al. 2009). If a subset of the decision variables in the first and second stages are required to satisfy the integrality restrictions, the problem is referred to as a two-stage stochastic mixed-integer program (Sen 2005).

The multi-stage model is a natural extension of the two-stage one. Under the multi-stage setting, ξ_1 is known at the first stage, the uncertain parameters $\xi_2, \dots, \xi_T$ are revealed gradually over time and the decisions should be determined accordingly. The decision process has the following form.

$$\begin{aligned} &\text{decision } (x_1) \sim \text{observation } (\xi_2) \sim \text{decision } (x_2) \sim \dots \\ &\dots \sim \text{observation } (\xi_T) \sim \text{decision } (x_T). \end{aligned}$$

The sequence $\xi_t, t = 2, \dots, T$, can be viewed as a stochastic process, i.e., a sequence of random variables with a specified probability distribution. The values of the decision vector x_t at stage t depend on the history of the stochastic process ξ_t up to time t , but not on the those later than t . A T -stage stochastic programming model can be written in the following nested formulation, where $x_t \in X_t \subseteq R^{n_t}, t = 1, \dots, T$, are decision variables.

$$\min_{x_1 \in X_1} f_1(x_1) + E\left[\min_{x_2 \in X_2(x_1, \xi_2)} f_2(x_2, \xi_2) + E[\dots + E[\min_{x_T \in X_T(x_{T-1}, \xi_T)} f_T(x_T, \xi_T)]]\right], \quad (2.3)$$

$$X_t(x_{t-1}, \xi_t) = \{x_t | T_t(x_{t-1}, \xi_t) + W_t(x_t, \xi_t) = h_t(\xi_t), x_t \in R^{n_t}\}.$$

$$\forall t \in \{2, \dots, T\} \quad (2.4)$$

2.3.1.2 Chance-Constrained Models

In some problems, certain constraints do not necessarily hold absolutely. Instead, they can hold within some probability level. The constraints in (2.2) become probabilistic or chance constraints models if they have the following form (Birge and Louveaux 2011).

$$P\{T_i(x, \xi) + W_i(y, \xi) \geq h_i(\xi)\} \geq p_i, y \in Y \subseteq R^m. \quad (2.5)$$

where $0 < p_i < 1$ and $i \in I$, I includes a set of constraints that must hold jointly. In addition, the expression $T_i(x, \xi) + W_i(y, \xi) \geq h_i(\xi)$ can have single or multiple rows (constraints). If it is a single constraint, then the probabilistic constraint $P\{T_i(x, \xi) + W_i(y, \xi) \geq h_i(\xi)\} \geq p_i$ is a single chance constraint. Otherwise, it is called a joint chance constraint (Jiang and Guan 2016, King and Wallace 2012). The models with probabilistic or chance constraints are called chance-constrained models.

Chance-constrained models remain challenging to handle. The current way is to develop the deterministic equivalent form of the probabilistic constraints based on derived properties and solve the models by applying the algorithms for deterministic models (Birge and Louveaux 2011). However, deterministic equivalent is difficult to obtained due to two reasons: (1) chance constraints are non-convex in general, (2) the constraints' probability can be difficult to compute since it requires multi-dimensional integral (Jiang and Guan 2016).

2.3.1.3 Simulation-Based Optimization

The simulation-based optimization is not a modeling approach but a computational framework. It consists of two components, including an optimization component and a simulation component (Keskin et al. 2010). The optimization component consists of a mathematical model that provide solutions

for the simulation component to evaluate. The simulation component depicts a complex stochastic scenario and is commonly seen as a discrete-event system in the literature (Chen, et al. 2012, Gosavi 2003, Keskin, et al. 2010). Large number of independent samples are generated to represent the stochastic elements and used to evaluate the performance of the system under uncertainty. The simulation results work as a feedback to the optimization component, which will provide a new solution to be evaluated in the simulation component. This procedure is repeated iteratively until certain predefined termination criteria are met. The termination criteria can be the solution convergence, the maximal number of iterations, the solving time limit, acceptable objective value, etc. (Nguyen et al. 2014).

2.3.2 Modeling Approaches in Stochastic Programming Models

2.3.2.1 Sample Average Approximation

With known continuous probability distributions of uncertain parameters, closed forms of expressions and solutions for stochastic programming problems are rarely available (Shapiro, et al. 2009). The sample average approximation (SAA) method (Santoso et al. 2005, Verweij et al. 2003) is an approach which makes it practical to solve stochastic optimization problems. The Monte Carlo sampling approach (Birge and Louveaux 2011, Kleywegt et al. 2002, Shapiro, et al. 2009) is used to randomly generate a number of samples (scenarios) from the probability distributions of the uncertain parameters. Each sample (scenario) represents a possible realization of the uncertain parameters in the entire planning horizon. As a result, the expected objective function is approximated by a sample average estimate derived from those random samples (Kleywegt, et al. 2002,

Verweij, et al. 2003). The stochastic constraints are transformed into deterministic ones and the resulted sample average approximation problem can take advantages of existing deterministic optimization algorithms.

Let $\xi^1, \dots, \xi^N$ be independently and identically distributed (i.i.d.) random samples of the random vector ξ , then the two-stage model (2.1) can be estimated by the following sample average approximation problem:

$$\min f_1(x) + Q_N(x), s. t. x \in X \subseteq R^n, \quad (2.6)$$

$$Q_N(x) = \frac{1}{N} \sum_{j=1}^N Q(x, \xi^j). \quad (2.7)$$

(2.1) and (2.6) are referred to as the “true” and SAA problems respectively. Let v^* and v_N represent the optimal values of the corresponding problems. Due to the law of large numbers, $Q_N(x)$ and v_N converge to $E[Q(x, \xi)]$ and v^* respectively with probability 1 as $N \rightarrow \infty$ (increases to infinity). The optimal solution of the SAA problem also converges to the exact optimal solution of the true problem with probability approaching 1 exponentially fast with the increase of N (Kleywegt, et al. 2002).

As proved by Kleywegt, et al. (2002), to find an ε -optimal solution of the true problem with probability at least $1 - \alpha$ in the SAA problem, the sample number N should have the lower bound below, where $\alpha \in (0,1)$, the absolute optimality gap δ has $0 \leq \delta < \varepsilon$, $\sigma_{\max}^2$ is the maximal variance of certain function differences, and $|S|$ is the size of the feasible set of the first stage decision variables.

$$N \geq \frac{3\sigma_{\max}^2}{(\varepsilon - \delta)^2} \log\left(\frac{|S|}{\alpha}\right)$$

The lower bound of N obtained in the above inequality may be far too conservative for practical applications (Verweij, et al. 2003). With the

consideration of the tradeoff between the solution quality and the computational effort required, Kleywegt, et al. (2002) proposed an SAA algorithm to find a moderate number of scenarios which is applicable in practice.

This algorithm estimates the optimality gap and its variance of a feasible solution of the true problem by those in the SAA problem. The solution's objective value of the true problem can be estimated by that of the SAA problem with a large number of scenarios. The optimal objective value of the true problem is estimated by the expected objective value of a number of SAA problems with a moderate scenario number. The two estimates work as the upper and lower bounds respectively of the optimal objective value of the true problem. The feasible solution's optimality gap and gap's variance can be obtained with the two estimates and their variances (Kleywegt, et al. 2002, Santoso, et al. 2005). If the gap and its variance are less than a specified tolerance, the moderate scenario number is sufficient to represent the uncertainty in this problem (Kleywegt, et al. 2002).

According to the following procedure, the algorithm chooses the sample size dynamically.

- (1) Generate M independent samples of size N and solve the corresponding SAA problem (2.6). Denote the optimal objective value by v_N^m and the optimal solution by x_N^m , $m = 1 \dots M$.
- (2) Compute the average of all optimal values of SAA problems, $\bar{v}_{N,M}$ and its variance $\sigma_{\bar{v}_{N,M}}^2$.

$$\bar{v}_{N,M} = \frac{1}{M} \sum_{m=1}^M v_N^m$$

$$\sigma_{\bar{v}_{N,M}}^2 = \frac{1}{(M-1)M} \sum_{m=1}^M (v_N^m - \bar{v}_{N,M})^2$$

As discussed by Norkin et al. (1998) and Mak et al. (1999), the interchange of minimization and mathematical expectation operators in (2.1) gives a lower bound of the problem (2.1), i.e., $\min f_1(x) + E[Q(x, \xi)] \geq E[\min f_1(x) + Q(x, \xi)]$. Hence, the average objective value $\bar{v}_{N,M}$ provides a statistical lower bound for the optimal value of the problem (2.1).

- (3) Pick a feasible first-stage solution $\bar{x}$ for problem (2.1), e.g. one of the solutions x_N^m . With that solution, estimate the objective value of the problem (2.1) using a reference sample size N' .

$$f_1(\bar{x}) + Q_{N'}(\bar{x}) = f_1(\bar{x}) + \frac{1}{N'} \sum_{j=1}^{N'} Q(\bar{x}, \xi^j).$$

By setting the sample size N' as a large number, the estimator $f_1(\bar{x}) + Q_{N'}(\bar{x})$ serves as an upper bound of the optimal objective value. The variance of the upper bound is calculated below.

$$\sigma_{N'}^2(\bar{x}) = \frac{1}{(N'-1)N'} \sum_{j=1}^{N'} (Q(\bar{x}, \xi^j) - Q_{N'}(\bar{x}))^2$$

- (4) Compute the estimators of the optimality gap and its variance.

$$gap_{N,M,N'}(\bar{x}) = Q_{N'}(\bar{x}) - \bar{v}_{N,M}$$

$$\sigma_{gap}^2 = \sigma_{\bar{v}_{N,M}}^2 + \sigma_{N'}^2(\bar{x})$$

- (5) If $gap_{N,M,N'}(\bar{x})$ and σ_{gap}^2 are sufficiently small, choose the best solution from all candidate solutions x_N^m and stop; otherwise, increase the sample size M , and/or N , and/or N' , and return to Step 1.

This algorithm has been applied in some stochastic programming models (Santoso, et al. 2005, Schütz et al. 2009, Verweij, et al. 2003). Numerical results

show that a limited sample size (e.g. dozens of scenarios) can guarantee small optimality gap and variance (Santoso, et al. 2005, Schütz, et al. 2009).

2.3.2.2 Comparison between Two-stage and Multi-stage Stochastic Models

When a problem consists of decisions in a number of periods of the planning horizon and the uncertain parameters are revealed gradually over time, it can be formulated as a multi-stage stochastic programming model (Shapiro, et al. 2009). Nickel et al. (2012) proposed a multi-stage stochastic supply chain network design problem where the location decisions can be made in each period. Ahmed et al. (2003) considered a multi-stage stochastic capacity expansion problem, where the capacity expansion decisions can be determined in each period. Kazemi Zanjani et al. (2010) discussed a multi-stage production planning problem, where the production plan can be determined in each period. Li et al. (2006) proposed a multi-stage stochastic water resource allocation problem, where the water resource can be allocated in each period. Körpeoğlu et al. (2011) discussed a multi-stage stochastic problem where the production quantity and scheduling decisions should be determined in each period.

The multi-period stochastic problems are also be formulated as two-stage stochastic programming models in literature. Alonso-Ayuso et al. (2003) proposed a two-stage stochastic programming model for the supply chain planning under uncertainty, where the strategic decisions (supply chain topology) were determined at the first stage and the tactical decisions over the time horizon are at the second stage of the model. Gebreslassie et al. (2012) proposed a two-stage stochastic programming model where the annual production capacity was determined in the first stage and the production transportation and inventory

levels were determined in the second stage. MirHassani et al. (2000) proposed a two-stage stochastic programming model where the first stage strategic decisions consisted of the opening/closing of sites and DCs and their capacity levels, the second-stage decisions were operational decisions, including the production quantities, packing quantities and transportation amounts.

Compared to a two-stage stochastic programming model, although better solutions can be obtained through a multi-stage one (Huang and Ahmed 2009, Kazemi Zanjani, et al. 2010), the total number of scenarios required to solve the true multi-stage problem with a reasonable accuracy grows exponentially with the increase of the number of stages (Shapiro, et al. 2009). While the SAA method shows good performance for solving two-stage problems, it could be practically inapplicable for multi-stage problems with a large (greater than four) number of stages. Hence, although multi-stage stochastic models have been addressed in the literature (Ahmed, et al. 2003, Kazemi Zanjani, et al. 2010, Li, et al. 2006), most of the researches fall into the categories of scale reductions (Kazemi Zanjani, et al. 2010, Körpeoğlu, et al. 2011) or heuristic algorithms (Ahmed, et al. 2003, Goh et al. 2007, Li, et al. 2006). It is still a challenge to solve multi-stage stochastic problems, especially when there are integer variables in stages other than the first (Sahinidis 2004). As a result, the solutions of multi-stage models may not be better than those of two-stage ones. The applications of multi-stage stochastic models are rare relative to the two-stage ones (Körpeoğlu, et al. 2011).

2.3.2.3 The Rolling Horizon Scheme

Two-stage stochastic model with the rolling horizon scheme is commonly used in practice to reduce the computational complexity (Campbell 2011, Cheung and Chen 1998, Gupta and Maranas 2000). The rolling horizon solution mechanism

is often seen in real-time or periodic review applications, such as process scheduling, production planning, inventory routing and allocation, where the decisions have to be made before demand is known and the decisions can be revised with the subsequent realized demand and updated forecast (Cheung and Chen 1998, Jeunet 2006, Meng and Zhou 2011).

The process of the rolling horizon scheme is shown in Figure 2-1. Denote the current stage (period) by T . The uncertainty in stages (periods) before T has been revealed. By considering the uncertainty in all later stages (periods), the decision makers will make the decision at the current stage (period) and carry it out immediately. As time goes forward, the parameters in stage (period) T are revealed. The probability distributions of the uncertain parameters in stages (periods) later than T might be updated accordingly. Decisions in stage (period) $(T + 1)$ are to be made subsequently. Such kind of sequential decision making process is called a rolling horizon scheme, which will not stop until the last stage (or the end of the planning horizon).

Figure 2-1. The rolling horizon scheme

2.3.3 Algorithms for Stochastic Programming Models

Since they combine the difficulty of stochastic programming with integer programming, stochastic mixed-integer programs (SMIPs) are challenging optimization problems (Ahmed 2010, Klein Haneveld and van der Vlerk 1999). Even though advanced theoretical results and efficient numerical methods have been developed for many optimization problems, such as linear programming,

quadratic programming, convex optimization and nonlinear optimization, no specific formulation plays the role of the generic stochastic programming (Shapiro, et al. 2009). So do SMIPs.

Studies about the properties and solution approaches for stochastic programming are very active (Laporte and Louveaux 1993, Schultz et al. 1996, Sherali and Zhu 2006). Since the problem size and computational complexity increase with the number of scenarios in stochastic programming models, decomposition and heuristic algorithms are widely studied as the solution approaches (Birge and Louveaux 2011). By dividing a stochastic model into a number of smaller and more manageable sub-problems, decomposition algorithms can provide practical approaches for models with a large number of scenarios (Sen 2005). Benders decomposition (Benders 1962, Birge 1985) and progressive hedging algorithm (Rockafellar and Wets 1991) are two commonly used solution approaches in stochastic programming models. They decompose the entire stochastic model by stage and scenario respectively. In the following sections, we will discuss the two algorithms in detail respectively.

2.3.3.1 Benders Decomposition

Benders decomposition is a popular algorithm for solving the two-stage stochastic mixed-integer problems (Alguacil and Conejo 2000, Cordeau et al. 2001, Geoffrion and Graves 1974). It is also known as the L-shaped method in stochastic programming (Van Slyke and Wets 1969).

It solves two-stage mixed-integer stochastic programming problems by partitioning the variables into two mutually exclusive subsets, including one master problem and multiple scenario sub-problems. The master problem is usually a mixed-integer problem, which is a relaxed version of the original

problem, including the first stage's decision variables and associated constraints. The scenario sub-problems are usually linear and independent. Each of them is composed of the second stage's decision variables, associated constraints, and the constraints connecting the decision variables in both stages in a scenario (Costa 2005). Due to their independence, scenario sub-problems can be solved in parallel (Li 2008, Martin 2012). The optimal solution can be obtained by solving the master problem and scenario sub-problems interactively and iteratively. The solution process of Benders decomposition is illustrated in Figure 2-2.

As shown in Figure 2-2, the original stochastic problem is decomposed into a master problem and a set of sub-problems. The notation v represents the iteration number and the variables with the index v represent the variables' solutions in the v th iteration. u_s and π_s respectively represent the dual variables of the corresponding constraints in sub-problems. The solving process starts from the master problem. If the master problem is not solvable (bounded and feasible), this stochastic problem has no bounded solution. Otherwise, the obtained solution is substituted into the dual sub-problems. Depending on the solutions of the dual sub-problems, different cuts are added back to the master problem. If a dual sub-problem is unbounded, a feasibility cut will be added to the master problem. Otherwise, if it has feasible solution, the algorithm can either find the optimal solution of β_s or identify an optimality cut depending on the following rules. If the solution of β_s is no less than the objective value of the corresponding dual sub-problem, the solution is optimal and no further iterations are needed. Otherwise, an optimality cut should be added to the master problem.

Figure 2-2. Benders decomposition

Benders decomposition solves a relaxed problem of the original problem. The master problem is actually a relaxed problem of the original problem. According to the Duality Theorem, all extreme points and extreme rays of the dual sub-problems should satisfy the constraints of feasibility cuts and optimality cuts, which should be added to the master problem (Geoffrion 1970). However, adding so many constraints at once will make the master problem a bulky one and very difficult to solve. Hence, in the working mechanism of Benders decomposition, only a few feasibility or optimality cuts are generated and added to the master problem in each iteration. The master problem is a relaxed problem which can be solved much easier. However, without all those cuts which should

be added, the master problem does not have sufficient information about the sub-problems and cannot give the optimal solution at once. Certain number of iterations are required to achieve the optimality. The number of iterations required may depend on the problem size and structure (Geoffrion 1970).

Although the sub-problems should be convex to apply Benders decomposition (Li 2008), researchers have extended the algorithm to solve stochastic problems with integer variables at the second stage. Laporte and Louveaux (1993) proposed an integer L-shaped method for two-stage stochastic integer problems with pure binary variables at the first stage and an easily solvable second stage. In addition, Carøe and Tind (1998) generalized the integer L-shaped method (Laporte and Louveaux 1993) to deal with problems with mixed-integer variables at both stages. Since the resulted master problem includes nonlinear (possibly discontinuous) cuts in this approach, no practical method for its solution is currently known (Sahinidis 2004).

2.3.3.2 Progressive Hedging Algorithm

Progressive hedging algorithm (PHA, (Rockafellar and Wets 1991)) is a typical scenario-wise decomposition method. It is based on the augmented Lagrangian and aims at solving nonlinear stochastic problems (Birge 1997). It is proved to converge theoretically to global optimal solutions in convex models and can be used as an effective heuristic in nonconvex ones (Fan and Liu 2010, Watson and Woodruff 2011). It is an effective method for solving multi-stage stochastic models, especially for those with integer decision variables in each stage (Gade et al. 2016).

It decomposes a stochastic model into a number of independent deterministic models (scenario sub-problems) and solves these scenario sub-

problems iteratively. Scenario sub-problems are obtained by fixing the random variables to the corresponding realizations in each scenario. All decision variables and constraints in the scenario sub-problems remain the same as those in the multi-stage stochastic models except the non-anticipativity constraints. According to the non-anticipativity constraints, the decisions at a certain period should be identical among a set of scenarios if the random variables in those scenarios are indistinguishable up to the specific period (Rockafellar and Wets 1991). Instead of these constraints, two penalty terms (a linear and a quadratic one) are added in the objective functions of the scenario sub-problems to gradually enforce the convergence of the decision variables' solutions restricted by the non-anticipativity constraints. Meanwhile, according to the solutions obtained in each scenario sub-problem, the coefficients of the linear term are updated in each iteration. This iterative solving process will not stop until the termination criteria are met. Parallel computing can also be applied in the PHA to reduce the computational time significantly (Fan and Liu 2010, Gade, et al. 2016, Ryan et al. 2013).

The basic procedures of the PHA can be stated as follows. At the initial step, the scenario sub-problems are solved independently, where the vectors χ_s and χ_s^v represent all decision variables and the corresponding solutions in scenario s of the multi-stage stochastic model. c_s is a vector representing the corresponding parameters in the objective function. With the average value $\bar{\chi}_n^v$ at node n in the v -th iteration, we calculate the errors (e.g. Euclidean distance) of the solutions in all scenarios and check if convergence is achieved, where p_s is the probability of scenario s . If the error is no greater than the predetermined tolerance threshold ϵ , it is regarded that the convergence is achieved, i.e., the

non-anticipativity constraints are fulfilled. Otherwise, two penalty terms (a linear and a quadratic one) are added in the objective function, where w_s^v and $\rho/2$ are the corresponding coefficients. In practice, the termination criteria can also depend on the total elapsed computational time due to the possible non-convergence (Løkketangen and Woodruff 1996). This procedure is repeated for more iterations and w_s^v is updated in each iteration. Since both penalty terms depend on the parameter ρ , the algorithm's performance can be sensitive to ρ 's value (Watson and Woodruff 2011).

Table 2-1. The progressive hedging algorithm

Progressive Hedging Algorithm	
1.	$v = 0, w_s^v = 0;$
2.	$\chi_s^v = \arg \min_{\chi_s} (c_s \chi_s), \text{ for } \forall s \in S;$
3.	$\bar{\chi}_n^v = \sum_{s \in S, \text{ if } n \in s} p_s \chi_s^v / p_n, \text{ for } \forall n \in N;$
4.	based on the scenario tree structure, the node-based value $\bar{\chi}_n^v$ is transformed into the scenario-based value $\bar{\chi}_s^v$ for $\forall s \in S;$
5.	$d^v = \sum_s p_s \ \chi_s^v - \bar{\chi}_s^v\ ;$
6.	if $d^v \leq \epsilon$, then terminate; otherwise, go on;
7.	$v = v + 1;$
8.	$w_s^v = w_s^{(v-1)} + \rho(\chi_s^{(v-1)} - \bar{\chi}_s^{(v-1)}), \text{ for } \forall s \in S;$
9.	$\chi_s^v = \arg \min_{\chi_s} (c_s \chi_s + w_s^v \chi_s + \rho/2 \ \chi_s - \bar{\chi}_s^{(v-1)}\ ^2), \text{ for } \forall s \in S;$
10.	go to step 3.

Cycling behavior may be detected in the algorithm for stochastic mixed-integer programs (Watson and Woodruff 2011). Table 2-2 illustrates the cycling behavior of the solution of a decision variable in the first stage of a four-scenario stochastic problem. As shown in the table, this decision variable's solution is not identical in all scenarios. In each scenario, the obtained solution from the 28th iteration to the 38th iteration will come again after every 11 iterations. As a result, this decision variable can hardly converge.

Table 2-2. An example of the cycle behavior in progressive hedging algorithm

Iteration	Scenario1	Scenario2	Scenario3	Scenario4	Average
...	...	...	...	...	...
28	48	49	49	49	48.75
29	48	49	49	49	48.75
30	48	49	47	49	48.25
31	48	48	49	48	48.25
32	48	48	47	48	47.75
33	48	47	47	47	47.25
34	48	47	47	47	47.25
35	48	47	47	47	47.25
36	48	47	49	47	47.75
37	48	49	46	49	48
38	48	48	51	48	48.75
39	48	49	49	49	48.75
40	48	49	49	49	48.75
41	48	49	47	49	48.25
42	48	48	49	48	48.25
43	48	48	47	48	47.75
44	48	47	47	47	47.25
45	48	47	47	47	47.25
46	48	47	47	47	47.25
47	48	47	49	47	47.75
48	48	49	46	49	48
49	48	48	51	48	48.75
50	48	49	49	49	48.75
51	48	49	49	49	48.75
52	48	49	47	49	48.25
53	48	48	49	48	48.25
54	48	48	47	48	47.75
55	48	47	47	47	47.25
56	48	47	47	47	47.25
57	48	47	47	47	47.25
58	48	47	49	47	47.75
59	48	49	46	49	48
60	48	48	51	48	48.75
61	48	49	49	49	48.75
62	48	49	49	49	48.75
63	48	49	47	49	48.25
64	48	48	49	48	48.25
65	48	48	47	48	47.75
66	48	47	47	47	47.25
67	48	47	47	47	47.25
68	48	47	47	47	47.25
69	48	47	49	47	47.75
70	48	49	46	49	48
71	48	48	51	48	48.75
72	48	49	49	49	48.75
73	48	49	49	49	48.75
74	48	49	47	49	48.25
75	48	48	49	48	48.25
76	48	48	47	48	47.75
77	48	47	47	47	47.25
78	48	47	47	47	47.25

79	48	47	47	47	47.25
80	48	47	49	47	47.75
81	48	49	46	49	48
82	48	48	51	48	48.75

Watson and Woodruff (2011) proposed a mechanism to detect and avoid cycles based on w_s^v . w_s^v is the coefficient of the linear term added to the objective function. It is updated in each iteration according to the solutions obtained in previous iterations. Due to the integrality restrictions, the decision variables' solutions and average values may stagnate for many iterations. As a result, the variation of w_s^v is much easier to capture than that of its corresponding decision variable's solution χ_s^v . Hence, w_s^v is used in cycle detection. Assume $w_{s,i}$ and $\chi_{s,i}$ are the i -th element of the vectors w_s and χ_s respectively. For computational efficiency, the values $w_{s,i}^v$ of all scenarios are jointly considered to generate a hash value for the variable $\chi_{s,i}$ by applying a simple hashing scheme proposed in the literature (Woodruff and Zemel 1993). If a variable's hash value in an iteration is identical to that obtained in a previous iteration, a cycle is detected. The solution of $\chi_{s,i}$ is then fixed to $\max_{s \in S} \chi_{s,i}^v$ to break the cycle.

2.3.3.3 Comparison

Benders decomposition can guarantee the rigorous optimality and converge more naturally than the progressive hedging algorithm. Since Benders decomposition is based on the duality, rigorous optimality can be guaranteed. In contrast, progressive hedging algorithm gets the solutions by relaxing and then enforcing the non-anticipativity constraints, it finds feasible but not necessarily optimal solutions. Moreover, while the working mechanism makes Benders decomposition converge naturally and simply, the convergence of progressive

hedging algorithm depends largely on the specified parameters and is difficult to predict.

The progressive hedging algorithm is more practical than Benders decomposition, especially in multi-stage stochastic programming models. Since the dual information of the programs in stages other than the first is applied in Benders decomposition, the convexity property is required in those stages (Carøe and Tind 1998). In addition, although Benders decomposition can be applied in multi-stage stochastic models via a nested approach, the implementation is quite cumbersome in practice (Sherali and Fraticelli 2002). On the other hand, with the fully scenario decomposition, the progressive hedging algorithm is much more efficient to get solutions than Benders decomposition, especially when the number of scenarios is huge (Kim and Wu 2013). In addition, even if the predefined optimality criteria are not achieved, the obtained solution is feasible and can work as a reasonable solution.

2.3.4 Algorithms for Multi-objective Models

The multi-objective optimization approaches can be categorized into classical methods and meta-heuristic algorithms (Deb 2001, Jones et al. 2002). The weighted sum method (Marler and Arora 2010, Zadeh 1963) and ϵ -constraint method (Haimes et al. 1971, Ramezani, et al. 2013) are two typical classical methods. In such methods, multi-objective problems are converted to single objective ones and solved with conventional scalar optimization algorithms. Certain parameters, such as the orders of importance of the multiple objectives and the indication of target optimal objective values, are required in the conversion (Ngatchou et al. 2005). In each converted single objective problem with specified parameters, we can only get a single Pareto-optimal solution

(Konak et al. 2006, Ngatchou, et al. 2005). On the other hand, meta-heuristic algorithms, such as the genetic algorithm (Altıparmak, et al. 2006, Deb 1999), simulated annealing (Kirkpatrick 1984, Pasandideh, et al. 2013) and tabu search (Belfares et al. 2007), deal with a set of possible solutions simultaneously to find all Pareto-optimal solutions in a single run (Coello 1999).

2.4 Summary of Research Gaps

The clinical trial supply chain is a supply chain for a product in R&D phase and differs greatly from those for commercial products. It has a number of characteristics, including the limited (non-replenished) inventory, non-stationary stochastic demand, and transshipment options. In addition, the clinical trial duration is counted in the limited patent life of drugs. The leftover clinical drugs cannot be used elsewhere and should be disposed of after an unsuccessful trial.

In this chapter, we review the relevant studies in pharmaceutical supply chains and the problems of inventory allocation, network design, and production planning respectively in clinical trial supply chains. The research gaps between our study and the literature are also discussed for each of the three problems. They are summarized as follows.

The inventory allocation problem is widely studied in the literature. Most researchers discussed the stationary inventory policies in other applications, which are proven optimality at a single retailer with stationary stochastic demand. When they are applied jointly with transshipment options, the optimal policies in a multi-echelon supply chain can only be obtained within a single period, deterministic demand, few retailers (e.g. two or three), or multiple retailers with identical cost parameters at every retailer. Non-stationary inventory policies are studied by only a limited number of researchers. Some researchers also studied

the inventory allocation in supply chains with the limited inventory. However, the problem settings are restricted to either single period or deterministic demand. As a result, none of the researchers has studied the inventory allocation in clinical trial supply chains or other applications with the consideration of all the features (including the limited inventory, non-stationary stochastic demand, and transshipment options).

The supply chain structure should also be designed elaborately to improve the efficiency of clinical trials. More than one performance measures are usually considered in network design problems. The duration and cost are two important measures in clinical trial supply chains. Few researchers have studied the duration minimization in clinical trials or the cost in clinical trial supply chain design. Some researchers studied the time-cost tradeoff in other applications. However, the clinical trial duration is different from the time in these applications. While the time to be minimized in these applications is the total time required to complete a predetermined number of jobs with limited resources and fixed processing times, the clinical trial duration is the time required to achieve the predetermined patient accrual goal in the supply chain with stochastic demand. It is not restricted by the available resources or processing times but the stochastic patient arrivals at clinical sites.

The production planning in clinical trial supply chains has been studied in the literature. Some researchers studied it with jointly the distribution planning and assumed stationary inventory policies at each retailer. The production planning is also studied with the consideration of clinical drugs' failure rates and demand realizations. The production and distribution planning problem is widely studied in other applications. Most of researchers consider commercial products,

which are usually characterized by production and storage capacities, salvage cost, and immediate stock-out cost. Hence, they mainly minimize the stock-out cost but do not consider the overage cost. On the contrary, we aim to reduce the overage and only consider the stock-out implicitly in clinical trial supply chains. When there exists no stock-out in clinical trial supply chains, the shortest clinical trial duration can be simply obtained. However, since a clinical trial stops immediately when it reaches its patient accrual goal, the stock-outs after the clinical trial completion will not increase the clinical trial duration. In addition, the overage is a more important concern in clinical trial supply chains. Since the drugs are becoming more expensive nowadays and leftover drugs should be disposed of after an unsuccessful trial, overproduction should be reduced to improve the supply chain's efficiency.

None of the studies in the literature considers problems similar to ours. The inventory policy is also analytically intractable due to the characteristics of clinical trial supply chains. Hence, a numerical approach through stochastic programming is proposed to formulate the inventory allocation problem. Since the network design and production planning problems also consider the delivery and holding costs in inventory allocation, they are formulated as stochastic programming models as well. With the corresponding algorithms discussed in Section 2.3, we can find the optimal decisions of these problems. In the following chapters, we will present each problem's mathematical model, properties, and numerical results respectively.

Chapter 3 Inventory Allocation in Clinical Trial Supply Chains

3.1 Introduction

The clinical trial plays a key role in the competitiveness of a pharmaceutical company and its efficiency can be improved through the efficient coordination of the inventories among DCs and clinical sites in the supply chain. With the globalization of clinical trials, clinical drugs are packed and labelled in multiple languages with country-specific contents (Cusworth 2017). As shown in Figure 3-1, the delivery can be longitudinal (cross echelons) or lateral (within the same echelon). The longitudinal delivery can be direct or reverse replenishment. Direct replenishment means that the inventory orders at clinical sites are fulfilled by a DC. In a clinical trial supply chain, sometimes a DC may recall the inventory from the clinics with excess inventory and then redistribute it to those who are facing (or going to face) stock-outs. This activity is called *reverse replenishment* in this study. By convention, the lateral delivery is called lateral transshipment and is referred to as the directly shipping the inventory at a clinic to some other clinics. Lateral transshipment and/or reverse replenishment may occur at the end of a selling season when the DC has low or zero inventory, even if they take considerable lead time and/or cost. Due to the double-blind requirement and various regulations at different countries, reverse replenishment is commonly seen but lateral transshipment only exists between a few clinical sites within the same country.

Figure 3-1. The delivery routes in a clinical trial supply chain

When transshipments reduce the stock-outs, they inevitably increase the delivery cost. To avoid the transshipments, the inventory should be held at the DC and the DC has to deliver more frequently with smaller shipping quantities. The delivery cost will be increased as well. Hence, in the inventory allocation problem, the lateral transshipment and reverse replenishment are applied jointly with the direct replenishment to optimize the system performance. For simplicity, lateral transshipment is called transshipment in the following parts.

Since inventory policies are not applicable in this problem according to the discussion in Section 2.2.1, a numerical approach through the two-stage stochastic programming is applied to model the inventory allocation problem in clinical trial supply chains. All types of deliveries are jointly considered in the model. Benders decomposition is applied as the solution approach to find out the optimal inventory allocation decisions and some acceleration techniques are identified to improve the solving efficiency. In numerical experiments, the performance of the proposed acceleration techniques is checked by comparing the required iteration number and computational time under different settings. The model's performance is also identified by comparing the total cost between our approach and the (s, S) policy.

The rest of this chapter is organized as follows. The model development is given in Section 3.2, including the model assumptions, formulation, and analysis.

The acceleration techniques and numerical results are presented in Sections 3.3 and 3.4 respectively. The conclusion is given in the end.

3.2 Model Development

Since clinical trial supply chains differ fundamentally from commercial supply chains, classical inventory policies cannot be applied directly. In this section, we develop a rolling-based two-stage stochastic programming model to minimize the total inventory allocation cost in clinical trial supply chains.

3.2.1 Assumptions

Our model considers a specific drug candidate in Phase III of the clinical trial. The model assumptions are summarized below.

- (1) The demand distribution at each clinical site is given and follows a non-stationary Poisson distribution.
- (2) The total inventory in the planning horizon is equal to the given initial inventory at the DC.
- (3) Delivery cost is independent of the shipping quantity.
- (4) Transshipment can be lateral or longitudinal.
- (5) Lateral transshipment and reverse replenishment are more expensive than the direct replenishment, and the stock-out cost is more expensive than the delivery cost.
- (6) Patients are fulfilled with the same clinical drugs in the entire treatment period, and the stock-out cost is identical for the new and existing patients.
- (7) The dropout rate is assumed to follow a uniform distribution and the forecast demand for patients for follow-up treatments is rounded down if it is not an integer.

The demand for clinical drugs is stochastic at each clinical site in the clinical trial supply chain due to the patient recruitment uncertainty. Its distribution can be estimated from the track record of patients visiting the site and/or the previous trial data for the same disease (Fleischhacker, et al. 2015). It is commonly assumed to follow a non-stationary Poisson distribution at each site (Chen, et al. 2012). Drugs usually have a shelf life of several years, which is able to cover the entire planning horizon of a clinical trial. The clinical drug production is featured by long lead time and notable changeover cost. As a result, drugs are usually produced in large batches and we assume production is not conducted during clinical trials. Hence, the total inventory in the planning horizon is equal to the given initial inventory at the DC, which is a variable determined at the production planning level.

Since the volume of drug packs is small, each shipment can contain many packs with the same shipping cost in practice. The delivery cost is assumed to be independent of the shipping quantity. In a clinical trial supply chain, the DC and clinics are geographically dispersed. At the beginning of a replenishment cycle, the DC has sufficient inventory and clinics are replenished by the DC directly. However, in the later periods of a replenishment cycle, the inventory at the DC is low or depleted and the clinics with excess inventory have to replenish the clinics with insufficient inventory. To minimize the cost of the supply chain, both the reverse replenishment and transshipment should be considered jointly with the direct replenishment. Since lateral transshipment and reverse replenishment are done by clinics rather than the DC, their costs are higher than the cost of direct replenishment. Due to the various regulations at different countries, lateral transshipment only exists between a few clinical sites, but reverse replenishment

is commonly seen, especially for expensive drugs which have little overage to hedge the demand uncertainty.

All new and existing patients for the follow-up treatments are fulfilled with the same clinical drugs in a clinical trial. Since the treatment period (consisting of several weeks) is much shorter than the clinical trials duration (up to several years), the costs of losing a new or an existing patient (due to stock-outs or dropout) will not differ much. Furthermore, when facing inventory stock-outs, a clinic will try to fulfill the existing patients first. Hence, it is reasonable to assume the stock-out cost is identical for the new and existing patients. In addition, when we make a forecast of the numbers of existing patients who will return for follow-up treatments, due to the dropout rate, these numbers may not be integers. They should be rounded up or down to guarantee their integral nature in the mathematical formulation. Without loss of generality, we round them down in the following section. Similar to the assumption in the literature (Chen, et al. 2012), we assume the dropout rate follows a uniform distribution and has identical values in each period. In practice, it may have smaller values in the later periods of the treatment period since a patient who does not drop out early on is more likely to stay later.

3.2.2 Mathematical Model

We formulate the multi-period inventory allocation problem in clinical trial supply chains as a rolling-based two-stage stochastic programming model. The objective is to allocate the limited amount of inventory to satisfy the stochastic demand with the minimal total stock-out, replenishment, reverse replenishment, transshipment, and holding costs. To facilitate the solving process of this problem with the existing deterministic optimization algorithms, we discretize the

uncertain demand at clinical sites into a set of possible future scenarios. The variables and parameters in the model are defined as follows.

Table 3-1. Summary of the notations

Notation	Description
Sets:	
R	Set of demand scenarios, $R = \{r r = 1, \dots, m\}$
L	Planning horizon of a clinical trial (number of time periods), $L = \{t t = 1, \dots, l\}$
I	Set of clinical sites, $I = \{i i = 1, \dots, n\}$
Parameters:	
w_i	Replenishment cost from DC to clinical site i
rw_i	Reverse replenishment cost from clinical site i to DC
$w_{i,j}$	Transshipment cost from clinical site i to site j
u	Stock-out cost at each clinical site
h_{DC}	Holding cost at DC in each period
h_i	Holding cost at clinical site i in each period
l_i	Delivery time between DC and clinical site i
$l_{i,j}$	Delivery time from clinical site i to site j
Ca_i	Capacity of clinical site i
ID_0	Inventory level of DC at the beginning of period 0
$d_{r,t,i}$	Forecast new arrival patients at clinical site i in period t in scenario r
$D_{t,i}$	Realized new arrival patients at clinical site i in period t
q	The number of periods in a treatment period
ρ	The average dropout rate in each period
p_r	Probability of scenario r
T	The current period, moving from 1 to l
M	A sufficiently large number (e.g. one may define $M = ID_0$)
Decision variables:	
$a_{t,i}$	1 if clinical site i orders from DC at the beginning of period t , 0 otherwise
$sa_{r,t,i}$	1 if clinical site i orders from DC at the beginning of period t in scenario r , 0 otherwise
$b_{t,i}$	1 if DC is replenished from clinical site i at the beginning of period t , 0 otherwise
$sb_{r,t,i}$	1 if DC is replenished from clinical site i at the beginning of period t in scenario r , 0 otherwise
$c_{t,i,j}$	1 if there exists transshipment from clinical site i to site j at the beginning of period t , 0 otherwise
$sc_{r,t,i,j}$	1 if there exists transshipment from clinical site i to site j at the beginning of period t in scenario r , 0 otherwise

$x_{t,i}$	Order quantity of clinical site i from DC at the beginning of period t
$sx_{r,t,i}$	Order quantity of clinical site i from DC at the beginning of period t in scenario r
$y_{t,i}$	Reverse replenishment quantity of clinical site i at the beginning of period t
$sy_{r,t,i}$	Reverse replenishment quantity of clinical site i at the beginning of period t in scenario r
$z_{t,i,j}$	Transshipment quantity from clinical site i to site j at the beginning of period t
$sz_{r,t,i,j}$	Transshipment quantity from clinical site i to site j at the beginning of period t in scenario r
ID_t	Inventory level of DC at the beginning of period t after decision making
$sID_{r,t}$	Inventory level of DC at the beginning of period t in scenario r after decision making
$IC_{t,i}$	Inventory level of clinical site i at the beginning of period t after decision making
$sIC_{r,t,i}$	Inventory level of clinical site i at the beginning of period t in scenario r after decision making
$LC_{t,k,i}$	Stock-out amount in their k th period of the treatment period for patients who first arrived at clinical site i at period t , $k \leq q$
$sLC_{r,t,k,i}$	Stock-out amount in their k th period of the treatment period for patients who first arrived at clinical site i at period t in scenario r , $k \leq q$
$SD_{t,k,i}$	The number of patients who first arrived at clinical site i at period t and have already been satisfied in the clinical trial for k periods, $k \leq q$
$sSD_{r,t,k,i}$	The number of patients who first arrived at clinical site i at period t in scenario r and have already been satisfied in the clinical trial for k periods, $k \leq q$
θ	A vector representing all decision variables at period T , $\theta = (x, y, z, IC, ID)$
η	A vector representing all scenario binary decision variables in periods after T , $\eta = (sa, sb, sc)$

The inventory allocation problem in clinical trial supply chains is formulated as the following two-stage stochastic programming model, where (P1-MP) is the first-stage problem and (P1-SP) is the second-stage problem for each forecast demand scenario. Due to the positive delivery lead time and fixed shipping cost (which is independent of the shipping quantity), the inventory allocation decisions in all periods of the planning horizon should be considered

jointly to minimize the total cost in the supply chain. The decisions at the first stage consist of those to be determined at the beginning of the current period T , when the demands in period T are not known. These decisions are identical for all scenarios. The decisions at the second stage consist of those in all periods later than T . They are determined based on the forecast demand in each scenario and work as a corrective or recourse action of the first-stage decisions to optimize the performance in each scenario (e.g. minimize each scenario's total cost). They may have distinct solutions in different scenarios.

Since the inventory allocation can be conducted in each period and both the inventory level and forecast demand can be updated after the demand realization, the model is solved under the rolling horizon scheme shown in Figure 2-1. By minimizing the total costs at the current period and all scenarios' expected total costs in the following periods, we can find the optimal inventory allocation decisions in the current period and all following periods for all scenarios. The decisions at the current period are implemented immediately in the supply chain. When the demand is realized at the end of the current period, the inventory levels and forecast demand can be updated accordingly, and the forecast can be more accurate if they are closer to the current period. Hence, the forecast demand scenarios are generated by assuming that the variance of the forecast demand is smaller if the period is closer to the current period. The model is solved subsequently to make the decisions in the period $T + 1$. This procedure will continue until the end of the planning horizon to find the optimal inventory allocation decisions in each period. Note that the decision variables in all periods before T are deterministic in the model of period T , since they have already been implemented in the models at prior periods under the rolling scheme.

Period T : for $T = 1:l$

$$\begin{aligned} \text{(P1-MP): } \min \quad & \sum_i (w_i a_{T,i} + r w_i b_{T,i} + \sum_{j \in I, j \neq i} w_{i,j} c_{T,i,j} + h_i IC_{T,i}) + \\ & h_{DC} ID_T + \sum_r p_r Q(\theta, r) \end{aligned} \quad (3.1)$$

$$\text{s.t. } ID_T = ID_{(T-1)} - \sum_i x_{T,i} + \sum_i y_{(T-l_i),i}, \quad (3.2)$$

$$IC_{T,i} = x_{T,i}, \quad \forall i \in I, \text{ if } T = 1 \quad (3.3)$$

$$\begin{aligned} IC_{T,i} = IC_{(T-1),i} - \sum_{t=\max(T-q,1)}^{T-1} SD_{t,(T-t),i} + x_{(T-l_i) \setminus \{1\},i} \\ - y_{T,i} - \sum_{j \in I, j \neq i} (z_{T,i,j} - z_{(T-l_j,i),j,i}), \forall i \in I, \text{ if } T > 1 \end{aligned} \quad (3.4)$$

$$IC_{t,i} \leq C a_i, \quad \forall i \in I, \forall t \in L \quad (3.5)$$

$$x_{T,i} \leq M a_{T,i}, \quad \forall i \in I \quad (3.6)$$

$$y_{T,i} \leq M b_{T,i}, \quad \forall i \in I \quad (3.7)$$

$$z_{T,i,j} \leq M c_{T,i,j}, \quad \forall i, j \in I \quad (3.8)$$

$$\begin{aligned} x_{t,i} \geq 0, y_{t,i} \geq 0, z_{t,i,j} \geq 0, ID_t \geq 0, IC_{t,i} \geq 0, SD_{t,i} \geq 0, \\ \forall i, j \in I, \forall t \in L \end{aligned} \quad (3.9)$$

$$y_{T,i} = 0, z_{T,i,j} = 0, IC_{(T-1),i} = 0, \quad \forall i, j \in I, \text{ if } T = 1 \quad (3.10)$$

$$a_{t,i}, b_{t,i}, c_{t,i,j} \in \{0,1\}. \quad \forall i, j \in I, \forall t \in L \quad (3.11)$$

The objective function of the first-stage problem (3.1) is to minimize the total delivery and holding costs at the current period and the expectation of $Q(\theta, r)$, which represents the optimal objective value of the second-stage problem in scenario r . Constraints (3.2) to (3.4) are the inventory balance constraints for the DC and clinical sites at the current period. Constraint (3.5) defines the inventory capacity constraints at clinical sites. Constraints (3.6) to (3.8) enforce the fixed cost of replenishment, reverse replenishment and transshipment. Constraints (3.9) to (3.11) are non-negativity, zero, or binary constraints for variables.

Constraints (3.3) and (3.10) indicate the special case at the period $T = 1$. Before the start of clinical trials, all drugs are located at the DC and there is no prior replenishment, reverse replenishment, transshipment, or inventory at clinics. To hedge the potential stock-outs at clinical sites during the nonnegative lead time between the DC and clinical sites, the replenishment at the period $T = 1$ is assumed to arrive at each clinical site immediately. (In practice, the inventory allocation decisions at the first period can be determined and implemented in advance to guarantee the inventories arrive and there exist initial inventories at clinical sites before the start of clinical trials.) Obviously, the replenishment at the period $T = 1$ will not arrive again after the positive lead time.

$$\begin{aligned}
 \text{(P1-SP): } Q(\theta, r) = \min \sum_i [& u \sum_{k=1}^q \sum_{t=T}^l sLC_{r,t,k,i} + \sum_{t=(T+1)}^l (w_i sa_{r,t,i} + \\
 & rw_i sb_{r,t,i} + \sum_{j \in I, j \neq i} w_{i,j} sc_{r,t,i,j} + h_i sIC_{r,t,i})] + \sum_{t=(T+1)}^l h_{DC} sID_{r,t}
 \end{aligned} \tag{3.12}$$

$$\begin{aligned}
 \text{s.t. } sID_{r,t} &= sID_{r,(t-1)} - \sum_i sx_{r,t,i} + \sum_i sy_{r,(t-l_i),i}, \\
 &\forall t \in \{(T+1), \dots, l\}
 \end{aligned} \tag{3.13}$$

$$sID_{r,T} = ID_T, \tag{3.14}$$

$$\begin{aligned}
 sIC_{r,t,i} &= sIC_{r,(t-1),i} - \sum_{\tau=t-q}^{t-1} sSD_{r,\tau,(t-\tau),i} + sx_{r,(t-l_i) \setminus \{1\},i} \\
 &- sy_{r,t,i} - \sum_{j \in I, j \neq i} (sz_{r,t,i,j} - sz_{r,(t-l_{j,i}),j,i}), \\
 &\forall i \in I, \forall t \in \{(T+1), \dots, l\}
 \end{aligned} \tag{3.15}$$

$$\begin{aligned}
 sIC_{r,(l+1),i} &= sIC_{r,l,i} - \sum_{\tau=l-q+1}^l sSD_{r,\tau,(l+1-\tau),i} \\
 &+ sx_{r,(l-l_i) \setminus \{1\},i} + \sum_{j \in I, j \neq i} sz_{r,(l-l_{j,i}),j,i}, \quad \forall i \in I
 \end{aligned} \tag{3.16}$$

$$sIC_{r,T,i} = IC_{T,i}, \quad \forall i \in I \tag{3.17}$$

$$sSD_{r,t,1,i} + sLC_{r,t,1,i} = d_{r,t,i}, \quad \forall i \in I, \forall t \in \{T, \dots, l\} \tag{3.18}$$

$$sSD_{r,t,k,i} + sLC_{r,t,k,i} \leq (1 - \rho) sSD_{r,t,(k-1),i},$$

$$\forall i \in I, \forall t \in \{T, \dots, l\}, \forall k \in \{2, \dots, q\} \quad (3.19)$$

$$sSD_{r,t,k,i} + sLC_{r,t,k,i} > (1 - \rho)sSD_{r,t,(k-1),i} - 1,$$

$$\forall i \in I, \forall t \in \{T, \dots, l\}, \forall k \in \{2, \dots, q\} \quad (3.20)$$

$$sSD_{r,\tau,k,i} + sLC_{r,\tau,k,i} \leq (1 - \rho)sSD_{r,\tau,(k-1),i},$$

$$k = t + 1 - \tau \ \& \ k \leq q,$$

$$\forall i \in I, \forall \tau \in \{(T - q + 1), \dots, (T - 1)\}, \forall t \in \{T, \dots, l\} \quad (3.21)$$

$$sSD_{r,\tau,k,i} + sLC_{r,\tau,k,i} > (1 - \rho)sSD_{r,\tau,(k-1),i} - 1,$$

$$k = t + 1 - \tau \ \& \ k \leq q,$$

$$\forall i \in I, \forall \tau \in \{(T - q + 1), \dots, (T - 1)\}, \forall t \in \{T, \dots, l\} \quad (3.22)$$

$$sSD_{r,t,(T-t),i} = sD_{t,(T-t),i},$$

$$\forall i \in I, \forall t \in \{(T - q + 1), \dots, (T - 1)\} \quad (3.23)$$

$$sIC_{r,t,i} \leq Ca_i, \quad \forall i \in I, \forall t \in \{(T + 1), \dots, (l + 1)\} \quad (3.24)$$

$$sx_{r,t,i} \leq Msa_{r,t,i}, \quad \forall i \in I, \forall t \in \{(T + 1), \dots, l\} \quad (3.25)$$

$$sy_{r,t,i} \leq Msb_{r,t,i}, \quad \forall i \in I, \forall t \in \{(T + 1), \dots, l\} \quad (3.26)$$

$$sz_{r,t,i,j} \leq Msc_{r,t,i,j}, \quad \forall i, j \in I, \forall t \in \{(T + 1), \dots, l\} \quad (3.27)$$

$$sx_{r,t,i} \geq 0, sy_{r,t,i} \geq 0, sz_{r,t,i,j} \geq 0, sID_{r,t} \geq 0,$$

$$\forall i, j \in I, \forall t \in \{(T + 1), \dots, l\} \quad (3.28)$$

$$sIC_{r,t,i} \geq 0, \quad \forall i \in I, \forall t \in \{(T + 1), \dots, (l + 1)\} \quad (3.29)$$

$$sSD_{r,t,k,i} \geq 0, sLC_{r,t,k,i} \geq 0,$$

$$\forall i \in I, \forall t \in \{T, \dots, l\}, \forall k \in \{1, \dots, q\} \quad (3.30)$$

$$sa_{r,t,i}, sb_{r,t,i}, sc_{r,t,i,j} \in \{0, 1\}.$$

$$\forall i, j \in I, \forall t \in \{(T + 1), \dots, l\} \quad (3.31)$$

The objective function of the second-stage problem (3.12) minimizes the total delivery, holding and stock-out costs in scenario r . New and existing

patients returned for the follow-up treatments are fulfilled with the same clinical drugs. Constraints (3.13) to (3.17) are the inventory balance constraints for the DC and clinical sites in periods later than T . Constraints (3.18) to (3.22) guarantee the sum of the served demand and stock-out amount equals the demand, where the treatment period, dropout rate, and the integral nature of decision variables are all considered. $(1 - \rho)sSD_{r,t,k,i}$ at the right hand sides of constraints (3.19)~(3.22) are used to calculate the demand of the existing patients for their follow-up treatments. Due to the decimal dropout rate ρ between 0 and 1, $(1 - \rho)sSD_{r,t,k,i}$ may not be integers. Since the patient arrivals follow Poisson distributions and should be integers, we apply the constraints (3.19)~(3.20) and (3.21)~(3.22) to round down $(1 - \rho)sSD_{r,t,k,i}$. As a result, both the demand and the left hand side (i.e., the sum of the served demand and stock-out amount) are integers. Constraints (3.14), (3.17) and (3.23) represent the non-anticipativity constraints for the known inventory levels and satisfied demand at the current period. Constraint (3.24) defines inventory capacity constraints at clinical sites. Constraints (3.25) to (3.27) enforce the fixed cost of replenishment, reverse replenishment and transshipment. Constraints (3.28) to (3.31) are non-negativity or binary constraints for variables.

After solving the problem at the current period T , the inventory levels at DC and all clinical sites are updated. The realized demands in period T are either fulfilled or lost, including the new arrival patients and the existing patients for the follow-up treatments. The satisfied demand can be updated with the following formulations.

$$\min \sum_i \sum_{t=\max(T-q+1,1)}^T LC_{t,(T-t+1),i} \quad (3.32)$$

$$\text{s.t. } SD_{T,1,i} + LC_{T,1,i} = D_{T,i}, \quad \forall i \in I \quad (3.33)$$

$$SD_{t,k,i} + LC_{t,k,i} \leq (1 - \rho)SD_{t,(k-1),i},$$

$$k = T - t + 1, \forall i \in I, \forall t \in \{(T - q + 1), \dots, (T - 1)\} \quad (3.34)$$

$$SD_{t,k,i} + LC_{t,k,i} > (1 - \rho)SD_{t,(k-1),i} - 1,$$

$$k = T - t + 1, \forall i \in I, \forall t \in \{(T - q + 1), \dots, (T - 1)\} \quad (3.35)$$

$$\sum_{t=\max(T-q+1,1)}^T SD_{t,(T-t+1),i} \leq IC_{T,i}, \quad \forall i \in I \quad (3.36)$$

$$SD_{t,k,i} \geq 0, LC_{t,k,i} \geq 0,$$

$$\forall i \in I, \forall t \in \{(T - q + 1), \dots, T\}, \forall k \in \{1, \dots, q\} \quad (3.37)$$

3.2.3 Model Analysis

Even if we consider single period, only deterministic demand, zero holding cost, and no transshipment or reverse replenishment, the clinical trial inventory allocation problem is NP-hard. If the total demand is greater than the initial inventory at the DC and the stock-out cost is a very large number, all inventory at the DC will be distributed to clinical sites to achieve the minimal stock-outs in the supply chain. The total shipping cost is minimized as well. As a result, the clinical trial inventory allocation problem can be reduced to a min-knapsack problem as follows:

$$\min \left\{ \sum_i w_i a_{1,i} : s. t. \sum_i D_{1,i} a_{1,i} \geq ID_0, a_{1,i} \in \{0,1\} \forall i \in I \right\}.$$

Since the min-knapsack problem is NP-hard (Csirik et al. 1990), unless P = NP, it is not possible to develop an algorithm that can run in polynomial time.

In this research, we apply Benders decomposition to solve the proposed model. Benders decomposition is commonly referred to as the L-shaped method when applied to stochastic problems (Rahmaniani et al. 2016). It can decompose a large-scale model into a master problem and multiple scenario sub-problems. (P1-MP) and (P1-SP) are actually the master problem and a scenario sub-problem

respectively. Since the sub-problems are independent, they can be solved in parallel (Li 2008, Martin 2012).

Although Benders decomposition enjoys above advantages, to apply it, the sub-problems must be convex (Li 2008). However, the sub-problems (P1-SP) are not convex since the delivery decisions, $sa_{r,t,i}$, $sb_{r,t,i}$ and $sc_{r,t,i}$, are binaries. As discussed in Section 2.3.2.1, although researchers have extended Benders decomposition to solve the stochastic problems with integer variables at the second stage, the proposed approaches cannot be applied in our problem. To facilitate the application of Benders decomposition, we propose to model all scenario binary variables (including $sa_{r,t,i}$, $sb_{r,t,i}$ and $sc_{r,t,i}$ for $\forall r, i, j$ and $\forall t \in \{T+1, \dots, l\}$) in the master problem instead of the sub-problems. The modified master problem is

$$\begin{aligned}
 \text{(P2-MP): } \{ \min \quad & \sum_i (w_i a_{T,i} + r w_i b_{T,i} + \sum_{j \in I, j \neq i} w_{i,j} c_{T,i,j} + h_i IC_{T,i}) + \\
 & h_{DC} ID_T + \sum_r p_r \beta_r + \sum_r p_r \sum_i \sum_{t=(T+1)}^l (w_i sa_{r,t,i} + r w_i sb_{r,t,i} + \\
 & \sum_{j \in I, j \neq i} w_{i,j} sc_{r,t,i,j}) : s. t. (3.2) - (3.11), (3.31) \}. \quad (3.38)
 \end{aligned}$$

where β_r is a variable which represents the lower bound of the objective value of the sub-problem in scenario r . The expectation of β_r represents the lower bound of expected stock-out and holding costs of all scenarios. The corresponding sub-problem in scenario r is presented as follows.

$$\begin{aligned}
 \text{(P2-SP): } \{ \min \quad & u \sum_i \sum_{k=1}^q \sum_{t=T}^l sLC_{r,t,k,i} + \sum_{t=(T+1)}^l (\sum_i h_i sIC_{r,t,i} + \\
 & h_{DC} SID_{r,t}) : s. t. (3.13) - (3.30) \}. \quad (3.39)
 \end{aligned}$$

Since (P2-SP) is a linear programming model, Benders decomposition can be applied to (P2). However, solving (P2) using Benders decomposition is inefficient due to the following issues.

- (1) Solving an iteration of (P2-MP) can be more time-consuming than solving (P1-MP). After adding the scenario binary variables, $sa_{r,t,i}$, $sb_{r,t,i}$ and $sc_{r,t,i,j}$, the solving time of the master problem may be increased.
- (2) The convergence rate of (P2) can be slow. All scenario binary variables relocated to the master problem (including $sa_{r,t,i}$, $sb_{r,t,i}$ and $sc_{r,t,i,j}$ for $\forall r, i, j$ and $\forall t \in \{T + 1, \dots, l\}$) are unconstrained at the initial step. To find the optimal solution of these binary variables, a large number of iterations may be required. This is because only a limited number of optimality or feasibility cuts are added to the master problem in each iteration of Benders decomposition.

3.3 Proposed Methodologies

We propose two techniques to reduce the model's solving time. The first technique is a reformulation method, which derives an equivalent model by reformulating the master problem and the sub-problems. The second one is a row generation strategy to generate the optimality cuts elaborately and selectively. We will explain the reformulation method and row generation strategy in Sections 3.3.1 and 3.3.2, respectively.

3.3.1 Reformulation Method

Even if two models are mathematically equivalent, their solving time through Benders decomposition can be considerably different (Koenig and Holte 2003, Liberti 2004). The quality of cuts can be improved by better formulation (Geoffrion and Graves 1974, Magnanti and Wong 1981, Saharidis and Ierapetritou 2010). In this section, we reformulate (P2) to reduce the solving time.

The stock-out cost is only considered through β_r in the objective function

of (P2-MP). In other words, Benders decomposition first searches for the solution mainly focusing on the minimization of the total delivery and holding costs. Since the stock-out cost is only considered implicitly in the master problem via the optimality cuts, the searching process can be time-consuming.

To reduce the solving time, we reformulate (P2-MP) by adding an estimation function of the stock-out cost in its objective function. Let $sd_{r,t,i}$ represent the total forecast demand of clinic i in period t of scenario r , including the new and existing patients for follow-up treatments. In addition, denote the estimated total stock-out amount of clinic i in scenario r by $sUSD_{r,i}$. It is actually the stock-out amount without future deliveries in the clinical trial supply chain.

$$sUSD_{r,i} = \begin{cases} 0, & \text{if } \sum_{t=T}^l sd_{r,t,i} \leq IC_{T,i}, \\ \sum_{t=T}^l sd_{r,t,i} - IC_{T,i}, & \text{Otherwise.} \end{cases} \quad (3.40)$$

$$sd_{r,t,i} = \sum_{\tau=t-q+2, \tau < T}^{t-1} (1 - \rho)^{(t+1-T)} SD_{r,\tau,(t+1-\tau),i} + \sum_{\tau=t-q+1, \tau \geq T}^t (1 - \rho)^{(t-\tau)} d_{r,\tau,i}. \quad \forall i \in I, \forall t \in \{T, \dots, l\}, \forall r \in R \quad (3.41)$$

Without loss of generality, clinic i 's total forecast demands $\sum_{t=T}^l sd_{r,t,i}$ are ordered ascendingly by scenarios. Let $ssd_{p,i}$ be the p th smallest total forecast demand, UD_i be the expected estimated stock-out amount at clinic i . We have

$$UD_i = \sum_r p_r sUSD_{r,i} = \begin{cases} \sum_r p_r ssd_{r,i} - IC_{T,i}, & \text{if } IC_{T,i} \leq ssd_{1,i}, \\ \sum_{r=p+1}^m p_r ssd_{r,i} - IC_{T,i} \sum_{r=p+1}^m p_r, & \text{if } ssd_{p,i} < IC_{T,i} \leq ssd_{p+1,i}, \\ 0, & \text{if } ssd_{m,i} \leq IC_{T,i}. \end{cases} \quad (3.42)$$

Based on (3.42), UD_i is a piecewise non-increasing linear function of $IC_{T,i}$. By adding this term, the master problem is reformulated to (P3-MP) below. The new formulation has two advantages: (1) the estimation function UD_i is a piecewise linear function and can be modeled without adding any binary variable

in the master problem, and (2) UD_i is scenario and time independent so the problem size will not increase with respect to the number of scenarios and time periods.

$$\begin{aligned}
 \textbf{(P3-MP): } & \{ \min \sum_i (w_i a_{T,i} + r w_i b_{T,i} + \sum_{j \in I, j \neq i} w_{i,j} c_{T,i,j} + h_i IC_{T,i}) + \\
 & h_{DC} ID_T + \sum_r p_r \sum_i \sum_{t=(T+1)}^l (w_i s a_{r,t,i} + r w_i s b_{r,t,i} + \sum_{j \in I, j \neq i} w_{i,j} s c_{r,t,i,j}) \\
 & + u \sum_i UD_i + \sum_r p_r \beta_r : s. t. (3.2) - (3.11), (3.31), (3.40) - (3.42) \}. \quad (3.43)
 \end{aligned}$$

To ensure the equivalence between (P2) and (P3), the sub-problems are reformulated as follows.

$$\begin{aligned}
 \textbf{(P3-SP): } & \{ \min u \sum_i (\sum_{k=1}^q \sum_{t=T}^l s LC_{r,t,k,i} - s USD_{r,i}) + \\
 & \sum_{t=(T+1)}^l (\sum_i h_i s IC_{r,t,i} + h_{DC} s ID_{r,t}) : s. t. (3.13) - (3.30) \}. \quad (3.44)
 \end{aligned}$$

Note that the optimality cuts in (P2) and (P3) are different. Let β_r^1 and β_r^2 represent β_r in (P2) and (P3) respectively. Let v be the iteration number, (θ^v, η^v) be the optimal solution in the v th iteration. As proved in Appendix 1, the objective function of the dual sub-problems and the optimality cuts in (P2) can be simplified as follows.

$$\max \pi_1^T (d_3 - A_2 \theta^v) + M \pi_2^T \eta^v + \pi_3^T d_5, \quad (3.45)$$

$$\beta_r^1 \geq (\pi_1^v)^T (d_3 - A_2 \theta) + M (\pi_2^v)^T \eta + (\pi_3^v)^T d_5, \quad (3.46)$$

where π_1 , π_2 and π_3 are dual variables corresponding to their respective constraints. Since $\beta_r^1 = u \sum_i \sum_{k=1}^q \sum_{t=T}^l s LC_{r,t,k,i} + \sum_{t=(T+1)}^l (\sum_i h_i s IC_{r,t,i} + h_{DC} s ID_{r,t})$ and $\beta_r^2 = u \sum_i (\sum_{k=1}^q \sum_{t=T}^l s LC_{r,t,k,i} - s USD_{r,i}^v) + \sum_{t=(T+1)}^l (\sum_i h_i s IC_{r,t,i} + h_{DC} s ID_{r,t})$,

we have $\beta_r^2 = \beta_r^1 - u \sum_i s USD_{r,i}^v$. In addition, $s USD_{r,i}^v$ is a value obtained in the master problem. As a result, the objective function of the dual sub-problems and the optimality cuts in (P3) should be as follows.

$$\max \pi_1^T (d_3 - A_2 \theta^v) + M \pi_2^T \eta^v + \pi_3^T d_5 - u \sum_i sUSD_{r,i}^v, \quad (3.47)$$

$$\beta_r^2 \geq (\pi_1^v)^T (d_3 - A_2 \theta) + M (\pi_2^v)^T \eta + (\pi_3^v)^T d_5 - u \sum_i sUSD_{r,i}. \quad (3.48)$$

Although the reformulation method is proposed for the inventory allocation in clinical trial supply chains, it can also be applied to optimization models in other applications. For example, a problem is formulated as a two-stage stochastic programming model and the decisions at the first stage solely depend on those at the second stage (i.e., each decision variable at the first stage has the solution 0 at the initial step in a minimization problem and these variables' solutions can only be improved by those generated cuts in Benders decomposition).

3.3.2 Row Generation Strategy

Since all scenario binary variables η in the master problem (P3-MP) are unconstrained at the initial step, a large number of iterations and optimality cuts are needed to find the optimal solution. If a better cut is derived earlier, the number of iterations may be reduced. In this section, we develop some strategies in generating optimality cuts to reduce the number of iterations and improve the solving efficiency.

The optimality cut in (3.48) has a special structure that the coefficients of all scenario binary variables η are big-Ms. Denote the i -th variable in η and its solution in η^v by η_i and η_i^v , the corresponding dual variable in π_2 and its solution in π_2^v by $(\pi_2)_i$ and $(\pi_2^v)_i$ respectively. We have the following theorem.

THEOREM 1.

- a) If $\eta_i^v = 0$, then $(\pi_2^v)_i \leq 0$.
- b) If $\eta_i^v = 1$, then $(\pi_2^v)_i = 0$.

Please see Appendix 2 for the proof. By THEOREM 1, we have the following conclusion.

THEOREM 2. *If $(\pi_2^v)_i < 0$, $\eta_i^w = 1, w \geq v + 1$ will make the optimality cut (3.48) in the v -th iteration valid.*

If the coefficient $(\pi_2^v)_i$ of a scenario binary variable η_i is negative in the optimality cut (3.48) generated in the v -th iteration (in other words, the variable η_i exists in the cut), $M(\pi_2^v)_i \eta_i$ and the left hand side of the optimality cut (3.48) are negative infinite when the variable η_i has the solution 1 in later iterations (i.e., $\eta_i^w = 1, w \geq v + 1$). As a result, the cut will be always valid.

Due to the big-M coefficients in the optimality cut (3.48), the convergence of Benders decomposition can be slow in solving the problem (P3), especially in the optimization of θ and β_r . According to THEOREM 2, an optimality cut can be always valid if any variable η_i in the cut is equal to 1. These cuts may not be strong in optimizing the variables θ and β_r . The coefficient of each η_i is the product of the scenario probability and delivery cost. It may be much less than the cost increase (including the total delivery and holding costs as well as the estimated stock-out cost) incurred by a slight change of θ 's solution. In addition, the improvement of β_r 's lower bound depends on the optimality cut (3.48). As a result, when the objective function of (P3-MP) is minimized subject to the cut (3.48), it is possible that some binary variable in the cut has the solution 1 to make the cut valid. The solution of θ and β_r may keep unchanged until either of the following two situations happens: (1) the solution change of the variables η will incur a higher cost, (2) it cannot validate all cuts generated in the new and all previous iterations. In the worst case, the solutions of θ and β_r will not be improved before evaluating all possible combinations of the variables in η , i.e.,

at most $2^{|\eta|}$ iterations are required.

The convergence can be improved if the big-M coefficients in the optimality cut (3.48) are removed temporarily. By setting all binary variables in η as 1, the optimality cut only has the variables θ and β_r . Since there does not exist any constraint for η in (P3-MP) or the generated cuts, the solution space of (P3-MP) only includes that of θ and β_r . The convergence of θ and β_r can be improved. Since the optimal solution of η changes corresponding to θ , the problem's solving efficiency can be improved correspondingly. In addition, when the scenario binary variables are set as 1, the objective value of (P3-SP) is minimized with θ 's solution obtained in (P3-MP). The minimal solution of β_r can be obtained in this process.

Although no big-M coefficient in (3.48) can improve the convergence of Benders decomposition, the scenario binary variables η in the optimal solution may not be always equal to 1. Two strategies are proposed to improve the solving efficiency of η .

The first strategy is called the RM (**R**educe big-**M**) strategy, which reduces the number of big-M coefficients to 1 in the optimality cut. To make the cut valid, the solution should be either of the following two results: (1) this big-M's corresponding variable η_i has the solution 1, (2) the solution of θ or β_r is changed. These cuts should be generated for each variable in η . If the current solution of θ and β_r is global optimal and a variable η_i has the value 1 in the global optimal solution, then the variable can be forced to 1 by such optimality cuts. Otherwise, at most $|\eta|$ cuts are required to make the solution of θ and β_r change. In the two cases, the convergence of η or θ and β_r can be improved respectively. This strategy can be realized by setting one variable in η to 0 and

all the others to 1. We apply this strategy to each variable in η to maximally improve the solving efficiency.

In the other strategy, (P1-SP) is applied to efficiently search for the optimal solution of η . As shown in Section 4.1, (P1) is equivalent to (P3). Since each iteration only evaluates one solution of η , finding the optimal solution of η in (P3-MP) may require large number of iterations. (P1-SP), on the other hand, can find the optimal solution of η at once for a given θ^v . Thus, we have the following result, which can be proved by the definition of Benders decomposition.

THEOREM 3. *With a given current decision θ^v , the upper bound of (P1) and (P3) can be obtained from the objective value of (P1-SP).*

Since the upper bound solution can be a good candidate for the optimal solution, we will search the solution space near the upper bound solution and call this method the NS (Near the Solution) strategy. For a given current decision θ^v , we can find the corresponding optimal solution of the second stage decisions η^v in (P1-SP) and the upper bound of (P1) as well as (P3). The optimal solution of η in (P1-SP) can be used to generate the optimality cuts in (P3). If the variables with big-M coefficients in an optimality cut only include those with the solution 1 in (P1-SP), then some of these variables will have the solution 1 in (P3-MP) if the solutions of θ and β_r are unchanged in subsequent iterations. As a result, the solution in these iterations will approach to the solution of (P1-SP) and the problem's upper bound. By THEOREM 1, if $\eta_i^v = 1$, the coefficient of η_i in the optimality cut (3.48) must be zero. Hence, this strategy can be conducted by setting the variables η_i with the solution 0 in (P1-SP) to the value 1 in (P3-MP).

The detailed procedure of the row generation strategy is shown in Figure 3-2. The strategy has two phases. While the decisions of θ and β_r are

considered in both phases, η is only solved in the second phase. The minimal solution of β_r is obtained in the first phase. In each iteration of phase 1, θ^v , η^v , β_r^v , and the problem's lower bound P_{lb} are obtained in (P3-MP). θ^v is then substituted into (P1-SP) to get the upper bound P_{ub} of (P3). If it is not the optimal solution, all η_i^v are set to 1 to find the minimal solution of β_r . α_1 and α_2 are small positive numbers less than one and represent the acceptable optimality gaps. We will start the second phase when the minimal solution of β_r is found. In the second phase, the RM and NS strategies are used. In the RM strategy, each generated cut has at most one scenario binary variable η_i . In the NS strategy, the solution of η in (P1-SP) is used in (P3-SP) to generate the optimality cuts. The generated optimality cuts are added to (P3-MP) and the solving process will continue until optimality.

Figure 3-2. The procedure of the row generation strategy

This row generation strategy is proposed based on the special structure of

the optimality cuts, where all coefficients of the scenario binary variables η are big-Ms. Hence, when other problems have optimality cuts with similar structures, e.g. the constraint (3.48), this strategy might be applicable as well.

3.4 Numerical Experiments

In this section, we report a numerical study and apply the proposed methodologies to solve the inventory allocation problem in clinical trial supply chains. In Section 3.4.1, we show the parameters and implementation details in the numerical study. In Section 3.4.2, we compare different acceleration techniques. Finally, the total inventory allocation costs are compared with those in an identical supply chain, where the non-stationary (s, S) policy is applied with transshipment jointly.

3.4.1 Numerical Study

We consider a numerical study of the inventory allocation problem between distribution centers and clinical sites in clinical trial supply chains. As distribution centers tend to be geographically dispersed, the inventory allocation is planned independently among a distribution center and all its downstream clinical sites. Clinical drugs can be shipped between the DC and each clinical site. Due to the different regulations between countries and regions, transshipments are only allowed among several groups of clinical sites within the same region.

The demand uncertainty is modeled using non-stationary Poisson distribution in this numerical study. Similar assumptions can be found in the literature (Chen, et al. 2012, Li, et al. 2014, Tang et al. 2012). The new patient arrival rate varies at different periods. Patients are usually advised (by friends or advertisements, etc.) to participate in clinical trials and the arrival rate increases

in the later periods of clinical trials (Chen, et al. 2012). The demand also depends on the length of the treatment period and the dropout rate. The treatment period usually includes three to six weeks and a patient is required to take the trial once a week and three to six weeks in total. Some patients may drop out during the treatment. The trial demand in each week consists of both new patients and the existing patients for the follow-up treatments.

The sample average approximation method (Santoso, et al. 2005, Verweij, et al. 2003) is applied to model the demand uncertainty in this problem. Please see the detailed description and procedure in Section 2.3.1.1. We present the optimality gap estimate and its 90% confidence interval for different m in Table 3-2. Each SAA problem is solved for forty replications to obtain the lower bound. The estimated optimality gap is equal to the difference of the lower and upper bounds. With the gap and its standard deviation, we can get the gap estimate's confidence interval. As shown in the table, with the increase of m , the 90% confidence interval for the optimality gap is getting narrower. Both the lower and upper limits of the gap are within 5% of the lower bound when m increases to forty. Thus, the SAA method with forty scenarios can provide solutions with reasonable quality to the true stochastic problem with certain demand distributions at clinical sites. Hence, we set the number of scenarios at forty and run forty replications in the numerical study.

Table 3-2. Statistics in the SAA method

m	Lower bound		Upper bound		Estimated optimality gap			90% Confidence interval			
	Average	σ_{LB}	Average	σ_{UB}	Average	%	σ_{gap}	Min	%	Max	%
20	7308.04	225.62	7545.08	230.78	237.04	3.24	322.74	-293.88	-4.02	767.95	10.51
30	7370.84	159.19	7538.93	172.19	168.09	2.28	234.50	-217.67	-2.95	553.84	7.51
40	7458.03	113.60	7574.07	91.63	116.04	1.56	145.95	-124.05	-1.66	356.12	4.78

The method is implemented in C#. CPLEX 12.5 is used as a callable library to solve the mixed-integer master problem and the linear sub-problems in

Benders decomposition. The method is executed on a computer with an Intel® Core™ i7-2600 CPU 3.4GHz processor and a Windows 7 system.

3.4.2 Performance Analysis

The performances of the acceleration schemes are demonstrated in Table 3-3 and Table 3-4. The CPU seconds and the number of iterations are compared under different problem settings. The acceleration schemes are denoted as follows: (1) the reformulation method (R), and (2) the row generation strategy's phase 1 (G1) and phase 2 (G2), including the strategy RM (G2_RM) and NS (G2_NS). The row "None" represents the cases applying Benders decomposition without any acceleration technique, and the row "All" represents the cases applying Benders decomposition with all acceleration schemes.

Table 3-3. Comparison of all accelerating techniques

Method	(n,m)=(5,5)		(n,m)=(5,10)		(n,m)=(10,5)		(n,m)=(10,10)		Method	(n,m)=(5,5)		(n,m)=(5,10)		(n,m)=(10,5)		(n,m)=(10,10)	
	Iter	CPU	Iter	CPU	Iter	CPU	Iter	CPU		Iter	CPU	Iter	CPU	Iter	CPU	Iter	CPU
None	31	97	-	-	-	-	-	-	R	19	11	2	2	4	13	11	2880
G1	33	99	14	75	63	988	54	8188	R+G1	19	11	2	2	7	17	12	1864
G2_RM	8	40	9	93	12	79	23	3170	R+G2_RM	3	12	3	24	4	22	5	59
G2_NS	22	117	13	120	23	494	>32	>10800	R+G2_NS	3	12	3	23	5	24	10	30
G1+G2_RM	16	53	12	79	44	159	42	4187	R+G1+G2_RM	2	3	2	7	5	21	6	54
G1+G2_NS	19	63	15	91	52	189	50	3529	R+G1+G2_NS	2	3	2	6	8	27	12	2640
G2	19	112	9	93	12	73	18	1798	R+G2	3	12	3	24	4	22	4	42
G1+G2	26	123	12	79	44	163	43	3422	All	2	3	2	7	5	20	6	56

Notes. (1) '-' means the CPU seconds required to solve the instance is much greater than all the other solving methods in the table. (2) The values in bold are the minimal results among all solving methods.

The results in Table 3-3 show R and G1 can improve the solving efficiency effectively. The methods containing R have the shortest CPU seconds and the minimal number of iterations in each instance. When compared with the best results without R, the computational time has been reduced by at least 82%, and the number of iterations has been reduced by at least 66%. Hence, the method R is effective in improving the problem's solving efficiency. As shown in Table 3-3, after G1 is implemented, the computational time can be reduced in the instances

with larger problem sizes, but may increase in smaller cases.

When both R and G1 are adopted, the results in Table 3-3 cannot illustrate the performances of G2_RM and G2_NS clearly. Since G2_RM and G2_NS are both developed to deal with problems with a large number of scenario binary variables, instances with larger problem sizes are solved to compare the performances of G2_RM and G2_NS.

The results in Table 3-4 show that both G2_RM and G2_NS can improve the solving efficiency. Note that Table 3-4 has the same notes under Table 3-3. The results in “All” always have the best performance in each instance. Furthermore, G2_RM and G2_NS can reduce the computational time and the number of iterations individually in each instance. Hence, the two strategies can provide directions to improve the solving efficiency of the two-stage mixed-integer programming problems.

Table 3-4. Comparison of two acceleration strategies

Method	(n,m)=(15,10)		(n,m)=(15,15)		(n,m)=(20,10)		(n,m)=(20,15)	
	Iter	CPU	Iter	CPU	Iter	CPU	Iter	CPU
R+G1	43	9407	-	-	305	444160	-	-
R+G1+G2_RM	4	33	67	143337	5	76	307	592655
R+G1+G2_NS	17	3344	62	132140	84	68042	>208	>600000
All	4	33	60	126889	5	56	198	331060

3.4.3 Cost Comparison with the (s, S) Policy

We compare the total costs obtained in our proposed approach with those obtained in the clinical trial supply chain when the non-stationary (s, S) policy (Porteus 1990) is applied at each clinical site. For a site with non-stationary stochastic demand, if the ordering cost is constant, the time-varying (s, S) policy is its optimal inventory ordering policy (Clark and Scarf 1960, Scarf 1959). Since the replenishment cost in our problem is fixed, we compare the costs of the supply chain with our approach and a time-varying (s, S) policy at each clinical site. The

reverse replenishment and transshipment are also applied when the inventory at the DC is depleted under the (s, S) policy. Without loss of generality, all clinical sites are divided into two identical groups, where the patient arrival rates in the two groups are set at λ_1 and λ_2 , respectively. Two non-stationary patient arrival rates are constructed in Table 3-6 based on the demand pattern described by Chen, et al. (2012). In addition, with the parameter settings for the comparison shown in Table 3-5, the optimal (s, S) policy for each patient arrival rate is presented in Table 3-6.

Table 3-5. Parameter settings

Parameter	Value
The number of periods	10
The number of clinical sites	10
The replenishment cost of each clinical site	4
The reverse replenishment cost of each clinical site	4
The transshipment cost between each two clinical sites	6
The stock-out cost	10
The holding cost at DC in each period	0
The holding cost at each clinical site in each period	0.01
The delivery time between DC and each clinical site	1
The delivery time between each two clinical sites	1

Table 3-6. Non-stationary patient arrival rates and their corresponding optimal (s, S) policy

Period	1	2	3	4	5	6	7	8	9	10
λ_1	1	3	4	5	5	4	5	5	4	1
s_1	3	7	8	10	10	9	11	9	6	1
S_1	48	47	44	40	35	30	25	19	12	5
λ_2	1	4	6	7	7	6	7	7	4	2
s_2	3	8	11	13	13	12	13	12	7	3
S_2	62	62	58	52	45	38	31	23	14	8

The total cost of our approach or the (s, S) policy depends on both the demand realization and the initial inventory level at the DC. We consider thirty

cases with different demand realizations. The total realized demand in these cases ranges from 90% to 110% of the total patient arrival rate, which is 440 according to the patient arrival rates in the table above. The initial inventory is set at 10%, 20%, or 30% more than the total patient arrival rate. When the initial inventory is less than the inventory required under the (s, S) policy, the max-min fair resource allocation method (Bertsimas et al. 2011), reverse replenishment and transshipment are jointly used in supply chains applying the (s, S) policy.

Table 3-7 compares the average cost between the proposed approach and the time-varying (s, S) policy with transshipment when the initial inventory takes different values. Our approach gives a lower average total cost than the (s, S) policy in terms of the delivery, holding, stock-out, and sum in all but one case. When the initial inventory is 572, our approach has a greater average total delivery cost. This is because the inventory levels at clinical sites in our approach are less than those in the (s, S) policy (this can be proved by the smaller holding cost in our approach) and more delivery cost is required to avoid the stock-out.

Table 3-7. The average total cost comparison between our approach and the (s, S) policy with transshipment

Average total cost	Our approach			(s, S) policy with transshipment		
	$ID_0 = 484$	$ID_0 = 528$	$ID_0 = 572$	$ID_0 = 484$	$ID_0 = 528$	$ID_0 = 572$
Delivery	76.93	56.80	57.13	101.60	59.87	46.13
Holding	27.71	30.32	33.67	28.08	32.88	35.80
Stock-out	61.67	10.00	2.33	132.67	56.67	13.67
Sum	166.31	97.12	93.13	262.35	149.41	95.60

Table 3-8 shows the detailed cost comparison between the proposed approach and the time-varying (s, S) policy with transshipment in the thirty cases. The stock-out cost of our approach is no greater than that of the (s, S) policy in

most cases. When the stock-out cost of the (s, S) policy is greater than zero, applying our approach can reduce it by up to 100%. Since the inventory may be held at the DC and allocated later due to the demand uncertainty at clinical sites, the holding cost of our approach is less than that of the (s, S) policy in all cases. Due to the reverse replenishment and transshipment applied in both our approach and the (s, S) policy, the delivery cost may be less or greater in our approach.

Applying our approach can reduce the total cost of the (s, S) policy by up to 65%. Based on the discussion above, the holding and/or stock-out costs are generally reduced in our approach. The delivery cost may be less in the supply chain applying the (s, S) policy. When the initial inventory is low, the stock-out cost reduction can hardly be offset. Up to 65% of the total cost can be reduced by applying our approach. When the stock-out cost decreases to zero in both our approach and the (s, S) policy, our approach may have a greater total cost due to the higher delivery cost. Hence, when we have only a limited number of inventory at the DC, our approach gives a less total cost than the (s, S) policy with transshipment.

Table 3-8. Cost comparison between our approach and the (s, S) policy with transshipment

Case number	Total realized demand	Total cost							
		Our approach, $ID_0 = 484$				(s, S) policy with transshipment, $ID_0 = 484$			
		Delivery	Holding	Stock-out	Sum	Delivery	Holding	Stock-out	Sum
1	396	85	29.92	0	114.92	72	30.26	20	122.26
2	400	53	29.48	0	82.48	44	30.28	0	74.28
3	416	71	28.75	0	99.75	84	29.13	50	163.13
4	418	65	28.66	0	93.66	68	29.23	40	137.23
5	419	83	28.91	10	121.91	84	29.28	110	223.28
6	420	77	28.71	0	105.71	76	29.22	60	165.22
7	426	71	28.85	10	109.85	68	29.52	0	97.52

Inventory Allocation in Clinical Trial Supply Chains

8	430	71	28.31	0	99.31	96	28.36	160	284.36
9	432	77	28.4	10	115.4	64	29.2	30	123.2
10	434	71	27.56	130	228.56	84	27.71	140	251.71
11	436	71	28.06	110	209.06	96	28.19	120	244.19
12	437	83	27.33	0	110.33	104	27.79	50	181.79
13	440	77	28.44	10	115.44	92	28.73	70	190.73
14	440	71	27.88	10	108.88	76	28.19	30	134.19
15	443	83	27.59	30	140.59	100	28.03	140	268.03
16	444	71	26.88	30	127.88	120	27.01	60	207.01
17	444	83	27.43	130	240.43	120	27.64	240	387.64
18	446	83	27.25	0	110.25	96	27.68	140	263.68
19	447	83	28.23	40	151.23	112	28.26	170	310.26
20	448	83	28.19	30	141.19	84	28.68	180	292.68
21	452	77	26.49	110	213.49	120	26.58	170	316.58
22	452	77	27.65	50	154.65	120	27.69	170	317.69
23	456	89	26.92	30	145.92	132	27.28	110	269.28
24	457	77	27.46	140	244.46	108	27.79	300	435.79
25	459	83	26.22	40	149.22	108	26.85	70	204.85
26	470	83	27.06	170	280.06	152	27.11	250	429.11
27	474	83	26.02	160	269.02	152	26.12	280	458.12
28	476	77	26.33	210	313.33	148	26.51	240	414.51
29	484	85	25.16	180	290.16	136	26.71	300	462.71
30	486	65	27.07	210	302.07	132	27.38	280	439.38

Case number	Total realized demand	Our approach, $ID_0 = 528$				(s, S) policy with transshipment, $ID_0 = 528$			
		Delivery	Holding	Stock-out	Sum	Delivery	Holding	Stock-out	Sum
1	396	48	32.86	0	80.86	52	35.13	0	87.13
2	400	48	31.18	0	79.18	40	35.02	0	75.02
3	416	48	30.76	0	78.76	52	34	10	96
4	418	48	30.6	20	98.6	52	33.91	0	85.91
5	419	48	31.66	0	79.66	64	33.85	20	117.85
6	420	48	30.79	0	78.79	48	34.04	10	92.04
7	426	56	30.92	10	96.92	40	34.58	10	84.58
8	430	48	31.02	0	79.02	56	33.11	30	119.11
9	432	52	30.97	0	82.97	48	33.73	0	81.73
10	434	52	30.05	10	92.05	52	32.7	80	164.7
11	436	60	30.73	10	100.73	64	32.94	20	116.94
12	437	56	30.08	0	86.08	44	32.9	10	86.9
13	440	52	30.89	0	82.89	44	33.93	20	97.93
14	440	52	30.12	0	82.12	40	33.5	0	73.5
15	443	56	30.35	0	86.35	56	32.72	0	88.72

Inventory Allocation in Clinical Trial Supply Chains

16	444	52	30.83	0	82.83	60	32.76	70	162.76
17	444	64	30.8	10	104.8	68	32.68	130	230.68
18	446	58	29.86	0	87.86	52	32.5	40	124.5
19	447	52	31.06	0	83.06	64	33.27	70	167.27
20	448	52	30.93	0	82.93	56	33.57	110	199.57
21	452	62	29.43	10	101.43	84	30.73	20	134.73
22	452	68	30.34	0	98.34	72	32.4	70	174.4
23	456	70	29.92	0	99.92	72	31.75	60	163.75
24	457	64	30.02	10	104.02	56	32.56	90	178.56
25	459	60	28.74	0	88.74	60	31.32	30	121.32
26	470	68	29.5	50	147.5	64	31.9	120	215.9
27	474	70	29.69	10	109.69	88	31.25	160	279.25
28	476	68	28.61	30	126.61	84	30.79	130	244.79
29	484	64	27.56	90	181.56	96	30.79	270	396.79
30	486	60	29.32	40	129.32	68	31.98	120	219.98

Case number	Total realized— demand	Our approach, $ID_0 = 572$				(s, S) policy with transshipment, $ID_0 = 572$			
		Delivery	Holding	Stock-out	Sum	Delivery	Holding	Stock- out	Sum
1	396	52	36.6	0	88.6	40	37.81	0	77.81
2	400	48	34.88	0	82.88	40	37.22	0	77.22
3	416	56	34.91	0	90.91	44	36.66	0	80.66
4	418	56	34.53	0	90.53	40	36.47	0	76.47
5	419	56	34.03	0	90.03	44	36.83	20	100.83
6	420	52	34.09	0	86.09	44	36.71	0	80.71
7	426	52	33.6	0	85.6	40	36.78	0	76.78
8	430	48	34.37	0	82.37	44	36.42	0	80.42
9	432	52	34.29	0	86.29	40	36.31	0	76.31
10	434	52	33.02	0	85.02	48	35.38	30	113.38
11	436	52	34.2	0	86.2	56	36.21	30	122.21
12	437	56	32.53	0	88.53	40	35.29	0	75.29
13	440	52	34.78	0	86.78	40	36.34	0	76.34
14	440	56	33.61	0	89.61	40	35.7	0	75.7
15	443	60	33.6	0	93.6	44	35.59	0	79.59
16	444	56	33.8	0	89.8	44	35.9	10	89.9
17	444	68	33.22	0	101.22	44	36.34	50	130.34
18	446	52	32.83	0	84.83	44	35.36	0	79.36
19	447	60	34.19	0	94.19	44	36.52	0	80.52
20	448	56	33.22	0	89.22	48	36.13	60	144.13
21	452	68	32.86	0	100.86	56	34.51	0	90.51
22	452	62	34.37	0	96.37	48	36.05	0	84.05
23	456	66	32.95	0	98.95	56	34.83	40	130.83

Inventory Allocation in Clinical Trial Supply Chains

24	457	56	33.31	0	89.31	48	35.46	10	93.46
25	459	60	32.12	0	92.12	44	34.29	0	78.29
26	470	52	33.32	40	125.32	48	35.06	10	93.06
27	474	74	33	0	107	64	34.67	40	138.67
28	476	68	32.88	0	100.88	48	34.22	0	82.22
29	484	64	31.91	30	125.91	64	34.15	110	208.15
30	486	52	33	0	85	40	34.74	0	74.74

3.5 Conclusion

This study proposes an inventory allocation model at the operational level of a clinical trial supply chain using the rolling-based two-stage stochastic programming through Benders decomposition. The model gives the optimal inventory allocation at the distribution center and clinics at each time period through balancing the delivery and holding costs against the potential stock-out cost.

In literature, inventory policies, such as (Q, r) , (s, S) and $(S-I, S)$ policies, are often used at the retailers in a multi-echelon supply chain. This approach is justifiable if the DC has infinite inventory and the transshipment cost is high, since each retailer acts independently in this situation. However, with a finite replenishment cycle at the DC, the ordering quantity of a retailer will have an influence on that of the other retailers, especially at the end of the replenishment cycle with limited on-hand inventory. Since the optimality of above inventory policies is not guaranteed in multi-echelon supply chains, a two-stage stochastic model is formulated to optimize the inventory allocation in our study.

Because this problem is NP-hard, the time required by our method based on Benders decomposition can be longer than acceptable when the number of clinics is large, especially during the third phase of a clinical trial. Hence, the reformulation method and row generation strategy are developed to improve the

solving efficiency of the problem in this study. As shown in the numerical results, the solving efficiency has been improved effectively in terms of the CPU seconds and the number of iterations required.

The proposed approach is also compared with the (s, S) policy in terms of the delivery, holding, stock-out and total costs. The stock-out and holding costs of our approach are generally no greater than those of the (s, S) policy respectively. Compared to the (s, S) policy, up to 65% of the total cost and up to 100% of the stock-out cost are reduced by applying our approach. Since reverse replenishment and transshipment are applied in both our approach and the (s, S) policy, the delivery cost may be greater or less in our approach.

Appendix

1. Derivation of the dual problems of (P2-SP) and the optimality cuts in (P2).

In the model (P2), (P2-MP) and (P2-SP) can be simply represented in the following formulations.

$$\{\min f^T \theta + g^T \eta + \sum_r p_r \beta_r : s.t. A_1 \theta \leq d_1; B_1 \eta \leq d_2, \eta \in \{0,1\}^{|\eta|}\},$$

$$\{\min h^T \mu : s.t. C_1 \mu \leq d_3 - A_2 \theta; C_2 \mu \leq M \eta; C_3 \mu \leq d_5; \mu \geq 0\},$$

where $|\eta|$ is the number of elements in η , each element in C_2 is either 0 or 1, each element in C_1 and C_3 is finite.

According to the duality theory, the dual problems of (P2-SP) and the optimality cuts in (P2) can be formulated as follows, where π_1 , π_2 and π_3 correspond to the first three constraints respectively.

$$\{\max \pi_1^T (d_3 - A_2 \theta^v) + M \pi_2^T \eta^v + \pi_3^T d_5 : s.t. C_1^T \pi_1 + C_2^T \pi_2 + C_3^T \pi_3 \leq$$

$$h; \pi_1, \pi_2, \pi_3 \leq 0\},$$

$$\beta_r \geq (\pi_1^v)^T (d_3 - A_2 \theta) + M (\pi_2^v)^T \eta + (\pi_3^v)^T d_5. \quad \text{Q.E.D.}$$

2. Proof of THEOREM 1.

The sub-problem (P2-SP) is a linear programming model. Its primal and dual problems are shown in Appendix I.

Due to the complementary slackness in linear programming, we have

$$\pi_2 (C_2 \mu - M \eta) = 0, \text{ i.e., } (\pi_2)_i ((C_2)_i \mu - M \eta_i) = 0 \text{ for } \forall i \in \{1, \dots, |\eta|\}.$$

This result has the following two possibilities:

$$(1) (\pi_2^v)_i = 0 \text{ and } (C_2)_i \mu^v - M \eta_i^v < 0,$$

$$(2) (\pi_2^v)_i \leq 0 \text{ and } (C_2)_i \mu^v - M \eta_i^v = 0.$$

Since $\mu \geq 0$ and each element in C_2 is nonnegative, $\eta_i^v = 1$ if $(\pi_2^v)_i = 0$ and $(C_2)_i \mu^v - M \eta_i^v < 0$.

In addition, due to the finite initial inventory in the supply chain, each element in μ is finite, including the scenario delivery quantities and inventory levels at each location. As a result, $(C_2)_i \mu^v$ cannot be equal to M . Hence, $\eta_i^v = 0$ if $(\pi_2^v)_i \leq 0$ and $(C_2)_i \mu^v - M \eta_i^v = 0$.

In conclusion, we have the following results.

(1) If $\eta_i^v = 1$, then $(\pi_2^v)_i = 0$,

(2) If $\eta_i^v = 0$, then $(\pi_2^v)_i \leq 0$.

Q.E.D.

Chapter 4 Network Design in Clinical Trial Supply

Chains

4.1 Introduction

Long duration and high cost are two key characteristics of clinical trials. Since the clinical trial duration is part of the limited patent life and determines the time-to-market, its reduction is of critical importance. While the duration is reduced through increasing the number of clinical sites in the supply chain, the total cost is increased simultaneously, including the total enrollment cost of clinical sites (i.e., the cost of recruiting these clinical sites to participate in clinical trials) and the total operational cost in the distribution of clinical drugs.

Since the cost increase is incurred simultaneously with the duration reduction by the scale increase of clinical trial supply chains, the time and cost required by a clinical trial become a tradeoff. A clinical trial supply chain configuration is called Pareto-optimal (Pareto-efficient) if no other configuration can have both shorter duration and less cost. In other words, for any given cost (or duration), the configuration with the shortest duration (or the minimal cost) must be Pareto-optimal. To have an efficient clinical trial supply chain, we need to find a Pareto-optimal configuration. However, identifying such configurations can be hard, especially in large-scale supply chains. The total cost and duration of a clinical trial depend on the cost structure and patient recruitment process, which may be different among clinical sites.

There may exist more than one Pareto-optimal configuration in this problem. If we can identify all Pareto-optimal configurations, we can construct an efficiency curve by connecting all Pareto-optimal tradeoffs, where the horizontal

axis represents the clinical trial's total cost and the vertical axis represents the duration. Each Pareto-optimal tradeoff on the curve corresponds to a specific order of importance of the cost and time. To facilitate site selections, the efficiency curve can give decision makers a clear understanding of the tradeoff between the time and cost of clinical trials.

In this chapter, we study the clinical site selection problem considering the time-cost tradeoff. The contributions are three-fold. First, we propose a multi-objective site selection model considering the tradeoff between the cost and duration of clinical trials, where the duration depends on the patient recruitment process. The weighted sum method is applied to solve this model by introducing a parameter named delay cost. Delay cost can be defined as the opportunity cost of a period's delay (Getz and De Bruin 2000, Roberts, et al. 2003), in terms of the lost market share and alternative investment possibilities (Colvin and Maravelias 2010). It can be approximated by the average revenue of a previous successful drug in each period (Getz and De Bruin 2000). Second, since exploring all Pareto-optimal configurations can be time-consuming, we identify several properties to improve the solving efficiency. Thirdly, our model can generate the efficiency curve of Pareto-optimal tradeoffs for decision makers to design their supply chains.

The rest of this chapter is organized as follows. The clinical site selection problem is formulated as a mathematical model in Section 4.2. Propositions and optimality cuts are developed in Section 4.3. Based on the propositions proposed in Section 4.3.1, the algorithm used to solve this problem is proposed in Section 4.4. The numerical results are discussed in Section 4.5. Concluding remarks and future works are presented in Section 4.6.

4.2 Model Development

The multi-objective site selection problem is formulated as a mathematical model in this section. We give the assumptions and notations in Section 4.2.1 and the model in Section 4.2.2.

4.2.1 Assumptions.

Since the patient accrual goal of clinical trials is predetermined, the patient recruitment process has a finite planning horizon (Getz and De Bruin 2000). The treatment period usually ranges from three to six weeks. That is, each patient is required to take one trial per week and three to six continuous trials in total. If patients cannot take the trials on time, they must drop out the whole treatment process. Since clinical drugs usually have a shelf life of several years, we assume they are effective during the entire planning horizon. Due to the long production time of clinical drugs (about 6 to 9 months) (Láinez, et al. 2012), drugs are usually produced in large batches. Hence, we assume production is not considered in this problem and the initial inventory is greater than the number of drugs required to reach the patient accrual goal with the dropout rate considered. The dropout rate is assumed to follow a uniform distribution and have identical values in each period (Chen, et al. 2012). Indeed, it may decrease in the treatment period since patients who do not drop out early on are more likely to stay later. In addition, some other assumptions of the problem are listed as follows.

- (1) A set of candidate clinical sites is given.
- (2) There exists only one distribution center (DC) and the location of the DC is known.
- (3) The shipping cost from the DC to each clinical site is independent of the shipping quantity.

- (4) A patient always chooses the closest site, and the patient arrivals at each clinical site are independent of those at other sites and follow non-stationary Poisson distributions.
- (5) The forecast of the patient arrivals is given.
- (6) The delay cost in every period is identical but unknown.

The candidate clinical sites are assumed given since participating clinical sites are usually selected from existing hospitals and doctors' offices (Fleischhacker, et al. 2015). We assume all of them share a common DC since the clinical trial supply chain usually has a predefined central warehouse (Fleischhacker, et al. 2015) and regional warehouses can be selected after the site selection. The location of the DC is known. In addition, when the clinical drugs are distributed from the DC to each clinical site, the shipping cost is assumed to be independent of the shipping quantity since the volumes of drug packs are small and each shipment can contain many packs with the same shipping cost.

The patient recruitment uncertainty is usually modeled as non-stationary Poisson distributions (Chen, et al. 2012, Li, et al. 2014, Tang, et al. 2012), where the means can be approximated by the track record of patient arrivals and/or the previous trial data of similar diseases at each clinical site (Fleischhacker, et al. 2015). The forecast of the patient arrivals can be estimated accordingly.

When clinical sites are geographically dispersed, their patient recruitment processes tend to be independent. On the other hand, the patient arrival rates can be dependent at some sites due to the close distances among them. They may have different rates at each site under different supply chain structures. For example, while there are 10 patients arriving at the selected site per week if either site A or B is selected in clinical trials, only 8 patients arrive at each site per week

if both sites are selected, since the two sites have overlapped patient recruitment regions. However, if we assume patients always choose the closer site, those patients from the overlapped regions can be classified into two mutually exclusive groups. Hence, even if the patient recruitment processes can be dependent among some sites, as long as the supply chain structure is determined, the patient arrival rate at each site can be assumed fixed and independent of those at all the other sites.

The strategic supply chain design problem considers the tradeoff between the time and cost required to reach the patient accrual goal in clinical trials. The cost includes the enrollment costs of all selected clinical sites as well as the shipping and holding costs. And the time (i.e., the clinical trial duration) can be captured by the total delay cost. Since the delay cost can be approximated by the average revenue of a successful drug in each period (Clemento 1999, Getz and De Bruin 2000), it is reasonable to assume identical delay cost in every period. Hence, this problem's objective becomes the minimization of the total cost, including the enrollment, shipping, holding and delay costs. The model is formulated in Section 4.2.2 with the notations in Table 4-1.

Table 4-1. Summary of the notations

Notation	Description
Sets:	
S	Set of patient recruitment scenarios
T	Set of time periods
I	Set of candidate clinical sites
Parameters:	
nc	The number of candidate clinical sites
f_i	The enrollment cost of clinical site i
f	Vector of clinical sites' enrollment costs
p_s	The probability of scenario s with $\sum_s p_s = 1$
c_i	The shipping cost from the DC to clinical site i
h_i	The holding cost at clinical site i

λ	The delay cost per period
l_i	The lead time between the DC and clinical site i
$d_{s,t,i}$	The number of forecast new arrival patients at clinical site i in period t in scenario s
ca_i	The inventory capacity at clinical site i
m	The number of periods in a treatment period
dr	The average dropout rate in each period
G	The patient accrual goal
M	A sufficiently large number (E.g. one may define $M = \sum_{v=0}^{m-1} (1 - dr)^v \max_t \sum_i d_{s,t,i}$)

Decision variables:

y_i	1 if clinical site i is selected, 0 otherwise
y	Vector of all clinical sites' selection variables
$r_{s,t}$	1 if the patient accrual goal has not been reached at period t in scenario s , 0 otherwise
$x_{s,t,i}$	1 if clinical site i orders from the DC at the beginning of period t in scenario s , 0 otherwise
$X_{s,t,i}$	The order quantity of clinical site i from the DC at the beginning of period t in scenario s
$IC_{s,t,i}$	The inventory level of clinical site i at the beginning of period t in scenario s
$SD_{s,t,v,i}$	The number of patients who newly arrived at clinical site i in period t and were fulfilled in the clinical trial for exact v periods in scenario s , $1 \leq v \leq m$

4.2.2 Mathematical Model

Stochastic programming approaches are commonly used in supply chain network design problems under uncertainty (Nickel, et al. 2012). When a problem consists of decisions in a number of periods of the planning horizon and the uncertain parameters are revealed gradually over time, the problem can be formulated as a multi-stage stochastic programming model (Shapiro, et al. 2009). It is also formulated as a two-stage stochastic programming model in the literature (Alonso-Ayuso, et al. 2003, Badri et al. 2017, Gebreslassie, et al. 2012, MirHassani, et al. 2000). Similar to our problem, those problems consist of determining the strategic level decisions at the beginning of the planning horizon and the operational level decisions in the following periods. As discussed in

Chapter 2, due to the significant complexity, the multi-stage models are rare applied relative to the two-stage ones.

In this chapter, the site selection problem in clinical trial supply chains is formulated as a two-stage stochastic mixed-integer programming model, where the stochastic patient recruitment process is represented as a number of scenarios at the second stage. The long-term decisions, i.e., selection of candidate clinical sites, should be determined in the first stage. The time of the clinical trial and the ordering decisions, which depend on future patient arrivals, should be modeled in each scenario of the second stage. This model can be formulated as follows.

$$(P1): \min \{ f^T y + E[Q(y, s)], s.t. y \in \{0,1\}^{nc} \}. \quad (4.1)$$

In model (P1), $f^T y$ represents the total enrollment cost of clinical sites associated with the first-stage decision variables $y \in \{0,1\}^{nc}$, and $E[Q(y, s)]$ denotes the expectation of $Q(y, s)$ under all scenarios. For given configuration y and scenario s , $Q(y, s)$ represents the minimal operational and delay costs. It can be obtained with the following formulation, where $QC(y, s)$ and $QT(y, s)$ denote the minimal operational cost and time of the clinical trial respectively.

$$(P2): Q(y, s) = \min QC(y, s) + \lambda QT(y, s) \quad (4.2)$$

$$s.t. \quad QC(y, s) = \sum_{t,i} (c_i x_{s,t,i} + h_i IC_{s,t,i}), \quad (4.3)$$

$$QT(y, s) = \sum_t r_{s,t}, \quad (4.4)$$

$$\sum_{t=1}^{\max(\tau-m+1,0)} \sum_i SD_{s,t,m,i} \geq G(1 - r_{s,\tau}), \quad \forall \tau \in T \quad (4.5)$$

$$\sum_{v=1}^m SD_{s,t,v,i} \leq d_{s,t,i} y_i, \quad \forall i \in I, \forall t \in T \quad (4.6)$$

$$SD_{s,t,v,i} \geq dr \sum_{u=v}^m SD_{s,t,u,i}, \quad \forall i \in I, \forall v \in \{1, \dots, m-1\}, \forall t \in T \quad (4.7)$$

$$IC_{s,1,i} = X_{s,1,i}, \quad \forall i \in I \quad (4.8)$$

$$IC_{s,t,i} = IC_{s,(t-1),i} + X_{s,(t-1-l_i),i} - \sum_{\tau=t-m}^{t-1} \sum_{v=t-\tau}^m SD_{s,\tau,v,i},$$

$$\forall i \in I, \forall t \in T \setminus \{1\} \quad (4.9)$$

$$IC_{s,t,i} \leq ca_i, \quad \forall i \in I, \forall t \in T \quad (4.10)$$

$$X_{s,t,i} \leq Mx_{s,t,i}, \quad \forall i \in I, \forall t \in T \quad (4.11)$$

$$x_{s,t,i} \leq y_i, \quad \forall i \in I, \forall t \in T \quad (4.12)$$

$$X_{s,t,i} \geq 0, IC_{s,t,i} \geq 0, SD_{s,t,v,i} \geq 0,$$

$$\forall i \in I, \forall v \in \{1, \dots, m\}, \forall t \in T \quad (4.13)$$

$$r_{s,t}, x_{s,t,i} \in \{0, 1\}. \quad \forall i \in I, \forall t \in T \quad (4.14)$$

The objective function (4.2) minimizes the sum of the shipping, holding and delay costs in each scenario. Constraint (4.5) checks if the clinical trial can be finished at period τ , i.e., if the total number of completed trials has reached the patient accrual goal within the first τ periods. Constraint (4.6) guarantees that the total number of patients fulfilled during the treatment depends on the clinical site's forecast patient arrivals and site selection decision. The dropout rate is considered in constraint (4.7). Constraints (4.8) and (4.9) are the inventory balance constraints at clinical sites. Constraints (4.10) and (4.11) define the inventory capacity and fixed shipping cost for each clinical site. Constraint (4.12) shows that each clinical site's shipping decision depends on its site selection decision. Constraints (4.13) and (4.14) enforce the non-negativity or binary restrictions on the decision variables in this model.

The two aspects in the tradeoff considered in this multi-objective problem can be represented as follows, where $f(y)$ and $g(y)$ denote the cost and time of the clinical trial respectively.

$$f(y) = f^T y + E[QC(y, s)] = f^T y + \sum_{s,t,i} p_s (c_i x_{s,t,i} + h_i IC_{s,t,i}) \quad , \quad (4.15)$$

$$g(y) = E[QT(y, s)] = \sum_{s,t} p_s r_{s,t}. \quad (4.16)$$

Thus, the model (P1) is equivalent to the following (P3).

$$(P3): \quad \min_y \{z(\lambda) = f(y) + \lambda g(y), s. t. (4.5) \sim (4.16)\}. \quad (4.17)$$

This clinical trial supply chain network design problem is NP-hard even if we only consider a single period, deterministic patient arrivals, zero shipping, holding and delay cost in the above model. Under these conditions, this problem can be reduced to a min-knapsack problem as follows.

$$\min \left\{ \sum_i f_i y_i : s. t. \sum_i d_i y_i \geq G, y_i \in \{0,1\} \forall i \in I \right\}.$$

Since the min-knapsack problem is NP-hard (Csirik, et al. 1990), unless $P = NP$, it is not possible to develop an algorithm that can run in polynomial time to solve this multi-period stochastic problem.

4.3 Propositions and Optimality Cuts

In this section, we propose some propositions to identify all Pareto-optimal configurations efficiently through solving the problem (P3) for only a few values of λ . In addition, some optimality cuts are derived to improve the solving efficiency of the problem (P3).

4.3.1 Pareto-optimal Propositions

In general, due to the uncertain clinical trial results (success, failure, or other effects) and potential development, it can be difficult to estimate a drug's delay cost with a single value or even a reasonably small range. A clinical drug which performs well enough in animal trials may fail in clinical trials. While some

clinical drugs fail in a clinical trial for its target disease, unexpected effects may be discovered and the drugs can be used for other diseases. Some drugs may also be found effective for a number of distinct diseases, e.g. Aspirin and Sildenafil. Aspirin has been used to treat pain, fever, and inflammation for at least 2,400 years in the form of willow tree leaves and sold around the world by Bayer as treatments for similar diseases since 1899. While its popularity declined after the rise of acetaminophen/paracetamol in 1956 and ibuprofen in 1962, it has been widely used as a preventive treatment for heart attacks and strokes and the sales has revived significantly since the last decades of last century due to its anti-clotting efficacy. In addition, while Sildenafil was initially studied for the treatment of hypertension and angina pectoris, the clinical trial results showed that the drug had little effect on angina but induced noticeable penile erections. Hence, the potential sales revenue and the delay cost may be predicted far away from the reality. In this research, we try to find all Pareto-optimal supply chain configurations so that the decision makers can select the appropriate configuration based on their preferences, such as the time and/or cost budget in the pharmaceutical companies. If a pharmaceutical company can estimate the delay cost as a single value or a small range, they also can find the optimal clinical trial supply chain configuration in the Pareto frontier.

Given a value of λ , we can find one Pareto-optimal configuration by solving the problem (P3). Hence, all Pareto-optimal configurations can be identified by evaluating all possible λ values, i.e., solving the problem (P3) for all possible λ values. However, since the value of λ may range from zero to infinity, evaluating all possible λ values is computationally infeasible. To find the

Pareto-optimal set efficiently, we propose propositions which can be used to minimize the number of λ values to be evaluated.

For a given λ , denote the Pareto-optimal solution obtained in the problem (P3) by Y_λ^* , i.e., $Y_\lambda^* = \arg \min_y z(\lambda)$. The set of all Pareto-optimal configurations is denoted by $\mathcal{P}$. For any two given λ values, denoted by a and b where $a \neq b$, one of the following three possible relations must hold:

Case (1): $Y_a^* = Y_b^*$, $f(Y_a^*) = f(Y_b^*)$ and $g(Y_a^*) = g(Y_b^*)$.

Case (2): $Y_a^* \neq Y_b^*$, but $f(Y_a^*) = f(Y_b^*)$ and $g(Y_a^*) = g(Y_b^*)$.

Case (3): $Y_a^* \neq Y_b^*$, and $f(Y_a^*) \neq f(Y_b^*)$ and $g(Y_a^*) \neq g(Y_b^*)$.

If $f(Y_a^*) = f(Y_b^*)$ and $g(Y_a^*) \neq g(Y_b^*)$, either Y_a^* or Y_b^* will be dominated by the other one. It cannot be the Pareto-optimal solution. The same rule can be applied if $f(Y_a^*) \neq f(Y_b^*)$ and $g(Y_a^*) = g(Y_b^*)$. As a result, only these three cases exist.

Y_a^* and Y_b^* have the same cost and time in both Cases (1) and (2). As discussed, the purpose of our work is to identify the Pareto-optimal configurations. Since one of Y_a^* and Y_b^* can represent the other one in terms of the Pareto-optimal tradeoff in the first two cases, it is sufficient to investigate only one of them. Without loss of generality, we assume $Y_a^* = Y_b^*$ if $f(Y_a^*) = f(Y_b^*)$ and $g(Y_a^*) = g(Y_b^*)$. The following Proposition 1 is applicable to the first two cases.

Proposition 1: For any given λ such that $a \leq \lambda \leq b$, if $Y_a^* = Y_b^*$ and $a < b$, then $Y_\lambda^* = Y_a^*$.

Proof: Since Y_a^* and Y_b^* have the minimal objective value in the problem (P3) for $\lambda = a$ and $\lambda = b$ respectively, we have

$$f(Y_a^*) + ag(Y_a^*) \leq f(y^*) + ag(y^*), \forall y^* \in \mathcal{P} \setminus Y_a^* \text{ and}$$

$$f(Y_b^*) + bg(Y_b^*) \leq f(y^*) + bg(y^*), \forall y^* \in \mathcal{P} \setminus Y_b^*.$$

If $Y_a^* = Y_b^*$, $a < b$, then with the linear combination of the above two inequalities, for $\forall \alpha \in [0,1], \forall y^* \in \mathcal{P} \setminus Y_a^*$,

$$f(Y_a^*) + (\alpha a + (1 - \alpha)b)g(Y_a^*) \leq f(y^*) + (\alpha a + (1 - \alpha)b)g(y^*).$$

Hence, Y_λ^* is equal to Y_a^* if $a \leq \lambda \leq b$.

Q.E.D.

By Proposition 1, if $Y_a^* = Y_b^*$, solving (P3) for any λ value between a and b will result in the same configuration. The solving process of (P3) for all λ values in this interval can be avoided.

However, if $Y_a^* \neq Y_b^*$ as in Case (3), it is not clear if there exist other Pareto-optimal configurations in the interval between a and b . According to Proposition 2, we can find it efficiently by evaluating the λ value where Y_a^* and Y_b^* have equal objective values. The specific λ value is called the break point of Y_a^* and Y_b^* . We denote it by $S_{a,b}$ and formulate it as follows.

$$S_{a,b} = (f(Y_b^*) - f(Y_a^*)) / (g(Y_a^*) - g(Y_b^*)). \quad (4.18)$$

Proposition 2: If $Y_a^* \neq Y_b^*$ and $Y_{S_{a,b}}^*$ is equal to Y_a^* or Y_b^* , $a < b$, then the following two results must hold.

(1) $Y_\lambda^* = Y_a^*$ if $a \leq \lambda < S_{a,b}$.

(2) $Y_\lambda^* = Y_b^*$ if $S_{a,b} < \lambda \leq b$.

Proof: According to (4.18), we have

$$f(Y_a^*) + S_{a,b}g(Y_a^*) = f(Y_b^*) + S_{a,b}g(Y_b^*). \quad (4.19)$$

If $Y_{S_{a,b}}^*$ is equal to Y_a^* or Y_b^* , then

$$(f(y^*) - f(Y_a^*)) + S_{a,b}(g(y^*) - g(Y_a^*)) \geq 0, \forall y^* \in \mathcal{P} \setminus \{Y_a^*, Y_b^*\} \quad (4.20)$$

and

$$(f(y^*) - f(Y_b^*)) + S_{a,b}(g(y^*) - g(Y_b^*)) \geq 0, \forall y^* \in \mathcal{P} \setminus \{Y_a^*, Y_b^*\}. \quad (4.21)$$

Assume there exists a value u such that $a < u < b$ and $Y_u^* \neq Y_a^*, Y_u^* \neq Y_b^*$.

Since $g(Y_\lambda^*)$ does not increase with λ , by definition of Pareto-optimality, we have

$$f(Y_a^*) < f(Y_u^*) < f(Y_b^*), \quad g(Y_a^*) > g(Y_u^*) > g(Y_b^*). \quad (4.22)$$

Thus,

$$u(g(Y_u^*) - g(Y_a^*)) > S_{a,b}(g(Y_u^*) - g(Y_a^*)), \text{ if } u < S_{a,b},$$

$$u(g(Y_u^*) - g(Y_b^*)) > S_{a,b}(g(Y_u^*) - g(Y_b^*)), \text{ if } u > S_{a,b}.$$

According to the inequalities in (4.20) and (4.21), we have

$$(f(Y_u^*) - f(Y_a^*)) + u(g(Y_u^*) - g(Y_a^*)) > (f(Y_u^*) - f(Y_a^*)) +$$

$$S_{a,b}(g(Y_u^*) - g(Y_a^*)) \geq 0, \quad \text{if } u < S_{a,b},$$

$$(f(Y_u^*) - f(Y_b^*)) + u(g(Y_u^*) - g(Y_b^*)) > (f(Y_u^*) - f(Y_b^*)) +$$

$$S_{a,b}(g(Y_u^*) - g(Y_b^*)) \geq 0, \quad \text{if } u > S_{a,b}.$$

No matter u is less or greater than $S_{a,b}$, the objective value of Y_u^* at u is always greater than that of either Y_a^* or Y_b^* , i.e., $f(Y_u^*) + ug(Y_u^*) > f(Y_a^*) + ug(Y_a^*)$ or $f(Y_u^*) + ug(Y_u^*) > f(Y_b^*) + ug(Y_b^*)$. Hence, the assumption of Y_u^* such that $Y_u^* \neq Y_a^*, Y_u^* \neq Y_b^*, a < u < b$ is invalid. Y_a^* and Y_b^* are the only two Pareto-optimal solutions in the interval $[a, b]$.

By (4.18) and (4.22), $S_{a,b}$ divides this interval into two parts, Y_a^* and Y_b^* are the Pareto-optimal solutions in the intervals $(a, S_{a,b})$ and $(S_{a,b}, b)$ respectively.

Q.E.D.

According to Proposition 2, by setting λ as $S_{a,b}$ and solving the problem (P3), we can either identify a new Pareto-optimal configuration or make sure there is no other Pareto-optimal configuration in the interval $[a, b]$. If a new

configuration is identified, the maximal (or minimal) λ value to get the configuration Y_a^* (or Y_b^*) should be less (or greater) than $S_{a,b}$. Otherwise, $S_{a,b}$ is indeed the separating point between the intervals of Y_a^* and Y_b^* , where the interval of Y_a^* (or Y_b^*) represents the interval of all λ values at which the Pareto-optimal solution is equal to Y_a^* (or Y_b^*).

4.3.2 Optimality Cut Properties

In this section, we will derive some optimality cuts to improve the solving efficiency of (P3) for each λ value. The optimality cuts can be obtained by analyzing the break points of Y_0^* , Y_∞^* and other Pareto-optimal configurations.

Y_0^* and Y_∞^* consist of the two extreme points of the Pareto-optimal tradeoffs between time and cost, i.e., the minimal cost with the maximal time and the minimal time with the maximal cost. We can simply get $f(Y_0^*) = 0$ and $g(Y_0^*) = l$, where l is the planning horizon. Let $t_m = \min\{g(y), s. t. (4.5) \sim (4.14)\}$ and $c_m = \min\{f(y), s. t. (4.5) \sim (4.14), g(y) = t_m\}$. We have $f(Y_\infty^*) = c_m$ and $g(Y_\infty^*) = t_m$. Denote the entire interval of λ (from zero to infinity) by $\mathcal{IV}$, the interval of a Pareto-optimal configuration Y_λ^* by $IV_{Y_\lambda^*}$. Any λ value not in the intervals of Y_0^* and Y_∞^* is denoted by $\hat{\lambda}$, i.e., $\hat{\lambda} \in \mathcal{IV} \setminus \{IV_{Y_0^*}, IV_{Y_\infty^*}\}$. Hence, for any $\hat{\lambda}$, the break point of $Y_{\hat{\lambda}}^*$ and Y_0^* (or Y_∞^*) can be formulated as follows.

$$S_{0,\hat{\lambda}} = f(Y_{\hat{\lambda}}^*) / (l - g(Y_{\hat{\lambda}}^*)), \quad \hat{\lambda} \in \mathcal{IV} \setminus \{IV_{Y_0^*}, IV_{Y_\infty^*}\}. \quad (4.23)$$

$$S_{\hat{\lambda},\infty} = (c_m - f(Y_{\hat{\lambda}}^*)) / (g(Y_{\hat{\lambda}}^*) - t_m), \quad \hat{\lambda} \in \mathcal{IV} \setminus \{IV_{Y_0^*}, IV_{Y_\infty^*}\}. \quad (4.24)$$

With these break points, we have the following optimality cut properties.

Optimality cut property 1: $(l - t_m)f(y) + c_m g(y) \leq c_m l$.

Proof. Since Y_0^* and Y_∞^* have the intervals with the minimal or maximal λ values respectively, a Pareto-optimal configuration $Y_{\hat{\lambda}}^*$'s break point with Y_0^* can be no greater than that with Y_∞^* , i.e.,

$$S_{0,\hat{\lambda}} \leq S_{\hat{\lambda},\infty},$$

which is equivalent to

$$f(Y_{\hat{\lambda}}^*)/(l - g(Y_{\hat{\lambda}}^*)) \leq (c_m - f(Y_{\hat{\lambda}}^*))/(g(Y_{\hat{\lambda}}^*) - t_m), \text{ then}$$

$$(l - t_m)f(Y_{\hat{\lambda}}^*) + c_m g(Y_{\hat{\lambda}}^*) \leq c_m l.$$

Since the above inequality is also valid for Y_0^* and Y_∞^* , it is valid for all Pareto-optimal configurations of this problem. Q.E.D.

Optimality cut property 2:

(1) If $\lambda \leq c_m/(l - t_m)$, then $f(y) + \lambda g(y) \leq \lambda l$.

(2) If $\lambda \geq c_m/(l - t_m)$, then $f(y) + \lambda g(y) \leq c_m + \lambda t_m$.

Proof. By Optimality cut property 1, $S_{0,\hat{\lambda}} \leq c_m/(l - t_m) \leq S_{\hat{\lambda},\infty}$. According to Proposition 2, for any λ value in $Y_{\hat{\lambda}}^*$'s interval, we have $S_{0,\hat{\lambda}} \leq \lambda \leq S_{\hat{\lambda},\infty}$, i.e.,

$$f(Y_{\hat{\lambda}}^*)/(l - g(Y_{\hat{\lambda}}^*)) \leq \lambda \leq (c_m - f(Y_{\hat{\lambda}}^*))/(g(Y_{\hat{\lambda}}^*) - t_m), \text{ then}$$

$$f(Y_{\hat{\lambda}}^*) + \lambda g(Y_{\hat{\lambda}}^*) \leq \lambda l \text{ and } f(Y_{\hat{\lambda}}^*) + \lambda g(Y_{\hat{\lambda}}^*) \leq c_m + \lambda t_m.$$

In addition, since $\lambda l \leq c_m + \lambda t_m$ if $\lambda \leq c_m/(l - t_m)$, we have

(1) if $S_{0,\hat{\lambda}} \leq \lambda \leq c_m/(l - t_m)$, $f(Y_{\hat{\lambda}}^*) + \lambda g(Y_{\hat{\lambda}}^*) \leq \lambda l$;

(2) if $c_m/(l - t_m) \leq \lambda \leq S_{\hat{\lambda},\infty}$, $f(Y_{\hat{\lambda}}^*) + \lambda g(Y_{\hat{\lambda}}^*) \leq c_m + \lambda t_m$.

Obviously, the inequalities in both (1) and (2) are also valid for Y_0^* and Y_∞^* respectively. Since Y_0^* and Y_∞^* are the corresponding Pareto-optimal configurations when $\lambda \leq S_{0,\hat{\lambda}}$ and $\lambda \geq S_{\hat{\lambda},\infty}$, the two inequalities are valid for all Pareto-optimal solutions obtained when λ is no greater or no less than $c_m/(l - t_m)$. Q.E.D.

For any given λ , there are two optimality cuts in this problem. While the optimality cut 1 is identical for all λ values, the optimality cut 2 is unique and depends on the λ value applied in the problem (P3). Meanwhile, it also provides an upper bound for the objective value of this problem.

Figure 4-1. The optimality cuts in the properties

As shown in Figure 4-1, the optimality cuts have eliminated part of the feasible region in this problem. As obtained at Y_0^* and Y_∞^* , $(0, c_m)$ and (t_m, l) represent the lower and upper bounds of the cost and time in the problem (P3) respectively. They jointly determine the feasible region of this problem. With the application of the optimality cuts, only the part under a dashed line is left as the solution space. The solving efficiency of the problem (P3) for each λ may be improved.

4.4 Algorithm

In this section, a recursive algorithm is developed to find all Pareto-optimal configurations of this problem. In this algorithm, a function is defined to call itself within the program text. Given different Pareto-optimal solutions at two λ

values, the break point of them is evaluated in this function to check if there exist other Pareto-optimal configurations in the interval between the two λ values. According to Proposition 2, if the Pareto-optimal solution is equal to either of the given ones, there will be no other Pareto-optimal solutions in this interval. Otherwise, the break point divides this interval into two parts, which will be further investigated. Since the λ value may range from zero to infinity, by finding Y_0^* and Y_∞^* , we can find all Pareto-optimal configurations of this problem through recursively calling this function.

This algorithm has a linear time complexity $T(n) = 2n - 3$, where n is the cardinality of the Pareto-optimal set. Since Y_0^* and Y_∞^* are obtained before the algorithm starts, the algorithm does not include the solving processes of the two configurations but the other $(n - 2)$ Pareto-optimal configurations in this problem. In addition, the algorithm starts at evaluating the breakpoint $S_{0,\infty}$. When a new Pareto-optimal configuration is found, two function calls are incurred where one λ value is evaluated in each function. Hence, the total number of λ values to be evaluated in this algorithm should be $T(n) = 1 + 2(n - 2) = 2n - 3$. Meanwhile, since the two intervals requiring further investigation in the function are mutually exclusive, the solving processes in the two intervals are independent and can be conducted in parallel. Hence, this algorithm is efficient in exploring appropriate λ values to identify all Pareto-optimal configurations of this problem. The pseudo code of the function in this algorithm is described as follows.

Table 4-2. The function of finding the Pareto-optimal set

Function Find_Pareto-optimal_Set (a, b, Y_a^*, Y_b^*)	
1.	{
2.	Set λ as $S_{a,b}$, solve (P3), then $Y_{S_{a,b}}^* = \arg \min_y z(S_{a,b})$;
3.	If ($Y_{S_{a,b}}^* = Y_a^* \parallel Y_{S_{a,b}}^* = Y_b^*$)
4.	Save ($Y_a^*, a, S_{a,b}$) and ($Y_b^*, S_{a,b}, b$);
5.	Else
6.	{
7.	Find_Pareto-optimal_Set ($a, S_{a,b}, Y_a^*, Y_{S_{a,b}}^*$);
8.	Find_Pareto-optimal_Set ($S_{a,b}, b, Y_{S_{a,b}}^*, Y_b^*$);
9.	}
10.	}

4.5 Numerical Experiments

In this section, we give a numerical study and apply the proposed algorithm and optimality cuts in the solving process. In Section 4.4.1, we describe the parameters and implementation details in the numerical study. In Section 4.4.2, we give the efficiency curve of the time-cost tradeoff in this multi-objective problem and show the effect of optimality cuts in reducing the solving time in Section 4.4.3.

4.5.1 Numerical Study

We consider a numerical study of the multi-objective site selection problem in clinical trial supply chains. As recorded on the website ClinicalTrials.gov, the number of participating clinical sites in a clinical trial ranges from several to more than one hundred. The average number is 27 based on the analysis of Roberts, et al. (2003). Hence, in our numerical study, 50 candidate clinical sites are considered in this site selection problem. Since each patient is required to participate in the clinical trial for three to six weeks, we assume one period in the model represents one week and the treatment period lasts for three weeks. The

planning horizon is assumed to have 20 periods in this numerical study. In addition, the randomly generated enrollment, shipping, and holding costs at each clinical site are assumed to range from one thousand to ten thousand, one to ten, and 0.1 to 1, respectively. The patient recruitment uncertainty is modeled as a number of scenarios based on the patient arrival rates at clinical sites according to the sample average approximation method (Santoso, et al. 2005, Verweij, et al. 2003).

The proposed problem and solving method in this chapter are implemented in C# and CPLEX 12.6 is used as a callable library to solve the mixed-integer problem. This program is executed on a windows server with two Intel® Xeon® E5-2620 v2 2.1GHz processors.

4.5.2 The Efficiency Curve

Since the patient recruitment uncertainty is modeled and captured by scenarios in the stochastic model in Section 4.2, we first apply the SAA method (in Section 2.3.1.1) to determine the number of scenarios required in this problem and then give the efficiency curve to show the tradeoff between the cost and time of a clinical trial.

Table 4-3 presents the true problem's upper bound in the SAA method. The 23 values of the delay cost in this table are evaluated to find all Pareto-optimal configurations of the problem (P3) with 20 scenarios according to the algorithm in Section 4.4. For each of these delay cost values, we can get a Pareto-optimal supply chain configuration of the problem (P3) with 20 scenarios. According to the SAA method, the true problem's upper bound can be obtained by substituting this configuration into the problem (P3) with 1000 scenarios and solving this problem. Meanwhile, we also get the corresponding cost and time in the upper

bound and the standard deviation for the objective value, cost and time.

Figure 4-2 illustrates the comparison between the lower and upper bounds in the SAA method. Based on the results in Table 4-3, the upper bounds and their standard deviations for all possible values of the delay cost are approximated according to Proposition 1 and Proposition 2. For each value of the delay cost, the standard deviation is within 0.6% of the upper bound. It is shown by error bars in the figure. As shown in this figure, the gap between the lower and upper bounds is quite small. Based on our calculation, the maximal optimality gap in this figure is 2.33% and the maximal standard deviation of the lower bound is 0.79% for all values of the delay cost. As a result, the maximal standard deviation of the optimality gap is less than 1.39%. Both the estimated optimality gap and its standard deviation are acceptable. Hence, 20 scenarios are sufficient to represent the patient recruitment uncertainty in this problem.

Table 4-3. The upper bound in the SAA method

No.	Delay cost	Upper bound			Standard deviation			Standard deviation (%)		
		ObjValue	Cost	Time	ObjValue	Cost	Time	ObjValue	Cost	Time
1	9527.59	190551.72	0.00	20.00	0.00	0.00	0.0000	0.000	NaN	0.000
2	10360.34	202653.05	67657.82	13.03	612.01	8.31	0.0594	0.302	0.012	0.456
3	11272.99	214318.09	71602.09	12.66	662.53	7.85	0.0590	0.309	0.011	0.466
4	11884.90	222064.93	71602.09	12.66	698.54	8.00	0.0590	0.315	0.011	0.466
5	13312.70	240140.87	71602.09	12.66	782.84	7.87	0.0590	0.326	0.011	0.466
6	13943.71	248129.41	71602.09	12.66	811.75	7.41	0.0584	0.327	0.010	0.461
7	14047.80	249447.24	71602.09	12.66	821.78	7.73	0.0586	0.329	0.011	0.463
8	14838.84	259461.80	71602.09	12.66	868.14	7.73	0.0586	0.335	0.011	0.463
9	14957.49	260963.98	71602.09	12.66	801.43	9.74	0.0540	0.307	0.014	0.426
10	21989.87	344385.75	90622.67	11.54	1181.16	9.66	0.0540	0.343	0.011	0.468
11	23205.12	356524.29	112638.47	10.51	1199.65	14.48	0.0522	0.336	0.013	0.497
12	27114.77	397614.72	112638.47	10.51	1403.78	14.48	0.0522	0.353	0.013	0.497
13	27249.20	399027.56	112638.47	10.51	1410.80	14.48	0.0522	0.354	0.013	0.497
14	30209.91	430706.58	117731.89	10.36	1470.07	13.07	0.0490	0.341	0.011	0.473
15	30632.87	435088.44	117731.89	10.36	1490.79	13.07	0.0490	0.343	0.011	0.473
16	48768.39	604900.17	148915.76	9.35	2295.69	13.51	0.0473	0.380	0.009	0.505
17	51593.20	629734.23	197899.15	8.37	2495.96	12.02	0.0485	0.396	0.006	0.580
18	76983.70	842252.71	197899.15	8.37	3728.01	12.00	0.0485	0.443	0.006	0.580
19	79857.50	866306.42	197899.15	8.37	3867.45	12.00	0.0485	0.446	0.006	0.580
20	80060.33	867980.47	212286.34	8.19	3151.19	11.57	0.0394	0.363	0.005	0.481
21	80161.75	868807.30	220298.74	8.09	2566.97	12.04	0.0321	0.295	0.005	0.397
22	111442.24	1121866.43	220298.74	8.09	3570.51	12.05	0.0321	0.318	0.005	0.397
23	120000.94	1188350.17	285943.13	7.52	6006.40	20.87	0.0502	0.505	0.007	0.668

Figure 4-2. The comparison between the lower and upper bounds in the SAA method

Table 4-4 presents the Pareto-optimal tradeoffs between the cost and time in this multi-objective problem. Each tradeoff is presented together with an interval defined by the “Start” and “End”. For any per-period delay cost value in each interval, the Pareto-optimal tradeoff obtained in the problem (P3) is identical to this interval’s corresponding tradeoff in this table. In addition, the 13 Pareto-optimal tradeoffs are obtained by solving the problem (P3) for each of the 23 delay cost values in Table 4-3. The time complexity is consistent with $T(n) = 2n - 3$ discussed in Section 4.4.

Table 4-4. The Pareto-optimal tradeoffs

No.	Cost	Time	Start	End
1	0.00	20.00	0.00	9527.59
2	67645.86	12.90	9527.59	11272.99
3	71591.40	12.55	11272.99	13312.70
4	73588.31	12.40	13312.70	14047.80
5	75695.48	12.25	14047.80	14957.49
6	90652.97	11.25	14957.49	23205.12
7	112697.84	10.30	23205.12	27249.20
8	116785.22	10.15	27249.20	30632.87
9	148949.74	9.10	30632.87	51593.20
10	197963.27	8.15	51593.20	79857.50
11	201956.15	8.10	79857.50	80161.75
12	209972.32	8.00	80161.75	120000.94
13	275972.84	7.45	120000.94	∞

Figures 4-3, 4-4 and 4-5 are obtained from Table 4-4. The efficiency curve in Figure 4-3 provides an explicit description of the time-cost tradeoff of the clinical trial. The time reduction rate is decreasing with the increase of the cost in this figure. In addition, the total cost and time required in the clinical trial are increasing and decreasing with the delay cost in Figure 4-4 and Figure 4-5, respectively. When the delay cost increases from zero to around 50 thousand, the time decreases rapidly with the increase of the cost. However, when the delay cost is greater than 50 thousand, although the cost required to reduce the time is largely increased, the time reduction is trivial.

Figure 4-3. The efficiency curve

Figure 4-4. Total cost's increase with the delay cost

Figure 4-5. Total time's decrease with the delay cost

To complete a clinical trial with both acceptable duration and cost, decision makers can make site selection decisions based on these figures and their preferences or other higher-level information, such as the estimated potential delay cost of a specific drug. When a specific drug's estimated potential delay cost is not so high (such as the values less than 50 thousand in our case), it is not

worth large expenses to enroll numerous clinical sites to minimize the time in this clinical trial. However, sometimes the drug's delay cost can be very high (such as the values greater than 150 thousand in our case), especially when the drug is urgently required to conquer certain intractable diseases seriously threatening people's lives, e.g. the drugs for cancer and AIDS. Even if tremendous cost is required, the time should be reduced as much as possible. As referred in the literature (Clemento 1999), a typical drug's daily delay cost can be more than \$1 million. According to the results in Table 4-4, all candidate clinical sites should be enrolled in the clinical trial supply chains in this case.

4.5.3 Performance Analysis

Table 4-5 presents the two optimality cuts' performances in reducing the problem's solving time. The optimality gap and the time limit are set as 0.01% and ten hours respectively in each run. The columns under "Solving time" record the computational time of the problem (P3) for each evaluated delay cost with the application of no cut, both cuts, and either only cut 1 or only cut 2. The columns under "Improvement" (or "Decline") indicate how many percents of the solving time in the no cut case has been reduced (or increased), i.e., the solving efficiency is improved (or declined), with the application of the corresponding cuts.

As shown in the table, the application of both cuts has improved the solving efficiency of the problem (P3) for most evaluated delay cost values. Except one problem, the solving efficiency in all problems reaching the time limit without any cut (i.e., the no cut case) has been largely improved. With the application of both cuts, these problems can get the optimal solution within the time limit. The solving time has been reduced by at least 76.06% to 95.45%. In addition, it has

also been reduced by 52.58% in the problem solved within more than 6 hours. Although the application of both cuts increases the solving time in some problems, compared with the improvement obtained above, the increase is quite trivial. The last row in the table shows that the application of both cuts has reduced 71.69% of the total solving time of the problem (P3) for all of the 23 evaluated delay cost values. Hence, the optimality cuts can be applied to improve this problem's solving efficiency.

Table 4-5. The performance of the optimality cuts

No.	Delay cost	Solving time (hh:mm:ss)				Improvement (%)			Decline (%)		
		No cut	Both cuts	Cut 1	Cut 2	Both cuts	Cut 1	Cut 2	Both cuts	Cut 1	Cut 2
1	21989.87	00:41:05	02:34:11	02:34:17	02:33:53				275.33	275.58	274.58
2	10360.34	10:00:41	00:56:36	10:00:26	10:00:26	90.58					
3	9527.59	10:00:59	10:01:10	10:00:40	10:00:18						
4	13943.71	10:00:32	01:28:42	02:16:23	03:43:39	85.23	77.29	62.76			
5	11884.90	10:00:20	01:21:57	05:17:23	10:00:22	86.35	47.13				
6	11272.99	10:00:58	02:23:53	10:00:33	10:13:59	76.06					
7	13312.70	10:01:40	01:58:44	10:00:50	11:38:32	80.27					
8	14838.84	06:38:23	03:08:54	02:44:00	10:01:57	52.58	58.83				51.10
9	14047.80	02:42:12	02:43:38	03:14:23	10:02:46				0.88	19.84	271.63
10	14957.49	10:08:14	01:46:37	10:01:27	10:30:10	82.47					
11	48768.39	10:01:21	00:36:50	00:25:20	00:28:46	93.87	95.79	95.22			
12	27114.77	10:01:30	01:46:54	10:00:59	10:00:36	82.23					
13	23205.12	10:00:55	02:07:50	10:03:46	10:00:47	78.73					
14	30209.91	10:01:30	02:03:55	10:00:53	10:01:01	79.40					
15	27249.20	10:09:28	02:04:40	10:01:23	10:00:48	79.54					
16	30632.87	10:04:24	02:15:38	10:02:36	10:01:57	77.56					
17	76983.70	00:30:43	00:30:47	01:37:49	02:33:08					218.44	398.49
18	51593.20	01:01:03	00:58:25	03:12:51	03:16:41					215.90	222.16
19	111442.24	00:13:32	00:18:05	01:17:23	01:18:00				33.63	471.66	476.14
20	80060.33	00:27:03	00:46:32	01:14:54	01:12:34				72.00	176.83	168.20
21	79857.50	00:22:03	00:34:42	01:32:44	01:29:37				57.43	320.74	306.55
22	80161.75	00:23:29	00:32:10	10:02:47	01:57:33				36.99	2467.78	400.72
23	120000.94	10:00:23	00:27:17	10:02:44	10:00:45	95.45					
Sum		153:32:27	43:28:08	145:46:31	161:08:14	71.69	5.06				4.95

In addition, Figure 4-6 demonstrates the corresponding optimality cuts 1 and 2 of two delay cost values on the efficiency curve. As discussed, the optimality cut 1 is applicable for all Pareto-optimal tradeoffs in the figure. By solving the problem (P3) for the two delay cost values, we get the tradeoffs at the points “A” and “B” in this figure. The dashed lines above the two points are actually their corresponding optimality cut 2, which can provide a tight upper bound for the objective value of this problem.

Figure 4-6. The optimality cuts on the efficiency curve

4.6 Conclusion

In this chapter, we consider a network design problem in clinical trial supply chains. Since the clinical trial duration is part of the limited patent life, its reduction brings more profit. To complete the clinical trial (i.e., reach the patient accrual goal) within less time, we can accelerate the patient recruitment process by increasing the number of participating clinical sites. However, the cost is increased simultaneously. To conduct the clinical trial efficiently, the clinical sites should be selected with the consideration of the tradeoff between the time and cost.

Although network design problems have been addressed in the literature, their settings are different from ours. While those researchers either only consider time minimization or assume fixed processing times at certain phases, we consider the tradeoff between time and cost of clinical trials, where the time is variable and depends on the stochastic patient recruitment process.

The network design problem in this chapter is formulated as a multi-

objective site selection model, which considers the tradeoff between the time and cost required by clinical trials. By introducing the delay cost, this multi-objective model is converted into a single objective one using the weighted sum method. To efficiently identify all Pareto-optimal configurations, we develop an algorithm based on the proposed propositions. In this algorithm, rather than all possible values of the delay cost, only a few are required to be evaluated. Two optimality cuts are also derived to improve this problem's solving efficiency in each evaluation. An efficiency curve showing the Pareto-optimal tradeoffs is provided for decision makers to support their supply chain design.

Chapter 5 Production Planning in Clinical Trial Supply Chains

5.1 Introduction

Due to the increasing complexity and cost of new drug development, pharmaceutical companies face declining profits. They strive to reduce the overproduction of clinical drugs, which is commonly seen in the pharmaceutical industry. When the overproduction is reduced in clinical trial supply chains, stock-outs may occur when the allocated inventory cannot meet the demand at some clinical sites. As a result, the clinical trial duration may increase. Since it is included in the limited patent life, its increase may reduce the drug's commercialization time under patent protection and the total sales revenue significantly. The duration should be minimized in clinical trials.

To reduce the overproduction without the duration increase in clinical trials, efficient inventory coordination between the DCs and clinical sites is crucial. As discussed in Chapter 3, the reverse replenishment and transshipment can be applied jointly with the direct replenishment in clinical trial supply chains. Due to the stochastic patient arrivals and positive lead time of regular deliveries, the clinical trial delay may not be avoided. Hence, in addition to the regular deliveries, the emergency delivery mode is also applicable in each delivery route of the supply chain. Compared with the regular one, the emergency delivery mode has negligible lead time but more expensive cost. The regular and emergency transshipments are equivalent to the proactive and reactive transshipments in the literature (Paterson, et al. 2011). While proactive transshipments are used to arrange the inventory distribution before the stochastic demand is realized,

reactive transshipments ship the inventory at a clinic with excess inventory to clinics facing or going to face stock-outs (Paterson, et al. 2011).

While transshipments and emergency deliveries avoid the clinical trial delay caused by stock-outs, they inevitably increase the delivery cost. However, to avoid the transshipments, the inventory should be held at the DC and the DC has to deliver more frequently with smaller shipping quantities, which will also increase the delivery cost. On the other hand, to avoid the emergency deliveries, enough inventory should be held at clinical sites. Due to the limited inventory in the supply chain and stochastic demand, stock-outs may happen at some clinics but there still exist inventories at other clinics.

The production planning problem in clinical trial supply chains is a multi-objective problem which minimizes the time and cost of clinical trials. The time is the clinical trial duration and the cost consists of the production and inventory allocation costs. Since clinical drugs have no salvage cost and all leftovers should be disposed of after an unsuccessful trial, the production cost should be considered in this problem. The inventory allocation results should also be considered in this problem, since they depend on the production quantity and influence both the duration and operational cost in clinical trials. In addition, since the clinical trial delay can be costly (for a typical drug, more than \$1 million revenue can be lost for a single day's loss in clinical trials (Clemento 1999)), the duration minimization has a higher priority than the cost minimization in clinical trial supply chains.

In this chapter, we consider the production planning problem in clinical trial supply chains. The contributions of this study are three-fold. First, we propose a multi-objective model to find the optimal production quantity through

minimizing the clinical trial duration and cost. Since the duration minimization is usually more important than the cost minimization in clinical trials, this multi-objective model can be solved by the lexicographic method (or the preemptive optimization, (Coello et al. 2007, Marler and Arora 2004)). In this method, all objectives are ranked in order of importance and the optimization is performed by considering one objective at a time, based on priorities. Second, due to the patient recruitment uncertainty at clinical sites, this model is formulated as a multi-stage stochastic programming model to capture the dynamic inventory allocation process in the supply chain. Since the computational complexity increases significantly with the increase of stage and scenario numbers in terms of the computational time and required memory space, the progressive hedging algorithm is applied as the solution approach. Thirdly, numerical experiments are conducted to study the algorithm's performance and find the optimal production quantity of clinical drugs in a numerical study. Sometimes the production cost of certain clinical drugs is too expensive that the pharmaceutical companies would rather save the cost with the sacrifice of the minimal clinical trial duration. Hence, the tradeoffs between the clinical trial duration and cost are also provided for decision makers in production planning.

The remainder of this chapter is organized as follows. With the assumptions in Section 5.2.1, the production planning model is given in Section 5.2.2 under deterministic and stochastic settings respectively. By applying the progressive hedging algorithm as the solution approach, numerical experiments are conducted in Section 5.3 to find the optimal production quantity of clinical drugs in a numerical study. We also discuss the influence of the parameter ρ in the algorithm on the algorithm's convergence and solution quality. Concluding

remarks and future works are given in Section 5.4.

5.2 Model Development

In this section, we present a mathematical formulation for the production planning problem in clinical trial supply chains. With the model assumptions given in Section 5.2.1, the problem is formulated as a deterministic model first and then extended to a stochastic programming model in Section 5.2.2.

5.2.1 Assumptions

Clinical trials are commonly conducted in a multi-echelon supply chain shown in Figure 3-1. Prior to their commencements, some implementation details should be determined. The model assumptions are summarized as follows.

- (1) The planning horizon of a clinical trial is finite.
- (2) The shelf life of clinical drugs is longer than the planning horizon.
- (3) Only one production is considered for each clinical trial.
- (4) The patient arrivals at each clinical site are independent and follow a non-stationary Poisson distribution.
- (5) Patients are fulfilled with the same clinical drugs in the whole treatment period.
- (6) The treatment period consists of several continuous periods and the dropout rate is identical in each period.
- (7) The forecast demand is rounded down if it is not an integer.
- (8) The delivery cost is composed of a fixed and a variable element in each route.
- (9) The lateral transshipment has a shorter lead time but a higher cost than the direct replenishment.

(10) The emergency delivery has negligible lead time but a higher cost than the regular delivery in each route.

(11) The holding cost is considered only at the clinical sites.

Since a clinical trial has a predetermined patient accrual goal, its planning horizon is finite. As shown on the website clinicaltrials.gov, the planning horizon is determined by a study start date and an estimated study completion date. The shelf life of drugs usually lasts for several years and can cover the entire planning horizon of a clinical trial. In addition to the long production time (about 6 to 9 months, (Laínez, et al. 2012)), clinical drugs are usually produced in large batches. We only consider one production for each clinical trial.

As discussed in the literature (Chen, et al. 2012, Fleischhacker, et al. 2015), patient arrivals follow an independent non-stationary Poisson process. The forecast demand at each clinical site can be approximated from the record of patients visiting the site and/or the previous data of clinical trials for the same disease (Fleischhacker, et al. 2015). Patients are required to take the same clinical drugs for the whole treatment period, which usually lasts for three to six weeks. In other words, the data collected on a patient is valid only if the patient has taken the clinical trial once a week and three to six continuous weeks in total. If patients fail to take the trials on time, they should drop out since their data is no longer valid. Similar to the assumption in the literature (Chen, et al. 2012), the dropout rate is assumed to follow a uniform distribution and have identical values in each period. In practice, it may decrease in the treatment period since patients who do not drop out early on are more likely to stay later. Due to the dropout rate, the forecast demand of existing patients for follow-up treatments may not be integers. To keep their integral nature, they are rounded down in the mathematical model

in the following section.

In this study, two elements (a fixed and a variable one) are included in the delivery cost for the sake of extension to general applications. In this problem, our objective is to achieve the minimal clinical trial duration with the minimal total production and inventory distribution costs. Due to the non-negligible lead time of the regular delivery and stochastic patient arrivals, transshipments and emergency deliveries are required to achieve the minimal clinical trial duration in certain cases. Compared with the direct replenishment and regular deliveries, they have a shorter or negligible lead time but a higher cost. Due to the differences of local regulations, transshipment is only applicable between a few clinical sites. The holding cost is usually greater at the clinical sites than at the DC in terms of the limited storage spaces. Their differences are considered here as the holding costs at the clinical sites for simplicity.

5.2.2 Production Planning Model

In this section, the production planning problem in a multi-period multi-retailer clinical trial supply chain is formulated as a mathematical model, which considers the production planning and inventory allocation decisions jointly. The inventory allocation decisions follow a sequential process. With the patient recruitment process revealed gradually over time, they are determined and implemented correspondingly in each period of the planning horizon (consisting of T periods). This decision process has the following form (Shapiro, et al. 2009).

$$\begin{aligned} &\text{decision } (\theta_1) \sim \text{observation } (d_2) \sim \text{decision } (\theta_2) \sim \dots \\ &\dots \sim \text{observation } (d_T) \sim \text{decision } (\theta_T). \end{aligned}$$

If the observations are deterministic (e.g. patients are recruited before the

clinical trial commencement and they will arrive by appointment), this decision process can be formulated as a deterministic model. Due to the patient recruitment uncertainty in clinical trials, the production planning problem here is a stochastic one and can be formulated as a multi-stage stochastic programming model. In the following sections, we first present the model under the deterministic setting and extend it to the stochastic setting subsequently.

5.2.2.1 A Deterministic Model

We consider a production planning problem in a clinical trial supply chain, where the planning horizon consists of $|T|$ time periods and the patient arrivals in each period are known. The objective is to determine the production quantity of clinical drugs with which the minimal clinical trial duration can be achieved with the minimal total cost. With the notations in Table 5-1, the model formulation is presented as follows.

Table 5-1. Summary of the notations in the deterministic model

Notation	Description
Sets:	
I	Set of clinical sites, $i \in I$
T	Set of time periods, $t \in T$ (Please see Appendix 1 for the explanation for the length requirement of the planning horizon.)
Parameters:	
r	The unit production cost of clinical drugs
αf_i	The fixed delivery cost between the DC and clinical site i
αv_i	The variable delivery cost between the DC and clinical site i
l_i	The delivery lead time between the DC and clinical site i
h_i	The holding cost at clinical site i
$d_{t,i}$	The number of forecast new arrival patients at clinical site i in period t
φ	The average dropout rate in each period
m	The number of periods in the treatment period
G	The patient accrual goal
M	A sufficiently large number (E.g. one may define $M = G \sum_{k=1}^m (1 - \varphi)^{k-1} / (1 - \varphi)^{m-1}$, please see Appendix 2 for explanation.)

Decision variables:

Q	The production quantity
f_t	1 if all clinical sites' total number of patients fulfilled for m periods until period t is less than the patient accrual goal, 0 otherwise
$a_{t,i}$	1 if clinical site i orders from the DC at the beginning of period t , 0 otherwise
$x_{t,i}$	The replenishment quantity of clinical site i at the beginning of period t
ID_t	The inventory level of the DC in period t
$IC_{t,i}$	The inventory level of clinical site i at the beginning of period t
$Sd_{t,k,i}$	The number of patients who first arrive at clinical site i in period t and are fulfilled in the clinical trial for at least k periods, $1 \leq k \leq m$.
θ	All decision variables in the model

$$\min Df_1(\theta) = \sum_{t \in T} f_t, \quad (5.1)$$

$$\min Df_2(\theta) = rQ + \sum_{t \in T, i \in I} (h_i IC_{t,i} + \alpha f_i a_{t,i} + \alpha v_i x_{t,i}), \quad (5.2)$$

$$\text{s. t. } \sum_{\tau=1}^{\max(t-m+1,0)} \sum_i Sd_{\tau,m,i} \geq G(1 - f_t), \quad \forall t \in T \quad (5.3)$$

$$Sd_{t,1,i} \leq d_{t,i}, \quad \forall i \in I, \forall t \in T \quad (5.4)$$

$$Sd_{t,k,i} \leq (1 - \varphi) Sd_{t,(k-1),i}, \quad \forall i \in I, \forall k \in \{2, \dots, m\}, \forall t \in T \quad (5.5)$$

$$\sum_{\tau=\max(t-m+1,1)}^t Sd_{\tau,(t-\tau+1),i} \leq IC_{t,i}, \quad \forall i \in I, \forall t \in T \quad (5.6)$$

$$ID_0 = Q, \quad (5.7)$$

$$ID_t = ID_{(t-1)} - \sum_i x_{t,i}, \quad \forall t \in T \quad (5.8)$$

$$IC_{t,i} = x_{t,i}, \quad \forall i \in I, t = 1 \quad (5.9)$$

$$IC_{t,i} = IC_{(t-1),i} - \sum_{\tau=\max(t-m,1)}^{t-1} Sd_{\tau,(t-\tau),i} + x_{(t-l_i) \setminus \{1\},i}, \quad \forall i \in I, \forall t \in T \setminus \{1\} \quad (5.10)$$

$$x_{t,i} \leq M a_{t,i}, \quad \forall i \in I, \forall t \in T \quad (5.11)$$

$$x_{t,i}, ID_t, IC_{t,i}, Sd_{t,k,i} \in \mathbb{N}, \quad \forall i \in I, \forall k \in \{1, \dots, m\}, \forall t \in T \quad (5.12)$$

$$f_t, a_{t,i} \in \{0,1\}. \quad \forall i \in I, \forall t \in T \quad (5.13)$$

The production planning model has two objectives. The clinical trial duration $Df_1(\theta)$ and the total production, holding and delivery costs $Df_2(\theta)$

are minimized in (5.1) and (5.2) sequentially. Under the deterministic settings, the minimal clinical trial duration can be achieved via the regular deliveries. Constraint (5.3) checks if the clinical trial's patient accrual goal has been reached within the first t periods. Constraints (5.4) to (5.6) guarantee the fulfilled patients at each clinical site in each period are no greater than the patient arrivals (new arrivals in (5.4) and existing patients for follow-up treatments in (5.5)) and the inventory level there. In Constraint (5.7) the initial inventory at the DC is set to the production quantity Q . Constraints (5.8) to (5.10) are the inventory balance constraints at the DC and clinical sites. Constraint (5.11) defines the delivery cost's fixed element each site. Constraints (5.12) and (5.13) enforce the non-negativity (natural numbers) or binary restrictions on the decision variables in the model.

Since the duration $Df_1(\theta)$ is more important than the total cost $Df_2(\theta)$ in clinical trials, this model can be solved by the lexicographic method. At first, we obtain the minimal clinical trial duration Df_1^* in the following program (5.14). By setting the clinical trial duration $Df_1(\theta)$ to its minimal value Df_1^* , we find the optimal production quantity through minimizing the total production and operational costs $Df_2(\theta)$ in (5.15).

$$Df_1^* = \min\{Df_1(\theta), s. t. (5.3) \sim (5.13)\}. \quad (5.14)$$

$$Q^* = \arg \min_Q \{Df_2(\theta), s. t. (5.3) \sim (5.13), Df_1(\theta) = Df_1^*\}. \quad (5.15)$$

The production planning problem is NP-hard. If we only consider single period, then f_1^* is equal to 0 to achieve the minimal duration in the first objective. Constraints (5.3) to (5.6) are equivalent to $\sum_i Sd_i \geq G$ and $Sd_i = \min(d_i, x_i)$. Due to Constraint (5.11), we have $Sd_i = \min(d_i, x_i) = \min(d_i a_i, x_i)$. Since

$\sum_i S d_i = \sum_i \min(d_i a_i, x_i) \leq \sum_i d_i a_i$, $\sum_i S d_i \geq G$ can be represented as $\sum_i d_i a_i \geq \sum_i S d_i \geq G$. If we further set all of the production, holding, and variable delivery costs to zero, then the model of the second objective can be reduced to a min-knapsack problem as follows.

$$\min\{\sum_i \alpha f_i a_i : s. t. \sum_i d_i a_i \geq G, a_i \in \{0,1\} \forall i \in I\}$$

The min-knapsack problem is NP-hard (Csirik, et al. 1990). Unless $P = NP$, it is not possible to develop an algorithm which can solve this multi-period problem in polynomial time.

5.2.2.2 Stochastic Programming Extensions

The patient recruitment in clinical trials is a stochastic process over the planning horizon. As discussed in Section 5.2.1, the patient arrivals follow an independent Poisson process. They have a finite number of possible realizations with specific probabilities in each period and can be represented as a scenario tree shown in Figure 5-1. This scenario tree consists of $|N|$ nodes, $|T|$ periods, and $|S|$ scenarios. Each scenario represents a full evolvement of the stochastic process over the entire planning horizon, i.e., a scenario consists of the path from the root node to a leaf node (Huang 2005). If the patient arrivals are independent in each period, then we have $|S| = \prod_{t=1}^{|T|} |S(t)|$, where $|S(t)|$ represents the number of patient arrivals' possible values in period t . Each node n represents a possible state of the stochastic patient arrivals with associated probability p_n in a specific period. The root node represents the initial state of the stochastic process. Node $a(n)$ is the direct ancestor of node n . A node's direct descendants are called its children. The set of all nodes along the path from the root node to the node n is denoted by $P(n)$. We say that $n \in s$ if node n belongs to scenario s . The

notations $t(n)$ and S_t represent the time period of node n and the set of all nodes in period t respectively. The nodes in the tree are usually indexed in the increasing order of the time period (Huang 2005), i.e., $t(1) \leq t(2) \leq \dots \leq t(|N|)$. The root node has an index of 1. The indices of the nodes in the same period are not specified. In addition, the sum of the probabilities of the nodes in each period is equal to 1, i.e., $\sum_{n \in S_t} p_n = 1, \forall t$. A node's probability is equal to the total probability of all its child nodes, i.e., $p_n = \sum_{m \in S_{t(n)+1}, a(m)=n} p_m, \forall n$.

Figure 5-1. A scenario tree

Table 5-2. Summary of the notations in the stochastic model

Notation	Description
Sets:	
N	Set of nodes in the scenario tree, $n \in N$
S	Set of forecast patient arrival scenarios, $s \in S$
Parameters:	
$\beta f_{i,j}$	The fixed regular transshipment cost between clinical site i and j , $i \neq j$
$\beta v_{i,j}$	The variable regular transshipment cost between clinical site i and j , $i \neq j$
$e \alpha f_i$	The fixed emergency delivery cost between the DC and clinical site i
$e \alpha v_i$	The variable emergency delivery cost between the DC and clinical site i
$e \beta f_{i,j}$	The fixed emergency transshipment cost between clinical site i and j , $i \neq j$
$e \beta v_{i,j}$	The variable emergency transshipment cost between clinical site i and j ,

	$i \neq j$
$lb_{i,j}$	The regular delivery lead time between clinical site i and j , $i \neq j$
$d_{n,i}$	The number of forecast new arrival patients at clinical site i at node n
p_n	The probability of node n with $\sum_{n \in S_t} p_n = 1$ for $\forall t \in T$
$d_{n,i}$	The number of forecast new arrival patients at clinical site i at node n
Decision variables:	
f_n	1 if all clinical sites' total number of patients fulfilled for m periods until node n is less than the patient accrual goal, 0 otherwise
$a_{n,i}$	1 if clinical site i orders from the DC at node n using the regular delivery, 0 otherwise
$b_{n,i}$	1 if the DC is replenished from clinical site i at node n using the regular delivery, 0 otherwise
$c_{n,i,j}$	1 if there exists a regular transshipment from clinical site i to j at node n , 0 otherwise
$ea_{n,i}$	1 if clinical site i orders from the DC at node n using the emergency delivery, 0 otherwise
$eb_{n,i}$	1 if the DC is replenished from clinical site i at node n using the emergency delivery, 0 otherwise
$ec_{n,i,j}$	1 if there exists an emergency transshipment from clinical site i to j at node n , 0 otherwise
$x_{n,i}$	The regular replenishment quantity of clinical site i at node n
$y_{n,i}$	The regular reverse replenishment quantity of clinical site i at node n
$z_{n,i,j}$	The regular transshipment quantity from clinical site i to j at node n
$ex_{n,i}$	The emergency replenishment quantity of clinical site i at node n
$ey_{n,i}$	The emergency reverse replenishment quantity of clinical site i at node n
$ez_{n,i,j}$	The emergency transshipment quantity from clinical site i to clinical site j at node n
ID_n	The inventory level of the DC at node n
$IC_{n,i}$	The inventory level of clinical site i at node n
$Sd_{n,k,i}$	The number of patients who first arrive at clinical site i at node n and are fulfilled in the clinical trial for at least k periods, $1 \leq k \leq m$.
θ	All decision variables in the model

With the patient recruitment process specified by a scenario tree, the production planning problem in clinical trial supply chains can be formulated as a multi-stage stochastic programming model. The production quantity is determined at the first stage, and the inventory allocation decisions can be determined in each of the multiple stages. In this model, we use $a^k(n)$ to

represent the k -th ancestor of node n . $a^0(n)$ is the node n itself. $a(1)$ does not exist. A patient who has finished the whole treatment period (consisting of m periods) at node n participated the clinical trial at the node $a^{(m-1)}(n)$ for the first time. With the notations in Table 5-2, the multi-stage stochastic model is presented as follows.

$$\min Sf_1(\theta) = \sum_{n \in N} p_n f_n, \quad (5.16)$$

$$\begin{aligned} \min Sf_2(\theta) = & rQ + \sum_{n \in N, i \in I} p_n (h_i IC_{n,i} + \alpha f_i(a_{n,i} + b_{n,i}) + \alpha v_i(x_{n,i} + \\ & y_{n,i}) + e\alpha f_i(ea_{n,i} + eb_{n,i}) + e\alpha v_i(ex_{n,i} + ey_{n,i}) + \\ & \sum_{j \in I, j \neq i} (\beta f_{i,j} c_{n,i,j} + e\beta f_{i,j} ec_{n,i,j} + \beta v_{i,j} z_{n,i,j} + e\beta v_{i,j} ez_{n,i,j})) \end{aligned}, \quad (5.17)$$

$$\text{s. t. } \sum_{\tau \in P(n), \tau \leq a^{(m-1)}(n)} \sum_i Sd_{\tau,m,i} \geq G(1 - f_n), \quad \forall n \in N \quad (5.18)$$

$$Sd_{n,1,i} \leq d_{n,i}, \quad \forall i \in I, \forall n \in N \quad (5.19)$$

$$Sd_{n,k,i} \leq (1 - \varphi) Sd_{n,(k-1),i}, \forall i \in I, \forall k \in \{2, \dots, m\}, \forall n \in N \quad (5.20)$$

$$\sum_{k=1}^m Sd_{a^{(k-1)}(n),k,i} \leq IC_{n,i}, \quad \forall i \in I, \forall n \in N \quad (5.21)$$

$$ID_0 = Q, \quad (5.22)$$

$$\begin{aligned} ID_n = ID_{(n-1)} - \sum_i (x_{n,i} + ex_{n,i}) + \sum_i (y_{a^{l_i}(n),i} + ey_{n,i}), \\ \forall n \in N \end{aligned} \quad (5.23)$$

$$IC_{n,i} = x_{n,i}, \quad \forall i \in I, n = 1 \quad (5.24)$$

$$\begin{aligned} IC_{n,i} = IC_{a(n),i} - \sum_{k=1}^m Sd_{a^k(n),k,i} + (x_{a^{l_i}(n)\setminus\{1\},i} + ex_{n,i}) - (y_{n,i} + \\ ey_{n,i}) + \sum_{j \in I, j \neq i} (z_{a^{l_{j,i}(n)},j,i} - z_{n,i,j} + ez_{n,j,i} - ez_{n,i,j}), \\ \forall i \in I, \forall n \in N \setminus \{1\} \end{aligned} \quad (5.25)$$

$$x_{n,i} \leq Ma_{n,i}, \quad \forall i \in I, \forall n \in N \quad (5.26)$$

$$ex_{n,i} \leq Mea_{n,i}, \quad \forall i \in I, \forall n \in N \quad (5.27)$$

$$y_{n,i} \leq Mb_{n,i}, \quad \forall i \in I, \forall n \in N \quad (5.28)$$

$$ey_{n,i} \leq Meb_{n,i}, \quad \forall i \in I, \forall n \in N \quad (5.29)$$

$$z_{n,i,j} \leq Mc_{n,i,j}, \quad \forall i, j \in I, i \neq j, \forall n \in N \quad (5.30)$$

$$ez_{n,i,j} \leq Mec_{n,i,j}, \quad \forall i, j \in I, i \neq j, \forall n \in N \quad (5.31)$$

$$ex_{n,i} = 0, y_{n,i} = 0, ey_{n,i} = 0, z_{n,i,j} = 0, ez_{n,i,j} = 0, \\ \forall i, j \in I, i \neq j, n = 1 \quad (5.32)$$

$$x_{n,i}, ex_{n,i}, y_{n,i}, ey_{n,i}, z_{n,i,j}, ez_{n,i,j}, ID_n, IC_{n,i}, Sd_{n,k,i} \in \mathbb{N}, \\ \forall i, j \in I, i \neq j, \forall k \in \{1, \dots, m\}, \forall n \in N \quad (5.33)$$

$$f_n, a_{n,i}, ea_{n,i}, b_{n,i}, eb_{n,i}, c_{n,i,j}, ec_{n,i,j} \in \{0,1\}. \\ \forall i, j \in I, i \neq j, \forall n \in N \quad (5.34)$$

Due to the patient recruitment uncertainty, this stochastic production planning model is constructed based on the scenario tree shown in Figure 5-1. All decision variables and constraints are indexed and presented for each node of the scenario tree. The two objectives (5.16) and (5.17) represent the minimization of the expected clinical trial duration and the total production and expected operational costs at all nodes of the scenario tree. Since the minimal clinical trial duration or cost may not be achieved with only the regular delivery under the stochastic setting, transshipments and emergency deliveries are also applied in the model.

5.3 Numerical Experiments

In this section, we present a numerical study of the production planning problem in clinical trial supply chains. Since the minimal clinical trial duration in each scenario sub-problem can be achieved as long as each patient arrival is fulfilled, it only depends on the patient arrivals and can be simply obtained in (5.14). As a result, only the cost minimization needs to be achieved by the progressive

hedging algorithm (PHA, please see the detailed procedure and discussion in Section 2.3.2.2).

Since the performance of the PHA depends on the parameter ρ , with the parameters and implementation details in Section 5.3.1, the model is solved with different ρ values in Section 5.3.2 to check ρ 's effects on the algorithm's convergence and the solution quality. Since cycling behaviors are detected in the numerical experiments, the cycle detection and avoidance mechanism is applied in this study. The technique of variable fixing (fixing the values of variables those have converged for a number of iterations) proposed in the literature (Watson and Woodruff 2011) is also applied in the numerical experiment to accelerate the convergence. In addition, we solve the problem as a whole by CPLEX and using the PHA respectively for different stage and scenario numbers. The solving efficiency is compared in Section 5.3.3 in terms of the computational time, memory space, and the optimality gap.

Furthermore, by setting the production quantity to a number of possible values, we evaluate the expected total inventory allocation cost and identify the production quantity value with the minimal total cost, where the clinical trial duration is minimized as well. Since the inventory distribution may become less critical and the cost may be reduced when the clinical trial delay is allowed, we also provide the tradeoffs between the clinical trial duration and cost for decision makers in production planning in Section 5.3.4.

5.3.1 Numerical Study

In this study, we consider the production and distribution planning problem in a multi-echelon clinical trial supply chain consisting of a DC and 20 clinical sites. The patient accrual goal is set as 200, i.e., 200 pieces of valid trial data are

required to complete the clinical trial. To get a valid trial data, a qualified patient should take the trial for three continuous periods (weeks). However, 20% of the patients may drop out in each period due to various reasons. The new patient arrivals at clinical sites are assumed to follow Poisson distributions (Chen, et al. 2012), and the arrival rates range from two to five in each period. The planning horizon consists of ten periods (the length requirement in Appendix 2 is satisfied). In addition, the lead times of regular deliveries from the DC to each clinical site and between eligible clinical sites are set as one period respectively. Emergency deliveries are assumed to have negligible lead time. The fixed and variable regular delivery costs are randomly generated from 20 to 60 and 2 to 6 respectively. The emergency delivery is assumed to have a double cost of the regular one in each route.

The program is implemented in C# and CPLEX 12.6 is used as a callable library to solve the mixed-integer problem. This algorithm is executed on a windows server with two Intel® Xeon® E5-2620 v2 2.1GHz processors and 120GB memory.

5.3.2 Effects of ρ on Convergence and Solution Quality

The progressive hedging algorithm's convergence and solution quality are sensitive to the choice of ρ (Fan and Liu 2010, Ryan, et al. 2013, Watson and Woodruff 2011). Poor choices can lead to non-convergence or many iterations' to convergence (Ryan, et al. 2013). Watson and Woodruff (2011) proposed an effective formula to determine the appropriate value of ρ . The formula $\rho(i) = \frac{c(i)}{\max((\sum_{s \in S} p_s |\chi_{s,i}^0 - \bar{\chi}_{s,i}^0|), 1)}$ sets ρ as a value which is proportional to the variable cost of the associated decision variable $\chi_{s,i}$. To check ρ 's effects on the

algorithm's convergence and solution quality, a small size case is studied in this section, where ρ is set as different values around the value defined by the formula. In this numerical study, the stochastic model consists of three stages and nine demand scenarios. Each of the second and third stages has three scenarios and the stochastic demands are independent between stages.

The algorithm's convergence and solution quality should be quantified to check ρ 's effects. The convergence can be described by the total iteration number and the errors of the solutions in all scenarios (Fan and Liu 2010). The solution quality can be represented by the associated expected total cost of all scenarios in this problem. In each iteration of the algorithm, the average solution of all scenarios is a feasible solution of the original multi-stage stochastic problem and its associated total cost works as an upper bound of the problem.

Figure 5-2 and Figure 5-3 illustrate the effects of ρ on the algorithm's convergence and solution quality respectively. When ρ is equal to 10 or 100, the error of the solutions in all scenarios decreases to zero in 17 and 6 iterations respectively. The convergence is achieved in the two cases. On the other hand, when ρ is reduced to 5 or 1, although the error decreases to a small number in only a few iterations, it does not decrease to zero even after quite a number of iterations. As shown in the figures, while fewer iterations are required with the increase of ρ 's value, the solution quality is getting worse (i.e., a greater upper bound is achieved in case of a greater ρ 's value). As a result, with the consideration of the tradeoff between the convergence and solution quality, an intermediate value of ρ should be preferred in the algorithm. This conclusion is consistent to that obtained in the literature (Fan and Liu 2010, Mulvey and Vladimirou 1991). ρ is set as 10 in the PHA in the following sections.

Figure 5-2. The errors of the solutions in all scenarios obtained in PHA

Figure 5-3. The expected total cost of the solution obtained in PHA

5.3.3 The Solving Efficiency Comparison

Although it is claimed that the computational complexity increases significantly with the increase of the stage and scenario numbers in general stochastic programming models, it is not clear if it is true for our problem. In this section, we investigate it by solving the problem as a whole by CPLEX for different stage

and scenario numbers. This solution process is referred to as the “direct solution approach” in this section. The problem is also solved by the PHA for comparison. The initial inventory is set as 850 in these cases.

Table 5-3 compares the solving efficiency between the direct solution approach and the PHA. In the direct solution approach, the solving time and memory required increase dramatically with the increase of the stage and scenario numbers. The optimality gap increases as well. While the problem with only 2 stages and 2 scenarios is easily solved in 50 seconds with negligible memory, the problem with more stages and/or scenarios can only be solved using the direct solution approach with a considerable optimality gap in ten hours. The direct solution approach also requires more than 107GB memory in this case. In other words, it is difficult for the direct solution approach to solve a problem with a large number of stages and scenarios within a reasonable time and the available memory. On the other hand, the PHA gives a better solution in a less solving time within the limited available memory, especially in large cases. Since the solution obtained in PHA is a feasible solution, the corresponding objective value is an upper bound of the problem. As shown in the table, although the PHA gives a larger upper bound in small cases, the objective value obtained in the PHA is much better than that obtained in the direct solution approach in large cases. As a result, in large cases, the feasible solution obtained in the PHA has a better optimality gap than that obtained in the direct solution approach. In addition, the PHA is not constrained by the available computational resources. The PHA decomposes the whole problem into multiple scenario problems. Each scenario problem can be solved independently, the memory required in the PHA is largely reduced. By applying the parallel computing in the algorithm, the computational

time in large cases can be reduced significantly. Even though more iterations and then a longer solving time may be required for larger problems to converge, the PHA can work well with the available server (with the limited memory) to provide a better solution than the direct solution approach. Hence, the progressive hedging algorithm is applied as the solution approach in this paper.

Table 5-3. The solving efficiency comparison

Stage number	Scenario number	Direct solution approach					PHA	
		Lower bound	Upper bound	Optimality gap	Solving time	Memory	Objective value	Solving time
2	2	198.99	211.00	5.69%	50 s	Negligible	504	1h47m
2	3	167.84	184.33	8.95%	~ 47 h	> 107 G	328.67	6h12m
3	4	201.34	277.25	27.38%	~ 17 h	> 90 G	352.75	7h25m
3	9	196.87	383.67	48.69%	~ 9 h	> 64 G	313.22	~ 9 h
4	8	124.3	302	58.84%	~ 21 h	> 107 G	296.38	~ 21 h
4	27	285.08	854.74	66.65%	~ 27 h	> 107 G	379.67	~ 27 h

5.3.4 Production and Distribution Cost

The optimal production quantity can be obtained by balancing the production cost against the expected total inventory allocation cost in clinical trial supply chains. The inventory allocation decisions and associated cost may vary greatly when the production quantity has different values. In this experiment, the production quantity is set to a number of values no less than Q_m , which is the minimal production quantity value required to complete the clinical trial in all scenarios. γ is the ratio of the production quantity value to Q_m . The unit production cost is set to 3 in this numerical study. Since the inventory allocation decisions also depend on the clinical trial duration, the experiment is conducted in two cases where the delay is allowed or not in clinical trials. The stochastic model consists of four stages and 27 demand scenarios, where each of the stages

later than the first has three scenarios and the stochastic demands are independent between stages.

Figure 5-4 and Figure 5-5 demonstrate the inventory distribution cost and the total production and distribution cost obtained in supply chains with different production quantity values, where the minimal clinical trial duration is achieved in each case. The minimal production quantity Q_m is 830 in this numerical study. As shown in Figure 5-4, the expected total inventory distribution cost decreases monotonically with the increase of the production quantity. Since the total production cost increases linearly with the production quantity, it can be balanced with the total inventory distribution cost to get the minimal total cost. Figure 5-5 shows the total minimal production and inventory distribution cost under different production quantity values. With the tradeoff between the production cost and inventory distribution cost, the production quantity with the minimal total cost can be identified in this figure. The optimal γ in this figure is equal to 1.05, the corresponding optimal production quantity is 871 in this numerical study.

Figure 5-4. The total inventory distribution cost

Figure 5-5. The production and distribution cost

When the clinical trial duration is not too crucial, prolonging it may save the production and distribution costs. The inventory distribution cost is evaluated under different values of the allowed clinical trial delay and production quantity. With the obtained costs in Table 5-4, we show the cost comparisons in the following figures. Figure 5-6 shows that the total expected inventory distribution cost decreases with the increase of the production quantity for each clinical trial delay value. As shown in Figure 5-7, it also decreases with the increase of the allowed clinical trial delay for each production quantity. As a result, we have the total production and distribution cost comparison in Figure 5-8 and Figure 5-9. Due to the decreasing inventory distribution cost and increasing production cost, the total cost is convex to the production quantity for each delay value in Figure 5-8. It also decreases monotonously with the increase of the clinical trial delay for each production quantity value in Figure 5-9.

Table 5-4. The production and inventory distribution cost

γ	Q	Production cost	Inventory distribution cost			Production and distribution cost		
			Delay = 0	Delay = 1	Delay = 2	Delay = 0	Delay = 1	Delay = 2
1	830	2490	580.5	180.5	65.75	3070.5	2670.5	2555.75
1.025	850	2550	292.375	8.5	0	2842.375	2558.5	2550
1.05	871	2613	157.75	6.5	0	2770.75	2619.5	2613
1.1	913	2739	99.125	0	0	2838.125	2739	2739
1.2	996	2988	72.25	0	0	3060.25	2988	2988

Figure 5-6. The distribution cost for each delay value

Figure 5-7. The distribution cost for each γ value

Figure 5-8. The total cost for each delay value

Figure 5-9. The total cost for each γ value

Figure 5-10 shows the tradeoff between the minimal total production and distribution costs and the allowed delay in clinical trials. As shown in the figure, the total cost decreases with the increase of the allowed clinical trial delay. When the allowed clinical trial delay is increased from 0 to 1, the total cost is reduced by about 7.67% and the production quantity with the minimal total cost also drops from 871 to 850. On the contrary, when the delay is increased from 1 to 2, the total cost is reduced by only 0.33% and the optimal production quantity keeps unchanged. This time-cost tradeoff in this figure can provide some guidance for decision makers to determine the production quantity. For example, in cases where the clinical trial duration is less critical than the cost (e.g. the budget is very tight in a pharmaceutical company), the production quantity and the total cost can be reduced with the sacrifice of a short delay in clinical trials.

Figure 5-10. The tradeoff between the total cost and clinical trial delay

5.4 Conclusion

This chapter considers the production planning problem in clinical trial supply chains. New drug development becomes increasingly complicated that the production cost of drugs can be significant nowadays. Since leftover clinical drugs can only be disposed of after an unsuccessful trial, overproduction should be minimized in clinical trials. When the production quantity is reduced, it becomes more challenging to distribute the inventory among clinical sites. An inappropriate inventory distribution may increase the operational cost or even the clinical trial duration. Since the clinical trial delay can be costly, the production quantity should be optimized such that the production and inventory allocation costs are minimized without increasing the clinical trial duration.

The production planning problem in clinical trial supply chains has been studied in the literature. Although some studies consider the balance between the production cost and the inventory allocation cost, they commonly assume a stationary inventory policy at each location in the supply chain. Since the patient

arrivals in clinical trial supply chains follow non-stationary distributions (Chen, et al. 2012), the solution can be far from optimal ones if stationary inventory policies are applied in the supply chains. Some researchers have also studied the production planning problem considering the failure rates of clinical trials. However, the clinical trial duration is considered in none of these researches. In addition, the widely discussed production and distribution problem in other applications has different problem settings from ours. While those studies neglect overage and mainly consider stock-out minimization, our problem minimizes the overproduction and only considers the stock-out minimization in terms of the clinical trial duration minimization.

In this chapter, we find the optimal production quantity in clinical trial supply chains through a multi-objective multi-stage stochastic programming model, where the clinical trial duration and total production and distribution costs are minimized sequentially. Due to the significantly increasing computational complexity in multi-stage stochastic models, the progressive hedging algorithm is applied as the solution approach. As discussed in the numerical experiments, the algorithm's convergence and solution quality depend on the parameter ρ significantly. In addition, the inventory distribution cost decreases monotonously with the increase of the production quantity.

By balancing it with the increasing production cost, we can find the production quantity with the minimal total cost in clinical trial supply chains. The tradeoffs between the total cost and clinical trial delay are provided as a guidance for decision makers to determine the production quantity, especially when the clinical trial cost is too high to afford for pharmaceutical companies. In this situation, the cost can be reduced significantly with the sacrifice of only a few

periods' clinical trial delay.

Appendix

1. The explanation for the length requirement of the planning horizon.

In the production and distribution problem, the planning horizon should be long enough to guarantee the patient accrual goal can be achieved in the planning horizon with the probability 100% under the settings of the stochastic patient recruitment process. We have the following equation,

$$P(X \geq G) = 100\%,$$

where G represents the patient accrual goal, X is a random number and represents the total number of patients who have finished the entire treatment until the end of the planning horizon. Due to the dropout rate φ in each period during the treatment period (consisting of m periods), X follows a Poisson distribution and has a mean $\sum_{t=1}^{l-m+1} \sum_i \lambda_{t,i} (1 - \varphi)^{m-1}$, where l represents the number of periods in the planning horizon and $\lambda_{t,i}$ represents the patient arrival rate at clinical site i in period t .

2. The explanation for defining the sufficiently large number $M = G \sum_{k=1}^m (1 - \varphi)^{k-1} / (1 - \varphi)^{m-1}$.

The defined value for the sufficiently large number M is actually the minimal production quantity of clinical drugs. For each clinical trial, the production quantity should be large enough to guarantee the clinical trial's patient accrual goal is reached. Due to the dropout rate φ in each period during the treatment period (consisting of m periods), only $(1 - \varphi)^{m-1}$ of the recruited patients would finish the entire treatment and deliver valid trial data. Hence, the number of recruited patients should be no less than $G / (1 - \varphi)^{m-1}$. In addition, $(1 - \varphi)^{k-1}$ of them would attend the clinical trial for at least k periods, $k \in$

$\{1, \dots, m\}$. As a result, to reach the patient accrual goal G , the production quantity should be at least $\sum_{k=1}^m (1 - \varphi)^{k-1}$ times of the number of patients required, i.e.,

$$Q \geq G \sum_{k=1}^m (1 - \varphi)^{k-1} / (1 - \varphi)^{m-1}.$$

Chapter 6 Conclusion and Future Work

This chapter summarizes the conclusions and contributions of this research. With the limitations in this research, some potential research topics are also proposed correspondingly to provide perspectives for future work.

6.1 Conclusion

The clinical trial plays a critical role in new drug development due to its long duration and high cost. On average, it takes 37% of the total expense and more than 70% of the time in the R&D phase. The proportions are still increasing due to the increasing complexity of the drug development nowadays. Since the clinical trial duration is counted in the patent life, its increase will reduce the specific clinical drug's commercialization time under patent protection. As a result, the profit margin may decline. Under the challenging circumstances, the pharmaceutical companies strive to increase the profit through reducing the total cost and duration in clinical trials. To improve the efficiency of clinical trial supply chains, we consider three problems in this study, including the inventory allocation, network design, and production planning problems.

With the background of clinical trials and clinical trial supply chains in Chapter 1, a detailed literature review of relevant studies is given in Chapter 2. After an overview of the researches in pharmaceutical supply chains, we make a discussion of the studies for each of three research topics in clinical trial supply chains. Due to the patient recruitment uncertainty in clinical trial supply chains, the optimization approaches for problems under uncertainty are presented and their applicability for this research is discussed respectively. The modeling approaches and corresponding algorithms are also presented for stochastic

programming models, which are applied as the optimization approach in this research. With the consideration of both the cost and duration in clinical trials, the algorithms for multi-objective models are discussed as well. In addition, with the analysis of the characteristics of clinical trial supply chains, the research gaps are identified between the literature and this study.

The inventory allocation problem in clinical trial supply chains is studied in Chapter 3. It considers the efficient coordination of clinical drugs among the distribution centers and clinical sites. Due to the specific characteristics of clinical trial supply chains (e.g. the non-stationary stochastic demand, the limited inventory and transshipment options), the inventory policy is analytically intractable and multiple delivery routes (the replenishment, reverse replenishment and transshipment) are applicable. When reverse replenishments and transshipments reduce the stock-outs, they inevitably increase the delivery cost. To avoid them, the inventory should be held at the DC and the DC should deliver more frequently with small volumes. As a result, the total delivery cost will be increased as well. Hence, in the inventory allocation problem, the reverse replenishment and transshipment should be applied jointly with the direct replenishment to optimize the system performance.

In the study of the inventory allocation problem, the contributions are three-fold. First, this problem is formulated as a rolling-based two-stage stochastic programming model, which minimizes the total delivery, holding and stock-out costs in clinical trial supply chains. The replenishment, the reverse replenishment and transshipment are applied jointly to find the optimal inventory allocation decisions in the supply chain. Second, two acceleration techniques are proposed to improve the solving efficiency, where Benders' decomposition is applied as

the solution approach. As shown in the numerical results, they can improve the solving efficiency of Benders decomposition in terms of the required number of iterations and total computational time under different settings. Thirdly, the total cost is compared between our model and the (s, S) policy and numerical results show that our model gives a lower cost than the (s, S) policy in most cases. Since the (s, S) policy is proved optimal for non-stationary stochastic demand, it is applied in an identical supply chain to compare with our model. Transshipments are also applied for the (s, S) policy in case of stock-outs at the DC.

In Chapter 4, we present a study of the network design problem in clinical trial supply chains. In this chapter, we identify the efficient supply chain network configurations. When the clinical trial duration reduction is achieved through the scale increase of clinical trial supply chains, the total cost is increased simultaneously. Hence, the time and cost become a tradeoff in clinical trials. To conduct a clinical trial efficiently, the clinical sites should be selected with the consideration of the tradeoff between the time and cost required by the clinical trial.

We have three contributions in the study of this network design problem. First, we formulate this problem as a multi-objective model which minimizes the clinical trial duration and cost. With the introduction of a parameter named delay cost, the model is converted into a single objective one using the weighted sum method. Second, we develop some propositions, optimality cuts and an algorithm to efficiently identify all Pareto-optimal supply chain configurations. A Pareto-optimal configuration can be obtained by evaluating a single value of the delay cost. Since the value of delay cost can range from zero to the infinity, it is computational infeasible to evaluate all possible values. Thirdly, an efficiency

curve showing the Pareto-optimal tradeoffs is provided for decision makers to support their supply chain design. For example, if a drug's estimated potential delay cost is not so high but the budget in this pharmaceutical company is a bit tight, it is not worth large expenses to enroll numerous clinical sites. However, sometimes the drug's delay cost can be very high, especially when the drug is required urgently to conquer certain intractable diseases seriously threatening people's lives, e.g. the drugs for cancer and AIDS. Even if tremendous cost is required, the time should be reduced as much as possible.

Chapter 5 presents the study of the production planning problem in clinical trial supply chains. Overproduction reduction of clinical drugs is considered in this problem. Due to the increasing complexity of drug development, the production cost of drugs can be significant nowadays. Since leftover clinical drugs should be disposed of after an unsuccessful clinical trial, pharmaceutical companies strive to reduce the overproduction. The stock-outs and clinical trial duration increase may hardly be avoided if the production quantity is simply reduced without transshipments to efficiently coordinate the inventories among clinical sites. Hence, the inventory distribution is considered jointly with the production planning in clinical trial supply chains.

The contributions of the study of this production planning problem are three-fold. First, this problem is formulated as a multi-objective multi-stage stochastic programming model, where the clinical trial duration minimization has a higher priority than the cost minimization. Second, due to the computational complexity of multi-stage stochastic models, the progressive hedging algorithm is applied as the solution approach. Its performance is studied in terms of the solving time and required memory resources in numerical experiments. Thirdly, numerical

experiments are also conducted to find the optimal production quantity by balancing the inventory allocation cost against the production cost in clinical trials. The tradeoffs between the clinical trial duration and cost are also provided for decision makers to select the appropriate production quantity, in situations that the production cost is so expensive that the pharmaceutical company would rather save the cost with the sacrifice the shortest clinical trial duration.

6.2 Future Work

In this study, we consider the inventory allocation, network design, and production planning problems in a two-echelon clinical trial supply chain respectively. In the future, we can extend this work in the following topics.

A direction for future research is to extend the network design problem by considering the standard deviations of the time and/or cost among all scenarios as supply chain configurations' robustness measures. Since currently we only consider the cost and time's expectations of all scenarios, their values may differ greatly in each scenario. Efficient and robust supply chain configurations can be identified through the joint consideration of the expectations and standard deviations.

The network design problem can also be extended to one which optimizes the supply chain structure and production quantity simultaneously. According to the discussion of Croxton and Zinn (2005), the number of clinical sites in a supply chain affects the optimal production quantity due to the square root law of inventory. Hence, the production quantity can be determined together with the supply chain configurations in clinical trial supply chains. Since the inventory allocation decisions are determined based on both the supply chain structure and production quantity, the inventory allocation, network design, and production

planning decisions can be determined jointly in a single model. This problem is indeed an integration of the network design and production planning problems in this thesis.

Furthermore, when facing a larger supply chain, more efficient solution approaches are to be developed. Since clinical trial supply chains are multi-echelon, they can be extended from two-echelon to three-echelon ones. When we select participating clinical sites from the candidate ones, regional DCs are selected from the participating clinical sites. In addition, more efficient solution approaches are to be developed, especially in larger supply chains. Although the solving efficiency can be largely improved by applying the progressive hedging algorithm in stochastic programming models, solving mixed-integer quadratic scenario sub-problems may still require significant computational resources (e.g. CPU and memory), especially for those multi-stage stochastic models with a large number of scenarios.

Publications

1. Hui Zhao, Kan Wu, and Edward Huang. Clinical trial supply chain design based on the Pareto-optimal trade-off between time and cost. *IIE Transactions* 50.6 (2018): 512-524.
<https://www.tandfonline.com/doi/abs/10.1080/24725854.2017.1395978>
2. Hui Zhao, Edward Huang, Runliang Dou, and Kan Wu. A multi-objective production planning problem with the consideration of time and cost in clinical trials. *Expert Systems with Applications*. In press. Available online 11 Jan 2019.
<https://www.sciencedirect.com/science/article/pii/S0957417419300387?via%3Dihub>

References

- Abdelkafi C, Beck BH, David B, Druck C, Horoho M (2009) Balancing risk and costs to optimize the clinical supply chain—a step beyond simulation. *Journal of Pharmaceutical Innovation*. **4**(3) 96-106.
- Agrawal V, Chao X, Seshadri S (2004) Dynamic balancing of inventory in supply chains. *European Journal of Operational Research*. **159**(2) 296-317.
- Ahmed S (2010) Two - Stage Stochastic Integer Programming: A Brief Introduction. *Wiley Encyclopedia of Operations Research and Management Science*.
- Ahmed S, King AJ, Parija G (2003) A multi-stage stochastic integer programming approach for capacity expansion under uncertainty. *Journal of Global Optimization*. **26**(1) 3-24.
- Alguacil N, Conejo A (2000) Multiperiod optimal power flow using Benders decomposition. *IEEE Transactions on power systems*. **15**(1) 196-201.
- Alonso-Ayuso A, Escudero LF, Garín A, Ortuño MT, Pérez G (2003) An approach for strategic supply chain planning under uncertainty based on stochastic 0-1 programming. *Journal of Global Optimization*. **26**(1) 97-124.
- Alshawhi S, Saez-Pujol I, Irani Z (2003) Data warehousing in decision support for pharmaceutical R&D supply chain. *International Journal of Information Management*. **23**(3) 259-268.
- Altıparmak F, Gen M, Lin L, Paksoy T (2006) A genetic algorithm approach for multi-objective optimization of supply chain networks. *Computers & Industrial Engineering*. **51**(1) 196-215.

References

- Archibald TW, Black D, Glazebrook KD (2009) An index heuristic for transshipment decisions in multi-location inventory systems based on a pairwise decomposition. *European Journal of Operational Research*. **192**(1) 69-78.
- Archibald TW, Black D, Glazebrook KD (2010) The use of simple calibrations of individual locations in making transshipment decisions in a multi-location inventory network. *Journal of the Operational Research Society* 294-305.
- Arntzen BC, Brown GG, Harrison TP, Trafton LL (1995) Global supply chain management at Digital Equipment Corporation. *Interfaces*. **25**(1) 69-93.
- Axsäter S (2000) Exact analysis of continuous review (R, Q) policies in two-echelon inventory systems with compound Poisson demand. *Operations research*. **48**(5) 686-696.
- Axsäter S (2007). *Inventory control*. Springer Science & Business Media.
- Badri H, Ghomi SF, Hejazi T-H (2017) A two-stage stochastic programming approach for value-based closed-loop supply chain network design. *Transportation Research Part E: Logistics and Transportation Review*. **105** 1-17.
- Badri MA, Mortagy AK, Alsayed CA (1998) A multi-objective model for locating fire stations. *European Journal of Operational Research*. **110**(2) 243-260.
- Beamon BM (1998) Supply chain design and analysis:: Models and methods. *International journal of production economics*. **55**(3) 281-294.

- Belfares L, Klibi W, Lo N, Guitouni A (2007) Multi-objectives Tabu Search based algorithm for progressive resource allocation. *European Journal of Operational Research*. **177**(3) 1779-1799.
- Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. *Numerische mathematik*. **4**(1) 238-252.
- Bertsimas D, Farias VF, Trichakis N (2011) The price of fairness. *Operations research*. **59**(1) 17-31.
- Birge JR (1985) Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research*. **33**(5) 989-1007.
- Birge JR (1997) State-of-the-art-survey—Stochastic programming: Computation and applications. *INFORMS journal on computing*. **9**(2) 111-133.
- Birge JR, Louveaux F (2011). *Introduction to stochastic programming*. Springer Science & Business Media.
- Bollapragada S, Morton TE (1999) A simple heuristic for computing nonstationary (s, S) policies. *Operations Research*. **47**(4) 576-584.
- Boudia M, Dauzere-Peres S, Prins C, Louly MAO (2006) Integrated optimization of production and distribution for several products, in Proceedings of the 2006 International Conference on Service Systems and Service Management, pp. 272-277.
- Boudia M, Louly MAO, Prins C (2008) Fast heuristics for a combined production planning and vehicle routing problem. *Production Planning and Control*. **19**(2) 85-96.
- Buzby G, Knox L, Crosby L, Eisenberg J, Haakenson C, McNeal G, Page C, Peterson O, Reinhardt G, Williford W (1988) Study protocol: a randomized

References

- clinical trial of total parenteral nutrition in malnourished surgical patients. *The American journal of clinical nutrition*. **47**(2) 366-381.
- Byrom B (2002) Using IVRS in clinical trial management. *Applied Clinical Trials*. **11**(10) 36-43.
- Campbell GM (2011) A two-stage stochastic program for scheduling and allocating cross-trained workers. *Journal of the Operational Research Society*. **62**(6) 1038-1047.
- Caro F, Gallien J (2010) Inventory management of a fast-fashion retail network. *Operations Research*. **58**(2) 257-273.
- Carøe CC, Tind J (1998) L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming*. **83**(1-3) 451-464.
- Chan FT, Chung SH (2004) A multi-criterion genetic algorithm for order distribution in a demand driven supply chain. *International Journal of Computer Integrated Manufacturing*. **17**(4) 339-351.
- Chandra P, Fisher ML (1994) Coordination of production and distribution planning. *European Journal of Operational Research*. **72**(3) 503-517.
- Chen A, Kim J, Lee S, Kim Y (2010) Stochastic multi-objective models for network design problem. *Expert Systems with Applications*. **37**(2) 1608-1619.
- Chen Y, Mockus L, Orcun S, Reklaitis GV (2012) Simulation-optimization approach to clinical trial supply chain management with demand scenario forecast. *Computers & Chemical Engineering*. **40** 82-96.

- Chen Y, Pekny JF, Reklaitis GV (2012) Integrated planning and optimization of clinical trial supply chain system with risk pooling. *Industrial & Engineering Chemistry Research*. **52**(1) 152-165.
- Cheng SK (2008). *The impact of delay: Assessing the early indicators of development time and accrual milestones on oncology clinical trial success*. Vanderbilt University.
- Cheung RK, Chen C-Y (1998) A two-stage stochastic network model and solution methods for the dynamic empty container allocation problem. *Transportation science*. **32**(2) 142-162.
- Chien TW, Balakrishnan A, Wong RT (1989) An integrated inventory allocation and vehicle routing problem. *Transportation Science*. **23**(2) 67-76.
- Choi TM, Chiu CH, To KMC (2011) A fast fashion safety-first inventory model. *Textile Research Journal*. **81**(8) 819-826.
- Clark AJ, Scarf H (1960) Optimal policies for a multi-echelon inventory problem. *Management science*. **6**(4) 475-490.
- Clemento A (1999) New and integrated approaches to successful accelerated drug development. *Drug information journal*. **33**(3) 699-710.
- Coello CAC (1999) A comprehensive survey of evolutionary-based multiobjective optimization techniques. *Knowledge and Information systems*. **1**(3) 269-308.
- Coello CAC, Lamont GB, Van Veldhuizen DA (2007). *Evolutionary algorithms for solving multi-objective problems*. Springer.

References

- Colvin M, Maravelias CT (2008) A stochastic programming approach for clinical trial planning in new drug development. *Computers & Chemical Engineering*. **32**(11) 2626-2642.
- Colvin M, Maravelias CT (2009) Scheduling of testing tasks and resource planning in new product development using stochastic programming. *Computers & Chemical Engineering*. **33**(5) 964-976.
- Colvin M, Maravelias CT (2010) Modeling methods and a branch and cut algorithm for pharmaceutical clinical trial planning using stochastic programming. *European Journal of Operational Research*. **203**(1) 205-215.
- Cordeau J-F, Stojković G, Soumis F, Desrosiers J (2001) Benders decomposition for simultaneous aircraft routing and crew scheduling. *Transportation science*. **35**(4) 375-388.
- Correa JR (2007). *Optimization of a fast-response distribution network*. Massachusetts Institute of Technology, City.
- Costa AM (2005) A survey on benders decomposition applied to fixed-charge network design problems. *Computers & operations research*. **32**(6) 1429-1450.
- Croxton KL, Zinn W (2005) Inventory considerations in network design. *Journal of Business Logistics*. **26**(1) 149-168.
- Csirik J, Frenk JBG, Labbé M, Zhang S (1990). *Heuristics for the 0-1 min-knapsack problem*. European Institute for Advanced Studies in Management Erasmus Universiteit Rotterdam.
- Cusworth M (2017). *Surprise package: How packaging can improve patient retention in clinical trials*. City.

- da Graça Costa M, Captivo ME, Clímaco J (2008) Capacitated single allocation hub location problem—A bi-criteria approach. *Computers & Operations Research*. **35**(11) 3671-3695.
- Deb K (1999) Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary computation*. **7**(3) 205-230.
- Deb K (2001). *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons.
- DiMasi JA, Hansen RW, Grabowski HG (2003) The price of innovation: new estimates of drug development costs. *J Health Econ*. **22**(2) 151-185.
- Dong L, Rudi N (2004) Who benefits from transshipment? Exogenous vs. endogenous wholesale prices. *Management Science*. **50**(5) 645-657.
- Dowlman N, Lang M, McEntegart D, Nicholls G, Bacon S, Star J, Byrom B (2004) Optimizing the Supply Chain Through Simulation. *Applied Clinical Trials*. **13** 40-47.
- Drennan KB (2002) Patient recruitment: the costly and growing bottleneck in drug development. *Drug discovery today*. **7**(3) 167-170.
- Erengüç ŞS, Simpson NC, Vakharia AJ (1999) Integrated production/distribution planning in supply chains: An invited review. *European journal of operational research*. **115**(2) 219-236.
- Fan Y, Liu C (2010) Solving stochastic transportation network protection problems using the progressive hedging-based method. *Networks and Spatial Economics*. **10**(2) 193-208.

References

- Farahani RZ, Elahipanah M (2008) A genetic algorithm to optimize the total cost and service level for just-in-time distribution in a supply chain. *International Journal of Production Economics*. **111**(2) 229-243.
- Farahani RZ, Rezapour S, Drezner T, Fallah S (2014) Competitive supply chain network design: An overview of classifications, models, solution techniques and applications. *Omega*. **45** 92-118.
- Farahani RZ, SteadieSeifi M, Asgari N (2010) Multiple criteria facility location problems: A survey. *Applied Mathematical Modelling*. **34**(7) 1689-1709.
- Federgruen A, Tzur M (1999) Time-partitioning heuristics: application to one warehouse, multiitem, multiretailer lot-sizing problems. *Naval Research Logistics*. **46**(5) 463-486.
- Federgruen A, Zipkin P (1984) Approximations of dynamic, multilocation production and inventory problems. *Management Science*. **30**(1) 69-84.
- Federgruen A, Zipkin P (1984) A combined vehicle routing and inventory allocation problem. *Operations Research*. **32**(5) 1019-1037.
- Fleischhacker A (2009). *An investigation of clinical trial supply chains*. Rutgers University-Graduate School-Newark, City.
- Fleischhacker A, Ninh A, Zhao Y (2015) Positioning Inventory in Clinical Trial Supply Chains. *Production and Operations Management*. **24**(6) 991-1011.
- Fleischhacker AJ, Zhao Y (2011) Planning for demand failure: A dynamic lot size model for clinical trial supply chains. *European Journal of Operational Research*. **211**(3) 496-506.
- Fleischhacker AJ, Zhao Y (2013) Balancing learning and economies of scale for adaptive clinical trials. *Operations Research for Health Care*. **2**(3) 42-51.

- Fumero F, Vercellis C (1999) Synchronized development of production, inventory, and distribution schedules. *Transportation science*. **33**(3) 330-340.
- Gade D, Hackebeil G, Ryan SM, Watson J-P, Wets RJ-B, Woodruff DL (2016) Obtaining lower bounds from the progressive hedging algorithm for stochastic mixed-integer programs. *Mathematical Programming*. **157**(1) 47-67.
- Garcia DJ, You F (2015) Supply chain design and optimization: Challenges and opportunities. *Computers & Chemical Engineering*. **81** 153-170.
- Gatica G, Papageorgiou L, Shah N (2003) Capacity planning under uncertainty for the pharmaceutical industry. *Chemical Engineering Research and Design*. **81**(6) 665-678.
- Gebreslassie BH, Yao Y, You F (2012) Design under uncertainty of hydrocarbon biorefinery supply chains: multiobjective stochastic programming models, decomposition algorithm, and a comparison between CVaR and downside risk. *AIChE Journal*. **58**(7) 2155-2179.
- Geoffrion AM (1970) Elements of large-scale mathematical programming Part I: Concepts. *Management Science*. **16**(11) 652-675.
- Geoffrion AM (1970) Elements of large scale mathematical programming part II: Synthesis of algorithms and bibliography. *Management Science*. **16**(11) 676-691.
- Geoffrion AM, Graves GW (1974) Multicommodity distribution system design by Benders decomposition. *Management science*. **20**(5) 822-844.

References

- Getz KA, De Bruin A (2000) Breaking the development speed barrier: Assessing successful practices of the fastest drug development companies. *Drug Information Journal*. **34**(3) 725-736.
- Goh M, Lim JY, Meng F (2007) A stochastic model for risk management in global supply chain networks. *European Journal of Operational Research*. **182**(1) 164-173.
- Gosavi A (2003) Simulation-based optimization. *parametric optimization techniques and reinforcement learning*.
- Grabowski HG, Vernon JM (2000) Effective patent life in pharmaceuticals. *International Journal of Technology Management*. **19**(1-2) 98-120.
- Graves SC (1996) A multiechelon inventory model with fixed replenishment intervals. *Management Science*. **42**(1) 1-18.
- Graves SC (1999) A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management*. **1**(1) 50-61.
- Graves SC, Willems SP (2008) Strategic inventory placement in supply chains: Nonstationary demand. *Manufacturing & Service Operations Management*. **10**(2) 278-287.
- Grossman M, Bates S (2008) Knowledge capture within the biopharmaceutical clinical trials environment. *VINE*. **38**(1) 118-132.
- Gupta A, Maranas CD (2000) A two-stage modeling and solution framework for multisite midterm planning under demand uncertainty. *Industrial & Engineering Chemistry Research*. **39**(10) 3799-3813.
- Gupta A, Maranas CD (2003) Managing demand uncertainty in supply chain planning. *Computers & Chemical Engineering*. **27**(8) 1219-1227.

- Gurnani H, Tang CS (1999) Note: Optimal ordering decisions with uncertain cost and demand forecast updating. *Management science*. **45**(10) 1456-1462.
- Hadjiconstantinou E, Klerides E (2010) A new path-based cutting plane approach for the discrete time-cost tradeoff problem. *Computational Management Science*. **7**(3) 313-336.
- Haimes YY, Ladson L, Wismer DA (1971). *Bicriterion formulation of problems of integrated system identification and system optimization*. IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC 345 E 47TH ST, NEW YORK, NY 10017-2394, City.
- Herer YT, Rashit A (1999) Lateral stock transshipments in a two - location inventory system with fixed and joint replenishment costs. *Naval Research Logistics (NRL)*. **46**(5) 525-547.
- Herer YT, Tzur M (2001) The dynamic transshipment problem. *Naval Research Logistics (NRL)*. **48**(5) 386-408.
- Herer YT, Tzur M (2003) Optimal and heuristic algorithms for the multi-location dynamic transshipment problem with fixed transshipment costs. *IIE Transactions*. **35**(5) 419-432.
- Huang K (2005). *Multi-stage stochastic programming models in production planning*. Thesis (PhD). School of Industrial and Systems Engineering: Georgia Institute of Technology.
- Huang K, Ahmed S (2009) The value of multistage stochastic programming in capacity planning under uncertainty. *Operations Research*. **57**(4) 893-904.

References

- Iida T (2001) The infinite horizon non-stationary stochastic multi-echelon inventory problem and near-myopic policies. *European Journal of Operational Research*. **134**(3) 525-539.
- Jain V, Grossmann IE (1999) Resource-constrained scheduling of tests in new product development. *Industrial & Engineering Chemistry Research*. **38**(8) 3013-3026.
- Jang Y-J, Jang S-Y, Chang B-M, Park J (2002) A combined model of network design and production/distribution planning for a supply network. *Computers & Industrial Engineering*. **43**(1) 263-281.
- Jeunet J (2006) Demand forecast accuracy and performance of inventory policies under multi-level rolling schedule environments. *International Journal of Production Economics*. **103**(1) 401-419.
- Jiang R, Guan Y (2016) Data-driven chance constrained stochastic program. *Mathematical Programming*. **158**(1-2) 291-327.
- Jones DF, Mirrazavi SK, Tamiz M (2002) Multi-objective meta-heuristics: An overview of the current state-of-the-art. *European journal of operational research*. **137**(1) 1-9.
- Kaitin KI, DiMasi JA (2011) Pharmaceutical innovation in the 21st century: new drug approvals in the first decade, 2000–2009. *Clinical pharmacology & therapeutics*. **89**(2) 183-188.
- Karimi B, Ghomi SF, Wilson J (2003) The capacitated lot sizing problem: a review of models and algorithms. *Omega*. **31**(5) 365-378.
- Kazemi Zanjani M, Noureldath M, Ait-Kadi D (2010) A multi-stage stochastic programming approach for production planning with uncertainty in the

- quality of raw materials and demand. *International Journal of Production Research*. **48**(16) 4701-4723.
- Keskin BB, Melouk SH, Meyer IL (2010) A simulation-optimization approach for integrated sourcing and inventory decisions. *Computers & Operations Research*. **37**(9) 1648-1661.
- Khanna I (2012) Drug discovery in pharmaceutical industry: productivity challenges and trends. *Drug discovery today*. **17**(19-20) 1088-1102.
- Kim G, Wu C-H (2013) Scenario aggregation for supply chain quantity-flexibility contract. *International Journal of Systems Science*. **44**(11) 2166-2182.
- Kimko H, Duffull SB (2002). *Simulation for designing clinical trials: a pharmacokinetic-pharmacodynamic modeling perspective*. CRC Press.
- King AJ, Wallace SW (2012). *Modeling with stochastic programming*. Springer Science & Business Media.
- Kirkpatrick S (1984) Optimization by simulated annealing: Quantitative studies. *Journal of statistical physics*. **34**(5-6) 975-986.
- Klein Haneveld WK, van der Vlerk MH (1999) Stochastic integer programming: General models and algorithms. *Annals of Operations Research*. **85** 39-57.
- Klerides E, Hadjiconstantinou E (2015) The Stochastic Discrete Time-Cost Tradeoff Problem with Decision-Dependent Uncertainty, in Handbook on Project Management and Scheduling Vol. 2, Springer, pp. 781-809.
- Kleywegt AJ, Shapiro A, Homem-de-Mello T (2002) The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*. **12**(2) 479-502.

References

- Koenig S, Holte RC (2003). *Abstraction, Reformulation, and Approximation: 5th International Symposium, SARA 2002, Kananaskis, Alberta, Canada, August 2-4, 2002, Proceedings*. Springer.
- Konak A, Coit DW, Smith AE (2006) Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*. **91**(9) 992-1007.
- Körpeoğlu E, Yaman H, Aktürk MS (2011) A multi-stage stochastic programming approach in master production scheduling. *European Journal of Operational Research*. **213**(1) 166-179.
- Kukreja A, Schmidt CP (2005) A model for lumpy demand parts in a multi-location inventory system with transshipments. *Computers & operations research*. **32**(8) 2059-2075.
- Láinez JM, Schaefer E, Reklaitis GV (2012) Challenges and opportunities in enterprise-wide optimization in the pharmaceutical industry. *Computers & Chemical Engineering*. **47** 19-28.
- Laporte G, Louveaux FV (1993) The integer L-shaped method for stochastic integer programs with complete recourse. *Operations research letters*. **13**(3) 133-142.
- Lee YH, Kim SH (2002) Production–distribution planning in supply chain considering capacity constraints. *Computers & industrial engineering*. **43**(1) 169-190.
- Li H, Wang L, Wei L, Quan H (2014) Sample size calculation for count data in comparative clinical trials with non-uniform patient accrual and early dropout. *Journal of biopharmaceutical statistics*(just-accepted).

- Li X, Wang Q (2007) Coordination mechanisms of supply chain systems. *European journal of operational research*. **179**(1) 1-16.
- Li Y (2008). *Decision making under uncertainty in power system using Benders decomposition*. Thesis (PhD). Iowa State University.
- Li Y, Huang G, Nie S (2006) An interval-parameter multi-stage stochastic programming model for water resources management under uncertainty. *Advances in Water Resources*. **29**(5) 776-789.
- Liberti L (2004). *Reformulation and convex relaxation techniques for global optimization*. Springer, City.
- Løkketangen A, Woodruff DL (1996) Progressive hedging and tabu search applied to mixed integer (0, 1) multistage stochastic programming. *Journal of Heuristics*. **2**(2) 111-128.
- Magnanti TL, Wong RT (1981) Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations research*. **29**(3) 464-484.
- Mak W-K, Morton DP, Wood RK (1999) Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*. **24**(1) 47-56.
- Mantrala MK, Rao S (2001) A decision-support system that helps retailers decide order quantities and markdowns for fashion goods. *Interfaces*. **31**(3_supplement) S146-S165.
- Maravelias CT, Grossmann IE (2002) Logic inference and a decomposition algorithm for the resource-constrained scheduling of testing tasks in

References

- development of new pharmaceuticals and agrochemicals. *Modeling for discrete optimization*.
- Maravelias CT, Grossmann IE (2004) Optimal resource investment and scheduling of tests for new product development. *Computers & chemical engineering*. **28**(6) 1021-1038.
- Marler RT, Arora JS (2004) Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization*. **26**(6) 369-395.
- Marler RT, Arora JS (2010) The weighted sum method for multi-objective optimization: new insights. *Structural and multidisciplinary optimization*. **41**(6) 853-862.
- Martin RK (2012). *Large scale linear and integer optimization: a unified approach*. Springer Science & Business Media.
- Martinez B, Goldstein J (2007) Big pharma faces grim prognosis. *Wall Street Journal A*. **1** 12.
- Meixell MJ, Gargeya VB (2005) Global supply chain design: A literature review and critique. *Transportation Research Part E: Logistics and Transportation Review*. **41**(6) 531-550.
- Melo MT, Nickel S, Saldanha-Da-Gama F (2009) Facility location and supply chain management—A review. *European journal of operational research*. **196**(2) 401-412.
- Meng L, Zhou X (2011) Robust single-track train dispatching model under a dynamic and stochastic environment: a scenario-based rolling horizon solution approach. *Transportation Research Part B: Methodological*. **45**(7) 1080-1102.

- MirHassani SA, Lucas C, Mitra G, Messina E, Poojari CA (2000) Computational solution of capacity planning models under uncertainty. *Parallel Computing*. **26**(5) 511-538.
- Miseta E (2013) Bring down the cost of clinical trials with improved site selection. *Clinical Leader, Dec.* **19** 2013.
- Morton TE, Pentico DW (1995) The finite horizon nonstationary stochastic inventory problem: Near-myopic bounds, heuristics, testing. *Management Science*. **41**(2) 334-343.
- Mula J, Peidro D, Díaz-Madroñero M, Vicens E (2010) Mathematical programming models for supply chain production and transport planning. *European Journal of Operational Research*. **204**(3) 377-390.
- Mullin R (2011). *Before the Storm*. AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 USA, City.
- Mulvey JM, Vladimirou H (1991) Applying the progressive hedging algorithm to stochastic generalized networks. *Annals of Operations Research*. **31**(1) 399-424.
- Ngatchou P, Zarei A, El-Sharkawi A (2005) Pareto multi objective optimization, in Proceedings of the Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems, pp. 84-91.
- Nguyen A-T, Reiter S, Rigo P (2014) A review on simulation-based optimization methods applied to building performance analysis. *Applied Energy*. **113** 1043-1058.

References

- Nickel S, Saldanha-da-Gama F, Ziegler H-P (2012) A multi-stage stochastic supply network design problem with financial decisions and risk management. *Omega*. **40**(5) 511-524.
- Norkin VI, Ermoliev YM, Ruszczyński A (1998) On optimal allocation of indivisibles under uncertainty. *Operations Research*. **46**(3) 381-395.
- Olsson F (2009) Optimal policies for inventory systems with lateral transshipments. *International Journal of Production Economics*. **118**(1) 175-184.
- Papageorgiou LG, Rotstein GE, Shah N (2001) Strategic supply chain optimization for the pharmaceutical industries. *Industrial & Engineering Chemistry Research*. **40**(1) 275-286.
- Pasandideh SHR, Niaki STA, Hajipour V (2013) A multi-objective facility location model with batch arrivals: two parameter-tuned meta-heuristic algorithms. *Journal of Intelligent Manufacturing*. **24**(2) 331-348.
- Paterson C, Kiesmüller G, Teunter R, Glazebrook K (2011) Inventory models with lateral transshipments: A review. *European Journal of Operational Research*. **210**(2) 125-136.
- Paterson C, Teunter R, Glazebrook K (2012) Enhanced lateral transshipments in a multi-location inventory system. *European Journal of Operational Research*. **221**(2) 317-327.
- Pedroso MC, Nakano D (2009) Knowledge and information flows in supply chains: A study on pharmaceutical companies. *International journal of production economics*. **122**(1) 376-384.

- Pérez-Rodríguez N, Holguín-Veras J (2015) Inventory-allocation distribution models for postdisaster humanitarian logistics with explicit consideration of deprivation costs. *Transportation Science*.
- Peterson M, Byrom B, Dowlman N, McEntegart D (2004) Optimizing clinical trial supply requirements: simulation of computer-controlled supply chain management. *Clinical Trials*. **1**(4) 399-412.
- Pontrandolfo P (1999) Global manufacturing: a review and a framework for planning in a global corporation. *International Journal of Production Research*. **37**(1) 1-19.
- Porteus EL (1990) Stochastic inventory theory. *Handbooks in operations research and management science*. **2** 605-652.
- Powell M (2010) Presentation by Senior Vice President, Pharmaceutical Development, Bristol-Myers Squibb. Argyle Executive Forum's 2010 Leadership in pharmaceuticals & biotechnology conference. June 22, 2010. New York, NY, in Proceedings of the Argyle Executive Forum.
- Profile PI (2010) Pharmaceutical Research and Manufacturers of America. *Washington DC, US, PhRMA March*.
- Rahmaniani R, Crainic TG, Gendreau M, Rei W (2016) The Benders Decomposition Algorithm: A Literature Review. *European Journal of Operational Research*.
- Ramezani M, Bashiri M, Tavakkoli-Moghaddam R (2013) A new multi-objective stochastic model for a forward/reverse logistic network design with responsiveness and quality level. *Applied Mathematical Modelling*. **37**(1) 328-344.

References

- Roberts TG, Lynch TJ, Chabner BA (2003) The phase III trial in the era of targeted therapy: unraveling the “go or no go” decision. *Journal of Clinical Oncology*. **21**(19) 3683-3695.
- Robinson LW (1990) Optimal and approximate policies in multiperiod, multilocation inventory models with transshipments. *Operations Research*. **38**(2) 278-295.
- Rockafellar RT, Wets RJ-B (1991) Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of operations research*. **16**(1) 119-147.
- Roundy R (1985) 98%-effective integer-ratio lot-sizing for one-warehouse multi-retailer systems. *Management science*. **31**(11) 1416-1430.
- Rowland C, McMahon J (2004) Clinical trials seen shifting overseas. *International Journal of Health Services*. **34**(3) 555-556.
- Rudi N, Kapur S, Pyke DF (2001) A two-location inventory model with transshipment and local decision making. *Management science*. **47**(12) 1668-1680.
- Ryan SM, Wets RJ-B, Woodruff DL, Silva-Monroy C, Watson J-P (2013) Toward scalable, parallel progressive hedging for stochastic unit commitment, in Proceedings of the Power and Energy Society General Meeting (PES), 2013 IEEE, pp. 1-5.
- Saharidis GK, Ierapetritou MG (2010) Improving Benders decomposition using maximum feasible subsystem (MFS) cut generation strategy. *Computers & chemical engineering*. **34**(8) 1237-1245.

- Sahinidis NV (2004) Optimization under uncertainty: state-of-the-art and opportunities. *Computers & Chemical Engineering*. **28**(6) 971-983.
- Santoso T, Ahmed S, Goetschalckx M, Shapiro A (2005) A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*. **167**(1) 96-115.
- Scarf H (1959) The optimality of (s, S) policies in the dynamic inventory problem. in K.J. Arrow, S. Karlin, and P. Suppes (editors), *Mathematical Methods in Social Sciences*, Stanford University Press.
- Schmidt CW, Grossmann IE (1996) Optimization models for the scheduling of testing tasks in new product development. *Industrial & Engineering Chemistry Research*. **35**(10) 3498-3510.
- Schultz R, Stougie L, Vlerk MH (1996) Two - stage stochastic integer programming: a survey. *Statistica Neerlandica*. **50**(3) 404-416.
- Schütz P, Tomasgard A, Ahmed S (2009) Supply chain design under uncertainty using sample average approximation and dual decomposition. *European Journal of Operational Research*. **199**(2) 409-419.
- Sen S (2005) Algorithms for stochastic mixed-integer programming models. *Handbooks in operations research and management science*. **12** 515-558.
- Shah N (2004) Pharmaceutical supply chains: key issues and strategies for optimisation. *Computers & chemical engineering*. **28**(6) 929-941.
- Shapiro A, Dentcheva D, Ruszczyński A (2009). *Lectures on stochastic programming: modeling and theory*. SIAM.
- Shapiro A, Philpott A (2007) A tutorial on stochastic programming. *Manuscript*. Available at www2.isye.gatech.edu/ashapiro/publications.html. **17**.

References

- Sherali HD, Fraticelli BM (2002) A modification of Benders' decomposition algorithm for discrete subproblems: An approach for stochastic programs with integer recourse. *Journal of Global Optimization*. **22**(1-4) 319-342.
- Sherali HD, Zhu X (2006) On solving discrete two-stage stochastic programs having mixed-integer first-and second-stage variables. *Mathematical programming*. **108**(2-3) 597-616.
- Sousa RT, Liu S, Papageorgiou LG, Shah N (2011) Global supply chain planning for pharmaceuticals. *chemical engineering research and design*. **89**(11) 2396-2409.
- Sundaramoorthy A, Evans JM, Barton PI (2012) Capacity planning under clinical trials uncertainty in continuous pharmaceutical manufacturing, 1: mathematical framework. *Industrial & engineering chemistry research*. **51**(42) 13692-13702.
- Tang G, Kong Y, Chang CCH, Kong L, Costantino JP (2012) Prediction of accrual closure date in multi - center clinical trials with discrete - time Poisson process models. *Pharmaceutical statistics*. **11**(5) 351-356.
- Thiers FA, Sinskey AJ, Berndt ER (2008) Trends in the globalization of clinical trials. *Nature Reviews Drug Discovery*. **7**(1) 13-14.
- Tunc H, Kilic OA, Tarim SA, Eksioglu B (2011) The cost of using stationary inventory policies when demand is non-stationary. *Omega*. **39**(4) 410-415.
- Van Arnum P (2009) Managing the Clinical-Trial-Materials Supply Chain. *Pharmaceutical Technology Sourcing and Management*. **5**(8)
- Van Arnum P (2011) Managing the Global Clinical-Trial Material Supply Chain. *Pharmaceutical Technology*. **35**(11)

- Van Slyke RM, Wets R (1969) L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*. **17**(4) 638-663.
- Verweij B, Ahmed S, Kleywegt AJ, Nemhauser G, Shapiro A (2003) The sample average approximation method applied to stochastic routing problems: a computational study. *Computational Optimization and Applications*. **24**(2-3) 289-333.
- Vidal CJ, Goetschalckx M (1997) Strategic production-distribution models: A critical review with emphasis on global supply chain models. *European Journal of Operational Research*. **98**(1) 1-18.
- Wang Q (2011) Control policies for multi-echelon inventory systems with stochastic demand, in Supply chain coordination under uncertainty, Springer, pp. 83-108.
- Waters S, Dowlman I, Drake K, Gamble L, Lang M, McEntegart D (2010) Enhancing control of the medication supply chain in clinical trials managed by interactive voice response systems. *Drug Information Journal*. **44**(6) 727-740.
- Watson J-P, Woodruff DL (2011) Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *Computational Management Science*. **8**(4) 355-370.
- Wee H-M, Yang P-C (2004) The optimal and heuristic solutions of a distribution network. *European Journal of Operational Research*. **158**(3) 626-632.
- Woodruff DL, Zemel E (1993) Hashing vectors for tabu search. *Annals of Operations Research*. **41**(2) 123-137.

References

- Xiao T, Jin J (2011) Coordination of a fashion apparel supply chain under lead-time-dependent demand uncertainty. *Production Planning & Control*. **22**(3) 257-268.
- Yang J, Qin Z (2007) Capacitated production control with virtual lateral transshipments. *Operations Research*. **55**(6) 1104-1119.
- Zadeh L (1963) Optimality and non-scalar-valued performance criteria. *IEEE transactions on Automatic Control*. **8**(1) 59-60.
- Zhang J (2005) Transshipment and its impact on supply chain members' performance. *Management Science*. **51**(10) 1534-1539.