Positioning Inventory in Clinical Trial Supply Chains

Adam Fleischhacker∗ Anh Ninh† Yao Zhao‡

September 13, 2010

Abstract

As a result of slow patient recruitment and high patient costs in the United States, clinical trials are increasingly going global. While recruitment efforts benefit from a larger global footprint, the supply chain has to work harder at getting the right drug supply, usually in the form of patient kits, to the right place, at the right time. Clinical trial supply chains are unique due to a fixed patient horizon, an inflexible supply process, a requirement to quickly satisfy all demand, and an inability to transfer drug supplies among testing sites. In this paper, we provide a new class of multi-echelon inventory models to address these unique aspects. The resulting mathematical program is nonlinear, and the stocking policies used at warehouses and demand sites cannot be decoupled; they must be solved jointly. Given this complexity, we develop a decomposition scheme which leads to a strongly polynomial optimization algorithm. We also provide faster and more intuitive algorithms to compute upper and lower bounds. The algorithms are leveraged to provide managerial insights into the optimal supply chain configurations for global clinical trials.

Keywords and Phrases: Clinical trial supply management, multi-echelon inventory model, distribution, finite patient horizon.

1 Introduction

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

∗Alfred Lerner College of Business and Economics, University of Delaware, Newark, DE 19716, ajf@lerner.udel.edu
†Rutgers Center of Operations Research, Rutgers University, Piscataway, NJ 08854, anhninh@rci.rutgers.edu
‡Dept. of Supply Chain Management & Marketing Sciences, Rutgers University, Newark, NJ 07102, yaozhao@andromeda.rutgers.edu

1Paul Loveday (CEO, ClinStar) at the Clinical Research in Emerging Countries Third Annual Marcus Evans Conference 21-22 July 2008, Washington DC, USA
At a cost of $17 per second, clinical trial delays are to be avoided. However, avoiding clinical trial delays is easier said than done. The biggest stumbling block is often patient recruitment. Clinical trials, as mandated by the United States Food and Drug Administration (FDA), seek to test the safety and effectiveness of a new drug or treatment prior to commercialization. It is only natural that potential patients may be leery of enrolling in these trials, especially when already approved treatment options exist.

According to Getz and de Bruin (2000), 80% of clinical trials fail to meet their patient recruitment deadlines. As a result of slow patient recruitment and also high patient costs in the United States, clinical trials are increasingly going global in their search for patients (Rowland 2004). In quantifying this globalization, Thiers, Sinskey and Berndt (2008) report the following data on the growth in the number of clinical trial sites (the doctors’ offices and hospitals that are used to enroll patients in a clinical trial) for various countries:

<table>
<thead>
<tr>
<th>Country</th>
<th>Annual Growth Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>47.0</td>
</tr>
<tr>
<td>Russia</td>
<td>33.0</td>
</tr>
<tr>
<td>Argentina</td>
<td>26.9</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>24.6</td>
</tr>
<tr>
<td>Mexico</td>
<td>22.1</td>
</tr>
<tr>
<td>United States</td>
<td>-6.5</td>
</tr>
</tbody>
</table>

Table 1: Growth in Number of Offshore Clinical Trial Sites

We see from Table 1, that clinical trial growth in emerging markets far outpaces that in the United States. In fact, the United States is experiencing declines in the number of trial sites being used. In addition to the relocation of trials, more clinical trials are seeking participants in multiple countries simultaneously. In 2005, 7.8% of the trials reported in three top medical journals were being conducted in 10 or more countries while in 1995, none of the articles in those journals reported such high levels of globalization (Glickman, McHutchison, Peterson, Cairns, Harrington, Califf, and Schulman 2009).

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

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

It now becomes clear that both the demand chain and the supply chain for clinical trials are in flux. The demand chain, in its effort to increase patient recruitment rates, is going global with an increasing number of investigative sites. The supply chain, which may have been adequate for domestic trials, is now struggling with meeting the increasingly unpredictable and geographically dispersed demands imposed by global trials. The question is, how to design and manage the supply chain to efficiently support a global clinical trial?

1.1 Clinical Trial Supply Chains

For large global clinical studies, the clinical trial supply chain can be highly complex. For a U.S. pharmaceutical/biotech company, a typical clinical trial supply chain starts with active pharmaceu-
tical ingredient (API) production either in the U.S. or overseas, formulation and packaging in the U.S., then distribution from a central warehouse in the U.S., to multiple country depots (regional warehouses) and finally to sites. Regional warehouses are often needed to ensure more reliable and timely shipments to clinical trial sites because of required regulatory clearance time and the time required to simply transport supplies over large distances.

From a modeling perspective, clinical trial supply chains resemble spare part supply chains in their network structure, their unpredictable demand, and the fact that demand is satisfied only at sites. However, clinical trial supply chains are unique in the following aspects:

- **Finite patient horizon**: Each study has a recruitment target which is the necessary sample size for the study. Once the target is reached, recruitment is closed and no more patients will be enrolled in the trial.

- **Inflexible Supply Process**: Due to the significant fixed costs and time required for both production and quality control of investigational drugs, companies often produce enough to cover the entire trial or at least enough to cover all demand during the drug’s shelf-life. In addition, with the regulatory and testing requirements imposed by FDA, it is desirable, when possible, to have a drug of uniform quality that was produced all in one batch or in one run, see Oncolytics Biotech (2003) for an example.

- **Non-rejection and high service levels**: According to Clemento (1999), every extra day of patent availability is worth $1 million for a typical drug. Since patient recruitment is the typical bottleneck in conducting clinical trials, a rejection of a patient due to shortages of clinical drug is considered unacceptable and one has to ensure that the system has enough inventory for all patients recruited and the service level based on immediate inventory availability is sufficiently high.

- **Inability of Cross-Shipping**: By regulation, site-to-site shipping is discouraged and must “remain the exception” (European Commission 2003) due to possible safety and trial quality issues. Cross shipping among regional warehouses (e.g., from one country depot to another country depot) is not recommended because of customs issues, uncertain lead times, the need for relabeling, and potential risks of disqualifying a trial due to poor adherence with regulations. As Andrews (2004) notes, “in practice, the delays and potential for mistakes in retrieving a batch of material from one site and re-releasing and/or transferring to another
can, and does, make this course of action impractical for some organizations in some cases.”

These unique attributes, combined with a complex logistics network and inherent demand uncertainty (due to random enrollments at sites and randomization of patients to receive either the investigational drug or a placebo), make a global clinical-trial supply chain difficult to manage. In practice, companies produce multiples of baseline forecasts as planned overage (McDonnell and Mooraj 2009). However, this overage is essentially a costly waste that ultimately leads to paying for the destruction of expensive materials. Thus, one of the main challenges in the clinical trial supply chain is to achieve required service levels with minimum inventory overage by efficiently configuring the supply chain and intelligently positioning inventory.

1.2 Summary of Results

In this paper, we develop a mathematical model to aid decision making in stock positioning and network configuration for clinical trial supply chains. We first present a new class of multi-echelon inventory models that incorporates the unique aspects of the clinical trial supply chain. We then develop a decomposition scheme and an optimization algorithm that is strongly polynomial in the patient horizon, the number of clinical trial sites, and the number of regional warehouses. We also provide simpler lower and upper bounds on the optimal objective function for this new class, and finally we leverage the model and algorithms to provide insights into the optimal supply chain configurations.

The remainder of this paper is organized as follows: After reviewing the relevant literature in §2, we introduce the basic modeling assumptions, the mathematical programming formulation, and some preliminary results in §3. In §4, we provide the optimization algorithm for stock positioning. In §5, we present upper and lower bounds on the objective function. In §6, we conduct a numerical study to test the effectiveness of the bounds and to quantify the impact of recruitment rate and lead times on the supply chain configuration. §7 concludes this work.

2 Literature Review

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

The management of inventory in these supply chains has attracted limited attention from the industry, and has attracted even less attention from the academia. In industry literature, it is usually advocated that inventory control policies in the clinical trial supply chain are to be selected using simulation, see Abdelkafi, Beck, David, Druck and Horoho (2009), Peterson, Byrom, Dowlman and McEntegart (2004) and references therein. In addition, it is often assumed within these simulations that an integrated voice response system (IVRS) is available so that inventory can be monitored continuously (McEntegar and O’Gorman 2005) and as inventory at a location falls below a specified trigger, more inventory is ordered. In real-world practice, despite the availability of these sophisticated tools, inventory policies (e.g., the trigger levels, safety-stock) in the clinical trial supply chain are still often set using experience and by looking at the patterns from previous clinical trials.

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

Within the context of multiechelon inventory research, our models are most closely related to work done for service parts where end-user demand is low and one-for-one base stock policies are employed (see reviews by Zipkin 2000, Muckstadt 2005 and Simchi-Levi and Zhao 2007). One of the seminal works in this stream of literature is Sherbrooke (1968). He approximated the distribution of backorders at the depot by its first moment in a two-echelon distribution system when one-for-one ordering policies are used. Improving on this approximation is the approximation by Graves (1985) who shows how to effectively approximate backorder and lead time demand by a negative binomial distribution. In this paper, we utilize the result of Graves (1985) to approximate
certain subsystems within a three-echelon supply chain that includes one central warehouse, multiple regional warehouses and multiple sites, where some sites are supplied directly by the central warehouse and others are supplied indirectly by a regional warehouse. Svoronos and Zipkin (1991) refines the approximation by Graves (1985) and extends it to evaluate multi-echelon distribution systems.

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

The stock positioning problem in clinical trial supply chains represents a new variation of the classical multi-echelon inventory models because it differs from the literature in three key aspects: 1) system performance concerns are only relevant until an adequate number of patients are recruited, 2) production is done prior to the commencement of the trial and 3) clinical trial supply chains cannot afford to reject patients due to supply shortages and therefore have to place enough stock in the system to satisfy all recruited patients up to the pre-determined limit on the number of subjects needed to complete the trial. These differences result in non-trivial modifications to the objective functions and constraints used in classical models, and lead to new models, solution algorithms, and insights.

3 Model and Formulation

In this section, we consider a clinical trial supply chain of a general distribution topology, depicted in Figure 1, where a central warehouse (CW), indexed as 0, supplies multiple regional warehouses (RWH), indexed by \( i = 1, 2, ..., I \), and some sites directly, which are indexed by \( 0j \) where \( j = 1, 2, ..., J_0 \). Each regional warehouse, \( i \), supplies a set of sites indexed by \( ij \) where \( j = 1, 2, ..., J_i \).

Throughout this paper, we make the following assumptions.

Assumption 1

(a) The clinical study has a finite patient horizon. (b) Each patient requires one
kit which is produced prior to the commencement of the trial. (c) Inventory (i.e., the drug kit) cannot be transferred among sites and among regional warehouses. (d) Patient recruitment occurs only at trial sites and follows an independent Poisson process. (e) Patients recruited at one site are unable to seek treatment from other sites. (f) In the event of a site-level stockout (but resupply is available), patients will still enroll in the trial and wait for resupply. (g) No patients are rejected from the trial because the system does not have resupply available. (h) Inventory at sites and regional warehouses are under a continuous-review base-stock policy. (i) The lead times (for shipping, regulatory clearance) between CW and RWH, CW and direct sites, and RWH and sites are constant.

The first four assumptions are justified in §1.1. The assumption of Poisson recruitment process (1d) is commonly used in industry literature, see Abdelkafi et al. (2009) for examples. Assumption 1e is a reflection of the geographical dispersement of many clinical trials. For example, it is unlikely that a patient enrolled at Mount Sinai Medical Center (New York) for testing the investigational drug CERE-120 (targeting at Parkinson’s Disease) will be redirected to other clinical trial sites in either North Carolina or Atlanta. The other locations are simply too far away for the patient and the financial incentive of the New York site is to keep that patient at their site even if time for

---

resupply is required\textsuperscript{3}.

In reality, Assumption 1f may or may not hold depending on the type of disease, treatment and trial country. In this paper, we consider the case where patients can and will wait for resupply if the site runs into a temporary drug shortage. Assumption 1g is justified in \S 1.1 which indicates that although patients may suffer from a temporary shortage of the trial drug at a site, a resupply (from upstream stages of the site) is always available. The assumption of a continuous-review base-stock policy (1h) at both the sites and the regional warehouses closely resembles the modernized clinical trial supply chains. Through the use of integrated voice response systems (IVRS), all doctors administering clinical trial drugs to patients are required to call in to a system for instructions on which of the patient kits on site are to be administered to the patient. Through this system, both real-time inventory information is maintained and automated shipments to replenish site-level inventory are triggered (Byrom 2002). While sometimes it will take multiple demands at a site to trigger an order, we make the equally realistic assumption, given the high value of investigational drugs, that an order is triggered each time a demand occurs. The assumption of constant lead times (1i) is for simplicity, extensions are discussed in \S 7.

In this paper, we consider the cases where the patient recruitment period is much longer than the lead times (required for shipping and regulatory clearance) for trial drugs. In practice, the recruitment period for a Phase II or Phase III clinical trial is typically measured in months and years while the lead times are typically in days or weeks (Kohner 2010).

**Assumption 2** The patient recruitment period is significantly longer than the lead times required for shipping and regulatory clearance, thus the performance of each direct site and each regional distribution system – the subsystem of the regional warehouse $i$ and the sites $ij$, can be measured by their corresponding steady-state counterparts.

We define the following notation:

- $S$: Patient horizon or the number of subjects needed to complete the trial.
- $\lambda_{ij}, \lambda_{0j}$: Rates of patient recruitment at sites $ij$ and $0j$ respectively. Let $\lambda_i$ be the aggregate recruitment rate of all sites $ij$ supplied by regional warehouse $i$.

\textsuperscript{3}Parkinson's is a slow-progressing disease where a patient can afford to wait for resupply prior to undergoing treatment.
• $L_i, L_{ij}, L_{0j}$: lead times between CW and RWH $i$, between RWH $i$ and site $ij$, and between CW and direct site $0j$ respectively.

• $D_i(L), D_{ij}(L), D_{0j}(L)$: Demand during lead time $L$ at regional warehouse $i$, sites $ij$ and $0j$ respectively.

• $s_i, s_{ij}, s_{0j}$: Base-stock levels at locations $i$, $ij$ and $0j$ respectively.

• $s_0$: The initial stock at the central warehouse (which does not have the availability of additional replenishments).

• $\zeta$: The patient service level at sites or alternatively, the probability of not enrolling a patient in the trial at a site due to insufficient drug resupply.

• $\zeta'$: The immediate fill-rate of demand upon arrival at a clinical trial site.

Our objective is to minimize the system-wide inventory subject to 100% patient service level ($\zeta = 100\%$) and a high immediate fill-rate ($\zeta'$), by setting the stock levels at each location appropriately. Thus, our decision variables are $s_0, s_i, s_{ij}, s_{0j}$. Since there is already inventory allocated to sites and regional warehouses downstream of the central warehouse, $s_0$ can be less than the required number of patient recruits, $S$. Planning for the worst-case scenario of demand only arriving through one site and given $s_i, s_{ij}$ and $s_{0j}$, we must set $s_0 = S - \min\{\min_i\{s_i + \min_j s_{ij}\}, \min_j s_{0j}\}$ to guarantee $\zeta = 100\%$. Given the above definitions, the problem ($P0$) of minimizing inventory investment subject to fill-rate constraints ($\zeta'$) can be formulated as follows,

\[
(P0) \quad P0^{OPT} = \min \left\{ s_0 + \sum_i [s_i + \sum_j s_{ij}] + \sum_j s_{0j} \right\}
\]

s.t. \quad \text{Immediate fill rate at site } ij \geq \zeta', \quad \forall ij,
\quad \text{Immediate fill rate at site } 0j \geq \zeta', \quad \forall 0j,
\quad s_0 = S - \min\{\min_i\{s_i + \min_j s_{ij}\}, \min_j s_{0j}\},
\quad s_0, s_i, s_{ij}, s_{0j} \text{ are non-negative integers}, \quad \forall i, j.

By Assumption 2, we can set $S$ as an upper bound for $s_{0j}$ for all direct sites $0j$ and $s_i + \min_j s_{ij}$ for all regional warehouses $i$, and thus $s_0$ is nonnegative.

Due to the finite patient horizon, it is logical to assume that a site $0j$ (or $ij$) will stop ordering when the number of patients left to be recruited is smaller than or equal to $s_{0j}$ ($s_{ij}$ respectively). Similarly, a RWH $i$ will stop ordering when the number of patients left to be recruited is less than or equal to $s_i + \min_j s_{ij}$. By the constraint on $s_0$ (in Problem $P0$), the CW will never stockout.
before a direct site or a RWH stops ordering because when the CW runs out stock, the constraint on $s_0$ ensures that each direct site/RWH subsystem alone has enough stock to cover all the patients left to be recruited. In view of this argument, it follows by Assumption 2 that we can write the fill rate constraints for site $ij$ in Problem $(P0)$ as follows (see, e.g., Simchi-Levi and Zhao 2007),

$$\Pr\{s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij})\} \geq \zeta', \quad \forall ij,$$

where $B_{ij}$ is the number of backorders that RWH $i$ owes site $j$ in steady state, it depends on $s_i$. We can also write the fill rate constraints for direct sites $0j$ as follows (see Simchi-Levi and Zhao 2007),

$$\Pr\{s_{0j} > D_{0j}(L_{0j})\} \geq \zeta', \quad \forall 0j.$$

Problem $(P0)$ differs from the classical multi-echelon inventory models of inventory investment minimization subject to fill-rate constraints because of the constraint on $s_0$. This constraint comes from the uniqueness of clinical trial supply chains – the finite patient horizon, the inflexible supply process, the non-rejection requirement and the inability to cross-ship. This constraint on $s_0$ effectively connects all regional warehouse subsystems and thus their base-stock levels must be determined jointly.

**Lemma 3** In the optimal solution to Problem $(P0)$, $s_{ij}$ depends on $s_i$ in such a way that it is the minimum value so that the fill rate constraint at site $ij$ is satisfied given $s_i$ and unlimited supply from the CW; $s_{0j}$ is the minimum value so that the fill rate at site $0j$ is satisfied given unlimited supply from the CW.

**Proof.** We first note that the CW never stocks out before RWH $i$ stops ordering. Thus we can treat CW as it has unlimited supply. Assuming the claim on $s_{ij}$ is not true. Then there must exist an optimal solution to Problem $(P0)$ such that the new solution of reducing $s_{ij}$ by one while keeping everything else unchanged still remains feasible for all fill rate constraints. To ensure the constraint on $s_0$, reducing $s_{ij}$ by one can lead to one of the following two updates to the new solution: (1) $S$ (or $s_i$) increases by one, or (2) all other base-stock levels remain unchanged. In case (1), the objective function of $(P0)$ does not change; while in case (2), the objective function of $(P0)$ reduces by one. Because the updated new solution is feasible, we now have a better solution which creates a contradiction to the optimality claim of the original solution. The proof is completed for $s_{ij}$.

For $s_{0j}$, we note that by the constraint of $s_0$, the CW will never stockout before a direct site stops ordering. By the same logic as of $s_{ij}$, we can prove the desired result for $s_{0j}$. 

\[\Box\]
Lemma 3 implies that \( s_{0j} \) can be solved independently of all regional warehouse subsystems, and for a regional warehouse subsystem, the decision variables \( s_i, s_{ij} \) are dependent. Clearly, Problem \( (P0) \) is nonlinear with integer decision variables.

4 Optimization Algorithm

In this section, we develop a strongly polynomial algorithm to solve Problem \( (P0) \). We first consider two special cases in §4.1 to generate insights. Then we develop the core optimization algorithm to solve regional-warehouse-only networks in §4.2. Finally, we design an algorithm to solve the general network (see Figure 1) in §4.3.

4.1 Two Special Cases

In this subsection, we focus on distribution networks which supply the sites exclusively by direct shipment or supply the sites exclusively through one regional warehouse. These two cases are shown in Figure 2.
For the direct shipment network, Problem (P0) is simplified to:

\[
\begin{align*}
\min & \quad \{ s_0 + \sum_j s_{0j} \} \\
\text{s.t.} & \quad \Pr\{ s_{0j} > D_{0j}(L_{0j}) \} \geq \zeta', \quad \forall 0j \\
& \quad s_0 = S - \min_j s_{0j} \\
& \quad s_0, s_{0j} \text{ are nonnegative integers,} \quad \forall j.
\end{align*}
\] (3)

Replacing \( s_0 \) in the objective function by its constraint yields,

\[
\begin{align*}
\min & \quad \{ S + \sum_j s_{0j} - \min_j s_{0j} \} \\
\text{s.t.} & \quad \Pr\{ s_{0j} > D_{0j}(L_{0j}) \} \geq \zeta', \quad \forall 0j \\
& \quad s_0, s_{0j} \text{ are nonnegative integers,} \quad \forall j.
\end{align*}
\] (4)

By Lemma 3, an optimal solution is to choose the smallest \( s_{0j} \) that achieves the desired fill rate at site 0\( j \) for all \( j \) given unlimited supply from the CW.

Now, we look at the network with one regional warehouse subsystem, Problem (P0) reduces to:

\[
\begin{align*}
\min & \quad \{ s_0 + s_i + \sum_j s_{ij} \} \\
\text{s.t.} & \quad \Pr\{ s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij}) \} \geq \zeta', \quad \forall j \\
& \quad s_0 = S - s_i - \min_j s_{ij} \\
& \quad s_0, s_i, s_{ij} \text{ are nonnegative integers,} \quad \forall j.
\end{align*}
\] (5)

To simplify further, we place the non-rejection constraint, \( s_0 = S - s_i - \min_j s_{ij} \), into the objective function, then we have an equivalent problem:

\[
\begin{align*}
\min & \quad \{ S + \sum_j s_{ij} - \min_j s_{ij} \} \\
\text{s.t.} & \quad \Pr\{ s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij}) \} \geq \zeta', \quad \forall j, \\
& \quad S - s_i - \min_j s_{ij} \geq 0 \\
& \quad s_i, s_{ij} \text{ are nonnegative integers,} \quad \forall j.
\end{align*}
\] (6)

By Problem (6), the optimal inventory policy is revealed: To minimize the system-wide inventory, one must minimize site level inventory. This result is summarized in the following proposition.

**Proposition 4** The optimal solution to the following problem,

\[
\begin{align*}
\min & \quad \{ S + \sum_j s_{ij} \} \\
\text{s.t.} & \quad \Pr\{ s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij}) \} \geq \zeta', \quad \forall j, \\
& \quad S - s_i - \min_j s_{ij} \geq 0 \\
& \quad s_i, s_{ij} \text{ are nonnegative integers,} \quad \forall j,
\end{align*}
\] (7)

is also optimal for Problem (6).
Proof. Denote the solution of Problem (7) to be $s'$. Suppose that the proposition is not true, then there must exist a solution of Problem (6), $s''$, such that
\[ \sum_j s'_{ij} - \min_j s'_{ij} > \sum_j s''_{ij} - \min_j s''_{ij}. \] (8)

By Lemma 3, we note that given $s_i$, $s_{ij}$ are uniquely determined by the fill rate constraints in both Problem (6) and (7). In addition, $s_{ij}$ is non-increasing in $s_i$ (this is clearly true as a higher stocked the regional warehouse only reduces the need of inventory at sites). Thus, Inequality (8) implies $s'_i < s''_i$. To see this, we define $j' = \text{argmin}_j \{s'_{ij}\}$, $j'' = \text{argmin}_j \{s''_{ij}\}$, and consider two cases:

- Case 1: $j' = j''$. Then inequality (8) implies $\sum_{j \neq j'} s'_{ij} > \sum_{j \neq j'} s''_{ij}$. This is only true when $s'_i < s''_i$.

- Case 2: $j' \neq j''$. Rewrite Inequality (8) as $\sum_{j \neq j'} s'_{ij} + s'_{ij''} > \sum_{j \neq j'} s''_{ij} + s''_{ij'}$.

Suppose $s'_i \geq s''_i$, then $s'_{ij} \leq s''_{ij}$ for all $j$, and by the above inequality $s'_{ij''} > s''_{ij'}$. Together with $s''_{ij'} \geq s''_{ij''}$ (by definition of $j''$), we must have $s'_{ij'} > s''_{ij'}$, which contradicts the fact that $s'_{ij} < s''_{ij}$ for all $j$. Thus we must have $s'_i < s''_i$.

$s'_i < s''_i$ implies $s'_{ij} \geq s''_{ij}$ for all $j$. If there is one $j$ such that $s'_{ij} > s''_{ij}$, this creates a contradiction to the assumption that $s'$ is optimal for Problem (7). If $s'_{ij} = s''_{ij}$ for all $j$, this creates a contradiction to Inequality (8). In conclusion, the optimal solution $s'$ for Problem (7) is also optimal for (6).  

Intuitively, Proposition 4 indicates that one should stock inventory at the regional warehouse $i$ as much as possible to reduce the needed stock at sites. Effectively, one can increase $s_i$ as long as $s_0$ is non-negative. This solution minimizes the total inventory in the system while satisfying the fill rate constraint at the sites and the non-rejection constraint for recruitment.

4.2 Regional Warehouse Only Network

Expanding the one regional warehouse case in §4.1 to include multiple regional warehouses (but not direct shipment), Problem (P0) becomes Problem (P1):

\[
(P1) \quad P1^{OPT} = \min \{s_0 + \sum_i [s_i + \sum_j s_{ij}] \}
\]

\[
s.t. \quad \Pr\{s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij})\} \geq \zeta', \quad \forall i,j \]

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

\[
s_0, s_i, s_{ij} \text{ are nonnegative integers,} \quad \forall i, j.
\]
To solve Problem \((P_1)\), we create a decomposition scheme which consists of solving Problem \((P_{2i})\) for \(i = 1, 2, ..., I\), and is formulated as follows:

\[
(P_{2i}) \quad P^{2\text{OPT}}_{2i} = \min \{ S + \sum_{i \neq i}[s_i + \sum_j s_{ij}] + \sum_{j \neq j} s_{ij} \}
\]

\[
\text{s.t.} \quad \Pr\{s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij})\} \geq \zeta', \quad \forall ij
\]

\[
s_i + \min_j s_{ij} \leq s_i + \min_j s_{ij}, \quad \forall i \neq i, \forall j
\]

\[
s_i + \min_j s_{ij} \leq S
\]

\[
s_i, s_{ij} \text{ are nonnegative integers}, \quad \forall i, j,
\]

where \(j = \arg\min_j\{s_{ij}\}\).

Define Problem \((P_2)\) such that \(P^{2\text{OPT}} = \min_i\{P^{2\text{OPT}}_{2i}\}\).

**Proposition 5** The optimization problems \((P_1)\) and \((P_2)\) are equivalent.

**Proof.** Because in Problem \((P_{2i})\), the condition \(s_i + \min_j s_{ij} \leq s_i + \min_j s_{ij}, \forall i \neq i \) and \(\forall j\) means that \(s_i + \min_j s_{ij} = \min_i\{s_i + \min_j s_{ij}\}\). In addition, the condition \(s_i + \min_j s_{ij} \leq S\) guarantees the nonnegativity of \(s_0\). Thus, any feasible solution to \((P_2)\) is also a feasible solution to \((P_1)\). Due to the two problems’ identical objective functions (one can easily see this by replacing \(s_0\) by its constraint in the objective function of Problem \((P_1))\), \(P^{1\text{OPT}} \leq P^{2\text{OPT}}\).

Now assuming that we found the optimal solution for \((P_1)\). Then, if we let \(i = \arg\min_i\{s_i + \min_j s_{ij}\}\), \((P_1)\) becomes \((P_{2i})\). Thus, \(P^{2\text{OPT}} \leq P^{1\text{OPT}}\). Combining these arguments, \((P_1)\) and \((P_2)\) are equivalent. \(\square\)

Because \((P_1)\) and \((P_2)\) are equivalent, we now propose an algorithm to solve \((P_2)\) instead. Thus, we first must solve Problem \((P_{2i})\), and this is done by enumerating \(s_i\) over all possible values. By Lemma 3, we note that given \(s_i, s_{ij}\) can be determined by the fill rate constraint at site \(ij\). Let \(c_i = s_i + \min_j s_{ij}\). Then, we have Problem \((P_{2i})\) for all regional warehouses and all possible \(s_i\) such that \(c_i \leq S\), formulated as follows:

\[
(P_{2i}) \quad P^{2\text{OPT}}_{2i} = \min \{ S + \sum_{i \neq i}[s_i + \sum_j s_{ij}] + \sum_{j \neq j} s_{ij} \}
\]

\[
\text{s.t.} \quad \Pr\{s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij})\} \geq \zeta', \quad \forall ij
\]

\[
c_i \leq s_i + s_{ij}, \quad \forall i \neq i, \forall j
\]

\[
s_i, s_{ij} \text{ are nonnegative integers}, \quad \forall i \neq i, \forall j.
\]

Problem \((P_{2i})\) can be further separated into multiple problems \((P_i(c_i))\), one for each regional warehouse sub-system \(i \neq i\). Moreover, we can discard \(S + \sum_{j \neq j} s_{ij}\) in the objective function since
it is constant.

\[
(P_i(c_i)) \quad P_i(c_i)^{OPT} = \min \{ s_i + \sum_j s_{ij} \}
\]

s.t. \( \Pr\{ s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij}) \} \geq \zeta', \forall j \)

\[ c_i \leq s_i + s_{ij}, \forall j \]

\[ s_i, s_{ij} \text{ are nonnegative integers}, \forall j. \]

Note that we get the solution for \((P2_{s_i})\) after solving \(I - 1\) sub-problems \((P_i(c_i))\) where \(I\) is the number of regional warehouses.

To solve \((P_i(c_i))\), we enumerate on \(s_i\) over its possible values such that \(c_i \leq S\). For a given \(s_i\), \(s_{ij}\) is the minimum value such that the fill rate constraint at site \(ij\) is satisfied (Lemma 3). To compute the fill rate constraint, we utilize the two-moment approximation of Graves (1985). The worst-case computational complexity of this algorithm for Problem \((P_i(c_i))\) is \(O(S^2 \ast J_i)\). This is true because there are at most \(O(S)\) possible values of \(s_i\), one needs to evaluate the fill rate constraint for each \(j = 1, 2, ..., J_i\), and enumerating \(s_{ij}\) to find the minimum that satisfies the fill rate constraint requires at most \(O(S)\) steps. Consequently, the worst-case computational complexity of Problem \((P2_{s_i})\) is \(O(I \ast S^2 \ast J)\) where \(J = \max_i J_i\). Finally, the computational complexity of Problem \((P2)\) is at most \(O(I^2 \ast S^3 \ast J)\), because there are \(I\) many \((P2_i)\) problems, and for each there are at most \(S\) many \((P2_{s_i})\) problems.

In summary, to solve the problem \((P2_i)\), we have to solve at most \(S + 1\) problems \((P2_{s_i})\) for \(s_i = 0, ..., S\). Then, the optimal value of \((P2_i)\) is the smallest value among all the optimal values of \((P2_{s_i})\): \(P_{2_i}^{OPT} = \min_{s_i} P_{2_{s_i}}^{OPT}\). Moreover, each of those \((P2_{s_i})\) problems can be solved by solving \(I - 1\) sub-problems \((P_i(c_i))\). Therefore, we have the following pseudocode:

**Pseudocode for solving Problem \((P2_i)\):**

1. Set \(s_i = 1\).

2. Given \(s_i\), find the minimum stock level \(s_{ij}\) such that the fill rate constraint at site \(ij\) is satisfied. Then, we compute \(\min_j s_{ij}\).

   If \(s_i + \min_j s_{ij} > S\), increase \(s_i\) by 1 and repeat Step 2.

   Else, solve Problem \((P2_{s_i})\) and save the optimal value \(P_{2_{s_i}}^{OPT}\). Increase \(s_i\) by 1 and repeat Step 2.

   Once \(s_i\) reaches the level \(S\), go to Step 3.
3. For all the possible values of \(s_i\), output the smallest optimal values \(P_{2s_i}^{OPT}\). It is indeed the optimal value of Problem \((P_{2i})\).

Problem \((P_{2s_i})\) can be decomposed into \(I - 1\) Problem \((P_i(c_i))\), we can easily find its optimal value once we solve the latter:

\[
P_{2s_i}^{OPT} = S + \sum_{j \neq i} s_{ij} + \sum_{i \neq i} P_i(c_i)^{OPT}.
\]

Pseudocode for solving Problem \((P_i(c_i))\):

1. Set \(s_i = 1\).

2. Given \(s_i\), find the minimum stock level \(s_{ij}\) such that the fill rate constraint is satisfied at site \(ij\). Then, we compute \(\min_j s_{ij}\).

   If \(s_i + \min_j s_{ij} \geq c_i\), save the value of the corresponding objective function: \(s_i + \sum_j s_{ij}\), increase \(s_i\) by 1 and repeat Step 2.

   Else, for that given \(s_i\), Problem \((P_i(c_i))\) is infeasible. Increase \(s_i\) by 1 and repeat Step 2.

   Once \(s_i\) reaches the level \(S\), go to Step 3.

3. Output the smallest value among all the objective function \(s_i + \sum_j s_{ij}\) found. It is indeed the optimal value of Problem \((P_i(c_i))\).

4.3 General Networks

In this section, we consider the general network as depicted in Figure 1 where we have both regional warehouses and direct sites. By Lemma 3, the optimal stock level at direct sites \(s_{0j}^*\), \(\forall j\) can be computed independently of all regional warehouses. Let \(c_0^* = \min_j s_{0j}^*\). Then Problem \((P0)\) reduces to,

\[
(P1') \quad P1^{OPT} = \min \left\{ s_0 + \sum_i \left[ s_i + \sum_j s_{ij} \right] + \sum_j s_{0j}^* \right\}
\]

\[
s.t. \quad \Pr\{s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij})\} \geq \zeta', \quad \forall i, j,
\]

\[
s_0 = S - \min \{\min_i \{s_i + \min_j s_{ij}\}, c_0^*\},
\]

\[
s_0, s_i, s_{ij} \text{ are non-negative integers}, \quad \forall i, j.
\]
By the same logic as in §4.2, we can decompose this problem into multiple problems. Define Problem \((P_2')_i\) for \(i = 1, 2, ..., I\) as follows,

\[
(P_2'_{i}) \quad P_{2i}^{OPT} = \min \{S + \sum_{i \neq 1}[s_i + \sum_{j} s_{ij}] + \sum_{j \neq j^*} s_{ij} + \sum_{j} s_{0j}^*\}
\]

\[s.t. \quad \Pr\{s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij})\} \geq \zeta', \quad \forall ij,\]

\[s_i + \min_j s_{ij} \leq s_i + \min_j s_{ij}, \forall i \neq \bar{i}, j\]

\[s_i, s_{ij} \text{ are non-negative integers,} \quad \forall i, j,\]

where \(\bar{j} = \arg \min_j s_{ij}\). We also define Problem \((P_2')_0\) as follows,

\[
(P_2'_{0}) \quad P_{20}^{OPT} = \min \{S + \sum_{i} [s_i + \sum_{j} s_{ij}] + \sum_{j} s_{0j}^* - c_0^*\}
\]

\[s.t. \quad \Pr\{s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij})\} \geq \zeta', \quad \forall ij,\]

\[c_0^* \leq s_i + \min_j s_{ij}, \forall i, j\]

\[s_i, s_{ij} \text{ are non-negative integers,} \quad \forall i, j.\]

Define Problem \((P_2')\) such that \(P_2^{OPT} = \min\{\min_i P_{2i}^{OPT}, P_{20}^{OPT}\}\). By the same logic of Proposition 5, the optimization problems \((P_1')\) and \((P_2')\) are equivalent. For simplicity, we omit the proof. To solve Problem \((P_2')\), we slightly modify the algorithm of §4.2 as follows:

1. Solve direct sites for \(s_{0j}^*, \forall j\) independently of all regional warehouses, and calculate \(c_0^* = \min_j s_{0j}^*\).

2. Solve \((P_2'_{i})\) for all \(i = 1, 2, ..., I\).

3. Solve \((P_2'_{0})\) which is equivalent to solving \((P_i(c_0^*))\) for all \(i = 1, 2, ..., I\).

4. Compare Steps 2 and 3 to obtain the minimum.

Following a similar analysis of §4.2, the computational complexity of Step 2 (or 3) of the above algorithm is at most \(O(I^2 \ast S^3 \ast J)\) (\(O(I \ast S^2 \ast J)\) respectively). Step 1 takes a time at most proportional to \(J_0 \ast S\). Thus the computational complexity of this algorithm is at most \(O(I^2 \ast S^3 \ast J)\) which is equivalent to that of the regional warehouse only network.

Before we conclude this section, we must point out that the solution found by the optimization algorithm here is “optimal” based on the two-moment approximation of Graves (1985). If we replace the approximation by an exact calculation (e.g., Axsater 1990), the above optimization algorithm shall find the true optimal solution.
5 Heuristics and Bounds

In this section, we develop computationally efficient heuristic algorithms to compute bounds for Problem \((P_0)\). One simple heuristic \((H_1)\) is to solve the following problem for each regional warehouse \(i = 1, 2, ..., I\),

\[
\begin{align*}
\min & \quad \{s_i + \sum_j s_{ij}\} \\
\text{s.t.} & \quad \Pr\{s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij})\} \geq \zeta', \quad \forall j \\
& \quad s_i, s_{ij} \text{ are nonnegative integers,} \quad \forall j.
\end{align*}
\]

(9)

Then we compute \(s_0 = S - \min\{\min_i\{s_i + \min_j s_{ij}\}, \epsilon_0^*\}\). The complexity of \((H_1)\) is at most \(O(I \cdot S^2 \cdot J)\). \((H_1)\) cannot guarantee to find the optimal solution because by Proposition 4, the objective of one regional warehouse subsystem is to minimize site inventory while in this heuristic, the objective is to minimize total inventory in each regional warehouse subsystem. We will test the effectiveness of \((H_1)\) in §6.

To develop better upper bounds and a lower bound, we consider the regional warehouse only network in §4.2 again. Replacing \(s_0\) by its constraint in the objective function of Problem \((P_1)\), and suppose that we know in the optimal solution \(i = \arg\min_i\{s_i + \min_j s_{ij}\}\) and \(j = \arg\min_j\{s_{ij}\}\), \((P_1)\) becomes:

\[
\begin{align*}
\min & \quad \{S + \sum_{i\neq i}[s_i + \sum_j s_{ij}] + \sum_{j\neq j} s_{ij}\} \\
\text{s.t.} & \quad \Pr\{s_{ij} - B_{ij}(s_i) > D_{ij}(L_{ij})\} \geq \zeta', \quad \forall ij \\
& \quad s_i, s_{ij} \text{ are nonnegative integers,} \quad \forall i, j.
\end{align*}
\]

(10)

This formulation shows that the problem is decomposable if we know \(i\) in advance. Although we do not know \(i\) before we solve the problem, we can enumerate over \(i = 1, 2, ..., I\) and design the following heuristic algorithms to identify lower and upper bounds for Problem \((P_1)\).

**Lower Bound Solution**

1. Solve Problem (9) for \(i = 1, 2, ..., I\), and save the optimal value of the objective functions.

2. Solve Problem (6) for \(i = 1, 2, ..., I\), and save the optimal value of the objective functions.

3. Assume \(i = \arg\min_i\{s_i + \min_j s_{ij}\}\) for \(i = 1, 2, ..., I\), compute the system inventory by Step 2 for \(i\) and by Step 1 for all other \(i\).

4. Enumerate over \(i = 1, 2, ..., I\) to find the one with minimum objective function.
This solution clearly provides a lower bound for Problem (P1). This is true because if the condition $\tilde{i} = \arg\min_i \{s_i + \min_j s_{ij}\}$ always holds, one cannot do better than this solution. However, the solution found may violate this condition. When this condition $\tilde{i} = \arg\min_i \{s_i + \min_j s_{ij}\}$ holds in the solution found, then solution is also feasible and thus indeed optimal. The complexity of this lower bound solution is $O(I \times S^2 \times J)$.

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

Upper Bound Solution

1. Solve Problem (9) for $i = 1, 2, ..., I$, and save the optimal value of the objective functions.

2. Assume $\tilde{i} = \arg\min_i \{s_i + \min_j s_{ij}\}$ for $\tilde{i} = 1, 2, ..., I$, compute the system inventory by Step 1 for all $i \neq \tilde{i}$. For $\tilde{i}$, we solve Problem (6) but with an additional constraint $\tilde{i} = \arg\min_i \{s_i + \min_j s_{ij}\}$.

3. Enumerate over $\tilde{i} = 1, 2, ..., I$ to find the one with minimum objective function.

This solution may not be optimal but it is indeed feasible and thus provides a upper bound for Problem (P1). The complexity of this upper bound solution is $O(I \times S^2 \times J)$.

For the general network (see Figure 1), we have the following generalized lower bound solution.

Generalized Lower Bound Solution

1. Solve for $s^*_{0j}$ for all direct sites $0j$ and compute $c_0^*$.

2. Solve Problem (9) for $i = 1, 2, ..., I$, and save the optimal value of the objective functions and optimal solutions.

3. Solve Problem (6) for $i = 1, 2, ..., I$, and save the optimal value of the objective functions and the optimal solutions.

4. If $c_0^* < \min_i \{s_i + \min_j s_{ij}\}$ where $s_i, s_{ij}$ are the solution of Step 2, then output this solution, Done. Otherwise, assume $c_0^* = \min\{\min_i \{s_i + \min_j s_{ij}\}, c_0^*\}$, compute the system inventory by Step 2 for all $i = 1, 2, ..., I$, and go to next step.

5. Assume $\min_i \{s_i + \min_j s_{ij}\} < c_0^*$ and $\tilde{i} = \arg\min_i \{s_i + \min_j s_{ij}\}$ for $\tilde{i} = 1, 2, ..., I$, compute the system inventory by Step 2 for all $i \neq \tilde{i}$ and by Step 3 for $\tilde{i}$.
6. Compare Steps 4 and 5 to find the one with minimum objective function.

Clearly, this solution provides a lower bound for Problem \((P0)\). The solution is indeed optimal if \(c^*_0 < \min_i \{s_i + \min_j s_{ij}\}\) in Step 4, or for a \(\tilde{i}\), \(s_i + \min_j s_{ij} = \min \{\min_i \{s_i + \min_j s_{ij}\}, \min_j s_{0j}\}\) in Step 5. The algorithm runs in polynomial time with a complexity at most of \(O(I \ast S^2 \ast J)\).

We can also construct a generalized upper bound solution as follows.

**Generalized Upper Bound Solution**

1. Solve for \(s^*_{0j}\) for all direct sites 0\(j\) and compute \(c^*_0\).

2. Solve Problem (9) for \(i = 1, 2, ..., I\), and save the optimal value of the objective functions and the optimal solutions.

3. Solve Problem (6) for \(i = 1, 2, ..., I\), and save the optimal value of the objective functions and the optimal solutions.

4. If \(c^*_0 < \min_i \{s_i + \min_j s_{ij}\}\) where \(s_i, s_{ij}\) are the solution of Step 2, then output this solution, Done. Otherwise, solve \((P2'_0)\) which is equivalent to solving \((P_i(c^*_0))\) for all \(i = 1, 2, ..., I\), and go to next step.

5. Assume \(\tilde{i} = \arg\min_i \{s_i + \min_j s_{ij}\}\) for \(\tilde{i} = 1, 2, ..., I\), compute the system inventory by Step 2 for all \(i \neq \tilde{i}\). For \(\tilde{i}\), we solve Problem (6) but with additional constraints \(\tilde{i} = \arg\min_i \{s_i + \min_j s_{ij}\}\) and \(s_i + \min_j s_{ij} < c^*_0\).

6. Enumerate over \(\tilde{i} = 1, 2, ..., I\) to find the one with minimum objective function.

This solution may not be optimal but is feasible and thus provides a upper bound for Problem \((P0)\). The complexity of this algorithm is at most \(O(I \ast S^2 \ast J)\).

6 Numerical Analysis

The objective of this section is two-fold: (1) To test the efficiency of the solution algorithms and quality of the bounds, (2) to study the impact of lead times and recruiting rates on supply chain configurations.
6.1 Solution Algorithm Effectiveness

In this section, we report our observations on the computational performance of the proposed heuristics against the optimization algorithm in §4. All the problems were solved on a Dell computer (Dell Precision T3400, Intel Core2 Extreme Processor with CPU X9650 3.00GHz, 8GB RAM). To test the effectiveness of our solution algorithms, we examine a three-country trial. In each of the three countries, we consider two options: either setting up a regional warehouse or shipping directly. In each country the recruitment rate of patients, $\lambda$, is disaggregated among four sites such that $\lambda_1/\lambda = 0.1$, $\lambda_2/\lambda = 0.2$, $\lambda_3/\lambda = 0.3$, and $\lambda_4/\lambda = 0.4$ (similar to the test scenario of Graves 1985). The patient recruitment rate, $\lambda$, of the three countries is 0.4, 0.7, and 1 patient arrivals per day respectively (i.e. one slow recruitment rate country like the United States, one medium recruitment rate country, and one fast recruitment rate country like India, Russia, or China).

Lead time to the regional warehouse, $L_i$, may be 1, 7, or 14 days, and the lead time from warehouse to clinical trial site, $L_{ij}$, can be 1, 3, or 5 days. Since the advantage of a regional warehouse is speed in resupply to the sites, we place restrictions on the combinations of regional warehouse lead times and warehouse-to-site lead times that we investigate. Specifically, if $L_i$ is 1 day, then the warehouse to the site lead time $L_{ij}$ can only be 1 day. If $L_i$ is 7 days, then $L_{ij}$ can be either 1 or 3 days. If $L_i$ is 14, then $L_{ij}$ can be 1, 3 or 5 days. For simplicity, we assume identical shipping time from a regional warehouse to all its sites. If we ship directly to sites in a country without a warehouse, we assume that the total lead time to a site in that country equals to the sum of the lead time to the warehouse in that country and half the lead time from the warehouse to the site. In other words, we assume that the lead time of getting direct shipments from the central warehouse to a site is shorter than the total lead time when using a regional warehouse configuration.

We assume that the patient horizon, $S$, is 150 patients – a typical number for phase II clinical trials. Thus, the expected patient recruitment time is about 70 days which is significantly longer than the maximum direct lead time, 16.5 days. In summary, each country has 12 possible combinations of lead times and supply chain configurations (6 cases for using a warehouse and 6 cases for direct shipping). Thus, we have totally 1728 (i.e. $12^3$) cases examined.

The observations on the computational performance of the proposed heuristics are summerized in Tables 2 and 3. In these tables, “LB” stands for the lower bound solution, “UB” stands for the upper bound solution, “(H1)” stands for the simple heuristic (H1), and “OPT” refers to the
optimal solution. Table 2 shows the absolute percentage error between the objective values (i.e., the total inventory in system) of the lower bound (LB), upper bound (UB) and \((H1)\) against the optimal solution (OPT) that we obtained using the optimization algorithm in §4. In these tables, we ignore scenarios with no regional warehouses as they are straightforward to solve (see §4.1).

<table>
<thead>
<tr>
<th># of Regional Warehouses</th>
<th>% Gap LB</th>
<th>% Gap UB</th>
<th>% Gap ((H1))</th>
<th># of Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.8</td>
<td>0</td>
<td>0.02</td>
<td>648</td>
</tr>
<tr>
<td>2</td>
<td>3.8</td>
<td>0</td>
<td>0.07</td>
<td>648</td>
</tr>
<tr>
<td>3</td>
<td>4.5</td>
<td>0.025</td>
<td>0.82</td>
<td>216</td>
</tr>
</tbody>
</table>

Table 2: Performance gaps between heuristics and optimal solutions

Table 3 compares the running times of the proposed heuristic lower bound, upper bound and \((H1)\) algorithms against the optimization algorithm.

<table>
<thead>
<tr>
<th># of Regional Warehouses</th>
<th>Average CPU time (seconds)</th>
<th># of Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LB</td>
<td>UB</td>
</tr>
<tr>
<td>1</td>
<td>5.18</td>
<td>6.44</td>
</tr>
<tr>
<td>2</td>
<td>32.92</td>
<td>34.29</td>
</tr>
<tr>
<td>3</td>
<td>53.63</td>
<td>52.89</td>
</tr>
</tbody>
</table>

Table 3: Running times of heuristics and optimization algorithms

From these two tables, we conclude that (1) the upper bound is much closer to the optimal solution in the objective values than the lower bound. (2) Although the simple heuristic \((H1)\) is not as good as the upper bound in terms of the objective values, it is faster and its objective values are surprisingly close to those of the upper bound. (3) The performance gaps between the heuristic solutions and the optimal solution tend to increase as the number of regional warehouse subsystems increases. (4) The optimization algorithm takes the longest time to solve while the heuristic \((H1)\) takes the least time. In summary, it seems that the heuristic \((H1)\) can be a good alternative (to the optimization algorithm) for solving practical problems because of its computational efficiency, its small performance gap from the optimal solution, and its programmatic simplicity.
6.2 Supply Chain Configuration

Given the effectiveness of our optimization algorithm, we now leverage these solutions to develop insight into the optimal configuration decisions. Specifically, we focus on the impact of recruitment rates and lead times on supply chain configuration (i.e., using a warehouse or shipping directly for a country). Using the same parameters as those in §6.1 we look at 216 possible country/site lead time combinations. In addition, for each country, we make a decision as to whether to use a regional warehouse in that country or to ship directly to the sites. Thus, for each country/site lead time combination, we solve eight (i.e. $2^3$) possible warehouse and direct configurations to find their optimal inventory decisions, and we choose the best configuration defined to be the one with the lowest total inventory. If there is a tie between direct shipping and using a warehouse for a country, we always choose direct shipping as it reduces the logistic cost of setting up and running a warehouse.

To quantify the impact of lead times, we define the following notation to measure the speed of resupply advantage that is afforded by using a regional warehouse as opposed to shipping direct.

$$\text{LT ratio} = \frac{\text{Lead time from CW to a country’s regional warehouse}}{\text{Lead time from CW to site under direct shipping option}}.$$ 

The numerical results are summarized in Figure 3 where the numbers on chart refer to the numbers of countries using a certain option.

From Figure 3, we see two general trends: (1) As recruitment rate increases, it is more likely to utilize the direct shipping option rather than a warehouse. (2) As the LT ratio increases, it is more likely to utilize a warehouse rather than the direct shipping option. While most of our numerical results confirm these intuitive trends, there is an outlier: For medium LT ratio, when the recruitment rates are 0.7 and 1, the warehouse/direct ratios are 86/22 and 97/11 respectively. The two general trends do not capture all the scenarios due to the interaction among different countries because the optimal inventory decisions at different regional warehouse subsystems are determined jointly rather than independently.

In practice, regional warehouses are often recommended for countries/regions far away from the central warehouse. An article in *Applied Clinical Trials* (June 1, 2009) offered this rule of thumb:

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

The insights generated by our numerical study agree with this rule of thumb and make it more precise by demonstrating how the LT ratio (rather than just the lead time to a country/region) affects the choice of a warehouse or direct shipping.

From a practical perspective, the impact of recruiting rate on choosing a supply chain configuration seems somewhat counter-intuitive. In the slow recruiting country, a warehouse is recommended for a majority of scenarios. For the fastest recruiting country, the opposite is true; a warehouse is less likely to be recommended in the optimal configuration. In practice, it is often recommended to consider putting a regional warehouse in countries with long lead times and the reason these countries are even considered to be included in the trial in the first place is often because of their fast recruitment. So interestingly, in countries where we’d expect a warehouse to almost definitely be included as part of the optimal configuration, the numerical study may indicate otherwise. This seemingly counter-intuitive result relies on the fact that long lead time alone does not warrant the choice of a warehouse. For countries with long lead time and fast recruitment rate, the decision on whether to ship directly or to use a warehouse depends on the intricate balance between the LT
ratio and recruitment rate, as well as interactions with other countries.

7 Summary Remarks

In this paper, we have illuminated the supply issues that spreading clinical trial demand over multiple sites can cause. Specifically, getting the right inventory, to the right place, at the right time becomes increasingly difficult as both countries and sites are added to a clinical study. We have constructed a mathematical model for the general clinical trial supply chain that allows one to optimally position inventory in support of the trial’s service level requirements. We develop a strongly polynomial algorithm to solve this model up to optimality. To aid implementation of this model, we also presented heuristic algorithms that reduce the complexity of the optimization algorithm without greatly sacrificing the performance. Using these algorithms and a representative example, we obtain some insights into configuring the supply chain.

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

1. Stochastic Lead Times: For the model presented here, finding mechanisms to incorporate stochastic lead times (either i.i.d. or sequential) and study their impact on the optimal supply chain configuration would be a valuable way to expand the scope of its utility. This is especially true when the reliability of shipments to clinical trial sites becomes an issue due to random regulatory clearance time and shipping time over large distances.

2. General Demand Patterns: It is of great interest to test the assumption of Poisson patient recruitment process in practice and generalize the model to handle random dropouts and non-Poisson recruitment processes.

3. Batch Ordering Policy: In this paper, we analyze the clinical trial supply chain under the assumption of one-for-one ordering policies. Expanding these models to include batch ordering policies and shipping costs is one other area worthy of future study.

4. Applications of Models: While the model presented in this paper illustrates enormous oppor-
tunities to provide value for clinical trial supply chains, applying this model in the real-world practice would help refine and validate it.

Acknowledgments

We thank Dr. Indranil Nandi from Sandoz Inc., A Novartis Company, Yevgeniy Krol from Schering Plough, and Dr. Yanshui Wang from Merck for their wealth of information provided for this study. The first author is supported in part by a research grant from the Lerner Center for Pharmaceutical Management Studies at Rutgers University.

References


