Research Article

Minimization of Drug Shortages in Pharmaceutical Supply Chains: A Simulation-Based Analysis of Drug Recall Patterns and Inventory Policies

Rana Azghandi 1,1 Jacqueline Griffin 1,1 and Mohammad S. Jalali 2

1 Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, USA
2 Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA

Correspondence should be addressed to Jacqueline Griffin; ja.griffin@northeastern.edu

Received 21 May 2018; Revised 12 September 2018; Accepted 28 October 2018; Published 2 December 2018

Copyright © 2018 Rana Azghandi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The drug shortage crisis in the last decade not only increased health care costs but also jeopardized patients’ health across the United States. Ensuring that any drug is available to patients at health care centers is a problem that official health care administrators and other stakeholders of supply chains continue to face. Furthermore, managing pharmaceutical supply chains is very complex, as inevitable disruptions occur in these supply chains (exogenous factors), which are then followed by decisions members make after such disruptions (internal factors). Disruptions may occur due to increased demand, a product recall, or a manufacturer disruption, among which product recalls—which happens frequently in pharmaceutical supply chains—are least studied. We employ a mathematical simulation model to examine the effects of product recalls considering different disruption profiles, e.g., the propagation in time and space, and the interactions of decision makers on drug shortages to ascertain how these shortages can be mitigated by changing inventory policy decisions. We also measure the effects of different policy approaches on supply chain disruptions, using two performance measures: inventory levels and shortages of products at health care centers. We then analyze the results using an approach similar to data envelopment analysis to characterize the efficient frontier (best inventory policies) for varying cost ratios of the two performance measures as they correspond to the different disruption patterns. This analysis provides insights into the consequences of choosing an inappropriate inventory policy when disruptions take place.

1. Introduction

1.1. Background. Between 2006 and 2016, drug shortages in the United States increased a whopping 120%, with the highest peak of 280% in 2011 [1]. In 2017, 146 new shortages were added to the list of nonresolved shortages from prior years, and by the end of 2017, there were 183 active shortages [1]. Drug shortages not only bring the cost of health care above $400 million annually [2] but can also be considered a public health threat since patient lives are put at risk. Despite all the efforts undertaken by health care administrators to ensure product availability, there are still drugs that have experienced shortages for several years [3].

One drug that consistently creates challenges for health care providers is saline (sodium carbonate 9%) which has experienced shortages since 2013. In the United States, saline has been widely used (more than 40 million bags per month [4]) for treating dehydration and for patients undergoing dialysis, surgery, and chemotherapy. In the past decade, shortages of saline have raised questions about the ability to manage the pharmaceutical supply chain. What makes saline interesting to study is its low price (approximately US $4 for a 250ml bag of saline [5]), simple ingredients (only salt and water), high demand, exclusive production by certain manufacturers, and in some cases nonsubstitutability [6]. The US Food and Drug Administration (FDA) announced an ongoing national shortage of saline in early 2014. Several disruptions occurred around this time in the saline supply chain, e.g., multiple recalls and spikes in demand due to flu seasons [7], and the FDA attempted to mitigate these shortages by advising manufacturers to collaborate with each other and importing saline from other countries such as...
Norway and Germany. Despite these efforts, shortages in the United States persisted for several months [7]. The failure of these two approaches can be explained by the fact that: (1) the FDA cannot dictate the production changes by manufacturers, and (2) the import of saline from new sources requires several months of regulatory processing. Compounding the latter challenge, outsourcing was limited since manufacturers in other countries have limited capacities to expand their production.

The saline shortage is getting worse, and its supply chain was not fully recovered when a sudden severe disruption in the production of saline occurred after Hurricane Maria, in Puerto Rico in 2017 [8]. The hurricane shut down two of the United States’ main manufacturers of saline for several months and created a major disruption in the supply chain. A combination of this disruption with the other preexisting shortages amplified the shortages even more. Following that, the shortage increased as another producer interrupted its production for maintenance purposes. Interestingly, this was announced several months in advance, yet there remained no clear plan to stockpile the inventories to bolster the supply chain. Additionally, due to high rates of influenza during the 2017 flu season, demand for saline increased and drove further shortages of saline in most of the United States [8].

Despite the direct effect of these shortages on patients, health care providers faced the challenge of deciding who should receive the limited supply of saline and how to procure more units of saline, if any were available.

Commonly identified causes of drug shortages include complex manufacturing processes, supply or demand uncertainties, regulatory actions, and the discontinuation of products [9]. Furthermore, quality issues or a lack of incentives for manufacturers to produce high-quality products are other known reasons for drug shortages [9]. However, 53% of the reasons behind the shortages of drugs in 2017 designated by the American Society of Health-System Pharmacists (ASHP) database are labeled as unknown [1]. In fact, these shortages may be explained by a combination of these reasons in conjunction with the effects of decision making by supply chain stakeholders.

In general, pharmaceutical supply chains mirror other supply chains that are exposed to the risk of disruptions. Previous research has indicated that because drug shortages are caused by one or multiple disruptions in the pharmaceutical supply chain, preventing these disruptions can be nearly impossible [6]. The reasons for recalls can vary including mislabeling, defective products, or defects in the container. When product recalls occur in a pharmaceutical supply chain, the inventory of some or all of the supply chain members is removed. While product recalls are common for pharmaceutical products, the characteristics and patterns of recalls can differ between products both in terms of the frequency of shortages and the size of each recall, or the total number of units affected. Beyond recalls, pharmaceutical supply chains can also experience temporary shutdowns of manufacturing plants for multiple months on end. Designing a supply chain that remains resilient to different types of disruptions is among the major challenges researchers in supply chain risk management must tackle.

One common solution for mitigating shortages is to keep more inventory [10], as this will ensure that more inventories are available to treat patients in the event that products are recalled. However, retaining more inventory results in added costs in the supply chain and may not be feasible for many health care centers due to storage and budget limitations. With the need to balance the advantages and costs of holding inventory, strategic decisions about inventory policies throughout the supply chain become critical.

### 1.2. Research Problem and Contribution

The importance and complexity of supply chain management in the pharmaceutical industry raises critical questions that are not addressed in the literature. In this paper, we focus on answering the following questions. What is the effect of altering inventory policies on the ability to withstand product recalls? How should inventory policies be chosen to address a variety of disruption patterns? Is there any inventory policy that is robust to all disruption patterns?

In order to address these questions, we evaluate the performance of inventory policies such that the policies defining order quantities and target inventory levels do not change over time. Further, we evaluate the performance under various disruption patterns. We model product recall disruptions in pharmaceutical supply chains under deterministic and stochastic settings which are more realistic for pharmaceutical supply chains. To demonstrate the complexity of the pharmaceutical supply chains under disruptions including interactions among key decision makers, we employ a mathematical simulation model accounting for the dynamic features of the system. Different inventory policies are implemented as decision rules in the simulation. As the most critical consequences of drug shortages are experienced at the health care centers with unmet patient demand, the performance metric used to distinguish between policies is the total cost of the health care center. We also employ an approach similar to data envelopment analysis to prune inefficient policies corresponding to different disruption patterns and the ratio of costs. Finally, we aim to demonstrate the effects of making unsuitable inventory policy decisions under different types of disruptions and explore if an inventory policy exists that is efficient for all types of disruptions.

While the present study focuses on saline shortages, it should be noted that the findings are applicable to a wider range of drugs that frequently experience shortages, such as chemotherapy drugs and antibiotics, among others. The results of the presented research reveal that an inventory policy that is optimal for one disruption pattern may be inefficient when an alternate disruption pattern occurs. Therefore, preparing for one intense disruption will not render a system robust if multiple small disruptions happen instead.

The remainder of this paper is organized as follows. In the next section, we provide a review of the literature. Our model formulations and assumptions are then presented in Section 3, followed by numerical analyses in Section 4. Finally, Section 5 presents the conclusion and future research directions.
2. Literature Review

In order to address our research questions, we divide the literature areas into three sections: pharmaceutical supply chain modeling, supply chain disruption mitigation, and simulation modeling for supply chains.

**Pharmaceutical Supply Chain Modeling.** In recent years there has been a noticeable focus on researching pharmaceutical supply chain management. A study by Lücke and Seifert [10], focuses on operational strategies for managing pharmaceutical supply chains under disruption. They model disruptions with features defining the probability of the occurrence and intensity in length. They analyze three operational risk mitigation strategies for pharmaceutical supply chains experiencing disruption risks, namely (i) Risk Mitigation Inventory (RMI), (ii) Dual Sourcing and (iii) Agility Capacity, with a focus on minimizing the total cost, as a function of product stock-outs and fulfilled demand, over only one cycle [10]. While Lücke and Seifert’s [10] study considers one type of drug, Uthayakumar and Priyan [11] model an inventory decision tool for a two-echelon pharmaceutical supply chain with multiple products without considering disruptions in the system. They use continuous ordering policies to optimize and define the lot size, lead-time, and the number of deliveries for a hospital to minimize the total cost for the supply chain with respect to the target cycle service level for the hospital or pharmaceutical companies.

Overall, research in pharmaceutical supply chain disruptions has neglected to examine the role of product recalls, a critical type of disruption in this domain [10–15]. We develop a model to capture the role of various recall disruptions in the pharmaceutical supply chain.

**Supply Chain Disruption Mitigation.** While the focus on disruption mitigation in pharmaceutical supply chain management research is limited in the scope of the types of disruptions examined, research on modeling supply chain disruptions and response strategies has been conducted for other industries. There are many different approaches for mitigation of disruption risks that have been examined. Common tactics for alleviating the effects of disruptions are financial mitigation, operation mitigation, and operational contingencies [16]. Mitigation tactics are those in which the firm takes an action in advance of a disruption and correspondingly incurs the cost of the action regardless of a disruption’s occurrence. Operational tactics, such as inventory management, multiple sourcing, and flexibility in production are also studied in the literature of supply chain disruptions [16].

Researchers in the supply chain disruption area, demonstrate that an optimal strategy to implement when coping with a disruption varies based on the characteristics of the disruption. Tomlin [16] suggests inventory, dual sourcing, and acceptance strategies for dealing with disruptions, and demonstrates that the optimal strategy changes as disruptions become longer or more frequent. In this research, disruptions are modeled as an uncertainty in supply or demand. Tomlin and Snyder [17] demonstrate how strategies change when a firm has advanced warning of an impending disruption.

Chopra et al. [18] evaluate the errors resulting from “bundling” disruptions and yield uncertainty for making inventory decisions. Atan et al. [19] present hybrid strategies for dealing with supply uncertainty or demand uncertainty in multiechelon supply chains. They consider disruptions which occur at a specific location and demonstrate that an optimal policy for addressing a supply uncertainty disruption is not the same as the optimal policy for addressing a demand uncertainty disruption.

A variety of different features have been used by researchers for characterizing and modeling supply chain disruptions. The majority of disruptions are modeled as follows: (i) all-or-nothing events (also known as on and off disruptions, in which a disruption at a node will deplete the entire capacity) or (ii) random yield disruptions [20–24]. Most studies focus on large disruptions that happen for a short time duration. Some studies look at the different ordering policies and frequency of the disruptions, e.g., frequent but short disruptions versus rare but long disruptions [16]. To the best of our knowledge, there are no studies that examine the sequence of the disruptions.

In this study, we examine operational mitigation tactics which characterize the role of inventory policies in the prevention of shortages by studying the relationship between the performance of these policies and the disruption pattern exhibited. Rather than all-or-nothing disruptions or random yield disruptions, this research is unique in the focus on removing a fraction of inventory from nodes throughout the network.

**Simulation Modeling for Supply Chains.** In pharmaceutical supply chain disruptions, and particularly in cases of product recalls, inventory management behaviors are believed to be key to the behavior of the entire system. Often theoretical models in Operations Research misrepresent the ‘behavioral’ aspects of supply chain stakeholders and focus solely on the supply chain mechanisms. Unlike many of these models, system dynamics (SD) simulation modeling is able to represent interactions, nonlinearities, time delays, and feedback among the members of the supply chain [25–27]. Sterman [28] develops a simulated inventory distribution system for the beer game distribution supply chain which contains multiple actors. He examines the interaction of individual decision makers who make irrational decisions with the objective of minimizing the total supply chain costs and demonstrates the resulting bullwhip effect in the system. Lee et al. [29] demonstrate that even if each agent acts rationally and uses optimal ordering policies, the supply chain is still at the risk of experiencing the bullwhip effect.

Some studies using SD as a tool to model food or drug supply chains include Minegishi and Thiel [30], Georgiadis, Vlachos, and Iakovou [31], and Strohecker and Größler [14, 15]. Minegishi and Thiel [30] study the complexities of the behavior of poultry supply chains. They examine different simulations to reveal the role of changes in the demand on the behavior of the supply chain. Georgiadis, Vlachos, and Iakovou [31] develop an SD model for a multiechelon food supply chain. In this study, they try to identify effective long-run policies for managing a food supply chain under
uncertainty of demand and transportation capacity using a periodic order-up-to-level policy (R,S,s) when demand has a normal distribution. Strohhecker and Größler [14] study the effects of severe production shutdowns on the performance of pharmaceutical products. In separate work, Strohhecker and Größler study the effect of severe but infrequent quality breakdowns, as disruptions, on the conditions for preventing stock-outs by adapting safety stock levels using a SD simulation [15]. Among the SD-based studies, there is only one study which examines the performance of the system when there are transportation disruptions at two different locations in the supply chain and it compares the results for a traditional supply chain with a vendor managed inventory supply chain [32].

Prior research on supply chain disruptions, including those focused on pharmaceutical supply chains, reveals the importance of distinguishing the disruption features and identifying operational mitigation policies that address these characteristics. Despite this, minimal research examines supply chain performance as it pertains to (i) pharmaceutical product recalls and the unique features of recalls, (ii) patterns of disruptions that account for disruptions affecting multiple nodes and multiple disruptions occurring over time, and (iii) the inventory management policies and behaviors that drive the dynamics of the system.

To address these features, we employ a mathematical simulation model that accounts for the dynamics underlying the system. While there are a limited number of models which analyze the complexities of supply chains with disruptions, this is the first simulation model that focuses on product recalls and the role of disruption patterns in pharmaceutical supply chain.

3. Mathematical Modeling

We integrate two analytical methods to examine the performance of inventory policies in pharmaceutical supply chains experiencing disruptions: simulation modeling and data envelopment analysis (DEA) approaches. To identify the optimal performance—which is a function of the inventory policy corresponding to the safety stock level, inventory holding costs, and shortage costs—for varying disruption patterns, we populate data points from the simulation results. Using the simulation results, an approach derived from DEA is used to identify the optimal inventory policy for various types of disruptions and relative values of shortage and inventory holding costs.

3.1. Model Structure. The pharmaceutical supply chain that is presented in this paper has a serial network structure consisting of four echelons: manufacturer, distributor, wholesaler, and health care center with a single pharmaceutical product being distributed across the system over a finite time horizon. The assumption of considering the pharmaceutical supply chain as including four echelons is in alignment with the main players identified by the FDA [32]. Additionally, this is similar to the assumptions made in other models of these supply chains [33, 34]. The formulation of this model is adapted from the beer distribution game simulated by Sterman [28] and Croson and Donohue [33].

The supply chain members, each corresponding to one echelon, are denoted as i = 1, . . . , 4 for the health care center, wholesaler, distributor, and manufacturer, respectively. In each time period t (t = 1,2,3,...,260) representing a week during a 5-year period, each supply chain member will receive product shipments from the supply chain member immediately upstream, resulting in increased inventories, and will receive orders for products from supply chain members immediately downstream. Since there is no upstream echelon for the manufacturer, the inventory of the manufacturer is increased by the quantity that is produced in the previous period. Moreover, there is no downstream node for the health care center. Hence the health care center should satisfy the patients’ demand in each period.

After satisfying orders, if there are not enough inventories available, the remaining orders are added to the backlog of each member, except for the health care center, and must be satisfied in the next period. The unsatisfied demand for the health care center is lost, resulting in a shortage. At the end of each period, each supply chain member, other than the manufacturer, will place an order for more products with the supply chain member immediately upstream. The manufacturer will determine how much to produce in the next period, similar to an order. In this model, we assume all members use a periodic review inventory model with zero fixed costs when determining how much to order in each period. This order quantity is dependent on two features: cycle stock (for satisfying the expected demand) and safety stock (as extra inventory to buffer against uncertainty). Cycle stock is calculated as the mean of lead-time demand, and safety stock is a function of the standard deviation of lead-time demand and the target cycle service level (CSL). The cycle service level can be defined as the expected percentage of order cycles in which no shortage occurs. Each member wants to keep safety stock to address the uncertainty related to demand fluctuations during the lead-time.

Additionally, information about disruptions in the supply chain due to product recalls may become available to the affected member during any time period. If there is a product recall, some products will be removed from the inventory of the health care center, wholesaler, and distributor (in Section 3.2.2 we discuss why we do not consider removals from the manufacturer). Each member’s remaining inventory which was not recalled can be used to satisfy orders or patient demand.

We assume that the members of the supply chain make their ordering decisions based on a forecast of uncertain, but stationary, demand in each period. Lee et al. [29] indicate that even a rational ordering decision—which is changed dynamically—can lead to the bullwhip effect or the phenomena in which orders placed by downstream nodes to upstream nodes are variable over time and variability is amplified at members further upstream from the customer. There are four elements considered for the bullwhip effect in a supply chain: demand signal processing, rationing game, order batching, and price variations. Due to the lack of fixed ordering costs, price variability, and rationing ability in the presented model of these elements, only demand signal processing would be expected to cause the bullwhip effect phenomena.
3.2. Model Formulation

3.2.1. Ordering Decisions. The model used in this paper assumes that members place orders based on a periodic review policy with zero fixed costs. We follow the same notation as presented by Sterman [28]. The indicated order rate for echelon \( i \) \((IO_{it})\) is based on the anchoring and adjustment heuristic. In our model, we assume that each member is a rational decision maker and uses an approximately optimal local policy to place orders. Therefore, the indicated order rate is a function of demand, lead-time, service level, inventory level, and the order amount that has ordered but has not been received (on-order). The anchoring corresponds to the base stock \((S_a)\) over time which is locally optimal for a periodic review policy under certain assumptions. Furthermore, adjustments can be equivalent to the inventory position \( (IP_{it}) \), which is a function of on-hand inventory, on-orders (orders that have been placed but not yet received), backlog (health care center does not have backlog), and disruptions (as recalls).

\[
IO_{it} = S_a + IP_{it} \tag{1}
\]

As mentioned earlier each member is assumed to be a rational optimizer and uses a periodic review policy with zero fixed costs to place orders. The periodic inventory policy with base stock (order-up-to-level) is optimal under these assumptions:

1. demands are stationary,
2. the lead-time is fixed and there is no limitation on production,
3. there is no fixed ordering cost and there are no changes in the purchase cost of the product over time.

When using a periodic review policy, the base stock level at time \( t \) is given by

\[
S_a = \mu_{it}^L + z_a \sigma_a^{L} \tag{2}
\]

where \( \mu_{it}^L \) and \( \sigma_a^L \) are the mean and standard deviation of the stationary demand for echelon \( i \) over the fixed lead-time, respectively. Furthermore, \( z_a \) is denoted as the \( \alpha^{th} \) fractal of a standard normal distribution, known as the cycle service level (CSL). Also, it can be interpreted as the likelihood of being able to satisfy demand assuming no disruptions. In return, this drives the level of safety stock that is held to buffer against uncertainties in demand.

Since the distribution of the demand is unknown and may change in each period, we use the moving average forecasting technique to estimate the mean and standard deviation of the lead-time demand. This estimation may change periodically and adjust based on the demand for the previous \( n \) time periods. This process of updating demand forecasts is known as demand signal processing and can result in the bullwhip effect [34, Ch. 10]. Assuming that the demand is independent of previous periods, then, as demonstrated by Snyder and Shen [34], the estimator for the mean lead-time demand \( (\hat{\mu}_{it}^L) \) is

\[
\hat{\mu}_{it}^L = L \left( \frac{\sum_{j=1}^{n} D_{it-j}}{n} \right), \tag{3}
\]

where \( L \) is lead-time and \( D_{it-j} \) is demand at time \( t \).

The forecast error for the lead-time demand is

\[
\hat{\sigma}_{it}^L = k \sqrt{\frac{\sum_{j=1}^{n} (e_{it-j})^2}{n}} \tag{4}
\]

where \( k \) is a constant and assumed to be \( \sqrt{L} \). And the forecast error in time \( t \) is assumed to be the one-period forecast error as

\[
e_t = D_t - \mu_{it}^L. \tag{5}
\]

Following from the definition above, the base stock level at each period is defined as

\[
S_{it} = \hat{\mu}_{it}^L + z_a \hat{\sigma}_{it}^{L}, \tag{6}
\]

where it is approximately optimal by using the defined estimators to forecast demand.

The other element necessary to calculate for the indicated order is the inventory position of member \( i \) in time \( t \) \((IP_{it})\), which is defined as follows:

\[
IP_{it} = IP_{it-1} + a_{it-1} - D_{it-1} - R_{it-1} - B_{it-1} + On_{it}, \tag{7}
\]

where \( a_{it-1} \) is the acquisition amount in time \( t-1 \), \( D_{it-1} \) is the demand in time \( t-1 \), \( R_{it-1} \) is the quantity of recalled products that is removed from the inventory of member \( i \) in time \( t-1 \), \( B_{it-1} \) is the quantity of back ordered units in time \( t-1 \), and \( On_{it} \) is the quantity of on-order products expected to arrive in time period \( t \).

If the health care center does not have enough inventories to satisfy all patient demand, then the unmet demand is lost and not transferred to the next period, and a shortage will result. Shortage is defined as

\[
Shortage_{it} = D_{it} - (D_{it} \times Fulfillment ratio_{it}), \tag{8}
\]

where the fulfillment ratio is the percentage of demand at the health care center which is satisfied. However, for the rest of the supply chain members, the excess amount is added to the backorder quantity and must be satisfied in the subsequent periods.

Finally, the order amount should be nonnegative such that

\[
O_{it} = \max \left( 0, IO_{it} \right). \tag{9}
\]

3.2.2. Disruption Modeling. In the model, we examine supply chain disruptions due to product recalls, which are modeled as exogenous shocks that can affect the availability of inventory at all echelons of the system. As discussed earlier, a product recall happens when there are concerns related to the safety of products produced in the past. Consequently, the manufacturer enforces a voluntary recall of the affected product that has been already distributed within the supply chain. Because this process usually begins after some time has passed, we only consider recalls from three echelons, the distributor, the wholesaler and the health care center. It is possible to have multiple disruptions, or product recalls,
Disruptions as external interruptions have been studied in supply chain risk management as deterministic and stochastic events [16, 24, 35, 36]. However, to the best of our knowledge, there are no studies which model the effects of product recalls. To address the lack of modeling of pharmaceutical shortages, we develop deterministic and stochastic models of these types of disruptions which account for the possibility of multiple events over time and varying sizes of the disruptive events.

In deterministic scenarios, the hypothetical disruptions are defined in advance. There are multiple different elements that are considered for characterizing disruption patterns. Adapted from [37], the elements that are considered for characterizing a disruption are disruption periodicity, disruption breadth, disruption length, and disruption size. Figure 1 demonstrates each element. The periodicity of disruptions refers to the duration of time between disruption events. Disruption breadth refers to whether the disruption happens once or if there are multiple distinct disruptions. In this paper, we examine the effect of a single disruption versus multiple disruptions. Additionally, the length of the disruption refers to the duration of the time period during which the disruption occurs. In our model, we assume this element is either brief but severe in intensity or prolonged but moderately sized. The disruption quantity lost or the size of disruption refers to the total loss of the system due to disruption occurrence.

In addition to the deterministic set of disruptions, we model disruptions as exogenous stochastic events that happen during the simulation time horizon. We assume that the time between disruptions is variable and uncertain, but the occurrence of disruptions is modeled as a Poisson distribution with an average arrival rate lambda (λ). For the Poisson process with a mean of 1/λ events occurring in a specified time period, the time between events are independent and identically distributed with a continuous exponential distribution of mean 1/λ. In order to generate the random interarrival times with the rate of arrival as λ in a Poisson distribution, we use the algorithm presented in Algorithm 1.

In our model, we assume that the duration of recalls is greater than or equal to the lead-time to make sure that no defective product remains in the transshipment. The current simulation model is employed over multiple random scenarios which facilitates the comparison of the effects of recall periodicity and the size of disruptions on the shortages and inventory costs with varying inventory policies. Furthermore, in order to compare the different patterns of disruptions, the expected total amount recalled for all scenarios is assumed to be fixed. Another challenge for modeling recalls in both deterministic and stochastic settings is the need to ensure that there are enough defective products in the inventory compared to the recall amount which is removed, such that the inventory cannot be negative. Correspondingly, we assume that the amount that we recall from each echelon is proportional to the available inventory of that echelon.

The objective is to minimize the total costs of the health care center, as presented in (10), which consists of the summation of the total cost of shortages and the total cost for
holding inventory at the health care center where \( h \) is the per unit holding cost and \( p \) is the per unit shortage cost. The total cost varies based on the product recall disruption pattern and the inventory policy, as defined by the target cycle service level (CSL).

\[
\text{minimize} \quad hIL_{1t} + p \text{ Shortage}_{1t}.
\]  

(10)

3.3. Data Envelopment Analysis. Accessing the true value of the inventory costs and the shortage costs is challenging because health care centers are unlikely to share this information with the public. Moreover, if these costs are available for a particular drug, it may vary substantially in comparison to other pharmaceutical products. In order to assess the performance of the system, without knowledge of the true underlying costs of the system, we employ an approach derived from the concepts of the data envelopment analysis (DEA) approach. DEA is a data-driven approach which seeks to assess the relative performance of multiple systems, referred to as decision-making units (DMUs), accounting for multiple inputs and outputs without exact information about the relative costs and values associated with the inputs and outputs, respectively. Thus, a system, or DMU, is defined as efficient, in comparison to the others under consideration, if there is some set of costs of inputs and values of outputs that leads it to have the best, or most efficient, performance. A set of optimization problems are solved in order to determine which DMUs are denoted as efficient under some combination of costs and values. These DMUs are defined as making up the ‘efficient frontier’ [38, 39].

We employ an approach similar to DEA to evaluate which policies, as defined by a CSL, are efficient for some set of inventory costs and shortage costs, under different disruption patterns. The data created from the simulation is used as input into this approach. Using an approach similar to DEA with constant-returns to scale [38], we identify which policies, defined by the CSL value, are efficient, or perform best for some relative value of inventory and shortage costs. Correspondingly, we also identify which CSL values never perform best for any set of inventory and shortage costs, and therefore are denoted as inefficient. When plotting the outcomes of the policies, with respect to both the level of shortage and inventory, the efficient frontier can be estimated by connecting the efficient policies with isoquant lines. This efficient frontier can be interpreted as the set of inventory and shortage levels that are efficient for some pair of inventory and shortage costs.

4. Analysis

4.1. Data. Using the saline demand patterns as a basis for the model, we assume that demand follows a normal distribution with a mean of 10,000 bags per week and a standard deviation of 100 bags per week. In all experiments, we use a common random number seed to generate demand data. This allows for the isolation of variations due to ordering behavior and inventory policies rather than variations due to different demand streams in the comparison of the results.

As mentioned in Section 3.1, the simulated supply chain has a serial structure with four echelons (manufacturer, distributors, wholesalers, and health care centers). We assume that there is no limit on the capacity of the manufacturer to meet the demand. The simulated model is defined by several parameters. These parameters are assumed to be consistent across the duration of the entire simulation run, or 5 years. However, in order to capture the effect of inventory policies on the ultimate goal of minimizing shortages and limiting inventory costs as the health center, the CSL value which defines the inventory policy may change. While the CSL may vary among simulation runs, we assume that all echelons will have the same CSL. Table 1 provides a summary of the parameters and their values.

4.2. Disruption Pattern Comparison. As the characteristics defining product recalls can vary based on multiple characteristics, the goal of this section is to provide insights into the behavior of the supply chain defined in Section 4.1 under different disruption patterns while assuming that the inventory policies, as defined by the CSL, remain the same. For this purpose, we create five disruption scenarios (A-E) to examine the effects of disruption characteristics, presented in Figure 1, on the health care center behavior and the performance metrics.

Table 2 summarizes the characteristics of each scenario. To allow for easy comparison of these scenarios, we assume that all disruptions start at the same time (week 50), and that the total recall quantity remains the same for all of the scenarios, except Scenario C, which has a greater recall size.
The total recall amount corresponds to the area under the rectangles in Figure 1. Therefore, we define the variable \( \lambda \) as the number of disruptions (breadth), \( \gamma \) as the duration of each disruption, and \( \theta \) as the intensity of the disruption. The intensity of the disruption corresponds to the quantity of product effected, or recalled, during each period of time during the disruption. Equation (11) shows the relationship among these factors. In all scenarios, we assume that the total recall amount is equal to a percentage of the total demand during the simulation duration, or the percentage lost. The total recall amount is defined as

\[
\text{Total recall} = \lambda \gamma \theta. \tag{11}
\]

**Length of Disruption.** To examine the effect of the length of the disruption, Scenarios A and B are compared. While both scenarios have the same amount of total recalls, corresponding to 1% of the total demand, Scenario A represents a disruption that has a higher quantity of product recalled per week which lasts for fewer weeks than Scenario B. All other features of the scenarios are the same. Graphical depictions of the disruption pattern and behavior at the health center including inventory levels, shortage levels, and order amounts for Scenarios A and B are presented in Figure 2. While both systems quickly recover to a state of no shortages within one week after the disruption occurs, the ultimate effects of the disruption scenarios are different. For example, as presented in Table 3, the total shortage is 19% greater in Scenario A than in Scenario B when demand is modeled stochastically. When demand is modeled deterministically, the shortage in Scenario A is 13% greater than in Scenario B. While the shortage amounts differ, the total inventory does not vary significantly between the scenarios. These results demonstrate that the effects of recalls are driven by more than just the total amount recalled. Instead, the rate of products recalled per week and the duration of time over which recalls occur will affect the total number of individuals that are unable to receive the needed product, referred to as the shortage amount.

**Breadth of Disruption.** To examine the effect of the number of disruptions, or the disruption breadth, we compare Scenarios B and D. While in Scenario B there is one disruption that lasts for a total of 24 weeks, in Scenario D there are two disruptions each with a length of 12 weeks and with 3 weeks passing between the disruptions. Both scenarios have the same intensity \( \theta \), defined as the number of products recalled each week during the disruption. The results for these scenarios with deterministic demand are shown in Figure 2. Scenario B and D have slightly different behavior. It takes one week more for Scenario D to return to steady state than Scenario B. Also, the total order amount in Scenario D is greater than in Scenario B. As presented in Table 3, the total shortage is 2% greater in Scenario D than in Scenario B when the system is modeled with stochastic demand and 1% greater when the demand is modeled deterministically. This implies that the system performs slightly better when there is a continuous disruption as opposed to having a short break between the shortages. This is further examined by studying the effects of periodicity of the disruptions.

**Periodicity of Disruption.** Scenarios D and E are compared in order to examine the effect of periodicity of the disruptions. All disruption parameters except for the time between the disruptions are the same. Unlike the 3-week duration that occurs in Scenario D, in Scenario E there is a 24-week break between the disruptions. Similar to the results seen through the comparison of Scenarios B and D, as the time between the recall disruptions increase, so does the level of shortage in the system. As presented in Table 3, the shortage in Scenario E is
In Figure 2, it is shown that the pattern is the same for the per week is five times that found in Scenario B. From examining Figure 2, it is shown that the pattern is the same for the two disruption scenarios, but the intensity differs. Similarly, the temporal patterns pertaining to the inventory, orders, and shortage at the health center are similar for the two scenarios, although the scale of the effects are different. As shown in Table 3, the total shortage in Scenario C is 470% of that in Scenario B when demand is modeled deterministically. Similarly, when demand is modeled stochastically, the total shortage in Scenario C is five times that found in Scenario B.

The higher rate of shortage in proportion to the increase in the recall rate is expected since it is assumed that the inventory policy, as defined by the CSL, is the same regardless of the recall pattern. In practice, it would be natural for a supply chain stakeholder to choose a higher CSL if it was expected that substantially greater recalls were likely. Similarly, if a particular product is more likely to exhibit short and intense disruptions rather than long and more mild disruptions this should inform the management decisions.

<table>
<thead>
<tr>
<th>Disruption Pattern</th>
<th>Healthcare Center Shortage Rate</th>
<th>Healthcare Center's Inventory</th>
<th>Healthcare Center's Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario A</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
<td>!(Graph](image3.png)</td>
</tr>
<tr>
<td>Scenario B</td>
<td><img src="image4.png" alt="Graph" /></td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
<tr>
<td>Scenario C</td>
<td><img src="image7.png" alt="Graph" /></td>
<td><img src="image8.png" alt="Graph" /></td>
<td><img src="image9.png" alt="Graph" /></td>
</tr>
<tr>
<td>Scenario D</td>
<td><img src="image10.png" alt="Graph" /></td>
<td><img src="image11.png" alt="Graph" /></td>
<td><img src="image12.png" alt="Graph" /></td>
</tr>
<tr>
<td>Scenario E</td>
<td><img src="image13.png" alt="Graph" /></td>
<td><img src="image14.png" alt="Graph" /></td>
<td><img src="image15.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Figure 2:** Inventory levels, shortage rates, and orders at health centers over time for disruption Scenarios A-E when demand is modeled deterministically.

10% greater than that in Scenario D for both deterministic and stochastic demand patterns. We hypothesize that the reason for this behavior is that when another disruption has recently occurred the system is already working to recover, with greater production and shipping levels. Thus it is more prepared for a subsequent disruption. As the time between disruptions increases the system is more likely to revert to a steady state which is not prepared to accommodate an additional disruption.

**Size of Disruption.** Finally, we examine the effect of the size of the disruption on the system by comparing Scenario B and Scenario C. In Scenario C, the number of products recalled per week is five times that found in Scenario B. From examining Figure 2, it is shown that the pattern is the same for the two disruption scenarios, but the intensity differs. Similarly,
Table 3: Performance metrics for the recall scenarios with deterministic demand and stochastic demand.

<table>
<thead>
<tr>
<th>Recall Scenarios</th>
<th>Total Recall</th>
<th>Demand constant 10000 bags/week</th>
<th>Demand Normal (10000,100) bags/week</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average Health Center's Inventory</td>
<td>Total Shortage</td>
</tr>
<tr>
<td>0/ No Recall</td>
<td>0</td>
<td>10,000</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>26,200</td>
<td>9,941</td>
<td>18,501</td>
</tr>
<tr>
<td>B</td>
<td>26,200</td>
<td>9,943</td>
<td>16,317</td>
</tr>
<tr>
<td>C</td>
<td>131,000</td>
<td>9,728</td>
<td>77,214</td>
</tr>
<tr>
<td>D</td>
<td>26,200</td>
<td>9,942</td>
<td>16,504</td>
</tr>
<tr>
<td>E</td>
<td>26,200</td>
<td>9,942</td>
<td>18,235</td>
</tr>
</tbody>
</table>
made by supply chain members. As demonstrated in the comparison of scenarios here, even with just one change in the features of the disruption and no changes in the inventory strategy or the total amount recalled, the performance of the system can vary significantly. Thus, it is important to consider potential disruption features when choosing an inventory strategy.

The results for the defined scenarios show that the system acts very differently based on the differing disruption characteristics. In the next section, we examine how different policies affect the performance of the underlying system.

4.3. Effect of Inventory Policies on the Performance. The driving parameter that defines the inventory management policy, and correspondingly the safety stock, at each echelon of the supply chain, is the cycle service level. In the previous section, it was seen that for different recall disruption scenarios (A-E) and a constant CSL, the performance of the system varied significantly. Next, we examine the effects of changes to the inventory policies as they correspond to changing the CSL for the different recall scenarios. While we allow for the policies to change, it is assumed that all echelons will have the same CSL value. Further, we examine the effects of these changes in inventory policies and disruption patterns jointly on two performance metrics, (i) total shortage and (ii) inventory at the health center.

To provide a baseline performance level for the inventory policies, Scenario 0 is defined which represents the system when no disruptions occur.

Figure 3 demonstrates the effects of different CSL values on the two performance metrics in Scenario 0. As this figure shows, by increasing the CSL, the total inventory level increases and the total shortage level decreases.

Identifying the optimal CSL among the policy options is dependent on the relative costs associated with the two performance measures of total shortage and health care center inventory. We evaluate the performance of the policy options for different relative costs using the approach discussed in Section 3.3. The efficient frontier in Figure 3 depicts the estimated levels of both shortages and health care center inventory that can be achieved, and would be optimal, for some relative value of shortage and inventory costs. Thus, all CSL values that do not fall precisely on this efficient frontier are classified as inefficient by the DEA approach. An alternate interpretation of inefficiency is provided by noting that there is a point on this efficient frontier that achieves fewer shortages or fewer inventories while keeping the same level with respect to the second measure. Thus, the outcomes estimated by the efficient frontier are strictly better than those achieved with the inefficient CSL value. As shown in Figure 3, and determined by applying the DEA approach, the policies corresponding to the CSLs of 0.86, 0.87, 0.91, and 0.94 are not on the corresponding efficient frontiers and therefore are never efficient when no shortages occur, regardless of the relative values of the inventory holding cost and the shortage cost. All efficient inventory policies among those simulated are on the efficient frontier in the graph in Figure 3.

To identify efficient inventory policies when the system is experiencing disruptions, we introduce a stochastic model of recalls to present a more realistic setting in which the number of disruptions and the time between disruptions vary. To ensure that the total expected number of products recalled is consistent across scenarios, we model the disruptions with a Poisson distribution with the parameter lambda (λ) corresponding to the expected number of disruptions, or the disruption breadth. Additionally, with the assumption that the duration of each recall disruption (γ) is the same, the average recall quantity per week (θ) is scaled to ensure that the total expected quantity recalled is the same. Due to the stochastic nature of the modeling, it is possible that simulation instances experience different numbers of disruptions, and therefore different quantities are recalled, but the expected number recalled remains the same.

In order to study the effect of different inventory policies on the performance metrics of the health center, we simulate each recall pattern for 16 different CSLs. Furthermore, we run the simulation for multiple disruption patterns defined by the value of λ, for λ = (5, 10, 15, 20), for a 5% of loss inventory, to examine the relationship between policies and recall patterns. For smaller values of λ there are fewer disruptions with higher intensity. More frequent disruptions with fewer recalls per week results from higher values of λ. Figure 4 shows the performance of these policies, defined by the CSL, for different recall patterns, defined by the λ value. From the results, it is clear that different inventory policies perform significantly different depending on the disruption pattern. Additionally, the incremental difference in inventory and shortage between policies varies for different disruption patterns. Thus, the choice of policy should vary with respect to the expected disruption characteristics and with the relative costs of one unit of shortage and one unit of inventory.

Similar to the identification of efficient inventory policies for the scenario with no recalls, we can identify inventory policies that are inefficient regardless of the relative valuation of the two performance measures. This can be achieved.
by using an approach similar to DEA. For example, when \( \lambda = 5, 15 \) and 20, inventory policies with CSL equal to 0.91, 0.86 and 0.96, respectively, are found to be inefficient using the mathematical models in the DEA approach. While an inventory policy may be identified as efficient it will not be optimal for all relative values of per unit shortage costs and per unit inventory costs. The optimal CSLs for a selection of relative costs and for different disruption patterns, defined by \( \lambda \), are provided in Table 4.

For different ratios of the unit cost of shortage and the unit cost of inventory, we can identify the optimal CSL for the inventory policy. Table 4 illustrates the result of this analysis. This analysis reveals that increasing the CSL, which is equivalent to choosing to hold more safety stock in inventory, is not always best and it varies based on the ratio of costs and the type of disruption that is expected to occur. For example, if the ratio of the costs is 1, a CSL equal to 0.98 is optimal only if there are frequent recalls, of approximately more than 15 over 5 years, or low intensity in this model. However, if it is expected that there will only be one recall per year (\( \lambda = 5 \)) of a high intensity, then choosing a CSL of 0.98 is no longer optimal and we should instead choose a CSL of 0.94.

Finally, the results in Table 4 demonstrate the sensitivity of the choice of inventory policy for a disruption pattern. For example, if the ratio of costs is 1.4, choosing an inventory policy with the assumption of five recalls over the five-year duration will leave the supply chain stakeholder very far from the optimal policy if in fact there are fifteen recalls and the same total quantity is recalled. This demonstrates the criticality of choosing a policy that correctly matches the expected recall disruption pattern. Further, this supports the value of stakeholders gathering information about historical patterns in recalls and forecasting future recalls. By achieving a better understanding of the likely recall scenarios, supply chain decision makers can adjust and ultimately reduce costs from inventory and shortages. These results support the consideration of the use of adaptive policies such that inventory decisions change over time either in anticipation of or in response to recalls. This ultimately may lead to reduced drug shortages and reduced supply chain costs. As discussed below, we suggest expanding the model to integrate adaptive inventory policies in future work.

### 5. Conclusion

We present a mathematical simulation model of the pharmaceutical supply chain to understand the behavior of the drug shortages under different disruption patterns. There are currently few models that study the entire pharmaceutical supply chain, and none that focus on recalls and disruption patterns. To provide insights into this critical relationship we examine (i) the role of the disruption characteristics and (ii) the role of the inventory management policies within a simulation model inspired by the saline supply chain. Sterile solutions, such as saline, constitute a large portion of the pharmaceutical recalls and shortages that occur each year.

Through analysis of this supply chain, we find that small differences in the disruption patterns lead to very different system performance when the inventory management policies are held consistent. Further, this relationship holds for changes to either the (i) length, (ii) breadth, (iii) periodicity, or (iv) size of the disruptions. Additionally, we demonstrate the sensitivity of the optimal inventory policy to these features of the disruptions. Thus, rather than a traditional approach of using a consistent cycle service level (i.e., holding more safety stock in inventory), our results support the need for evolving inventory management strategies that account for the features of the disruption.

In the case of saline, frequent small recalls occurred in the system prior to 2017. If the inventory policies were designed to instead be resilient against major disruptions or large recalls, the inventory management policy and levels
of safety stock would not be efficient. Thus, choosing a policy that best responds to the likely disruptions is critical. Further, with an understanding of the effects of frequent small disruptions, one option for stakeholders in the saline supply chain would be to invest in manufacturing equipment and institute practices that would decrease the likelihood of this disruption pattern. Additionally, instituting policies that allow for inventory policies to agilely adapt based on current and expected future circumstances would improve the overall system performance.

A key challenge to implementing policies that adapt to disruptions is the critical unknown features of disruptions that will occur in the future. But while full knowledge of future events is not possible, the results of this work support the need for supply chain managers to examine and identify the best inventory management policies for a variety of types of disruptions that are probable. In return, understanding this relationship between the best policies and the disruption characteristics can inform implementation of policies that are robust to a variety of types of disruptions. Further, as disruptions occur and estimates of their features, such as length, breadth, size, and periodicity, are updated supply chain stakeholders can modify their inventory policies accordingly. Lastly, these results support the importance of supply chain stakeholder investing in gaining information about these disruption characteristics both in anticipation of or during a disruption. This additional information will be valuable as it allows the decision makers to adapt policies in real-time.

There are several limitations and extensions which require further investigation. First, this paper considers that there is no limitation on the manufacturing capacity. A more realistic model with inclusion of limited capacities and a sensitivity analysis examining the role of manufacturing capacity may be informative for future consideration as the capacity is expected to significantly influence the shortage size and the time for the system to return to a steady state. Second, we assume that all the echelons use the same inventory policy and have the same CSL. This representation may be violated for pharmaceutical supply chains in which downstream nodes tend to hold more safety stock to reduce the likelihood of unmet patient demand. Third, defining the total cost of the system as the inventory holding cost and shortage cost of the health care center is a limiting feature. While we only consider the inventory of health care centers, this can be considered to serve as a proxy that represents the total inventory in the supply chain. Additionally, with a focus on addressing the critical societal problem of drug shortages, we focus on developing an understanding of the elements driving shortages. Future extensions of this model can account for additional system elements in the objective function.

We propose to address these limitations in future research. Additionally, since, in the real world, planners are often unsure of the type of disruption that will occur, in future research we aim to construct and examine the performance of adaptive inventory policies by building on the findings from this work. There are also opportunities to expand this study to investigate other types of pharmaceutical supply chain disruptions, including manufacturing shutdowns and surges in demand to compare and contrast with the effects of recall disruptions. Also, consideration of the cooccurrence of manufacturing shutdowns and product recalls can further inform operational policies for supply chain stakeholders seeking to prevent future drug shortages.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded in part by the National Science Foundation (Grant CMMI #1638302).

References


