Incentivizing Resilient Supply Chain Design to Prevent Drug Shortages: Policy Analysis Using Two- and Multi-Stage Stochastic Programs

Emily L Tucker, Mark S Daskin, Burgunda V Sweet & Wallace J. Hopp

To cite this article: Emily L Tucker, Mark S Daskin, Burgunda V Sweet & Wallace J. Hopp (2019): Incentivizing Resilient Supply Chain Design to Prevent Drug Shortages: Policy Analysis Using Two- and Multi-Stage Stochastic Programs, IISE Transactions

To link to this article: https://doi.org/10.1080/24725854.2019.1646441
Incentivizing Resilient Supply Chain Design to Prevent Drug Shortages: Policy Analysis Using Two- and Multi-Stage Stochastic Programs

Emily L Tucker1*, Mark S Daskin1, Burgunda V Sweet2, Wallace J. Hopp3

1University of Michigan, Industrial and Operations Engineering, Ann Arbor, United States
2University of Michigan, College of Pharmacy, Ann Arbor, United States
3University of Michigan, Ross School of Business, Ann Arbor, United States

*Corresponding author: Emily L Tucker Email: eltuck@umich.edu

Abstract

Supply chain disruptions have caused hundreds of shortages of medically-necessary drugs since 2011. Once a disruption occurs, the industry is limited in its ability to adapt, and improving strategic resiliency decisions is important to preventing future shortages. Yet, many shortages have been of low-margin, generic injectable drugs, and it is an open question whether resiliency is optimal. It is also unknown what policies would be effective at inducing companies to be resilient.

Keywords: drug shortages, stochastic program, Pharmaceutical, resiliency, incentives

1. Introduction

Over the past decade, the United States (US) has experienced unprecedented shortages of medically-necessary drugs. In 2015 alone, 427 drugs were unavailable (GAO, 2016). Shortages last 14 months on average and can have negative effects on patient safety, clinical outcomes, and health system costs (GAO, 2016; Tucker et al., n.d.). They are often caused by disruptions to non-resilient pharmaceutical supply chains. The question of how best to reduce their impact is pressing for patients and the US healthcare system.

Drugs that have been short span a variety of therapeutic classes including central nervous system (anesthetics), cardiovascular, anti-infective, and oncology agents (UUDIS, 2016). When shortages occur, treatment of patients may be delayed, changed, or cancelled entirely (Goldsack et al., 2014; McLaughlin et al., 2013). Shortages have been associated with patient deaths (Fox, Sweet, and Jensen, 2014; Vail et al., 2017), and managing them has been compared to dealing with a “natural disaster or national emergency… occur[ring] on a daily basis” (Fox, Sweet, and Jensen, 2014).
Costs associated with managing shortages are high. It has been estimated that traditional US health systems spend $216-359 million each year on labor costs to manage shortages (Kaakeh et al., 2011; Vizient, 2019) and $200 million annually to purchase substitute drugs and hold extra inventory (Fox, Sweet, and Jensen, 2014).

Shortages are caused by a variety of factors. Production may be delayed, companies may leave the market, or production lines may be contaminated. In some cases, Food and Drug Administration (FDA) inspections have uncovered quality concerns at manufacturing and raw material facilities that have led to extended shutdowns (Fox, Sweet, and Jensen, 2014; Palmer, 2016). Some drugs have been short multiple times because of intermittent manufacturing issues (UUDIS, 2016). Often companies do not report the direct cause of a shortage, but of those reported, 82% are caused by a supply chain disruption (GAO, 2016).

There has been substantial research on mitigating disruptions in non-pharmaceutical supply chains, and researchers have found that maintaining some degree of resiliency is often optimal (Snyder et al., 2016; Tomlin, 2006). However, shortages of drugs persist. This may be due, at least in part, to the unique challenges of the highly regulated pharmaceutical industry. Whenever a pharmaceutical company changes its supply chain, it must go through a lengthy FDA approval process (GAO, 2016). This requirement makes it difficult to adapt to disruptions. When this is combined with the complex manufacturing processes of sterile drugs, there are inherent vulnerabilities. A stark contrast can be seen with other industries. When a fire shut down the sole supplier of a critical part for Toyota years ago, they were able to ramp-up other suppliers and resume production within days (Nishiguchi and Beaudet, 1998). When Hurricane Maria shut down pharmaceutical manufacturing plants in Puerto Rico, drug shortages continued to affect the healthcare system for months (Gottlieb, 2018b; Thomas and Kaplan, 2017). The strategic resiliency decisions made prior to disruption become critical.

Currently, companies that manufacture drugs that are particularly vulnerable to shortages (e.g., sole-source generic injectables) hold little backup capacity and maintain little to no safety stock (Fox, Sweet, and Jensen, 2014; GAO, 2016; Woodcock and Wosinska, 2013). Profit margins are low (GAO, 2016), and there are few consequences if there is a shortage. Companies rarely pay penalties if they cannot supply a drug, and the risk of losing market share to a new competitor is small because of high barriers to entry (GAO, 2016; Haninger, Jessup, and Koehler, 2011; Jia and Zhao, 2017).

Yet, drugs affected by shortages are often medically-necessary and life-sustaining. The fundamental question becomes: how can we strengthen pharmaceutical supply chains to provide a reliable drug supply? We will answer this by considering two specific questions: i) is it optimal for companies to choose low resiliency? and ii) if so, what is the best way to induce companies to maintain resilient supply chains?

Several strategies have been proposed. These include regulatory changes, e.g., require companies to maintain redundancy (ASHP, 2013; Chabner, 2011; FDA, 2013; Gehrett, 2012; “Health Policy Brief: Drug Shortages,” 2014; Jaroslawski et al., 2017) or hold safety stock (ASHP, 2013; Gupta and Huang, 2013; Jaroslawski et al., 2017; Wiggins et al., 2014). Others include contractual changes such as strengthening failure-to-supply clauses (Conti, 2011; FDA, 2013; Haninger,
Jessup, and Koehler, 2011; “Health Policy Brief: Drug Shortages,” 2014; Jia and Zhao, 2017; Reed et al., 2016) and increasing prices (Chabner, 2011; Gatesman and Smith, 2011; “Health Policy Brief: Drug Shortages,” 2014; Link, Hagerty, and Kantarjian, 2012). Limited analyses of the potential effects of these proposals have been conducted despite calls from experts for such studies (FDA, 2013; Fox, Sweet, and Jensen, 2014; Fox and Tyler, 2013; ISPE and Pew Charitable Trusts, 2017; Roberts et al., 2012). It remains unclear whether market-based interventions or regulatory changes would be more effective.

In this paper, we seek to fill this gap to consider why pharmaceutical companies may set up either vulnerable or resilient supply chains and to analyze how the proposed policy changes would affect supply chain decisions. As different policies have different implementation costs, we analyze the social-efficiency – how to reduce shortages to a specified level for the lowest cost. We conduct our analysis focusing on generic, oncology drugs in two steps: first, we develop new pharmaceutical supply chain design models, and second, we change the underlying market conditions to analyze the effects of the proposed policies.

The remainder of this paper is organized as follows. In Section 2, we review the relevant literature and discuss contributions of our analysis. In Section 3, we present the base model that includes redundancy as a resiliency strategy, and in Section 4, we develop the extension that includes safety stock. We discuss how we solve the models in Section 5. In Section 6, we present case examples of two oncology drug supply chains. In Section 7, we discuss our results and policy implications and conclude.

2. Literature Review

Our work relates to several streams of literature, and we briefly review relevant studies that focus on pharmaceuticals, supply chain risk management, disruptions, and incentives.

2.1. Pharmaceutical Modeling

The operations research and management science community has only recently begun to study drug shortages. Kim and Scott Morton (2015) analyzed factors that contribute to shortages with a game theory model of two competing manufacturers of perfectly substitutable generic injectable drugs. They suggested that spare capacities may have been removed when prices dropped in the early 2000s, revealing underlying vulnerabilities that led to shortages. In one of the only papers to evaluate policy, Jia and Zhao (2017) developed a model of contracts between key stakeholders to analyze the effects of failure-to-supply clauses and price increases. At the beginning of the contracting period, the manufacturer allocates production capacity at a single echelon and decides its inventory policy under stochastic supply and demand. The authors used this framework to study case examples of fluorouracil, cytarabine, and bleomycin and found Pareto-improving contracts for each stakeholder. Others have studied inventory control for hospitals struggling with shortages (Saedi, Kundakcioglu, and Henry, 2016), inventory policies as a response to product recalls (Azghandi, Griffin, and Jalali, 2018), and inventory policies related to human behavior and shortages (Doroudi et al., 2018). Jacobson, Sewell, and Proano (2006) analyzed the size of the Strategic National Stockpile of pediatric vaccines.
Optimization is regularly applied more broadly in the pharmaceutical literature (Narayana, Pati, and Vrat, 2014; Shah, 2004) though work on supply chain design is uncommon. Exceptions include a multi-stage stochastic program which considered demand uncertainty (Guillén et al., 2006) and a four-echelon supply chain model under uncertainty in demand, cost, and desired safety stock levels (Mousazadeh, Torabi, and Zahiri, 2015). In our models, we consider supply chain design models where disruptions are a source of supply uncertainty.

2.2. Supply Chain Risk Management

Supply chain risk management (SCRM) is a large research area that considers how companies structure and operate their supply chains to provide products to customers in the presence of uncertainty. There have been several reviews of this literature (including Ho et al., 2015; Tang 2006; Tang and Musa 2011), and one of the early reviews identified four main domains of SCRM – managing supply, demand, products, and information (Tang, 2006). In a seminal paper, Chopra and Sodhi (2004) discussed several strategies to manage risk, including capacity, inventory, redundancy, and flexibility.

Since then, many authors have developed quantitative models to analyze strategies in different contexts. Objectives have varied from cost minimization to bi-objective frameworks that trade off profit and risk (Nagurney, 2006; Tomlin, 2006). Some researchers explicitly penalize unmet demand (Dada, Petruzzi, and Schwarz, 2007; Schmitt, Snyder, and Shen, 2010), and others consider robust approaches (O’Hanley and Church, 2011). Researchers have found that supply and demand risks should be managed differently (Schmitt et al., 2015; Snyder and Shen, 2006). When the risk is supply chain disruptions, Tomlin (2006) found that it is rarely optimal to passively accept risk; companies should nearly always select some level of resiliency. Yet pharmaceutical companies often passively accept risk for generic injectable drugs (Fox, Sweet, and Jensen, 2014; GAO, 2016; Woodcock and Wosinska, 2013). One open question this paper seeks to address is whether this choice is optimal.

2.3. Disruptions

Within the field of SCRM, many models have focused on disruptions as a source of supply-side risk. Snyder et al. (2016) presented an extensive review of this area. Common mitigation strategies include maintaining redundancy, holding inventory, or sourcing from multiple suppliers.

Redundancy or backup capacity decisions are often considered within the facility location literature. A number of studies have considered where to locate facilities at a single echelon given disruptions (e.g., Snyder and Daskin 2005). Fewer papers have considered decisions for multiple echelons (e.g., the robust approach of Peng et al. 2011).

Within the inventory literature, researchers have extended standard, single-supplier models to include disruptions. A survey of key models is available from Atan and Snyder (2012). In the Economic Order Quantity model with Disruptions (EOQD), the supplier is disrupted according to a continuous-time Markov chain (presented by Parlar and Berkin 1991; corrected by Berk and Arreola-Risa 1994). One extension includes the risk of disruptions at the retailer (Qi, Shen, and
Snyder, 2009). In the periodic-review framework, Song and Zipkin (1996) proved that a base-stock policy is optimal for a single echelon if the order costs are linear, and Schmitt, Snyder, and Shen (2010) derived the exact and approximate expected costs when there is stochastic demand.

Where there are multiple suppliers, inventory decisions can be made using extensions to the EOQ model (Gurler and Parlar 1997; Parlar and Perry 1996) or a network of queues (Song and Zipkin, 2009). Schmitt and Tomlin (2012) analyzed whether single- or multi-sourcing is optimal in different contexts, and Saghafian and Van Oyen (2012) studied the effects of the flexibility of the backup. Mak and Shen (2012) considered dynamic sourcing to mitigate both demand and supply uncertainty.

Companies may also consider multiple disruption mitigation strategies. Tomlin (2006) compared a firm’s decision to dual-source with holding inventory, rerouting, and passive acceptance. Others analyzed strategies for supply chain networks (Bundschuh, Klabjan, and Thurston, 2003; Hopp and Yin, 2006; Schmitt, 2011). These included separate models for multi-sourcing, safety stock, and meeting an expected service level (Bundschuh, Klabjan, and Thurston, 2003) and models that traded off backup capacity and safety stock (Hopp and Yin, 2006; Schmitt, 2011). Schmitt (2011) noted that inventory is helpful for shorter, more frequent disruptions, and backup capacity is better for less frequent, longer disruptions. MacKenzie, Barker, and Santos (2014) studied the decisions suppliers and firms make during and after a disruption, including whether to switch to an alternate facility. Where trade-offs were evaluated, the key metric was company cost or profit (e.g., Hopp and Yin 2006); we focus on the overall societal cost, i.e., social-efficiency.

Disruption models that included inventory often considered decisions over multiple time periods (e.g., Berk and Arreola-Risa 1994, Tomlin 2006), though location and design models generally did not. The latter often assumed the system returned to steady-state before another disruption occurred (e.g., Hopp and Yin 2006; Kim and Scott Morton 2015; Schmitt 2011) or implicitly considered a single time period (e.g., Bundschuh et al. 2003). The strategic design models that included time considered decisions for a single echelon (Fattahi, Govindan, and Keyvanshokooh, 2017; Losada, Scaparra, and O’Hanley, 2012) or single layer of arcs between two echelons (Mak and Shen, 2012). A recent review noted that in general very few supply chain design problems under uncertainty have been formulated as multi-stage stochastic programs (Govindan, Fattahi, and Keyvanshokooh, 2017), and those that exist have tended to be small; e.g., a three-stage model with nine scenarios (Almansoori and Shah, 2012).

2.4. Incentives and Policy

The models in the previous subsections generally took the perspective of a company that aims to improve resiliency to reduce costs, though there is a stream of literature that considers incentives and policies from external decision-makers. Among those that included uncertainty in supply, researchers have analyzed strategies to incentivize capacity restoration after disruptions and to improve recovery time (Hu, Gurnani, and Wang, 2013; Kim et al., 2010). As discussed, failure-to-supply clauses have been analyzed in the context of capacity allocation for drug shortages (Jia and Zhao, 2017). Tang, Gurnani, and Gupta (2014) studied subsidies and increased demand to
incentivize a more reliable supply. Researchers have also analyzed government policy incentives in other areas, e.g., tax incentives for renewable energy (Karimi, Ekşioğlu, and Khademi, 2018).

2.5. Contributions

In this analysis of strategic resiliency decisions and policies to reduce drug shortages, we make the following contributions.

- We study resiliency decisions for the highly regulated supply chains of generic injectable drugs that have low profit margins and a limited ability to adapt if disruptions occur.
- We develop supply chain design models that incorporate a new combination of characteristics. They consider time; disruptions may occur at multiple echelons; multiple components may be concurrently unavailable; and the company may select multiple mitigation strategies (facility redundacy and safety stock).
- We introduce constraints to enforce a replenishment rule, and the non-anticipativity property of the multi-stage stochastic program is induced. This allows us to solve a large thirteen-stage model.
- We evaluate policies to induce resiliency and reduce drug shortages (mandatory redundancy, mandatory inventory, failure-to-supply penalties, pricing changes, and the combination of price increases and other interventions). We analyze the social-efficiency of these interventions.

3. Base Model (SCDD)

To begin to study strategic resiliency, we develop a two-stage stochastic program – the Supply Chain Design under Disruption (SCDD) model. In the first stage, the company selects the optimal configuration of the supply chain that will be fixed for the remainder of the time horizon. There is uncertainty about which components may be working in future periods. In each subsequent period, the uncertainty is realized, and the company decides the quantities of raw materials to order and finished goods to produce.

We consider a three-echelon supply chain for a single drug that is comprised of Active Pharmaceutical Ingredient (API) suppliers, manufacturing plants, and manufacturing lines. Manufacturers of sterile injectable drugs typically hold little safety stock (GAO, 2016), and we reflect this decision in the base analysis.

3.1. Background

A sample supply chain configuration is presented in Figure 1. It includes two suppliers, two plants, and one line in each plant. Plants may receive raw materials from either API supplier, but lines are associated with specific plants. Each candidate plant \( k \in K \) has a set of candidate lines \( l \in L_k \). In this example, \( L_1 = \{1,2,3\} \) and \( L_2 = \{4,5,6\} \). The set of all lines \( L \) is the union of the sets of lines in each plant, i.e., \( L = \bigcup_{k \in K} L_k = \{1,2,3,4,5,6\} \). While pharmaceutical companies have additional partners in practice (e.g., packaging and non-active raw materials), we model...
only the critical steps (cf. Bundschuh et al. 2003) and consider all echelons that contributed to shortfall categories in a recent Government Accountability Office report (GAO, 2016).

The objective is to maximize the expected profit under uncertainty in the status of the supply chain components. In the first stage, the company selects the supply chain configuration. They may choose to not market the drug and select no components. In the second stage, in each period, the company selects production and order quantities after uncertainty about component availability is realized. Demand may be met with production or unmet (a shortage), and both demand and price are constant over the time horizon.

There are fixed costs to select components, and these include the costs to maintain suppliers, plants, and lines as well as the government-mandated user fees for generic drug production (i.e., Generic Drug User Fee Amendments (GDUFA) fees; FDA 2018b). The GDUFA facility fees are incurred as fixed costs for each supplier and plant, and the GDUFA program fee is incurred if the company is in the market. There are variable costs to order raw materials and to produce the drug, and revenues come from sales.

Any component in any echelon may become disrupted, and in each period, the status of each component is either available \(\{1\}\) or disrupted \(\{0\}\). We model this uncertainty using discrete scenarios \(\omega \in \Omega\), where the random variable \(\xi_{nt}^\omega\) in scenario \(\omega \in \Omega\) represents the status of candidate component \(n \in N\) in period \(t \in T\) \((\xi_{nt}^\omega) \in \{0,1\}^{|N| \times |T| \times |\Omega|}\). The notation for SCDD is presented in Figure 2.

The lines have capacity limits to be consistent with the model that includes inventory, though these are not limiting for the SCDD model.

3.2. Model Formulation

The formulation of SCDD is as follows:

\[
\text{Maximize} \quad - \frac{\|T\|}{t} \left[ (c^{API} + f^{API}) \sum_{j \in J} x_j + (c^{Plant} + f^{Plant}) \sum_{k \in K} y_k + c^{Line} \sum_{l \in L} z_l + f^{Program} x_1 \right] + E_{\Omega}[Q(x, y, z)]
\]

(1)

Subject to:

\[ z_l \leq y_k \quad \forall k \in K, l \in L_k \]  
\[ x_j \geq x_{j+1} \forall j \in J \setminus \{|J|\} \]  
\[ y_k \geq y_{k+1} \quad \forall k \in K \setminus \{|K|\} \]  
\[ z_l \geq z_{l+1} \quad \forall l \in L_k \setminus \{|L_k|\}, k \in K \]  

(2)  
(3a)  
(3b)  
(3c)
\( x_j \in \{0,1\} \forall j \in J \quad (4a) \)
\( y_k \in \{0,1\} \forall k \in K \quad (4b) \)
\( z_l \in \{0,1\} \forall l \in L \quad (4c) \)

\[
E_\Omega[Q(x,y,z)] = \max_{u,v,\theta} \sum_{\omega \in \Omega} p^\omega d \sum_{t \in T} \left[ q \theta_t^\omega - c^{raw} \sum_{j \in J} u_{jt}^\omega - c^{prod} \sum_{l \in L} v_{lt}^\omega \right] 
\]

Subject to:

\[
u_{jt}^\omega \leq \xi_{jt}^\omega |L| g^{Line} x_j \quad \forall j \in J, t \in T, \omega \in \Omega(6)
\]
\[
v_{lt}^\omega \leq \xi_{kt}^\omega \xi_{lt}^\omega |L| g^{Line} z_l \quad \forall l \in L, k \in K, t \in T, \omega \in \Omega(7)
\]
\[
\sum_{l \in L} v_{lt}^\omega \leq \sum_{j \in J} u_{jt}^\omega \quad \forall t \in T, \omega \in \Omega(8)
\]
\[
\theta_t^\omega \leq \sum_{l \in L} v_{lt}^\omega \quad \forall t \in T, \omega \in \Omega(9)
\]
\[
\theta_t^\omega \leq 1 \quad \forall t \in T, \omega \in \Omega(10)
\]
\[
u_{jt}^\omega \geq 0 \quad \forall j \in J, t \in T, \omega \in \Omega(11a)
\]
\[
v_{lt}^\omega \geq 0 \quad \forall l \in L, t \in T, \omega \in \Omega(11b)
\]
\[
\theta_t^\omega \geq 0 \quad \forall t \in T, \omega \in \Omega(11c)
\]

The objective function (1) maximizes the expected profit. The annual fixed costs include, respectively, the cost per API supplier, the GDUFA fee per API supplier, the cost per plant, the GDUFA fee per plant, the cost per line, and the GDUFA program fee. Expected ordering and production costs and revenues are incurred in the second stage. Constraints (2) ensure that the selected lines are in selected plants. Constraints (3) require components to be selected in numerical order and are used to reduce alternative optima. Constraints (4) are standard binary constraints.

In the second stage, the company makes operational decisions each period after uncertainty is realized. The objective function (5) maximizes the expected profit in the second stage. Revenues are accrued based on sales, and costs include raw materials and production. Constraints (6) limit orders of raw materials to selected, available suppliers. Constraints (7) limit finished goods production to the capacity of selected, available lines in available plants. Constraints (8) limit production to the amount of raw material ordered. Constraints (9-10) ensure the fraction of demand met is not greater than the finished goods available and the amount customers demand, respectively. Constraints (11) enforce non-negativity.

3.3. Structural Property
It follows from the formulation of SCDD that demand is either fully met or fully unmet each period. We present this formally in Lemma 1.

Lemma 1: \( \theta_t^\omega \in \{0,1\}, \forall t \in T, \omega \in \Omega \). Proof: Provided in appendix.

3.4. Assumptions

To identify factors that contribute to supply chain vulnerability and resiliency, we make several simplifying assumptions. We do not consider transportation time or cost. While these are clearly present in practice, the time to ship from common API supplier locations to the US is often one month or less (SeaRates, 2018; US Department of Commerce, 2018), which is smaller than the periods we consider in our analyses. We assume production occurs throughout the year and suppliers are uncapped. We exclude some other operational dynamics such as costs and time of product changeover that are often present in other papers within the pharmaceutical literature (Lakhdar and Papageorgiou, 2008; Marques et al., 2017). We do not consider discounting because we focus on the realized costs and revenues of limited-term contracts. If a disruption occurs, we do not consider the cost of recovery. At an API facility, this cost would be incurred by the supplier, and at a plant the cost would be spread across multiple drugs, though this is a limiting assumption. We assume constant demand over the time horizon, consistent with the fairly stable demand of most drugs (Fox, Sweet, and Jensen, 2014). If demand is not met for the drugs considered in our case examples, in practice a clinical decision is generally made to switch to an alternative treatment. Treatment delays are less common, and as a simplifying assumption, we assume all demand is lost rather than backordered due to delays. We do not consider competition, which is justified by the fact that drugs affected by shortages are often sole-source, particularly those that are injectable (IMS, 2011; UUDIS, 2016). The drugs we consider have a single manufacturer.

We do not allow the company to make changes to the supply chain structure within the time horizon. Disruptions and recovery at each component occur independently and are exogenous to the model. Each candidate component within an echelon, e.g., all candidate lines, have identical disruption profiles and capacities. This could be relaxed by subscripting the capacity parameters by the components. Furthermore, following the work of others, we do not consider location decisions; rather we focus on resiliency strategies and policies (Hopp and Yin, 2006; Jia and Zhao, 2017).

4. Extension to Include Inventory (SCDD-I)

To extend our analyses to consider safety stock as a resiliency strategy, we introduce a second model. We call this the Supply Chain Design model under Disruption with Inventory (SCDD-I) and formulate it as a multi-stage stochastic program. The model relaxes the initial decision that no inventory is held, and we can use the model to study why companies may or may not use inventory.

4.1. Background
In the first stage, the company chooses the supply chain configuration and a target amount of safety stock to hold each period (i.e., stage). Neither the supply chain design nor the target safety stock level may be changed throughout the time horizon. As in the SCDD model, the company may choose no components and not produce the drug. In each of the subsequent stages, uncertainty in the component statuses is realized, and the company selects production and order quantities. Demand may be met through production or safety stock. If production exceeds demand, inventory is replenished. Unmet demand is lost, not backordered.

To model practice realistically, multi-stage stochastic programs must impose the non-anticipativity property. It requires decision-makers to make the same decision for each scenario that is identical up to that point. This prevents them from anticipating realizations of future uncertainty. The non-anticipativity property is typically enforced either through constraints or implied via the construction of the scenario tree, but as the number of stages increases, the number of constraints substantially increases, and the problem quickly becomes intractable to standard solution methods such as Sample Average Approximation (SAA). To avoid this intractability, we impose a safety stock replenishment rule: the manufacturer must meet demand when possible; they may only deplete safety stock if production capacity is unavailable; and they must replenish deficit safety stock if excess capacity is available. That is, given the component statuses and variables from the previous stage, the rule predetermines the sales and inventory decisions; the decision variables will be the same regardless of future realizations of uncertainty. This induces the non-anticipativity constraints to hold without including them in the model. The additional notation for SCDD-I is presented in Figure 3.

4.2. Model Formulation

The SCDD-I model includes constraints (2-4, 6-8, 10-11) from the SCDD model, and the revised objectives and additional constraints are as follows.

\[
\text{Maximize} \quad - \frac{[T]}{t} \left[ (c^{API} + f^{API}) \sum_{j \in J} x_j + (c^{Plant} + f^{Plant}) \sum_{k \in K} y_k + c^{Line} \sum_{l \in L} z_l + f^{Program} x_1 \right] + E_R[Q(x, y, z, l_0)]
\]

\[
\text{Subject to:}
\]

\[l_0 \leq o^{max} \tilde{z}_{11}\]  

\[
\tilde{z}_{jl} \geq x_j + z_l - 1 \quad \forall j \in J, l \in L
\]  

\[
\tilde{z}_{jl} \leq x_j \quad \forall j \in J, l \in L
\]  

(13)  

(14a)  

(14b)
\[ \tilde{z}_{jl} \leq z_l \quad \forall j \in J, l \in L \]  
(14c)

\[ I_0 \geq 0 \quad (4d) \]

\[ \tilde{z}_{jl} \in \{0, 1\} \quad \forall j \in J, l \in L \quad (4e) \]

\[ E_\Omega [Q(x, y, z, I_0)] = \max_{u,v,\theta,I} \sum_{\omega \in \Omega} P^\omega d \sum_{t \in T} [q \theta_t^\omega - c^{raw} \sum_{j \in J} u_{jt}^\omega - c^{prod} \sum_{l \in L} v_{lt}^\omega - h I_t^\omega] \]
(15)

Subject to:

\[ \theta_t^\omega \leq \sum_{l \in L} v_{lt}^\omega + I_t^{\omega - 1} \quad \forall t \in T, \omega \in \Omega(16) \]

\[ I_t^\omega = I_{t-1}^\omega + \sum_{l \in L} v_{lt}^\omega - \theta_t^\omega \quad \forall t \in T, \omega \in \Omega(17) \]

\[ I_t^\omega \leq I_0 \quad \forall t \in T, \omega \in \Omega(18) \]

\[ \theta_t^\omega \geq C_t^\omega \quad \forall t \in T, \omega \in \Omega(19) \]

\[ \theta_t^\omega \geq \delta_t^{Avail,\omega} \quad \forall t \in T, \omega \in \Omega(20) \]

\[ I_{t-1}^\omega - I_t^\omega \leq 1 - C_t^\omega \quad \forall t \in T, \omega \in \Omega(21) \]

\[ C_t^\omega \geq \xi_{jt}^\omega \xi_{kt}^\omega \xi_{lt}^\omega z_{jl} \quad \forall t \in T, \omega \in \Omega, j \in J, k \in K, l \in L_k \]  
(22a)

\[ C_t^\omega \leq \sum_{j \in J} \sum_{k \in K} \sum_{l \in L_k} \xi_{jt}^\omega \xi_{kt}^\omega \xi_{lt}^\omega z_{jl} \quad \forall t \in T, \omega \in \Omega(22c) \]

\[ \tilde{C}_t^\omega = \sum_{j \in J} \sum_{k \in K} \sum_{l \in L_k} g^{\text{Line}} \xi_{jt}^\omega \xi_{kt}^\omega \xi_{lt}^\omega z_{jl} - C_t^\omega \quad \forall t \in T, \omega \in \Omega(23) \]

\[ \sum_{l \in L} v_{lt}^\omega - C_t^\omega = \min(C_t^\omega, I_0 - I_{t-1}^\omega) \quad \forall t \in T, \omega \in \Omega(24-\text{nonlin}) \]

\[ \sum_{l \in L} v_{lt}^\omega - C_t^\omega \geq \tilde{C}_t^\omega - |L| g^{\text{Line}} \delta_t^{\text{Suffic,\omega}} \quad \forall t \in T, \omega \in \Omega(24a) \]

\[ \sum_{l \in L} v_{lt}^\omega - C_t^\omega \geq I_0 - I_{t-1}^\omega - (o^{\max} + \epsilon)(1 - \delta_t^{\text{Suffic,\omega}}) \quad \forall t \in T, \omega \in \Omega(24b) \]

\[ 1 - \delta_t^{\text{Avail,\omega}} \geq 1 - I_{t-1}^\omega \quad \forall t \in T, \omega \in \Omega(25a) \]

\[ (o^{\max} + \epsilon)\delta_t^{\text{Avail,\omega}} \geq I_{t-1}^\omega \quad \forall t \in T, \omega \in \Omega(25b) \]

\[ I_0^\omega = I_0 \quad \forall \omega \in \Omega(26) \]
\[ I_t^{\omega} \geq 0 \forall t \in \{0\} \cup T, \omega \in \Omega \quad (27a) \]

\[ C_t^{\omega}, \tilde{C}_t^{\omega} \geq 0 \quad \forall t \in T, \omega \in \Omega \quad (27b) \]

\[ \delta_t^\text{Stk}, \delta_t^\text{Avail,} \in \{0,1\} \quad \forall t \in T, \omega \in \Omega \quad (27c) \]

The objective function (12) is the same as (1) except it adds the target inventory as an argument to the expected cost of the subsequent stages. In the first stage, constraint (13) enforces two safety stock conditions. It requires a complete supply chain to be selected if safety stock is held, via the binary variable, \( z_{1,1} \), and it limits the number of periods of safety stock that can be held to an upper bound, \( o^{max} \). Constraints (14) define variables to indicate which combinations of suppliers and lines are selected. These are used in combination with constraints (22-23) to define the total capacity available in each period. Constraints (4d-e) enforce the domains of the new first stage decision variables.

In the subsequent stages, the objective function (15) maximizes expected profit. Revenues come from demand met. The variable costs are incurred through raw material orders; production of finished goods; and held safety stock, respectively. Constraints (16) ensure that demand can only be met from production and safety stock. Constraints (17) provide safety stock balance across the time periods. The safety stock remaining at the end of a period is equal to the amount held-over from the previous period plus the finished goods produced minus the amount used to meet demand. Constraints (18) prevent the manufacturer from holding more safety stock than the selected target level.

Constraints (19-20) enforce the rule that the company must meet demand, if possible. Constraints (19) require demand to be satisfied when there is production capacity, and constraints (20) require demand to be met when there is safety stock. We note that the capacity-to-meet-demand variable \( C_t^{\omega} \) is implied to be binary, proven via Lemma 2 in Section 4.3. Constraints (21) only allow safety stock to be depleted if there is no available capacity.

Constraints (22-23) define the two capacity-related variables: capacity-to-meet-demand and excess capacity, respectively. Constraints (22a-b) ensure the capacity-to-meet demand variables are 1 if a complete supply chain is working and selected, and constraints (22c) require them to be 0, if not. More specifically, in constraints (22a), the coefficient \( \tilde{\xi}_{jlt}^{\omega} \) represents the status (available or disrupted) of a complete supply chain \((j, k, l)\) of a given supplier \( j \in J \) and line \( l \in L_k \) in plant \( k \in K \) in period \( t \in T \). For a complete supply chain to be available, each component in the configuration must be available, i.e., \( \tilde{\xi}_{jlt}^{\omega} = \tilde{\xi}_{klt}^{\omega} = \tilde{\xi}_{lt}^{\omega} = 1 \). The variable \( \hat{z}_{jl} \) designates whether the complete supply chain that includes supplier \( j \in J \) and line \( l \in L \) is selected. If there is a complete supply chain selected and available, the right-hand side of constraint (22a) will be 1 for at least one combination, and the capacity-to-meet demand variable will be forced to at least 1; it is limited to 1 via constraints (22b). In constraints (22c), the term \( \sum_{j \in J} \sum_{k \in K} \sum_{l \in L_k} \hat{z}_{jlt} \) sums the statuses of candidate complete supply chains, and if there is not a selected and available supply chain, the capacity-to-meet demand variable is forced to 0. Constraints (23) define the excess capacity available each period. It is calculated as the total available capacity minus the capacity-to-meet demand.
Constraints (24-nonlin) enforce the requirement that the company must replenish safety stock when possible, up to the target level. As these are nonlinear, we reformulate them using binary indicator variables and implement constraints (24a-b) rather than constraints (24-nonlin). The values of the left-hand sides of constraints (24a-b) represent the amount of drug produced over the amount used to meet demand. The constraints require that these values must be at least equal to the excess capacity (24a) or the safety stock deficit (24b).

Constraints (25) indicate whether safety stock is available to meet demand. Constraints (25a) force the availability indicator to 0 if no safety stock is held-over from the previous period, and constraints (25b) force it to 1 if there is. We note that \( I^\omega_t \forall t \in \{0\} \cup T, \omega \in \Omega \) will not take on fractional values, as stated by Lemma 4 in Section 4.3. Constraints (26) set the safety stock levels at the beginning of the time horizon to the selected target level. Constraints (27) are standard non-negativity and domain constraints.

4.3. Structural Properties

In this section, we present key structural properties of the SCDD-I model. In Lemma 2, we determine that the capacity-to-meet demand variables are implied to be binary, and Lemma 3 states that the excess capacity variables are implied to be integer. Lemma 4 indicates that the number of periods of inventory held are implied to be integer. Lemma 5 states that demand is either fully met or fully unmet each period.

Lemma 2: \( C^\omega_t \in \{0,1\}, \forall t \in T, \omega \in \Omega \). Proof: Provided in appendix.

Lemma 3: \( \bar{C}^\omega_t \in \mathbb{Z}^+, \forall t \in T, \omega \in \Omega \). Proof: Provided in appendix.

Lemma 4: \( I_0, I^\omega_t \in \mathbb{Z}^+, \forall t \in \{0\} \cup T, \omega \in \Omega \). Proof: Provided in appendix.

Lemma 5: \( \theta^\omega_t \in \{0,1\}, \forall t \in T, \omega \in \Omega \). Proof: Provided in appendix.

We further use these lemmas and the corresponding safety stock replenishment rule to establish the implied non-anticipativity of SCDD-I through Theorem 1. The manufacturer’s decisions are only based on variables for the previous stage and the realization of uncertainty at the current stage; they do not consider uncertainty that will subsequently be revealed. For each period, it is optimal to make the same decisions for each of the scenarios that have identical realizations of uncertainty up to that period. We define \( S^\omega_t \) as the set of scenarios that have paths that are indistinguishable from scenario \( \omega \in \Omega \) in period \( t \in \{0\} \cup T \).

Theorem 1: The following relationships are implied by SCDD-I.

\[
\begin{align*}
C^\omega_t &= C'^\omega_t & & \forall \omega' \in S^\omega_t, t \in T, \omega \in \Omega \\
\bar{C}^\omega_t &= \bar{C}'^\omega_t & & \forall \omega' \in S^\omega_t, t \in T, \omega \in \Omega
\end{align*}
\] (28a) (28b)
\[ I_t^\omega = I_t^{\omega'} \quad \forall \omega' \in S_t^\omega, t \in T, \omega \in \Omega \] (28c)

\[ \theta_t^\omega = \theta_t^{\omega'} \quad \forall \omega' \in S_t^\omega, t \in T, \omega \in \Omega \] (28d)

\[ \sum_{i \in I} v_{it}^\omega = \sum_{i \in I} v_{it}^{\omega'} \quad \forall \omega' \in S_t^\omega, t \in T, \omega \in \Omega \] (28e)

\[ \sum_{j \in J} u_{jt}^\omega = \sum_{j \in J} u_{jt}^{\omega'} \quad \forall \omega' \in S_t^\omega, t \in T, \omega \in \Omega \] (28f)

\[ \delta_t^{\text{Available}, \omega} = \delta_t^{\text{Available}, \omega'} \quad \forall \omega' \in S_t^\omega, t \in T, \omega \in \Omega \] (28g)

\[ \delta_t^{\text{Sufficient}, \omega} = \delta_t^{\text{Sufficient}, \omega'} \text{ except case: } \tilde{C}_t^\omega = I_0 - I_{t-1}^{\omega'} \forall \omega' \in S_t^\omega, t \in T, \omega \in \Omega \] (28h)

**Proof:** Provided in appendix.

### 4.4. Additional Assumptions

SCDD-I is subject to the non-inventory-related assumptions of SCDD that are discussed in Section 3.4. In addition, because both the raw materials and finished form of the drugs are perishable, we make two assumptions: we do not allow the company to hold raw material inventory, and we apply an exogenous limit to the amount of finished goods inventory that may be held. In our conversations with a pharmaceutical manufacturer, these are consistent with practice. We assume that inventory is not destroyed if a facility is disrupted. Finally, we require the capacity of each line as a fraction of per-period demand to be integer-valued and at least equal to 1, and we evaluate different values in scenario analyses.

### 5. Solution Methods

The problems presented in Sections 3 and 4 represent a two- and a multi-stage stochastic program, respectively. There are \(2^{|I|}\) combinations of possible statuses for the candidate components, and over the entire time horizon, this produces a full scenario set of \(\left(2^{|I|}\right)^{|T|}\). For 10 candidate components and 4 time periods, this produces a set of \(1.1 \times 10^{12}\) scenarios and for 12 time periods produces a set of \(1.3 \times 10^{36}\) scenarios. As this is large, we approximate the optimal value using the SAA algorithm (Kleywegt, Shapiro, and Homem-de-Mello, 2002).

Our implementation is presented in Figure 4. The algorithm is comprised of three key steps – optimization, solution evaluation, and bound calculation. \(SCDD_{\text{SAA}}\) represents either the SCDD or SCDD-I model, as appropriate, where the complete uncertainty set \(\Omega\) is replaced by a set of sampled scenarios. In the optimization step, the set of \(\tau\) sampled scenarios is \(\tilde{\Omega}_r, \forall r \in \{1, R\}\). In the evaluation step, the set of \(\tilde{r}\) sampled scenarios is \(\tilde{\Omega}\). The optimal value of each solution evaluation step, \(\tilde{V}_r, \forall r \in \{1, R\}\), is an approximate lower bound on the true optimal value, \(V^*\). To set the lower bound, we select \(r'\) to be the index of the median value of \(\tilde{V}_r\).
6. Case study: Two Generic Oncology Drugs

To evaluate potential policy effects, we consider two drugs as case examples – vinblastine sulfate and vincristine sulfate. Both are generic, injectable drugs produced by single manufacturers that have been subject to recent shortages in the US (UUDIS, 2016). Vinblastine is used to treat various cancers including testicular cancer and lymphomas, and vincristine is used to treat leukemias and lymphomas (“Drugs.com,” 2018). Both are curative for some conditions, and we selected them based on conversations with an oncology pharmacist and a review of the literature.

6.1. Data

Table 1 presents data that are used for both analyses. Table 2 presents data that are specific to each drug, including demand and costs. These are derived from the available literature and conversations with subject matter experts.

We estimate the total US demand of each drug based on Medicare Part B data (CMS, 2018a, 2018b), and we apply prices from the Red Book (IBM Micromedex, 2018). Raw material costs are estimated from conversations with suppliers and available data online (PharmaCompass, 2018). Some of the costs are proprietary, but conversations with an industry expert estimated the full cost to produce a drug are 20-60% of the drug price, consistent with the values used by Jia and Zhao (2017). Using this range and other cost values, we calculated the production costs for each drug and non-fee-related fixed costs. Details are included in the supplementary materials, and we tested these values in sensitivity analyses. Based on conversations with an industry expert, we allow up to 2 years of finished goods inventory to be held, and we assume capacities for the suppliers and lines.

The distributions of time to disruption are estimated based on FDA data on drug approval dates and the start dates of shortages reported by University of Utah Drug Information Service (UUDIS) (FDA, 2018a; UUDIS, 2016). The distributions of time to recover are estimated from UUDIS on shortage length (UUDIS, 2016). Based on the data, we apply geometric distributions for both disruption and recovery, and further detail is available in the supplementary material. We consider a two-year time horizon based on conversations with the procurement office at a large academic health system and apply two-month time periods to be sufficiently granular while maintaining feasible run times. For the SCDD model, we apply 30 replications (R), 600 optimization scenarios (τ), and 1,200 evaluation scenarios (τ). For SCDD-I, we apply 40, 100, and 1,500, respectively. These were calibrated to consistently produce optimality gaps of 1% (SCDD) and 2% (SCDD-I). When the parameter values are at their baseline values, the SCDD model required approximately 340 seconds to run, and the SCDD-I required approximately 3,000 seconds.

6.2. Analysis Results

Using the two models, we can analyze how companies design their supply chains under different conditions and the associated impact on shortages and profit. All of the analyses were conducted
with both the SCDD and SCDD-I versions of the model, except for the safety stock analysis. When inventory is not selected in the optimal solution for either drug, we present SCDD results.

For each policy analysis, we present figures with 3 or 4 panels for each drug. These are: the optimal supply chain configuration; the target number of periods of inventory to hold (if selected); the expected shortage, and the percent difference in profit versus baseline. In some cases, there is apparent variability within a given cluster of points; this is largely due to the fact that the SAA method does not guarantee exact optimality. In the text, we round the values for unmet demand to the nearest percent and profit to the nearest $1,000. The algorithm and model were programmed in AMPL and solved using CPLEX 12.7 (Fourer, Gay, and Kernighan, 2002; IBM, 2017). We conducted analyses on a PC with a 2.3 GHz Intel Core i7 and 16 GB of RAM.

**No Intervention (Baseline)**

In the base-case, no policies are imposed. The manufacturer of vinblastine selects 2 suppliers, 1 plant, 1 line, and no safety stock. The expected percent of demand that is not satisfied (shortage) is 6% with a corresponding expected annual profit of $686,000. For vincristine (a higher cost, lower demand drug), the manufacturer selects 1 supplier, 1 plant, and 1 line with no safety stock, and the expected annual profit is $93,000. The expected shortage is 11%.

**Redundancy**

One proposal to increase resiliency is to require a company to maintain multiple components at a single echelon. This has been noted in the FDA Strategic plan (FDA, 2013), a report from the Drug Shortages Summit (ASHP, 2013), a joint letter to Congress from major health organizations (AHA et al., 2017), and other literature (Chabner, 2011; Gehrett, 2012; “Health Policy Brief: Drug Shortages,” 2014; Jaroslawski et al., 2017). To test the effects of redundancy regulation, we add the following variables and constraints.

**New decision variables**

\[ \tilde{\delta}_k = \begin{cases} 1 & \text{if at least one line in plant } k \in K \text{ is selected} \\ 0 & \text{otherwise} \end{cases} \]

**Constraints**

\[ \sum_{j \in J} x_j \geq 2x_1 \]  
\[ \sum_{k \in K} y_k \geq 2y_1 \]  
\[ \sum_{l \in L} z_l \geq 2z_1 \]  
\[ \tilde{\delta}_k - z_l \geq 0 \quad \forall l \in L_k, k \in K \]  
\[ \tilde{\delta}_k \leq \sum_{l \in L_k} z_l \quad \forall k \in K \]
\[ \sum_{k \in K} \delta_k \geq 2y_1 \] (34)

These constraints mandate that the company have multiple components at the given echelon(s) if they choose to be in the market. Constraints (29-31) require two suppliers, two plants, and two lines to be selected given that one component is selected, respectively. Constraints (32-34) require lines to be selected in multiple plants. Constraints (32-33) assign variables to indicate whether a line is selected in each plant. Constraint (34) requires lines to be selected in at least two plants if any plants are selected. To require the company to have multiple API suppliers, we include constraint (29). To require multiple plants, we include constraints (32-34), and to require multiple lines, we include constraint (31). To enforce redundancy at all levels, we include constraints (29-31).

We present results for this analysis in Figure 5. For both drugs across all regulations, the company selects no inventory and adds exactly as much redundancy as is required, unless it is unprofitable. For vinblastine, the company continues to maintain a second supplier, even when it is not mandatory, as in the baseline analysis. Any level of required redundancy is profitable for vinblastine, and the resiliency decisions lead to shortages of 1-6%, varying by echelon. Redundancy at all levels reduces the shortage to 1% but is the most costly, the expected annual profit decreases by 8%.

For vincristine, when redundancy at a single level is mandatory, the expected shortage drops to 5-9% of demand. Shortages are lowest (5% of demand) when a backup supplier is required, and expected profit is 21% lower than at baseline. If redundancy is required at every echelon, it would be unprofitable to make the drug, and the company chooses not to produce it (an expected shortage of 100%).

For both drugs, requiring a second plant causes substantial declines in expected profit (8% for vinblastine; 98% for vincristine). The costs and fees to maintain an additional plant and line are high relative to the baseline profits, and the increases in revenue from providing more of the drugs do not fully cover them. In general, redundancy regulations affect the difference in expected profit of vincristine more than vinblastine; this occurs because the baseline profit of vincristine is substantially lower. For a given regulation, the expected shortages for vincristine are generally higher than vinblastine because the vincristine supply chain does not include a backup supplier unless mandated.

**Mandatory Inventory Levels**

Some have proposed requiring manufacturers to hold minimum levels of inventory (e.g., ASHP 2013, FDA 2013, Gupta and Huang 2013). To run these analyses, we add a new parameter, \( \bar{\bar{I}} \), to represent the minimum level of target safety stock if the manufacturer is in the market. We also add constraint (35) to require that the target safety stock level be at least the minimum if any plants are selected. These analyses were run using the SCDD-I model.

\[ I_0 \geq \bar{\bar{I}}y_1 \] (35)
For both drugs, the manufacturer holds exactly the amount of inventory required up to a threshold at which it becomes unprofitable (Figure 6). For vinblastine, the company holds up to 20 months of inventory and for vincristine, up to 8 months. When at least 4 months are required, the vinblastine manufacturer does not maintain a backup supplier and uses inventory as the sole resiliency strategy. At 6 months of inventory, the expected shortages of vinblastine and vincristine are 5 and 4%, respectively, with drops in profit vs. baseline of 25% and 62%.

As the amount of mandatory inventory increases, the expected profits decrease. The expected shortages also generally decrease, up until the point where the drugs are not produced. The one exception occurs for vinblastine when the inventory requirement is increased from 2 months to 4 months. At 4 months of inventory, the company no longer maintains a backup supplier (Figure 6; panel A). This occurs because they are optimizing for expected profit, rather than for shortages, and with four months of inventory, it is more profitable to maintain a single supplier than multiple (even though 4 and 6 months of inventory provide less protection against expected shortages than a backup supplier and 2 months of inventory).

**Failure-to-Supply Penalties**

Pharmaceutical contracts typically do not include strong penalties if the manufacturer cannot supply the drug. If penalties are included, contracts are often written to require reimbursement for the additional cost to purchase the same drug from a different manufacturer (Haninger, Jessup, and Koehler, 2011; Jia and Zhao, 2017). However, frequently the drug is not available elsewhere, and these penalties are rarely paid. Several researchers have suggested that strengthening failure-to-supply clauses may induce resiliency (Conti, 2011; FDA, 2013; Haninger, Jessup, and Koehler, 2011; “Health Policy Brief: Drug Shortages,” 2014; Jia and Zhao, 2017; Reed et al., 2016). In this analysis, we apply a failure-to-supply penalty for each unit of unmet demand and add the term $-c_{\text{short}}(1 - \theta^\omega)$ to the objective functions (5) and (15). We present results for the SCDD-I model (Figure 7).

As it becomes more costly to not meet demand, the companies choose to add resiliency, and the expected shortages decrease. Resiliency is added at thresholds of the failure-to-supply penalties. Between these thresholds, as the penalty values increase, there is no change in the resiliency decisions, nor by extension, in the expected shortages. For example, when the failure-to-supply penalty for vincristine is $3.89 (70% of price), the company adds a backup supplier, and the shortages decrease from 12% to 5%. At the next threshold, $11.10 (200% of price), the company chooses to hold 2 months of safety stock and shortages drop to 2%. Between these thresholds, as the penalty is increased, the expected profit declines though the expected shortages stay fairly consistent (Figure 7, panels G and H). It becomes unprofitable to produce vincristine when the penalty is at least $22.20 (400% of price).

The thresholds at which failure-to-supply penalties change the resiliency decisions for vinblastine are $2.16 (50% of the unit price) when the company adds a backup line, and at $6.47 (150% of price) when the company adds a backup plant. These additions reduce the expected shortages to 3% and 1%, respectively.

**Pricing**
Some experts have pointed to the low prices of certain types of drugs as a primary driver of drug shortages. There have been corresponding calls for higher prices (e.g., Chabner 2011; Frakt 2016; Gatesman and Smith 2011; “Health Policy Brief: Drug Shortages,” 2014; Link et al. 2012). In this analysis, we vary the prices of each drug (Figure 8).

As the prices increase, the manufacturers of vinblastine and vincristine add more resiliency and expected shortages decline. As the price increases, the opportunity cost for not providing the drug during periods of shortage increases; at certain thresholds, it becomes more profitable to invest in resiliency and to be able to provide the drug more often. For vinblastine, when the price is 2 times baseline, the company adds a second line, and at 2.5 times baseline, it adds a second plant; the expected shortages are 3% and 1%, respectively. The corresponding expected profits increase 189% and 286% vs. baseline. For prices between 2.5 and 10 times baseline, the company does not add resiliency though expected profit continues to increase nearly linearly.

For vincristine, the price thresholds at which the company changes its supply chain are 1.75 times baseline (adds backup supplier); 2.5 times baseline (adds 2 months of inventory); and 9 times baseline (adds backup plant and removes inventory). The corresponding expected shortages are 5%, 3%, and 1%, and the difference in expected profits are 344%, 759%, and 4,214% vs. baseline, respectively.

We can also analyze the potential effects of price declines. If the price of vinblastine drops to 70% of baseline, the company does not maintain a second supplier, and the expected shortage increases from 5% to 11%. For vincristine, if the price drops to 75% of baseline, the company does not produce the drug.

**Social-Efficiency**

Many of the proposed policies increase cost, and prices could be concurrently increased to mitigate the effects on company profit. In this section, we analyze how much prices would need to increase to maintain expected profits at approximately baseline levels and calculate the societal costs to achieve target shortage levels. In Table 3, we present the policies that would lead to expected shortages of at most 5% and 2%, meaning the drug is available 95% and 98% of the time, respectively. We calculate the societal costs of each policy as the extra amount paid annually due to price increases, i.e., the product of the baseline drug price, annual demand, and the percentage price increase. For each policy analysis, we incrementally increased the price until the expected profit was approximately the baseline level.

For vinblastine, shortages at baseline are approximately 5%, and no intervention is necessary to reach this threshold. To achieve expected shortages of at most 2%, the following policies in combination with price increases would be effective: requiring a backup plant, failure-to-supply penalties of 150% of price, or 12 months of inventory; a price increase of 150% without other intervention would also be effective. The societal costs of each are $136,000 to $2 million, varying by policy.

To achieve expected vincristine shortages of at most 5%, requiring a backup supplier and increasing the price by 10% has the lowest societal cost. For expected shortages of 2%, price
increases of 30% in combination with requiring redundancy at all levels or 12 months of inventory have the lowest societal costs.

6.3. Sensitivity, Scenario, and Validation Analyses

To analyze the sensitivity of the results to changes in the parameter values, we conducted a one-way sensitivity analysis using the SCDD-I model. We varied each value by 20%. In each of these analyses, the optimal solution remained the same as in the baseline analysis. The variation in expected profit is available in the supplementary material. For both vinblastine and vincristine, the unit and fixed cost values are most influential on expected profit. In particular, the unit cost of production has the largest effect on expected profit. The disruption and recovery distribution parameters have less of an impact.

We also conducted scenario analyses on the other parameters. These included the lengths of the time horizon and periods, annual demand, and production capacity. The optimal solution did not change for either drug though there were minor differences in expected profit. Further detail is available in the supplementary material.

To validate our models, we compared our results with available data in the literature. The solutions at baseline also follow the lean supply chains and low inventory seen in practice (Fox, Sweet, and Jensen, 2014; GAO, 2016; Woodcock and Wosinska, 2013). Our failure-to-supply results are qualitatively confirmed with Jia and Zhao (2017). The baseline shortages of 6% for vinblastine and 11% for vincristine are similar to the percentage of drugs short each day reported by the drug shortage staff at a large academic health system. The other results have face validity; as price increases, resiliency increases, and the results mimic the dynamics of higher-margin, branded drugs.

7. Discussion

7.1. Modeling

Drug shortages are concerning because they are widespread, harmful, and persistent. To study why pharmaceutical companies may make supply chain decisions that contribute to shortages, we develop two new supply chain design models: SCDD and SCDD-I. These models combine features previously considered separately to provide a framework for understanding the effects of disruptions over time and for evaluating policies. They incorporate the multi-period aspect of inventory models under disruption with the facility selection decisions of location models to consider multiple mitigation strategies. In addition, disruptions may occur at multiple echelons and concurrently. These models allow us to approximate the strategic decisions pharmaceutical companies make for fixed-term contracts.

The baseline model, SCDD, is relatively simple and could be easily extended to include additional echelons, location decisions, or correlations between component disruptions. It is appropriate for settings in which inventory either cannot be held or is very expensive. The extended model, SCDD-I adds inventory as a resiliency strategy. This feature complicates the model, though we impose a replenishment rule that implies the non-anticipativity property holds
in the optimal solution. This substantially reduces the computational burden and allows us to use SAA to solve thirteen-stage stochastic programs within tight optimality gaps (i.e., thirteen based on the initial stage and 12 subsequent periods). Without this rule, the problem would require specialized algorithms to solve.

7.2. Is Low Resiliency Optimal?

Using these models, we consider the case examples of the supply chains of the oncology drugs vinblastine and vincristine. While resiliency is often optimal in other contexts (Tomlin, 2006), pharmaceutical companies may find instead that passive acceptance of risk is optimal for certain drugs, i.e., low-margin products with long, infrequent disruptions. Our results suggest that with a profit-maximizing objective and no intervention, it would be best to have no resiliency in the supply chain of vincristine, a low volume drug, leading to an expected shortage of 11%. For the higher volume drug, vinblastine, it is beneficial to have a backup supplier, and the expected shortage is 6%. For other stakeholders in the healthcare system, these levels are untenable.

7.3. How Can We Induce Resiliency and Reduce Drug Shortages?

Given that it is in society’s best interest to reduce shortages of life-supporting drugs, the question becomes which strategies to induce resiliency would be best. We use the metric of social-efficiency (i.e., lowest total cost to meet specified expected shortage levels) to evaluate proposed options. We studied the legislative policies of mandating redundancy and safety stock and contractual policies of failure-to-supply penalties and price increases. For each, we evaluated the prices needed to maintain company profits at baseline levels. We remind the reader that these results are presented within the context of low-profit margin, sole-source, generic, injectable oncology drugs.

We find that the most efficient policy depends on the desired shortage level. For shortages of 5% or less (i.e., expected to be available at least 95% of the time), no intervention is needed for vinblastine, and for vincristine, it is most efficient to require multiple suppliers and increase prices by 10%. A failure-to-supply penalty of 70% would equivalently induce a backup supplier but would require a greater price increase, 20%. These results suggest that maintaining multiple suppliers is an effective way to reduce shortages, though the societal cost to induce them would depend on the decision maker.

Shortages of at most 2% could be achieved by requiring a backup plant with a 10% price increase (vinblastine) or requiring redundancy at all levels with a 30% price increase (vincristine). In both cases, the outcome is a supply chain with a backup at each echelon. An alternative policy to induce the same supply chain for vinblastine would be a failure-to-supply penalty of 150% with a price increase of 10%. For vincristine, an alternative would be to mandate the company hold one year of safety stock in combination with a 30% price increase.

In general, requiring safety stock is a relatively expensive policy option. This may be because the average disruption length is long and holding inventory of injectable drugs is costly. For vinblastine to have at most 2% shortages, a safety stock mandate is three times as costly as mandating multiple plants or adding a 150% failure-to-supply penalty (price increases of 30% vs.
10%). For vincristine, it is two times as costly to mandate sufficient inventory as it is to require multiple suppliers for a 5% shortage level (price increases of 20% vs. 10%). If safety stock were held, it could either be maintained at the manufacturer or in a stockpile similar to the Strategic National Stockpile of pediatric vaccines (Jacobson, Sewell, and Proano, 2006); the analysis would be the same in either case.

While shortages may be driven by low profit margins, in no analysis are price increases alone the most efficient policy. For vinblastine, at the 2% shortage level, price increases are 15 times more costly than the most efficient policy (150% increase vs. 10%). For vincristine at the 5% shortage level, price increases are 7 times more costly (70% increase vs. 10%) than the most efficient option, and they are 27 times more costly (800% increase vs. 30%) at the 2% shortage level. These pricing results are consistent with an analysis from Jia and Zhao (2017) that found that adding failure-to-supply clauses in combination with moderate price increases would be more efficient than price increases alone.

These results suggest that legislative action to mandate redundant components in combination with price increases would have the lowest societal cost for both drugs. For vinblastine, sufficient failure-to-supply penalties in combination with price increases would also have the lowest societal cost; for vincristine, this is a costlier option. Though legislative change is difficult, it may be possible. A 2012 law changed reporting requirements for shortages, and in 2018, the FDA and other agencies initiated a Drug Shortage Task Force to provide new recommendations to Congress (“FDASIA, Public Law 112–144,” 2012; Gottlieb, 2018a). Contractual changes could be negotiated by Group Purchasing Organizations or other procurement officials. Price increases would likely be incurred by Medicare for Part B recipients and passed on to private payers for patients with private insurance.

7.4. Limitations

The results of this paper are tempered by its limitations, and readers should be careful to interpret analyses within the appropriate scope. Our models are subject to a variety of assumptions, discussed in Sections 3.4 and 4.4. The analyses assume a stationary market share, i.e., other companies do not enter the market. Given the high utilization of existing manufacturing capacity, most firms make decisions for a portfolio of products, rather than individual drugs. These analyses assume the manufacturer does not choose to use the capacity for a more profitable drug. Finally, we are limited by available data as pharmaceutical data are frequently proprietary. We have taken strides to estimate reasonable parameter values and conduct sensitivity analyses. In particular, profit results should not be taken as exact projections but rather as indications of the magnitude of policy effects. For this reason, we have focused on the change vs. baseline rather than absolute numbers. The optimal supply chain configurations and target inventory levels do not vary as parameter values are varied by 20%, and analyses of demand indicate that the solutions do not change within wide ranges.

7.5. Conclusions

Strategic supply chain decisions have contributed to major drug shortages, and we find that for certain types of drugs with low profit margins, pharmaceutical companies may find it optimal to
maintain vulnerable supply chains. In this analysis, we seek to align the interests of for-profit companies with the public good of a stable drug supply. Experts have suggested that regulation may be required to reduce shortages. Our results provide evidence that redundancy regulations would be at least as efficient as market-based solutions. If legislation is pursued, additional analysis would be necessary to determine the particular characteristics of medically-necessary drugs to which it should be applied. In the absence of expanded regulation, group purchasing organizations and other contract-makers could negotiate failure-to-supply clauses in combination with modest price increases to reduce shortages. Price increases alone could also be effective but would cost substantially more. These models provide a framework to consider disruptions in strategic design decisions. Future work could consider improving quality as a resiliency strategy, which could lead to less frequent disruptions or faster recovery. We also plan to study the effects of price increases and competition.

References


UUDIS. (2016). *Fox Data Summary.xlsx*.


**Acronyms**

API = Active Pharmaceutical Ingredient

EOQD = Economic Order Quantity model with Disruptions

FDA = Food and Drug Administration

GDUFA = Generic Drug User Fee Amendments

SAA = Sample Average Approximation
SCDD = Supply Chain Design under Disruption
SCDD-I = Supply Chain Design under Disruption with Inventory
SCRM = Supply Chain Risk Management
US = United States
UUDIS = University of Utah Drug Information Service
WAC = Wholesale Acquisition Cost

Acknowledgements

We would like to thank Erin R. Fox, PharmD of UUDIS for providing shortage data and her perspective. We appreciate Deanna Handley, Hannah Strat, Alexander Lopez, and the Institute for Healthcare Policy Innovation at the University of Michigan for supporting additional data collection and analysis. The authors are grateful for helpful discussions with the Drug Shortage Staff at the FDA and participants of the INFORMS Annual Meetings (2016, 2017) and INFORMS Healthcare 2017. We appreciate discussions with Barbara Higgins, PharmD, Iris Wang, PhD, and Allan Pfitzenmaier, MBA.

We would also like to thank the reviewers for their valuable comments.

Biosketches

Emily L. Tucker is a PhD Candidate in the Department of Industrial and Operations Engineering at the University of Michigan. She received her MSE in IOE from Michigan and her BS in Industrial Engineering from NC State. Prior to graduate school, she worked as a Research Health Economist for RTI International. She is a recipient of the NSF Graduate Research Fellowship and was a finalist for the 2017 Bonder Scholarship for Applied OR in Health Services. She has served as President of the Student Leadership Board in IOE and an editor of OR/MS Tomorrow, the INFORMS student magazine. Her research interests include the application of operations research to healthcare policy, operations, and supply chain decisions.

Mark S. Daskin is a professor in the Department of Industrial and Operations Engineering at the University of Michigan. He holds the Clyde W. Johnson Collegiate Chair in the College of Engineering. He is a fellow of both INFORMS and IIE and has received the IIE David F. Baker Distinguished Research Award. He is a member of the National Academy of Engineering. He has served as the Editor-in-Chief of both Transportation Science and IIE Transactions and as the president of INFORMS. He has also served as the chair of the IOE Department at Michigan and of the Industrial Engineering and Management Sciences Department at Northwestern University. His primary research interests lie in the application of operations research to supply chain problems, location models, and healthcare operations. He is the author of two books: Network and Discrete Location: Models, Algorithms, and Applications and Service Science, both published by Wiley.
Burgunda V. Sweet, PharmD, FASHP, is a clinical professor of pharmacy and director of curriculum assessment at the University of Michigan College of Pharmacy. She earned her PharmD degree at the University of the Pacific College of Pharmacy, completed a clinical pharmacy residency at the VA Hospital in San Diego, California, and an advanced practice residency in drug information at the University of Michigan. Dr. Sweet served as the director of the drug information and investigational drug services at Michigan Medicine for over 20 years where she interfaced with clinicians from different healthcare disciplines on a daily basis to resolve complex health-system problems, including managing drug shortages. She continues to practice as a clinical specialist in medication use policy at the Michigan Medicine health system. She has authored numerous articles and book sections in her area of specialty.

Wallace J. Hopp is the Prahalad Distinguished University Professor of Business and Engineering at the University of Michigan. His research on manufacturing and supply chain systems, innovation processes, and health care systems has received a number of awards including the 1985 Nicholson Prize, 1990 Scaife Award, 1998 IIE Joint Publishers Book-of-the-Year Award, 2005 IIE Technical Innovation Award, 2011 INFORMS Pietsch Award for Best Paper in Health Care Management Science, 2016 M&SOM Journal Best Paper Award, and 2016 MSOM Service Management SIG Best Paper Award. Hopp is a member of the National Academy of Engineering, and a Fellow of IIE, INFORMS, MSOM, POMS and SME. He has served as President of the Production and Operations Management Society and Editor-in-Chief of Management Science.

Table 1. Values for all analyses

<table>
<thead>
<tr>
<th>Input</th>
<th>Supplier</th>
<th>Plant</th>
<th>Line</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual fixed costs</td>
<td>$33,000</td>
<td>$65,000</td>
<td>$32,500</td>
<td>Rudge (2012) and assumptions</td>
</tr>
<tr>
<td>Annual GDUFA fees</td>
<td>$1,169</td>
<td>$4,401</td>
<td>n/a</td>
<td>Calculated based on FDA (2018b)</td>
</tr>
<tr>
<td>Capacity as a fraction of per-period demand</td>
<td>n/a</td>
<td>n/a</td>
<td>2</td>
<td>Assumed</td>
</tr>
<tr>
<td>Average time to disruption, in years</td>
<td>17.3</td>
<td>28.2</td>
<td>8.5</td>
<td>Calculated based on FDA (2018a) and UUDIS (2016)</td>
</tr>
<tr>
<td>Average time to recovery, in years</td>
<td>1.2</td>
<td>0.8</td>
<td>0.08</td>
<td>Calculated from UUDIS (2016)</td>
</tr>
</tbody>
</table>

API = Active Pharmaceutical Ingredient; GDUFA = Generic Drug User Fee Amendments; SAA = Sample Average Approximation

Costs in 2018 US dollars

Table 2. Drug-specific data

<table>
<thead>
<tr>
<th>Input</th>
<th>Vinblastine</th>
<th>Vincristine</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual demand, in ml</td>
<td>315,000</td>
<td>90,000</td>
<td>Estimated from CMS (2018a, b) and National Cancer Institute (2018)</td>
</tr>
</tbody>
</table>
ml = milliliter; WAC = Wholesale Acquisition Cost

§Costs in 2018 US dollars

Table 3. Summary of policy costs

<table>
<thead>
<tr>
<th>Drug</th>
<th>Shortage Upper Bound</th>
<th>Policy</th>
<th>Price Increase</th>
<th>Annual Societal Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vinblastine</td>
<td>0.05</td>
<td>Not applicable§</td>
<td>0%</td>
<td>$0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Require multiple plants</td>
<td>10%</td>
<td>$136,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150% failure-to-supply penalty</td>
<td>10%</td>
<td>$136,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Require 12 months inventory</td>
<td>30%</td>
<td>$407,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Price increase</td>
<td>150%</td>
<td>$2,036,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Require multiple suppliers</td>
<td>10%</td>
<td>$50,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70% failure-to-supply penalty</td>
<td>20%</td>
<td>$100,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Require 6 months of inventory</td>
<td>20%</td>
<td>$100,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Price increase</td>
<td>70%</td>
<td>$350,000</td>
</tr>
<tr>
<td>Vincristine</td>
<td>0.05</td>
<td>Require multiple redundancy at all levels</td>
<td>30%</td>
<td>$150,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Require 12 months of inventory</td>
<td>30%</td>
<td>$150,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200% failure-to-supply penalty</td>
<td>50%</td>
<td>$250,000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Price increase</td>
<td>800%</td>
<td>$3,996,000</td>
</tr>
</tbody>
</table>

§The results at baseline approximately achieve the 5% threshold (using the SCDD model: 6%; SCDD-I model: 5%).

Figure 1. Example supply chain
Figure 2. Notation for SCDD

<table>
<thead>
<tr>
<th>Sets</th>
<th>Decision Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J$</td>
<td>First Stage</td>
</tr>
<tr>
<td>$K$</td>
<td></td>
</tr>
<tr>
<td>$L$</td>
<td></td>
</tr>
<tr>
<td>$L_k$</td>
<td></td>
</tr>
<tr>
<td>$N$</td>
<td></td>
</tr>
<tr>
<td>$T$</td>
<td></td>
</tr>
<tr>
<td>$\Omega$</td>
<td></td>
</tr>
<tr>
<td>$x_j := \begin{cases} 1 &amp; \text{if API supplier } j \in J \text{ is selected} \ 0 &amp; \text{otherwise} \end{cases}$</td>
<td></td>
</tr>
<tr>
<td>$y_k := \begin{cases} 1 &amp; \text{if manufacturing plant } k \in K \text{ is selected} \ 0 &amp; \text{otherwise} \end{cases}$</td>
<td></td>
</tr>
<tr>
<td>$z_l := \begin{cases} 1 &amp; \text{if line } l \in L \text{ is selected} \ 0 &amp; \text{otherwise} \end{cases}$</td>
<td></td>
</tr>
</tbody>
</table>

**Second Stage**

As a fraction of demand in period $t \in T$ in scenario $\omega \in \Omega$:

- $w_{jt}^\omega$: Raw material purchased from supplier $j \in J$
- $t_{jt}^\omega$: Finished goods produced on line $l \in L$
- $\theta_{jt}^\omega$: Demand met

**Parameters**

- $p^\omega$: Probability of scenario $\omega \in \Omega$
- $e_n^\omega$: If component $n \in N$ is available in period $t \in T$ in scenario $\omega \in \Omega$
- $r$: Quantity of drug demanded each period
- $d$: Sales price per unit of drug
- $e_{\text{raw}}$, $e_{\text{prod}}$: Unit cost of raw materials and finished good production
- $c_{API}$, $c_{\text{Plant}}$, $c_{\text{Line}}$: Annual fixed costs for each supplier, plant, and line, respectively
- $f_{API}$, $f_{\text{Plant}}$: Annual GDUA fees for each supplier and plant, respectively
- $f_{\text{Program}}$: Annual GDUA fee for drug program
- $g_{\text{Line}} \in \mathbb{Z}^+$: Line capacity as a fraction of total demand, if line is available
- $\bar{f}$: Number of periods per year

Figure 3. Additional Notation for SCDD
Figure 4. Implementation of SAA

1. Optimization and approximation
   a. For $r = 1..R$, sample $r$ independently and identically distributed scenarios $\bar{\beta}^r$ from the complete uncertainty set $\Omega$
      i. Define the set of these scenarios to be $\tilde{\beta}^r$
      ii. Solve $SCDD_{SL}$ with uncertainty set $\tilde{\beta}^r$
      iii. Record the optimal value, $\bar{\nu}$, and the first stage decision variables $x^r, y^r, z^r, I^r$

2. Evaluation
   a. Generate $r$ independently and identically distributed samples of $\bar{x}_{\omega}^r$ from the complete uncertainty set $\Omega$
      i. Define the set of these scenarios to be $\bar{\beta}$
      ii. For $r = 1..R$,
         1. Fix $x = x^r, y = y^r, z = z^r, I = I^r$
         2. Solve $SCDD_{SL}$ with uncertainty set $\bar{\beta}$
         3. Record the optimal value, $\bar{\nu}$

3. Computation of the lower and upper bounds
   a. Select replication index $r'$, and set $LB := \bar{\nu}^{r'}$
   b. Set $UB := \frac{1}{R} \sum_{r=1}^{R} \bar{\nu}^r$
   c. Record solution $x^{r'}, y^{r'}, z^{r'}, I^{r'}$

Figure 5. Effects of redundancy regulations
Figure 6. Effects of safety stock requirements

Figure 7. Effects of failure-to-supply penalties
Figure 8. Effects of varying drug price