Planning for Demand Failure: A Dynamic Lot Size Model for Clinical Trial Supply Chains

Adam J. Fleischhacker[a], Yao Zhao[b]

[a] *Department of Business Administration, University of Delaware, Newark, DE 19716, USA*

[b] *Department of Supply Chain Management & Marketing Sciences, Rutgers Business School – Newark and New Brunswick, 1 Washington Park, Newark, NJ 07102, USA*

This paper examines the optimal production lot size decisions for clinical trial supply chains. One unique aspect of clinical trial supply chains is the risk of failure, meaning that the investigational drug is proven unsafe or ineffective during human testing and the trial is halted. Upon failure, any unused inventory is essentially wasted and needs to be destroyed. To avoid waste, manufacturers could produce small lot sizes. However, high production setup costs lead manufacturers to opt for large lot sizes and few setups. To optimally balance this tradeoff of waste and destruction versus production inefficiency, this paper generalizes the Wagner-Whitin model (*W-W model*) to incorporate the risk of failure. We show that this stochastic model, referred to as the *failure-risk model*, is equivalent to the deterministic W-W model if one adjusts the cost parameters properly to reflect failure and destruction costs. We find that increasing failure rates lead to reduced lot sizes and that properly incorporating the risk of failure into clinical trial drug production can lead to substantial cost savings as compared to the W-W model without the properly adjusted parameters.

1. Introduction

For every new drug that reaches a pharmacy's shelf, roughly 5,000 to 10,000 other potential medicines have failed to achieve commercialization (Pharmaceutical Research and Manufacturers of America 2009). For a pharmaceutical or bio-tech company attempting to create a new medicine or treatment, failure is not a surprise, but rather an event to be planned for. In this paper, we analyze the impact of failure during clinical trials on the production-inventory decisions for investigational drugs and discover that an extension of the Wagner-Whitin model (Wagner and Whitin 1958) can greatly improve efficiency in the clinical trial supply chain.

One of the most important hurdles prior to the U.S. Food and Drug Administration's (FDA) approval of a new drug is the testing of a drug candidate in clinical trials. Three phases of clinical trials are usually required to test both safety and efficacy of a potential treatment in human subjects. Typically, Phase I involves 50 to 100 healthy individuals, Phase II recruits a few hundred potential patients, and Phase III seeks to test the drug candidate in a few thousand patients. While we may know how many patients are needed in each phase of the clinical trial, there is an inherent uncertainty associated with each trial:

the risk of failure. Indeed, only 21.5% of drug candidates entering clinical trials actually achieve FDA approval (DiMasi et al. 2003). Many of these drug candidates that fail to pass through the clinical trial hurdle are well documented in the financial press. Below is just one example from the *New York Times* (Berenson 2006):

> The news came to Pfizer's chief scientist, Dr. John L. LaMattina, as he was showering at 7 a.m. Saturday: the company's most promising experimental drug, intended to treat heart disease, actually caused an increase in deaths and heart problems. Eighty-two people had died so far in a clinical trial, versus 51 people in the same trial who had not taken it.
>
> Within hours, Pfizer, the world's largest drug maker, told more than 100 trial investigators to stop giving patients the drug, called torcetrapib. Shortly after 9 p.m. Saturday, Pfizer announced that it had pulled the plug on the medicine entirely, turning the company's nearly $1 billion investment in it into a total loss.

The small success rate of clinical trials is painful to a pharmaceutical company's balance sheet because of the enormous amounts of time, labor, and materials required to perform a clinical trial. On average, 37% of the $100 billion R&D spending by pharmaceutical companies is spent on the clinical trial process (Cutting Edge Information 2004, Thomson CenterWatch 2007). Annual *s*upply chain spending for drugs under clinical trials can be substantial, e.g., accounting for 20% or more of a company's research and development spending.[1] For just one drug candidate, a company can spend millions of dollars every quarter to produce supplies for just one clinical trial. When failure in a clinical trial occurs, every dollar spent on manufacturing, packaging, and distribution of unused clinical trial supplies is wasted and in most cases, unused material must be returned to a proper disposal facility for destruction (English and Ma, 2007). For example, Cotherix Inc., estimated $126,000 in destruction costs for an obsolete drug that was valued at $1.5 million (Cotherix 2006).

It would be unfair of us to label all post-failure drug supply as waste. Inventory is needed to ensure that as patients are recruited to participate in the study, drug supply is available. Any delays in this phase of testing become one less day of patent protection available to the drug. According to Clemento (1999), every extra day of patent availability is worth $1 million for a typical drug. Since patient recruitment is the typical bottleneck in conducting clinical trials, a shortage of clinical drug is considered an unacceptable delay and our model

[1] As an example, 21.3% of Allos Theraputics R&D spending from 1995 - 2006 went towards clinical trial manufacturing-related activities (Allos Therapeutics Inc. 1999-2007).

assumes no backlogging of demand. That being said, one would usually be economically foolish to produce enough supply to support all three phases of a clinical trial at once.

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

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

- We demonstrate how every failure-risk model is equivalent to a corresponding deterministic W-W model if one adjusts the cost parameters properly to reflect failure risk and destruction costs, so many classic results of the W-W model still apply. Most interestingly, the planning horizon theorem indicates that in the failure-risk model, updating failure probabilities as the clinical trial proceeds does not affect the optimal supply decisions under certain conditions.

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

The remainder of this paper is organized as follows. We review the related literature in §2. The model and analysis are presented in §3, and their extensions are discussed in §4. The

potential benefits of properly accounting for failure are shown in an illustrative example in §5. A more thorough numerical study to test the effectiveness of the model under real-world scenarios is performed in §6 with implications within industry discussed in §7. Finally, we summarize the paper and discuss future research directions in §8.

2. Literature Review

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

Qualitative investigations of drug supply decisions made within the clinical trial process are found in the medical and pharmaceutical literature. For example, George (2005) presents common issues encountered during clinical trial supply management and proposes coordination and flexibility as keys to success. A more thorough description of clinical material manufacturing practices is provided by Bernstein and Hamrell (2000). In their paper, the authors advocate coordinating the disciplines of manufacturing and clinical programs to achieve efficient execution of drug development. Their study is conceptual and qualitative; in contrast, our approach leverages mathematical modeling and numerical studies to yield insights.

Quantitative research that deals with uncertainties specific to clinical trials and their production decisions has also been conducted. Shah (2004) provides a survey for this line of research. Colvin and Maravelias (2008) highlights more recent advances and presents a stochastic programming approach to scheduling clinical trials for a portfolio of drug candidates that share limited resources. Notable extensions of this work to accommodate outsourcing possibilities (Colvin and Maravelias, 2009) and more efficient algorithms (Colvin and Maravelias, 2010) have been studied. Gatica, et al. (2003) and Levis and Papageorgiou (2004) also leverage stochastic programming approaches to simultaneously determine the optimal capacity and production decisions for multiple clinical trial drugs. In each case, the underlying problem is a large-scale multi-stage stochastic program with integer and continuous variables. In addition to stochastic programming approaches, simulation has also been used to deal with production scheduling issues given the uncertainty in outcome of clinical trials for a portfolio of drugs (see for example Subramanian, 2003 and Blau, et al, 2004).

This paper differs from the previous quantitative work on clinical trial supply chains by its focus. We study a simpler model with only one drug candidate and aim at deriving structural results which provide managerial insights and enable efficient solution algorithms.

In contrast to scheduling production in a resource-constrained environment for a portfolio of drug candidates (see relevant review in Verderame, 2010), the heart of our model is balancing production efficiency for a single product versus the potential for waste (as a result of a clinical trial failing) in an environment where an expensive end product is often destroyed as opposed to consumed. Given this, our work is also closely related to the dynamic economic lot size (DEL) models and stochastic inventory models studied in the operations management literature.

There is a long lasting interest and huge body of literature on DEL models for production-inventory systems with time-varying but known demand. Wagner and Whitin (1958) proposes the basic model (referred to as the *W-W model* hereafter) and which was named as one of the "Ten Most Influential Titles of Management Science's First 50 Years" (Hopp, 2004). The paper characterizes several important system properties and develops a polynomial solution algorithm. Since then, many extensions and variations of this model have been studied. For more efficient solution algorithms, see Aggarwal and Park (1993), Federgruen and Tzur (1991) and Wagelmans, et al. (1992). For DEL models with various capacity constraints, see, e.g., Florian, et al. (1980) and Shaw and Wagelmans (1998). For more general cost functions, see Eppen, et al. (1969), Veinott (1963) and Zangwill (1969). More recently, Chu et al. (2005) study a lot sizing problem with general economies of scale cost functions. Realizing the problem is NP-hard, they develop approximation solutions and perform a worst-case analysis. Zipkin (2000) provides a thorough review of models and solution techniques on this topic.

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

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

W-W model) and new insights (e.g., impact of failure risk).

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

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

> Generally, we find that obsolescence does (or should) have a substantial impact in the way inventories are managed. The nature of these effects, moreover, is fairly intricate. It appears that obsolescence cannot be captured in a simpler model through parameter adjustments.

Leveraging the deterministic demand assumption, we can formulate the failure-risk model into the simpler W-W model where the adjusted cost parameters incorporate the costs of failure. This result connects the failure-risk model with the vast literature of the W-W models,

and thus, many results of the latter directly apply here. In addition, adjusting parameters of the W-W model is a simple way to include failure into production planning and thus, is more likely to be implemented than more complex obsolescence models. In conversations with industry professionals, they often comment that complicated models may prove too sophisticated as their supply planners often have strong pharmaceutical backgrounds, but not equally strong quantitative skill sets. Lastly, we believe the demand failure assumption to be tenable to practitioners who can often, but certainly not always, accurately predict demand (assuming the trial's success). According to a recent survey of clinical supply managers conducted by Bearing Point, 75% of Phase I and roughly 50% of Phases II-III SKU-level clinical supply forecasts are within 10% of actual demand (Kumar, 2008).

Our work is also related to the literature on R&D supply chains. Most of this literature focuses on supply chain design to support a product entering the market for the first time. However, much less attention has been devoted to the actual development supply chain (Krishnan and Ulrich 2001, Pisano 1997). At a pharmaceutical company, both the supply chain design for production ramp-up and the material supply during the development stage are important decisions. The focus of this paper is on creating a model for the latter. More recently, there is a growing interest in combining R&D and supply chain decisions. Hult and Swan (2003) provides a conceptual framework to analyze the interdependencies of product development and supply chain activities. Specific to the pharmaceutical world, Pisano (1997) presents strategic guidelines for effectively linking manufacturing strategy with the highly uncertain world of drug candidate development.

3. The Model

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

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

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

The system has the following parameters,

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

The estimates of failure probabilities in various therapeutic classes are readily available from the literature (see Gatica et al., 2003, DiMasi et al., 2010). It is possible that the failure probability of a trial does not depend on the results of previous trials if they are testing on different criteria, e.g., efficacy vs. safety. In this case, α_t is the unconditional probability of failure in the trial. It is also possible that the failure probabilities depend on the results of previous tests. For instance, during multiple trials for effectiveness, success in early trials can provide a strong indicator for success in on-going trials. In this case, α_t is effectively the failure probability conditioning on successes to date. Likewise, estimates of demand, d_t, also assume success to date as any failure in the trial results in the halting of a investigational drug's use.

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

$$
\begin{aligned}
f_t(0, I) &= r_t I, \quad 1 \leq t \leq N && (3.1)\\
f_t(1, I) &= \min_{\{x_t \geq 0,\ I + x_t \geq d_t\}} \{h_{t-1} I + \delta(x_t) s_t + c_t x_t + \alpha_t f_{t+1}(0, I + x_t - d_t) + \\
&\qquad \beta_t f_{t+1}(1, I + x_t - d_t)\}, \quad t = 1, 2, \ldots, N-1 && (3.2)\\
f_N(1, I) &= \min_{\{x_N \geq 0,\ I + x_N = d_N\}} \{h_{N-1} I + \delta(x_N) s_N + c_N x_N\}. && (3.3)
\end{aligned}
$$

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

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

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

$$\begin{aligned} g_t(I) &= \min_{\{x_t \geq 0,\ I+x_t \geq d_t\}} \{\frac{\alpha_{t-1} r_t + \beta_{t-1} h_{t-1}}{\beta_{t-1}} I + \delta(x_t) s_t + c_t x_t + \beta_t g_{t+1}(I + x_t - d_t)\}, \\ & \qquad t = 1, 2, \ldots, N-1 \\ g_N(I) &= \min_{\{x_N \geq 0,\ I+x_N = d_N\}} \{\frac{\alpha_{N-1} r_N + \beta_{N-1} h_{N-1}}{\beta_{N-1}} I + \delta(x_N) s_N + c_N x_N\}. \end{aligned}$$

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

$$\begin{aligned} s'_1 &= s_1 \\ s'_t &= s_t \cdot \Pi_{j=1}^{t-1} \beta_j, \quad 1 < t \leq N \\ c'_1 &= c_1 \\ c'_t &= c_t \cdot \Pi_{j=1}^{t-1} \beta_j, \quad 1 < t \leq N \\ h'_1 &= \alpha_1 r_2 + \beta_1 h_1 \\ h'_t &= (\alpha_t r_{t+1} + \beta_t h_t) \cdot \Pi_{j=1}^{t-1} \beta_j, \quad 1 < t < N. \end{aligned}$$

Hence,

$$g'_t(I) = \min_{\{x_t \geq 0,\ I+x_t \geq d_t\}} \{h'_{t-1} I + \delta(x_t) s'_t + c'_t x_t + g'_{t+1}(I + x_t - d_t)\},\ t = 1, 2, \ldots, N-1 \quad (3.5)$$

$$g'_N(I) = \min_{\{x_N \geq 0,\ I+x_N = d_N\}} \{h'_{N-1} I + \delta(x_N) s'_N + c'_N x_N\}. \quad (3.6)$$

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

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

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

1. *For each period t, $I_t \cdot x_t = 0$.*

2. $x_t = 0$ *or* $x_t = \sum_{j=t}^{k} d_j$.

3. *If* d_t *is satisfied by some* x_τ *for* $\tau < t$, *then* d_j, $j = \tau + 1, \ldots, t-1$ *is also satisfied by* x_τ.

4. *Given that* $I_t = 0$ *for period* t, *it is optimal to consider periods 1 through* $t-1$ *independent of other periods.*

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

$$\begin{aligned} g'_t(I) &= \min_{\{x_t \geq 0,\ I + x_t \geq d_t\}} \{(h'_{t-1} - c'_t)I + \delta(x_t)s'_t + c'_t(I + x_t - d_t) + c'_t d_t + g'_{t+1}(I + x_t - d_t)\}, \\ &\qquad\qquad t = 1, 2, \ldots, N-1 \\ g'_N(I) &= \min_{\{x_N \geq 0,\ I + x_N = d_N\}} \{(h'_{N-1} - c'_N)I + \delta(x_N)s'_N + c'_N d_N\}. \end{aligned}$$

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

$$\begin{aligned} G_t(I) &= c'_{t-1}I + g'_t(I) - [c'_t d_t + \sum_{n=t+1}^{N} (c'_n d_n \cdot \Pi_{j=t}^{n-1}\beta_j)], \quad t = 1, 2, \ldots, N-1 \\ G_N(I) &= c'_{N-1}I + g'_N(I) - c'_N d_N, \end{aligned}$$

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

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

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

where

$$\begin{aligned} S_t &= s'_t, \quad 1 \leq t \leq N \\ H_1 &= c_1 - c_2 + \alpha_1(c_2 + r_2) + \beta_1 h_1 \qquad (3.9) \\ H_t &= [c_t - c_{t+1} + \alpha_t(c_{t+1} + r_{t+1}) + \beta_t h_t] \cdot \Pi_{j=1}^{t-1}\beta_j, \quad 1 < t < N. \qquad (3.10) \end{aligned}$$

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

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

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

The backward formulation works as follows,

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

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

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

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

1. *If the optimal solution for $F(1,t)$ in Eq. (3.11) is $t^* \leq t$, then to solve $F(1,\tau)$ with $\tau > t$, one only needs to consider $F(t^*,\tau)$. In other words, if it is optimal to incur a set-up cost at period t^* when periods 1 through t are considered alone, then it is optimal to incur a set-up cost at period t^* in any τ-period model.*

2. *The optimal solution for periods 1 to t^* does not change even if we can update α_j for $j \geq t^*$ along with time.*

If $H_t < 0$, Theorem 2 may not hold, see Eppen, et al. (1969) for more discussion. Due to the high destruction cost and failure risk, H_t will be positive in clinical trial supply chains. Hence, we assume $H_t \geq 0$ for all $1 \leq t < N$ for the rest of the paper.

To avoid the costs of failure (i.e. wasted drug and destruction costs), more frequent production of smaller batches can be planned. At some point, failure becomes so likely, that production in every period becomes a prudent decision. By the following Theorem, we can mathematically characterize a threshold on our failure porbability, α_k, so that above which, it is optimal to produce in each period.

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

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

$$F(j, j+1) = \min\{S_j + F(j+1, j+1),\ S_j + H_j \cdot d_{j+1}\} \tag{3.13}$$
$$= \min\{S_j + S_{j+1},\ S_j + H_j \cdot d_{j+1}\}. \tag{3.14}$$

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

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

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

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

where x^*_t and I^*_t are the optimal production and inventory decisions.

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

Proof. See Appendix. □

C^* is generally not a monotonic function of α_t. Consider the special case of $\alpha_t \to 1$ for all t. C^* effectively reduces to a single-period cost function, which is clearly less than the multi-period cost function as $\alpha_t \to 0$ for all t. Proposition 1 gives a upper bound on the diminishing rate for C^* as α_t increases. On the other hand, if the optimal sequence of production times is to produce at period $N-1$ to cover demand in both $N-1$ and N, then $C^*(\alpha_{N-1} + \Delta) > C^*(\alpha_{N-1})$ can hold for sufficiently small Δ if $r_N > h_{N-1}$.

4. Extensions

In this section, we consider two extensions of the model in §3 to incorporate real-world situations: general concave cost functions and production/storage constraints.

4.1 Concave Cost Functions

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

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

$$\begin{aligned} f_t(0,I) &= r_t(I), \quad 1 \le t \le N & (4.1)\\ f_t(1,I) &= \min_{\{x_t \ge 0,\ I+x_t \ge d_t\}} \{h_{t-1}(I) + c_t(x_t) + \alpha_t f_{t+1}(0, I+x_t-d_t) + \\ &\qquad \beta_t f_{t+1}(1, I+x_t-d_t)\}, \quad t < N & (4.2)\\ f_N(1,I) &= \min_{\{x_N \ge 0,\ I+x_N = d_N\}} \{h_{N-1}(I) + c_N(x_N)\}. & (4.3) \end{aligned}$$

Similar to §3, define $g_t(I) = \frac{\alpha_{t-1}}{\beta_{t-1}} r_t(I) + f_t(1,I)$ for $t = 1, 2, \ldots, N$, where $\alpha_0 = 0$. Then

$$\begin{aligned} g_t(I) &= \min_{\{x_t \ge 0,\ I+x_t \ge d_t\}} \{\frac{\alpha_{t-1}}{\beta_{t-1}} r_t(I) + h_{t-1}(I) + c_t(x_t) + \beta_t g_{t+1}(I+x_t-d_t)\},\ t < N & (4.4)\\ g_N(I) &= \min_{\{x_N \ge 0,\ I+x_N = d_N\}} \{\frac{\alpha_{N-1}}{\beta_{N-1}} r_N(I) + h_{N-1}(I) + c_N(x_N)\}. & (4.5) \end{aligned}$$

Discounting cost functions in each period by β_t, we define,

$$\begin{aligned} c'_1(x) &= c_1(x)\\ c'_t(x) &= c_t(x) \cdot \Pi_{j=1}^{t-1} \beta_j, \quad t > 1\\ h'_1(I) &= \alpha_1 r_2(I) + \beta_1 h_1(I)\\ h'_t(I) &= [\alpha_t r_{t+1}(I) + \beta_t h_t(I)] \cdot \Pi_{j=1}^{t-1} \beta_j, \quad 1 < t < N. \end{aligned}$$

Finally,

$$\begin{aligned} g'_t(I) &= \min_{\{x_t \ge 0,\ I+x_t \ge d_t\}} \{h'_{t-1}(I) + c'_t(x_t) + g'_{t+1}(I+x_t-d_t)\}, \quad t < N & (4.6)\\ g'_N(I) &= \min_{\{x_N \ge 0,\ I+x_N = d_N\}} \{h'_{N-1}(I) + c'_N(x_N)\}. & (4.7) \end{aligned}$$

Note that the *effective* cost functions, $h'_{t-1}(I)$ and $c'_t(x)$, are still concave and increasing. By Eqs. (4.6)-(4.7), the stochastic failure-risk model is equivalent to the deterministic W-W model with general concave and increasing cost functions. By Zipkin (2000, Sections 4.3.3 and 4.4.6), the "Zero Inventory Property" (Theorem 1) still holds here. One can derive the forward and backward formulations in a similar way as Eqs. (3.11)-(3.12), for brevity, we omit the details. As Veinott (1963) and Aggarwal and Park (1993) point out, the W-W model with general concave cost functions can be solved using the forward formulation with complexity $O(N^2)$. However, Theorem 2 does not hold because of the general form of the concave production cost functions.

4.2 Additional Constraints

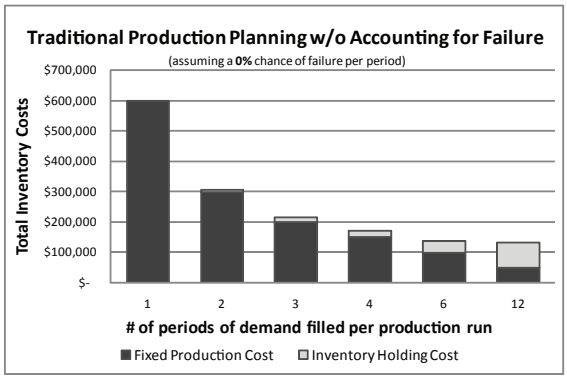
We discuss three types of constraints: the production capacity constraint, the inventory shelf-life constraint and the storage capacity constraint. The inventory shelf-life constraint specifies the number of periods that a unit can be carried in inventory, which limits the number of future periods that can be covered by a production batch. Thus, it is effectively a production capacity constraint. In fact, with any subset of these constraints, one can use the same technique as in §3 to reduce the stochastic failure-risk model to an equivalent deterministic W-W model with adjusted cost parameters and the same set of constraints. For the W-W model with production capacity, inventory shelf-life, and/or storage capacity constraints, one can find the solution using well established algorithms, see, e.g., Shaw and Wagelmans (1998).

5. An Illustrative Example

In this section, we demonstrate that accounting for demand failure when planning a production schedule can lead to substantial cost savings over using the Wagner-Whitin model ignoring the failure risk. To develop insight, we consider a special case of Phase II clinical trials with stationary data where $c_t = \$75$, $r_t = \$25$, $h_t = \$5$, $d_t = 250$ and $s_t = \$50,000$ $(t = 1, 2, ..., 12)$. Note that both h_t and d_t are defined per period where a period equals two months here. We consider a 12-period (two years) planning horizon and a 7% probability of failure in each period (i.e. approximately a 42% [2] chance of success for phase II trials). DiMasi and Grabowski (2007) provides averages for Phase II trial length of 26.0 months. DiMasi (2001) and DiMasi et al. (2010) estimate Phase II success rate as 41.2% and 45% respectively. Our illustrative example is consistent with these aforementioned averages. Although not required by our model, the risk of failure is assumed identical in every period so that the cumulative success rate is in line with industry averages. Further justification of the parameters chosen here is provided in the next section.

For stationary production variable costs, it makes sense to utilize Eqs. (3.7)-(3.8), which ignore the production variable costs except as included in the failure costs. We first consider the classic W-W model ignoring the risk of failure. Figure 1a shows the production-inventory costs, excluding the risk of failure, as a function of production schedule. From Figure 1a, we see that satisfying 12 periods of demand with just one production run, is both the optimal plan as calculated by the W-W model and also represents a typical heuristic of pharmaceutical manufacturers (Shah 2004).

[2] $(1 - 0.07)^{12} \approx 0.42$



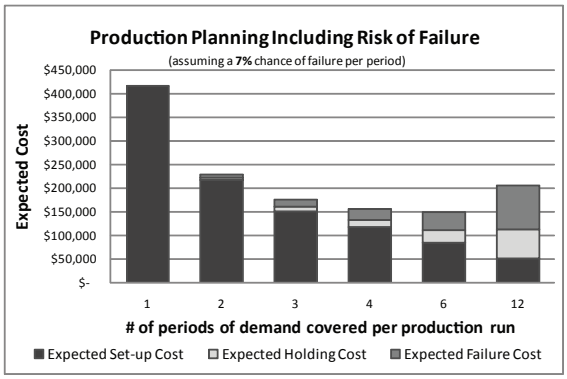


Figure 1: Production Costs Excluding & Including the Risk of Failure

Figure 1b shows the same example, but now the costs account for the 7% probability of failure in each period. The optimal plan is now to produce batches which satisfy six periods of demand. This plan, which minimizes total expected costs, calls for two batches during the planning horizon as compared with the one batch that is prescribed by the W-W model. From Figure 1b, one easily sees that using the plan of satisfying all 12 periods of demand with one production run as prescribed by the W-W model would lead to very high failure costs. This is a direct consequence from carrying large amounts of inventory that will likely be wasted due to demand failure. In fact, the optimal schedule generated by the failure-risk model is expected to be 28% less costly than the optimal plan of the W-W model. Even with the high fixed costs of this example, failure costs lead to reducing the optimal lot size and scheduling more frequent production runs.

6. Numerical Study of Potential Savings

In this section, we conduct a comprehensive numerical study to gauge the potential savings of incorporating failure risk into production planning by solving various environments that clinical trial manufacturers may face. From our discussions with industry professionals, most clinical supply managers plan for success despite knowing that failure is both likely and costly. Thus, given our assumptions, the best plan to use as as nchmark, would be the optimal plan as given by the Wagner-Whitin model. In addition, Friend, et al. (2001) deemed implementing the Wagner-Whitin model in a similar environment as "most beneficial" given that it is "consistently superior" over widely used industry heuristics and algorithms; this conclusion comes from a study using real data from the aircraft spare parts industry which is similarly characterized by high fixed costs and low production volumes. While other sophisticated models exist in the literature that account for failure uncertainty in clinical trials, specifically those using stochastic programming techniques, direct comparison with

those models is not possible as their focus is usually on scheduling the resource consumption of a portfolio of drug candidates ready for commencing clinical trials. In contrast, our model looks at just one drug candidate and optimally balances the forces advocating for larger batches (such as high fixed costs) against those which advocate for smaller batches (such as high failure probabilities, destruction costs, holding costs, and variable costs). This will be especially useful for the small pharmaceutical firms[3] that outsource their clinical supply manufacturing, are not capacity constrained, and are very focused on cost efficiencies. Our objective is to quantify the savings and identify conditions under which the savings of incorporating failure into production planning are likely to be substantial. We vary our key parameters, namely, production costs, holding costs, and failure probability based on our observations of the industry. Please note that we do not explicitly model destruction costs because including them in production costs is mathematically equivalent when production variable costs are constant (see Eqs. 3.7-3.10).

Productions costs in our model are the portion of total manufacturing costs that vary directly with the number of units produced. In our estimates of production costs, we estimate the variable cost of the active ingredient, the cost of packaging, the cost of distribution and tracking, and the cost of destruction as a percentage of the fixed set-up cost of production. In conversations with industry professionals, one clinical supply manager commented on a trial where the syringe used to administer the investigational drug was $400, a cost certainly included as a variable production cost in our model. As all of the aforementioned production costs can vary widely based upon the costs of active ingredient production, packaging requirements, formulation requirements, and dosing schedules; we employ a wide range of production costs in our numerical study. To get the magnitude of our estimates, we look to vaccine production, which, like clinical trial production, is less standardized and less predictable than typical commercial drug production (Institute of Medicine 2004). As a rough proxy for clinical trial production, we expect the total variable cost to fixed set-up cost ratio to be 3:5.[4] Assuming that we obtain this ratio when we are producing 10^3 treatments per batch, we get variable production cost per treatment of approximately 0.06% of the setup cost per batch. Using this estimate as a reference, we employ a wide range for simulating variable production costs per treatment of between 0.01% and 1.25% of the fixed set-up costs.

Holding costs are the per unit costs associated with storing inventory and having money tied up an inventory. Holding costs in the clinical trial supply chain are most likely higher

[3]For an example of this type of firm's manufacturing philosophy, see page 12 of Ariad Pharmaceuticals 2009 Annual 10-K Filing.

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

Parameter	Average Value	Lower Bound	Upper Bound
Production Cost Per Treatment (as % of Set-Up Cost)	0.63%	0.01%	1.25%
Annual Holding Cost (as % of Prod. Cost)	45.5%	11%	80%
Phase II Failure Probability	59%	45%	73%
Phase III Failure Probability	21.5%	13%	30%
Duration of Phase II Trial	2 Years	–	–
Duration of Phase III Trial	3 Years	–	–
Annual Phase II Demand	150 Treatments	–	–
Annual Phase III Demand	1,200 Treatments	–	–
Planning Period	Two Months	–	–

Table 1: Parameters for Phase II and Phase III Simulations

than incurred in typical pharmaceutical supply chains due to costly tracking and auditing of inventory levels. In addition, bio-tech molecules often require controlled storage environments which may add to the cost.[5] For a lower bound on annual holding costs, we take the conservative estimate of DiMasi et al. (2003) of 11% as the pharmaceutical industry's real cost-of-capital for money tied up in inventory. We choose an upper bound of 80% which may better reflect the potentially high costs of storing and tracking each treatment and the corresponding placebo.

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

Data Generation: We randomly generated 1,000 scenarios of potential cost structures for a Phase II trial and another 1,000 scenarios for a Phase III Trial. Within each phase, each scenario is identical with respect to planning horizon length, period length, demand

[5] An interesting note is how a whole industry has formed around clinical trial storage. In one corporate press release (August 4, 2010), titled "Fueled by Customer Demand, World Courier Expands Its Global Clinical Trial Storage and Distribution Network", it demonstrates how a whole industry is forming around the storage of clinical trial supply. (http://www.businesswire.com/news/home/20100804005073/en/Fueled-Customer-Demand-World-Courier-Expands-Global)

over the planning horizon, and fixed costs of production (which are normalized to \$1). For simplicity, demand is assumed equal in each period. Values for variable production costs, planning horizon failure probabilities, and holding costs are then chosen as realizations of uniformly distributed random variables bounded by the parameters in Table 3. Per period failure probabilities are then calculated so that the randomly chosen planning horizon failure probability is consistent with the cumulative effect of identical period failure probabilities. For example, to split an overall probability of failure equal to 45% over 12 periods, the per period failure probability would be $1 - (1 - 0.45)^{1/12} = 4.86\%$.

For each scenario, two optimal production plans were generated: 1) a Wagner-Whitin production plan and 2) a failure-risk production plan. Since our goal is to understand under what circumstances the failure-risk model is likely to outperform the Wagner-Whitin model, we calculate, for each scenario, the percentage cost reduction by the failure-risk model relative to the Wagner-Whitin model. We then plot the percentage savings against various system parameters to gain insight.

The Phase II Simulation: The results of our Phase II simulations are shown in the four graphs in Figure 2 with each diamond on the graphs representing one of the 1,000 scenarios. In the 419 scenarios where the two models yield different results, the failure-risk model led to savings with an average of 11%.

Since the traditional Wagner-Whitin model, like current industry practice, does not incorporate demand failure, it is quite intuitive that an increase in the likelihood of failure would reduce the W-W model's effectiveness. As shown in Figure 2a, the failure-risk model's maximum potential savings over the W-W model does increase in failure probability. Nonetheless, the presence of many scenarios with 0% savings at all failure probabilities demonstrates that even high failure probabilities do not always lead to different solutions by the two planning models. The solutions also depend on other system parameters, such as production costs and inventory holding cost.

Figure 2b shows the impact of the variable production cost to fixed setup cost ratio in our Phase II simulation. A threshold appears where the variable production costs, which in our model include destruction costs, need to be sufficiently high for savings to be realized. In particular, to achieve savings in our test scenarios, variable production costs must exceed 0.39% of the fixed set-up costs. This is true because when variable production costs are low, the costs of failure in the form of wasted production are also low. As the variable to fixed cost ratio increases beyond the threshold, the increased cost of failure leads to an increase in expected magnitude of savings, because the production frequency under the failure-risk model increases faster than that of the W-W model.

The impact of holding costs are studied in Figure 2c. It is interesting to note that in

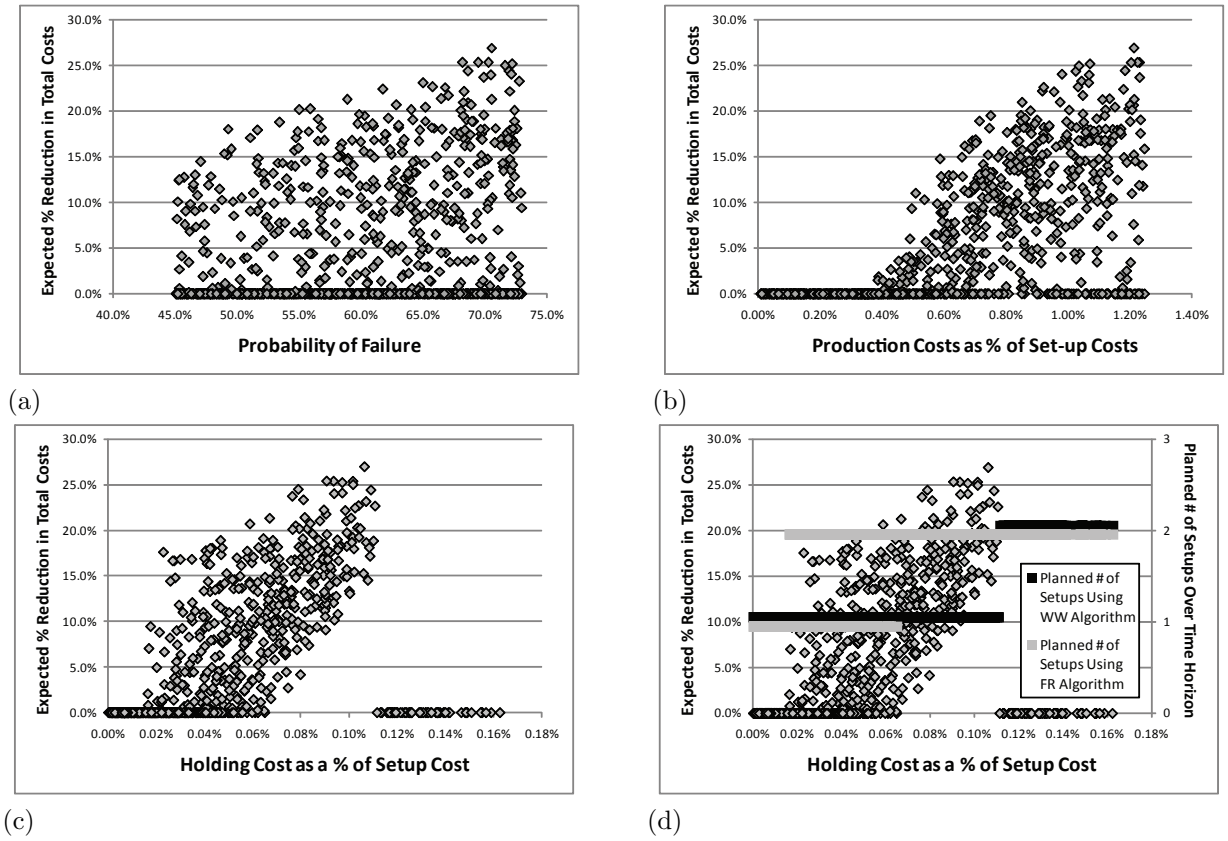


Figure 2: Expected Reduction in Costs Using F-R Algorithm in Phase II Simulations

our test scenarios, to achieve any savings, the ratio of holding costs to set-up costs must be within a range of 0.017% to 0.111%. To see what is behind this observation, we present the production frequency information in Figure 2d. On the secondary vertical axis, we present the planned number of setups that each model recommends for each of the 1,000 scenarios. Combining the information of production frequency and the percentage savings, we make the following observations about this Phase II simulation:

- Both models call for either one or two setups during the 2-year planning horizon.
- The failure-risk model is most beneficial when holding costs are not too low. When holding costs are too low, both the failure-risk and W-W models will both call for producing just one big lot.
- The failure-risk model is most beneficial when holding costs are not too high. When holding costs are too high, both models have the same plan of producing two times over the planning horizon.
- The range of holding to fixed cost ratios observed above (where savings are possible) approximately corresponds to the range within which the failure-risk model plans for

two setups while the Wagner-Whitin model calls for one setup. In this range of scenarios, the potential benefits of the failure-risk model increase with increasing holding costs. This is a simple reflection of the larger average inventory level, some of which is more likely to be wasted, under the plan of the W-W model.

Phase III Simulation: Phase III clinical trials differ from Phase II trials due to their longer duration, higher demand for treatments, but lower probability of failure. Our Phase III simulation adjusts these parameters accordingly. One result of these changes is that out of the 1,000 scenarios investigated, savings were achieved in 55.4% of the trials. This is more frequent than the 41.9% frequency in which savings were achieved in the Phase II simulation. However, Phase III savings, when they occurred, averaged about 2.85% which is significantly less than the average savings of 11% observed in Phase II. The first difference is due to the longer planning horizon, because with more periods, it is more likely that the two models yield different solutions. The second difference is largely due to the lower failure rate.

Similar to our Phase II investigation, we look to a graphical representation of the percentage savings against certain key parameters of the model. These graphs are shown in Figure 3. As seen in Phase II, the failure-risk model in Phase III has larger potential benefits as the probability of failure increases (see Figure 3a). However, in contrast to our Phase II results, Phase III has some noteworthy differences in regards to the potential for savings against other parameters.

First, as shown in Figure 3b, the larger demand for treatments and longer planning horizon substantially reduce the minimum production cost threshold. We see that savings are achievable at almost any level of production cost. We also see a saw-blade pattern in the diagram of these costs which is best explained by the data shown in Table 2. We see that the maximum potential savings are achieved when the Failure Risk (F-R) algorithm plans two setups while the W-W model plans only one setup. While production costs are not directly responsible for the drop in savings that we see when production costs reach about 0.47% of setup costs, they directly affect holding costs, which have been defined as a percentage of production costs. At production costs of about 0.47%, holding costs are driven sufficiently high such that the W-W model will perform a minimum of at least two setups over the planning horizon. As seen in Table 2, once the W-W model plans for more than one setup, the maximum potential savings drops significantly.

In the Phase II simulation, we saw that holding costs were required to be within a certain range for savings to occur. From our Phase III chart of holding costs (Figure 3c), we see several ranges of holding costs that have different effects on the savings achieved. In Figure 3d, we overlay the frequency of production that the two models call for on the second vertical axis. As we increase the holding costs, we notice a transition in the failure-

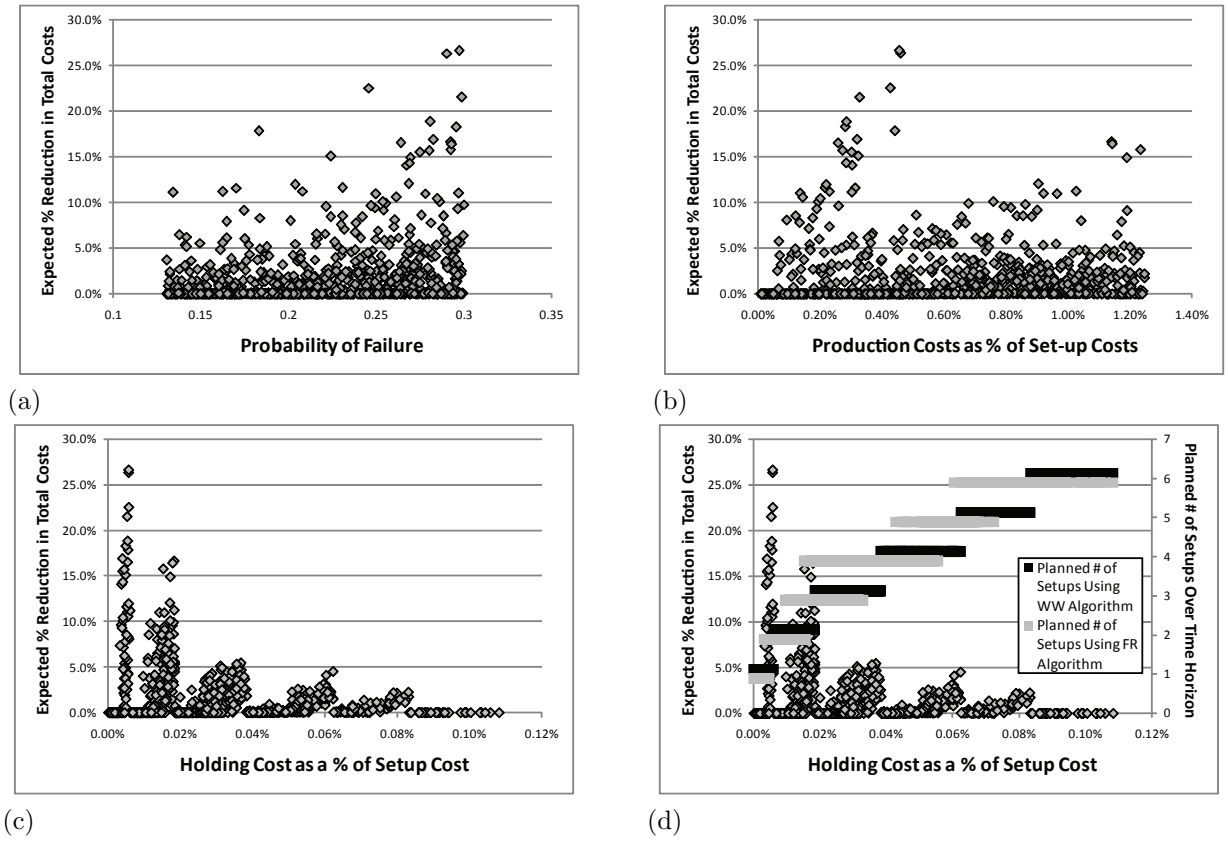


Figure 3: Expected Reduction in Costs Using F-R Algorithm in Phase III Simulations

risk production plans from one of less frequent setups to one of more frequent setups. This transition is then followed by a similar transition of the W-W plan to one of more frequent setups. As shown in both Table 2 and Figure 3d, savings are achieved every time the failure-risk model makes the jump to a production plan that has more frequent setups than the W-W model. Savings then return to zero once the W-W model transitions to the same schedule that the failure-risk model calls for.

7. Industry Notes

Mapping the results of our analysis to industry, we expect the failure-risk model to have the most significant impact for drugs that have a high probability of failure, sufficiently high production costs and relatively low inventory holding costs. Since it is hard to characterize holding and production costs for a certain clinical trial environment, we comment only on the probability of failure that is seen during clinical trials. In pharmaceutical and bio-tech industries, we see below-average success probabilities for drugs in the following therapeutic classes: antineoplastic, cardiovascular, central nervous system, immunologic, and respiratory

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

Table 2: Maximum Observed % Savings Versus Number of Setups During Phase III Simulation

Therapeutic Class	Phase I-II (%)	Phase II-III (%)	hase I-II (%)
Antineoplastic/Immunologic	71.8	49.0	55.3
Cardiovascular	62.9	32.4	64.3
Central Nervous System	59.6	33.0	46.4
GI/Metabolism	67.5	34.9	50.0
Musculoskeletal	72.4	35.2	80.0
Respiratory	72.5	20.0	85.7
Systemic anti-infective	58.2	52.2	78.6

Table 3: Phase transition probabilities by therapeutic class

medicines (DiMasi 2001, DiMasi et al. 2004). In a more recent study (DiMasi et al. 2010), phase transition rates (probability of a drug that enters one phase of clinical testing starting the subsequent phase of testing) are estimated as follows for various therapeutic classes: While clinical trials are often fraught with uncertainty, this model is applicable when the main uncertainty surrounding supply planning decisions stems from the potential for failure of the drug. This failure is most likely a result of information gathered about the safety or effectiveness of a drug as a particular phase of the trial is underway. Additionally, this failure may be a result of drug stability or even unforeseen financial considerations. No matter the cause of failure, this model is a useful tool for balancing production efficiencies against waste resulting from trial failure when uncertainty in actual drug requirements is minimal and drug portfolio considerations are less of a concern. Additionally, as a trial progresses and estimates of failure likelihood are updated, the planning horizon theorem (Theorem 2) demonstrates that updates to a scheduled production run's planned quantities may change, it will still be

optimal to produce in that period albeit possibly in a different amount. Thus, some learning about failure as the trial continues can be accommodated by the model and has been used by practitioners in rolling horizon models (Wagner, 2004).

As noted by Shah(2004), the need for sophisticated models in the pharmaceutical supply chain is great. Thus, the simple structure of the F-R model may not always accommodate the realities of a complex clinical trial supply chain. However, as noted by Federgruen and Tzur (1991), the W-W model arises as a subproblem in many hierarchical planning solutions including multi-item, multi-stage production systems (when "formulated as mixed integer programs and solved via Lagrangean relaxation") and multi-item capacitated lot sizing problems (e.g van Norden and van de Velde, 2005). Thus, by connecting the F-R model to the W-W model, failure uncertainty can be incorporated to an even more complex set of problems than just the examples presented here.

8. Conclusion

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

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

The huge potential of integrating clinical trial activities and the drug supply chain has recently been recognized both in academia and in industry. While there is ample work to be done, we suggest the following specific research directions: (1) *An Empirical Study*: This paper provides some empirical evidence of the magnitude of spending needed to support clinical trial supply chains. A more comprehensive empirical study is needed to verify the financial significance and determine the impacting factors. (2) *Multi-Product/Multi-echelon Optimization*: Drug supply chains often consist of multiple manufacturing steps that are done in geographically dispersed facilities, e.g., the active pharmaceutical ingredients (API) manufacturing, formulation and packaging (Bernstein and Hamrell 2000). Furthermore, companies may have multiple investigational drugs in clinical trials simultaneously. Thus, it is important to generalize the model to coordinate multiple drugs in multi-echelon clinical trial supply chains. (3) *Outsourcing Contracts*: While many large pharmaceutical companies produce investigational drugs in-house, most smaller companies outsource production to 3rd party manufacturers. Given the potential failure risk and the large costs of production, constructing efficient and fair outsourcing contracts is important to both clinical trial suppliers and pharmaceutical companies.

Appendix

Proof of Proposition 1. The proof of the first statement follows that of Zipkin (2000) problem 4.6. Briefly, for any fixed sequence of production times, we note that h'_j, s'_j and c'_j are either independent of α_t or linear functions of α_t. Therefore, given the sequence of production times, C^* is linear in every α_t. C^* is concave in α_t because C^* is the minimum cost over all possible sequence of production times.

To prove the second statement, we consider α_1 as a special case. For α_1^2, $s'_1 = s_1$ and $c'_1 = c_1$,

$$\begin{aligned} s'_t &= s_t \cdot \Pi_{j=1}^{t-1}\beta_j(1+\alpha_1),\ t>1 \\ c'_t &= c_t \cdot \Pi_{j=1}^{t-1}\beta_j(1+\alpha_1),\ t>1 \\ h'_1 &= \alpha_1^2(r_2-h_1)+h_1 \ \le\ [\alpha_1(r_2-h_1)+h_1](1+\alpha_1) \\ h'_t &= (\alpha_t r_{t+1}+\beta_t h_t)\cdot \Pi_{j=1}^{t-1}\beta_j(1+\alpha_1),\ t>1. \end{aligned}$$

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

Acknowledgments. This research is supported by the grant CMMI-0747779 from the National Science Foundation. We thank Dr. Indranil Nandi from Sandoz Inc., A Novartis Company, and Lee Resnick from Deloitte Consulting LLP for informative discussions on drug development processes and clinical trial practices. We also thank seminar participants for their feedback at the "Clinical Trials Supply Management for Pharmaceuticals" conference held in Philadelphia, June 23-25, 2008.

References

[1] Allos Therapeutics Inc. (1999-2007) Form 10-K for the Fiscal Years 1995-2006. Retrieved November 8, 2007, from http://ir.allos.com/.

[2] Aggarwal, A., J. K. Park (1993) "Improved algorithms for economic lot size problems", *Operations Research*, **41**: 549-571.

[3] Ariad Pharmaceuticals Inc. (2010) Form 10-K for the Fiscal Years 2009. Retrieved August 27, 2010, from http://www.ariad.com/.

[4] Berenson, A. (Dec. 4, 2006) "End of Drug Trial Is a Big Loss for Pfizer ", *New York Times*.

[5] Bernstein, D. F., M. R. Hamrell (2000) "Integrating drug supply issues with strategic preclinical and clinical development", *Drug Information Journal*, **34**: 909-917.

[6] Blau, G. E., J. F. Pekny, V. A. Varma, P. R. Bunch (2004) "Managing a portfolio of interdependent new product candidates in the pharmaceutical industry", *Journal of Product Innovation Management*, **21**(4): 227-245.

[7] Brown, G. W., J. Y. Lu, R. J. Wolfson (1964) "Dynamic modeling of inventories subject to obsolescence", *Management Science*, **11**: 51-63.

[8] Chu, L.Y., V.N. Hsu, Z.J. Shen (2005) "An economic lot-sizing problem with perishable inventory and economies of scale costs: Approximation solutions and worst case analysis", *Naval Research Logistics*, **52**: 536-548.

[9] Clemento, A. (1999) "New and integrated approaches to successful accelerated drug development", *Drug Information Journal*, **33**: 699-710.

[10] Colvin, M., C. T. Maravelias (2008) "A stochastic programming approach for clinical trial planning in new drug development", *Computers and Chemical Engineering*, **32**(11) 2626-2642.

[11] Colvin, M., C. T. Maravelias (2009) "Scheduling of testing tasks and resource planning in new product development using stochastic programming", *Computers and Chemical Engineering*, **33**(5) 964-976.

[12] Colvin, M., C. T. Maravelias (2010) "Modeling methods and a branch and cut algorithm for pharmaceutical clinical trial planning using stochastic programming", *European Journal of Operational Research*, **203**: 205-215.

[13] Cotherix Inc. (2006). Form 10-K, filed March 14, 2006, for the Fiscal Year Ended December 31, 2005. Retrieved November 8, 2007, from Lexis-Nexis Academic database.

[14] Cutting Edge Information (2004) ""Accelerating Clinical Trials: Budgets, Patient Recruitment and Productivity", *http://www.cuttingedgeinfo.com*.

[15] David, I., E. Greenshtrein, A. Mehrez (1997) "A dynamic-programming approach to continuous-review obsolescence inventory problems", *Naval Research Logistics*, **44**: 757-774.

[16] DiMasi, J. A. (2001) "Risks in new drug development: Approval success rates for investigational drugs", *Clinical Pharmacology & Therapeutics*, **69**(5): 297-307.

[17] DiMasi, J. A., R. W. Hansen, H. G. Grabowski (2003) "The price of innovation: new estimates of drug development costs", *Journal of Health Economics*, **22**:(4-5) 469-479.

[18] DiMasi, J. A., H. G. Grabowski, J. Vernon (2004) "R&D costs and returns by therapeutic category", *Drug Information Journal*, **38**: 211-223.

[19] DiMasi, J. A., H. G. Grabowski (2007) "The cost of biopharmaceutical R&D: is biotech different?", *Managerial and Decision Economics*, **28**: 151185.

[20] DiMasi, J. A., L. Feldman, A. Seckler, A. Wilson (2010) "Trends in risks associated with new drug development: Success rates for investigational drugs", *Clinical Pharmacology and Therapeutics*, **87**(3): 272-277.

[21] English, M., A. Ma (2007) "From cradle to grave", *Applied Clinical Trials*, **16** (8): 50-54.

[22] Eppen, G. D., F. J. Gould, B. P. Pashigian (1969) "Extensions of planning horizon theorem in the dynamic lot size model", *Management Science*, **15**: 268-277.

[23] Federgruen, A., M. Tzur (1991) "A simple forward algorithm to solve general dynamic lot sizing models with n periods in $O(nlogn)$ or $O(n)$ time", *Management Science*, **37**: 909-925.

[24] Florian, M., J. K. Lenstra, A. H. Rinnooy Kan (1980) "Deterministic production planning: algorithms and complexity", *Management Science*, **26**: 669-679.

[25] Friend, H. F., A. L. Swift and A. A. Ghobbar (2001) "A Predictive Cost Model in Lot- Sizing Methodology, with Specific Reference to Aircraft Parts Inventory: An Appraisal", *Production and Inventory Management Journal*, **42**(3/4) 24-33.

[26] Gatica, G., L. G. Papageorgiou, N. Shah (2003) "Capacity planning under uncertainty for the pharmaceutical industry", *Trans IChemE*, **81** (Part A): 665-679.

[27] George, L. (2005) "Investigational medicinal products Optimizing the supply chain", *Applied Clinical Trials*, **14**(4): 42-48.

[28] Hopp, W. (2004) "Ten Most Influential Titles of Management Science's First Fifty Years", *Management Science*, **50** (12) 1763.

[29] Hult, G., K. S. Swan (2003) "A research agenda for the nexus of product development and supply chain management processes", *Journal of Product Innovation Management*, **20**(5): 333-336.

[30] Iglehart, D., S. Karlin (1962) "Optimal policy for dynamic inventory process with nonstationary stochastic demands", In *Studies in Applied Probability and Management Science*. K. Arrow, S. Karlin, H. Scarf (eds.). Chapter 8, Stanford University Press, Stanford, CA.

[31] Institute of Medicine (2004) *Financing vaccines in the 21st century: Assuring access and availability.* National Academies Press.

[32] Jain, K., E.A. Silver (1994) "Lot sizing for a product subject to obsolescence or perishability", *European Journal of Operational Research*, **75**: 287-295.

[33] Johnson, G., H. Thompson (1975) "Optimality of myopic inventory policies for certain dependent demand processes", *Management Science*, **21**: 1303-1307.

[34] Katok E., A. Lathrop, W. Tarantino, S. H. Xu (2001) "Jeppesen uses a dynamic-programming-based DSS to manage inventory", *Interfaces*, **31**: 54-65.

[35] Katok, E., S. H. Xu (2001) "Managing inventory in the information age: a system with product lifecycle uncertainty and stochastic time-varying demand", Working Paper. Smeal College of Business Administration, The Pennsylvania State University. University Park, PA.

[36] Kumar, RS (2008) "Managing the Clinical Trials Supply Chain", Conference Presentation. Clinical Trials Supply Management for Pharmaceuticals conference. Philadelphia, PA. June 25, 2008.

[37] Krishnan, V., K. T. Ulrich (2001) "Product development decisions: A review of the literature", *Management Science*, **47**(1): 1-21.

[38] Levis, A., L. Papageorgiou (2004) "A hierarchical solution approach for multi-site capacity planning under uncertainty in the pharmaceutical industry", *Computers and Chemical Engineering*, **28**(5) 707-725.

[39] Mercer Management Consulting (2002) *Lessons learned: A new procurement strategy for vaccines: Final Report to the GAVI Board*

[40] Pharmaceutical Research and Manufacturers of America (2009) "Pharmaceutical Industry Profile 2009", PhRMA, Washington, D.C.

[41] Pisano, G. P. (1997) *Development factory: unlocking the potential of process innovation.* Harvard Business Press, Boston, MA.

[42] Pierskalla, W. P. (1969) "An inventory problem with obsolescence", *Naval Research Logistics Quarterly*, **16**: 217-228.

[43] Porteus, E. L. (2002) *Foundations of Stochastic Inventory Theory*. Stanford University Press, Stanford, CA.

[44] Sethi, P. S., F. Cheng (1997) "Optimality of (s, S) policies in inventory models with Markovian demand", *Operations Research*, **45**: 931-939.

[45] Shah, N. (2004) "Pharmaceutical supply chains: key issues and strategies for optimisation", *Computers and Chemical Engineering*, **28**: 929-941.

[46] Shaw, D. X., A. P. M. Wagelmans (1998) "An algorithm for single-item capacitated economic lot sizing with piecewise linear production costs and general holding costs", *Management Science*, **44**: 831-838.

[47] Song, J.-S., P. Zipkin (1993) "Inventory control in a fluctuating demand environment", *Operations Research*, **41**: 351-370.

[48] Song, J. S., P. Zipkin (1996) "Managing inventory with the prospect of obsolescence", *Operations Research*, **44**: 215-222.

[49] Subramanian, D. J.F. Pekny, G.V. Reklaitis, G.E. Blau (2003) "Simulation-optimization framework for stochastic optimization of R&D pipeline management", *AIChE Journal*, **49**(1) 96-112.

[50] Thomson CenterWatch (2007) "State of the Clinical Trials Industry 2007", *http://www.centerwatch.com*.

[51] van Norden, L. and S. van de Velde (2005) "Multi-product lot-sizing with a transportation capacity reservation contract", *European Journal of Operational Research*, **165**: 127-138.

[52] Veinott, JR., A. F. (1963) "Unpublished Class Notes". Program in Operations Research, Stanford University, Stanford, CA.

[53] Verderame, Peter M., J.A. Elia, J. Li, C.A. Floudas (2010) "Planning and Scheduling under Uncertainty: A Review Across Multiple Sectors", *Industrial & Engineering Chemistry Research*, **49** (9) 3993-4017.

[54] Wagelmans, A., S. Van Hoesel, A. Kolen (1992) "Economic lot sizing: an $O(nlogn)$ algorithm that runs in linear time in the Wagner-Whitin case", *Operations Research*, **40**: S145-S156.

[55] Wagner, H. M., T. M. Whitin (1958) "Dynamic version of the economic lot size model", *Management Science*, **5**: 89-96.

[56] Zangwell, W. I. (1969) "A backlogging model and a multiechelon model of a dynamic economic lot size production system – a network approach", *Management Science*, **15**: 506-527.

[57] Zipkin, P. (2000) *Foundations of Inventory Management*. McGraw Hill, Boston, MA.