

A Newsvendor Approach to Design of Surgical Preference Cards

Berk Görgülü,^a Vahid Sarhangian^{a,*}

^aDepartment of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada

*Corresponding author

Contact: bgorgulu@mie.utoronto.ca (BG); sarhangian@mie.utoronto.ca, <https://orcid.org/0000-0003-2633-6496> (VS)

Received: August 13, 2020

Revised: March 6, 2021; August 27, 2021

Accepted: November 29, 2021

Published Online in Articles in Advance:

February 14, 2022

<https://doi.org/10.1287/serv.2021.0298>

Copyright: © 2022 INFORMS

Abstract. Surgical procedures require a large number of consumable supplies that need to be kept in hospital inventory and transported to the operating rooms (OR) before the surgery. A surgical preference card (SPC) provides a list of items to be prepared for each surgery. For each item, a SPC also specifies how many should be taken to the OR (*fill* quantity). As the usage of most consumables in the OR is subject to uncertainty, the cards also specify how many of the filled items should be opened at the beginning of the surgery (*open* quantity). The fill and open quantities control the flow of consumables between the hospital inventory and the ORs and directly affect the wastage in ORs. In this work, we formulate the problem of determining the fill and open quantities on the preference cards as a stochastic optimization problem, where the objective is to minimize a weighted sum of the expected wastage and operational costs. We show that, as in the newsvendor problem, the optimal solution for the fill and open quantities takes the form of critical quantiles of the item usage distribution in the OR. The solution form together with historical usage data provide a data-driven approach to design of SPCs, as well as insights on the value of including an open decision. We demonstrate our approach using extensive numerical experiments and real usage data from a Canadian hospital. The results suggest a potential for significant reduction of wastage and operational costs in the ORs.

Funding: The research of V. Sarhangian was partially funded by The Ontario Ministry of Government and Consumer Services. The views expressed in the paper are the views of the authors and do not necessarily reflect those of the Province.

Supplemental Material: The online supplement is available at <https://doi.org/10.1287/serv.2021.0298>.

Keywords: healthcare operations • surgical supplies • newsvendor problem

1. Introduction

1.1. Background and Motivation

This study is concerned with the flow of surgical supplies between the hospital inventory and the operating rooms (ORs) and is motivated by a collaboration with the London Health Sciences Centre (LHSC), a network of hospitals in Ontario, Canada. Surgical procedures require a large number of supplies including instruments and consumable (disposable) items that need to be kept in hospital inventory and collected, prepared, and transported to the OR before each surgery. Our focus in this work is on consumable supplies (e.g., disposable instruments, sutures, staples, and catheters), which we refer to as items for brevity.

The items required for a surgery differ depending on the procedures involved in that surgery and the personal preferences of the surgeons. A *surgical preference card* (SPC) is a surgeon-specific document that lists the items required for each procedure. In addition to the required items, a SPC also specifies *fill* and *open* quantities. For each item, the *fill* quantity specifies how many should be taken to the OR, and the *open* quantity specifies the number of filled items that

should be opened at the beginning of the surgery (and how many should be kept on hold, namely *hold* quantity). Figure 1 illustrates the structure of a typical preference card at our partner hospital, but a similar structure is used in other hospitals in North America (Ben-Zvi 2014).

For items with no variability in usage in the OR, it clearly suffices to only specify the fill quantity on the SPC. The usage in the OR is, however, subject to uncertainty for many items, the source of which can be attributed to varying patient needs and/or unforeseen circumstances of the surgery. This uncertainty leads to mismatch between what is actually required in the OR and what is available and open at the beginning of the surgery, leading to operational and wastage costs. The fill and open decisions on the SPCs control the flow of items between the inventory and ORs and affect these costs, which we elaborate next.

Unused items that are taken to the OR need to be returned to inventory after the surgery and reshelved. This must be done in a timely manner to ensure the availability of the items for future surgeries. Furthermore, considering the large volume of daily surgeries,

Figure 1. Structure of a Typical SPC

Preference Card				
Date/Time: YYYY/MM/DD		OR: ####		Surgeon: Name Surname
Procedure:		Patient Age: ##		Patient Sex: M/F
Fill Qty	Open Qty	Hold Qty	Item Code	Description
1	1	0	Item 1	Description of item 1.
2	1	1	Item 2	Description of item 2.
6	3	3	Item 3	Description of item 3.

Notes. For each item, the card indicates how many should be taken to the OR and how many should be opened at the beginning of the surgery. The hold quantity is obtained by subtracting the open from fill quantity for each item.

this is not only time-consuming and resource intensive, but can also lead to misplacement of the items. Conversely, if an item is missing in the OR during the surgery, the item needs to be taken to the OR during the surgery. These so-called “OR runs” may lead to interruptions to the surgery or prolong its duration. The open decision directly affects wastage in the ORs. Items that are opened at the beginning of the surgery but end up not being used are wasted, incurring in many cases significant monetary cost. Finally, unopened but required items are undesired by surgeons and may also introduce delays to the surgery.

As the name suggests, SPCs are typically designed based on the surgeons’ preferences with little to no consideration for item wastage and operational costs. Although our partner hospitals did not collect data on wastage, our collaborators believed that wastage does happen and, in some cases, could be high. (We provide some evidence for high wastage in our case study in Section 5.2.) Indeed, the problem of high wastage in ORs is well documented in the medical literature. ORs in the United States produce more than 2,000 tons of waste per day (Wormer et al. 2013), the cost of which contributes to the high hospital spending associated with ORs. In a study of neurosurgical cases at a U.S. hospital, Zygorakis et al. (2017) find that 13.1% of the total surgical supply costs correspond to wastage. In another study of 50 procedures in a French hospital, Chasseigne et al. (2018) report that the wastage cost can constitute up to 20.1% of the total cost of the surgical items and estimate the annual savings associated with eliminating waste from three surgical departments to be €100,000.

We also conducted time studies on the number and time spent on returns and OR runs for four days in one of our partner hospitals and observed a considerable flow of material between the inventory and ORs during and after the surgeries. A large portion of disposable items (61%) taken to the OR were returned to the inventory and reshelved, with the time spent on returning each batch approximately linearly increasing in the number of items. During the surgeries, additional items were requested frequently. On

average, each surgery had 7.6 OR calls, with 21% of the calls asking for a disposable supply. Consistent with our observations, Chasseigne et al. (2018) also found that the circulating nurse spent 26.3% of operative time outside of the OR, mainly attending to the need for additional items.

Our partner hospitals recently invested in new scanning technology and information systems that allow for accurate tracking of usage (and wastage) data in the ORs. Although previous work has emphasized the importance of the impact of SPCs (Ben-Zvi 2014), to our knowledge a systematic approach to design of SPCs that could leverage the newly available usage data does not exist in the literature. As such, this work was initiated to develop practical analytical methods in support of determining the fill and open quantities on the SPCs and using the newly available usage data.

1.2. Summary of Results and Contributions

We formulate the problem of determining the fill and open quantities on a SPC as a parsimonious stochastic optimization problem, assuming that item usage is a discrete random variable. In practice, one can use the empirical usage distribution, obtained from historical usage data captured with an information system similar to that acquired by our partner hospitals. The objective is then to minimize the total expected cost of the surgery comprised of a weighted sum of expected returns, shortages, opening delays, and wastage incurred for all items. We then leverage our simple problem formulation to obtain both descriptive and prescriptive results on the design of SPCs and quantify the value of optimizing fill and open decisions. Our main contributions and results are outlined here.

- **Characterization of the optimal fill/open quantities:** Under the aforementioned assumptions (see Section 3 for details and Section 6 for a discussion), we show that the optimal quantities take the form of critical quantiles of the (marginal) usage distribution of each item, similar to the solution of the newsvendor problem. The solution structure enables a data-driven approach to determine the fill and open quantities

using the empirical usage distribution of the items. The simple solution form also allows for imputing the implied costs for given decisions. In addition, it provides insights on the value of including an open decision on the SPCs, in particular, on how the value depends on the cost parameters and properties of the usage distribution.

- **Service-level interpretation:** As in the newsvendor problem, the critical quantiles can be interpreted as service levels. When a higher service level is required for having all required items available in the OR than having them open at the beginning of the surgery, the solution prescribes a larger fill compared with the open quantity. Otherwise, it is optimal to open all items at the beginning of the surgery. Because missing an item in the surgery room (which triggers an OR run) is arguably less desirable than incurring delay in opening an item during the surgery, the solution form supports filling a larger quantity of items but opening a fraction of them at the beginning of the surgery. This observation is consistent with the current practice: around 35% of the consumables on all preference cards used at one of our partner hospitals had an open quantity smaller than the fill quantity. The service-level interpretation is practically appealing because specifying the cost parameters in practice, in particular, those associated with shortage and opening delays, could be hard. We further formalize this interpretation by proposing an alternative formulation that is cost-free and only requires specifying two target service levels for the shortage and opening delay probabilities. We show that the solution of the alternative formulation is closely related to the original cost-based formulation and leads to identical solutions in the practical parameter regime where a higher service level is required for the fill decision.

- **Numerical experiments and case study:** Finally, we illustrate the proposed approach and its benefits in terms of cost and wastage reduction using numerical experiments. In particular, in a case study conducted using historical usage data for a single-procedure surgery performed at one of our partner hospitals, we demonstrate how our proposed approach can be used in practice to revise SPCs and quantify the potential benefits compared with the status quo fill and open quantities used at the hospital. Our results illustrate the possibility of eliminating C\$6,212.78 worth of wastage and eliminating the shortage of 43 items in the OR over 127 cases considered in the case study. Given the large volume of surgeries conducted and the items used in hospitals (more than 2,000 unique items and 26,000 surgeries at our partner hospital), the aggregate benefits of revising SPCs using our proposed approach could potentially be very significant.

1.3. Organization of the Paper

The rest of the paper is organized as follows. After discussing related work in Section 2, we present the

problem formulation and our main results in Section 3. In Section 4 we present the service-level based formulation. Section 5 contains the numerical experiments including the case study using usage data from our partner hospital. Finally, we discuss the strengths and limitations of our work and provide directions for future work in Section 6. Proof of Lemma 2 is provided in the appendix. Proofs of all other results are relegated to the online supplement.

2. Related Literature

Preoperative supply chain management has received significant attention from researchers in the last decade (see Ahmadi et al. 2019 for a recent review). Most papers in this area focus on managing reusable items (instruments). For instance, van de Klundert et al. (2008) and Diamant et al. (2018) study inventory management of reusable instruments for hospitals with outsourced sterilization services. Reymondon et al. (2008) and Dobson et al. (2015) propose a mixed-integer program to optimize the composition of surgical instrument trays to minimize the associated operational costs.

A few studies have focused on disposable surgical supplies. Cardoen et al. (2015) provide an optimization model for creating a custom sterile packages for disposable items to be used in the surgeries. They propose a mixed-integer linear programming formulation and a linear programming-based heuristic to minimize the “point of touch” defined as the physical contact between staff and the medical materials. They conduct a case study based on data from a Belgian hospital and illustrate their solution approaches. Ben-Zvi (2014) provides a detailed description of the OR supply chain based on the operations of the Massachusetts General Hospital. The author investigates the problem of determining base-stock levels for disposable surgical supplies, calculates base stock levels for maintaining three days of sufficient inventory, and organizes the storage locations of each item based on their importance. The estimated potential savings is \$300,000 annually.

More relevant to our work are studies that are directly related to SPCs. Most works in this area are empirical studies from the medical literature that point out the importance of managing SPCs and their potential improvements. Huntley et al. (2018) note that preference cards are often under emphasized and poorly maintained and sustained, and Ahmadi et al. (2019) argue that SPC standardization is the most effective way of cost reduction in ORs. Ben-Zvi (2014) identifies the inaccuracies in the SPCs as a challenge in the management of disposable surgical supplies and discusses the importance of updating preference cards based on the real usage in the ORs. A few studies conduct ad hoc SPC standardization by manual

inspection and provide significant cost reduction by eliminating the waste even from a subset of procedures/items; examples include Simon et al. (2018), Williams (2007), Eiferman et al. (2015), and Harvey et al. (2017). Zygourakis et al. (2017) conduct an empirical investigation on 58 neurosurgical cases. They estimate the average cost of waste per surgery as \$653 that accounts for 13.1% of the overall surgical supply cost and approximately \$2.9 million per year in the neurological department of a single hospital. They suggest that this waste can be reduced by standardizing the SPCs through eliminating the unnecessary items and preventing items from being opened unless needed.

To the best of our knowledge, our work is the first to propose a systematic method for determining the open and fill quantities on the SPCs. The optimal quantities obtained based on our approach have the same structure as that of the newsvendor problem. The newsvendor model has been extensively studied, and its extensions have been applied to various application areas including revenue maximization (Petruzzi and Dada 1999), portfolio optimization (Carr and Lovejoy 2000), nurse staffing (Green et al. 2013), and call center staffing (Harrison and Zeevi 2005). Qin et al. (2011) provide a recent review. A large body of recent studies has focused on the newsvendor problem with unknown demand distribution. With historical demand data, a common approach is to use the empirical demand distribution, which is equivalent to a sample average approximation. Levi et al. (2007) provide bounds on the optimality gap of this approach. If empirical demand data are not available, various robust solutions can be used given limited information, for example, minimum, maximum, and mode of usage (see Perakis and Roels (2008) and Zhu et al. (2013) for min-max regret approaches). Ban and Rudin (2019) propose a data-driven newsvendor method that combines the estimation and optimization steps and solves the joint problem using machine learning algorithms in the presence of possibly a large number of covariates that are correlated

with demand. In our case study, we illustrate our approach using empirical distributions of item usage in the OR. Although we do not have access to patient-level data, we discuss the possibility of incorporating patient-level information into the design of personalized SPCs using the approaches proposed in Ban and Rudin (2019) in Section 6.

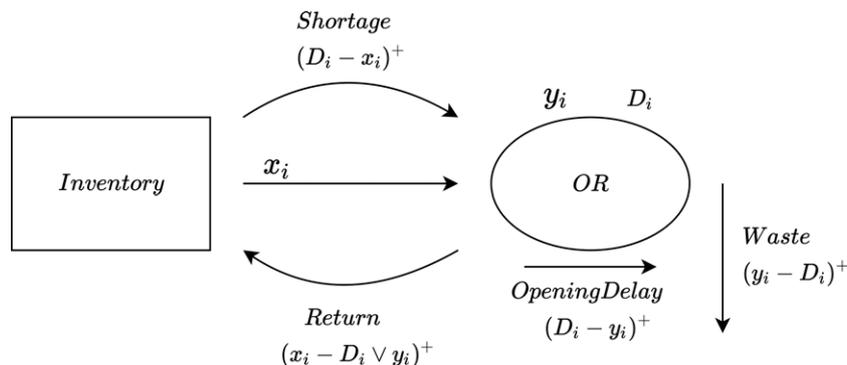
We further discuss imputing cost parameters from the historical decisions. A more general version of cost imputation using structural estimation is proposed in Olivares et al. (2008) in the presence of variability in the newsvendor decision. Unlike Olivares et al. (2008), here we are interested in imputing the costs for fixed decisions, that is, those on current SPCs, and under a discrete demand distribution.

3. Problem Formulation and Main Results

Consider a procedure requiring N consumable items (possibly more than one unit of each item). We assume that usage for item i is a discrete random variable D_i taking values on a finite bounded set. The minimum and maximum possible values of D_i are, respectively, denoted by A_i and B_i , where $B_i > A_i$. We denote the probability mass function of D_i by $p_i(\cdot)$ and its cumulative distribution function (cdf) by $F_i(\cdot)$. We do *not* require the usage random variables to be mutually independent. (The random variables are mutually independent if and only if $P(D_1 \leq d_1, \dots, D_N \leq d_N) = F_1(d_1) \dots F_N(d_N)$ for all d_1, \dots, d_N .) Finally, we use the notation $(a)^+ \equiv \text{Max}(a, 0)$ and $a \vee b \equiv \text{Max}(a, b)$.

Denote the fill quantity for item i by x_i and the open quantity by y_i . We begin by describing four *item-level* performance metrics affected by the fill and open decisions: (i) number of missing items that need to be brought to the OR during the surgery $(D_i - x_i)^+$; (ii) number of unused and unopened items taken to the OR that need to be returned and reshelved $(x_i - D_i \vee y_i)^+$; (iii) number of unopened but required items during the surgery $(D_i - y_i)^+$; and (iv) number of opened and unused (wasted) items $(y_i - D_i)^+$. The metrics are illustrated in Figure 2.

Figure 2. Flow of an Item Between the Inventory and OR and the Corresponding Metrics



We next formulate the problem of determining the open and fill quantities for all items as a stochastic optimization problem, where the objective is to minimize the total expected cost incurred for each instance of a procedure and comprised of the weighted sum of the previous metrics for all items of that procedure. As it will be evident shortly, however, it suffices to consider each item separately. Formally, the problem is to determine the optimal pair (x_i, y_i) satisfying $x_i \geq y_i$, $x_i \geq 0$ and $y_i \geq 0$ for all i to minimize

$$\sum_{i=1}^N u_1^i E[(D_i - x_i)^+] + o_1^i E[(x_i - D_i \vee y_i)^+] + u_2^i E[(D_i - y_i)^+] + o_2^i E[(y_i - D_i)^+], \quad (1)$$

where $u_1^i, o_1^i, u_2^i, o_2^i$ are cost parameters and can be interpreted as follows: u_1^i is the unit cost of bringing item i to the OR while the surgery is in progress; o_1^i is the unit cost of returning an unused item of type i back to the inventory; u_2^i is the unit cost associated with opening item i during the surgery; and o_2^i is the wastage cost (purchasing price) of item i . We assume that $o_2^i > o_1^i$ for all i , that is, the cost of returning an item is less than its value (otherwise, it would be optimal to waste the item in the OR rather than returning it). Furthermore, we assume that $u_1^i \geq u_2^i$ and $u_1^i \geq o_1^i$ for all i , implying that shortage is at least as costly as opening delay and returning the item, respectively.

Given the previous cost structure, it is easy to see that we can decompose the problem and optimize the fill and open quantity decisions for each item separately. Considering an arbitrary item and dropping the index i from the notation, we can state the item-level problem as follows:

$$\begin{aligned} \min_{(x,y)} \quad & C(x, y) \equiv u_1 E[(D - x)^+] + o_1 E[(x - D \vee y)^+] \\ & + u_2 E[(D - y)^+] + o_2 E[(y - D)^+], \quad (2) \\ \text{s.t.} \quad & y \leq x, \quad (3) \\ & x, y \geq 0. \quad (4) \end{aligned}$$

The following lemma allows us to rewrite the objective function in a more interpretable form and facilitates the characterization of the optimal solution to (2)–(4).

Lemma 1. *We have $(x - D \vee y)^+ = (x - D)^+ - (y - D)^+$ for all (x, y) satisfying $0 \leq y \leq x$, and all possible realizations of D .*

Using Lemma 1, we can rewrite the objective function (2) as follows:

$$C(x, y) = u_1 E[(D - x)^+] + o_1 E[(x - D)^+] + u_2 E[(D - y)^+] + (o_2 - o_1) E[(y - D)^+]. \quad (5)$$

The connection to the newsvendor problem is clear from (5). The objective is a linear combination of expected

underage and overage costs for the fill and open decisions. However, the two decisions are interconnected through the constraint $y \leq x$, requiring the number of items opened at the beginning of the surgery to be less than or equal to the fill quantity. Nevertheless, in the following we show that the optimal solution for each decision takes the same form as that of the newsvendor problem, although the solution differs depending on the cost parameters. Intuitively, we can think of the problem as a generalization of the newsvendor problem where, in addition to determining the number of items to be taken to the OR (the order quantity), one needs to also determine the number of items to be opened at the beginning of the surgery. This additional flexibility, introduced through the opening decision, allows for additional cost savings in certain cost parameter regimes as we show below and further examine in Section 3.2.

To present the optimal solution to Problems (2)–(4), we define the following cost ratios:

$$\beta_1 = \frac{u_1}{u_1 + o_1}, \quad \beta_2 = \frac{u_2}{u_2 + o_2 - o_1}, \quad \beta_3 = \frac{u_1 + u_2}{u_1 + u_2 + o_2}. \quad (6)$$

The following lemma is easy to verify using the definitions of the cost ratios.

Lemma 2. *Assume that $o_2 > o_1$. If $\beta_1 \geq \beta_2$, then $\beta_1 \geq \beta_3 \geq \beta_2$, otherwise, $\beta_2 > \beta_3 > \beta_1$.*

Denote by $F^{-1}(\cdot)$ the inverse cdf of the demand distribution interpreted as the generalized inverse function when D is a discrete random variable:

$$F^{-1}(q) = \inf\{z \geq 0; F(z) \geq q\}, \quad (7)$$

that is, the smallest integer quantity z such that $P(D \leq z) \geq q$.

Proposition 1. *The optimal solution of Problem (2)–(4), denoted by (x^*, y^*) , is characterized as follows. If the cost parameters satisfy $\beta_1 \geq \beta_2$, then $x^* = F^{-1}(\beta_1)$ and $y^* = F^{-1}(\beta_2)$, otherwise, $x^* = y^* = F^{-1}(\beta_3)$.*

Observe that in both cases the solution takes the form of critical quantiles of the usage distribution. In the first case ($\beta_1 \geq \beta_2$), it is optimal to use a larger quantile for the fill quantity compared with the open quantity. The ratios β_1 and β_2 have a similar structure to the critical quantile for the newsvendor problem, except β_2 uses the “net” overage cost ($o_2 - o_1$) because wasted items do not need to be reshelved. In the second case ($\beta_1 < \beta_2$), it is optimal to set the fill and open quantities equal to the same quantile determined by a cost ratio using “pooled” underage and overage costs.

When $\beta_1 \geq \beta_2$, the constraint $y \leq x$ is satisfied if we solve two independent newsvendor problems for each decision (fill/open) and using the corresponding cost ratios. Therefore, β_1 and β_2 can be interpreted as service levels corresponding to the fill and open quantity,

respectively. That is, β_1 is the optimal probability of incurring shortage in the OR, and β_2 is the optimal probability of not having all required items open at the beginning of the surgery.

In contrast, when $\beta_1 < \beta_2$, it becomes optimal to set $x = y$ and solve a single newsvendor problem with cost ratio β_3 , corresponding to an underage cost of $u_1 + u_2$ and overage cost of $o_1 + o_2 - o_1 = o_2$. The underage cost is the sum of shortage and opening delay costs, whereas the overage cost does not include the cost of return o_1 . Therefore, the implied service level for both fill and open decisions is β_3 which by Lemma 2 satisfies $\beta_1 < \beta_3 < \beta_2$.

Practically speaking, we expect most items to fall in the first case, that is, $\beta_1 \geq \beta_2$ where a higher service level is required for having all items in the OR than having all items open at the beginning of the surgery. A sufficient condition is for example to have $u_1 \geq u_2$ and $o_2 \geq 2o_1$, that is, shortage being costlier than opening delay and the item value being at least twice the cost of return. In this case, consistent with practice, the solution suggests filling a larger number of items (compared with the case with $\beta_1 < \beta_2$) but not opening all of them at the beginning of the surgery. As such, most of our analysis and numerical experiments focus on this parameter regime. Because the usage distribution is assumed to be discrete, equal fill and open quantities can also arise as the optimal solution when $\beta_1 \geq \beta_2$.

The following corollary provides explicit optimal solutions for certain cost parameters.

Corollary 1. *Assume that shortage is costlier than delay in opening, that is, $u_1 > u_2$. The following statements hold: (i) if $o_2 < o_1$, then $y^* = x^*$; (ii) if $u_2 = 0$ and $o_2 \geq o_1$, then $y^* = A$; (iii) if $u_2 = \infty$, then $y^* = x^* = B$; (iv) for sufficiently large $u_1, x^* = B$; (v) if $o_2 - o_1$ is sufficiently large and $o_2 > u_2$, then $y^* = A$.*

The results are intuitive: (i) implies that, if returning an item is costlier than wasting it, it is optimal to open every filled item so that no item is returned; (ii) states that if the opening delay cost is negligible, and the price of the item is higher than the cost of returning the item, it is optimal to set the open quantity to the minimum possible usage (A) of the item; (iii) states that if the item is critical, that is, no delays can be tolerated, it becomes optimal to fill the maximum possible usage for the item (B) and open them all at the beginning of the surgery; (iv) implies that if shortage in the OR cannot be tolerated, but opening delay can be tolerated, it is optimal to fill the maximum possible usage for the item (B) but possibly open a fraction of the filled items; finally, (v) states that if the item price is sufficiently larger than the return and shortage costs, it is optimal to open the minimum possible usage (A) and keep the remaining items unopened.

3.1. Cost Imputation

Proposition 1 further allows for imputation of the cost parameters of the model implied by the historical decisions. More specifically, given an open and fill decision, we can determine the cost parameters that would make that decision optimal. Because the overage costs (o_1, o_2) are easy to interpret, we focus on imputing the underage costs (u_1, u_2) that, respectively, correspond to the unit shortage and opening delay costs.

The following corollary presents two disjoint intervals for the underage cost parameters (u_1, u_2) that would make a given open and fill decision (x, y) optimal, assuming that the overage costs (o_1, o_2) are given and the cost parameters satisfy $\beta_1 \geq \beta_2$.

Corollary 2. *Assume that $\beta_1 \geq \beta_2$ and o_1 and o_2 are given. Then (x, y) is an optimal solution for any u_1 and u_2 satisfying*

$$\begin{aligned} u_1 &\in \left(o_1 \frac{F(x-1)}{1-F(x-1)}, o_1 \frac{F(x)}{1-F(x)} \right], \\ u_2 &\in \left((o_2 - o_1) \frac{F(y-1)}{1-F(y-1)}, (o_2 - o_1) \frac{F(y)}{1-F(y)} \right]. \end{aligned} \quad (8)$$

The ability to impute the implied underage costs for a given decision allows the decision makers to examine the current open and fill values on the preference cards and identify those in need of a revision. We will illustrate this further using an example in the numerical study of Section 5.

3.2. Value of Including an Open Decision

Intuitively, when $\beta_1 \geq \beta_2$, including an open quantity that is strictly smaller than the fill quantity can prevent incurring high wastage costs in return for smaller returning costs. Here, we quantify the value of this additional flexibility compared with the case with equal fill and open quantities and identify parameter regimes where this value is high. In the following proposition, we provide an exact expression, as well as a lower bound, for the expected total cost difference between the optimal solution of Problems (2)–(4) and the optimal solution when restricted to opening all items at the beginning of the surgery, that is, $x = y$. Denote the solution to the latter problem by (\hat{x}, \hat{x}) and note that Lemma 2 and Proposition 1 together imply that $x^* \geq \hat{x} \geq y^*$. Define,

$$\Delta C \equiv C(\hat{x}, \hat{x}) - C(x^*, y^*). \quad (9)$$

Proposition 2. *Assume that $\beta_1 \geq \beta_2$, then the expected value of including an open decision, ΔC , is given by,*

$$\Delta C = (u_1 + o_1) \sum_{z=1}^{x^* - \hat{x}} zp(z + \hat{x}) - (u_1 + o_1)(x^* - \hat{x})(F(x^*) - \beta_1) \quad (10)$$

$$\begin{aligned}
 & + (o_2 - o_1 + u_2) \sum_{z=0}^{\hat{x}-y^*-1} zp(\hat{x}-z) \\
 & + (o_2 - o_1 + u_2)(\hat{x} - y^*)(F(y^*) - \beta_2). \quad (11)
 \end{aligned}$$

Furthermore, ΔC satisfies,

$$\Delta C \geq \frac{(u_1 + o_1)p_{\min_{\beta_3, \beta_1}}}{2} \left[\frac{\beta_1 - \beta_3}{p_{\text{mode}_{\beta_3, \beta_1}}} - 1 \right] \left\lfloor \frac{\beta_1 - \beta_3}{p_{\text{mode}_{\beta_3, \beta_1}}} \right\rfloor \quad (12)$$

$$+ \frac{(o_2 - o_1 + u_2)p_{\min_{\beta_2, \beta_3}}}{2} \left[\frac{\beta_3 - \beta_2}{p_{\text{mode}_{\beta_2, \beta_3}}} - 1 \right] \left\lfloor \frac{\beta_3 - \beta_2}{p_{\text{mode}_{\beta_2, \beta_3}}} \right\rfloor, \quad (13)$$

where $p_{\min_{q_1, q_2}}$ and $p_{\text{mode}_{q_1, q_2}}$, respectively, denote the smallest and largest mass probability of usage between the q_1 th and q_2 th quantiles of the usage distribution.

The first term in (10) corresponds to the expected cost reduction because of filling a larger quantity $x^* \geq \hat{x}$, and the second term in (10) is because of the discrepancy between $F(x^*)$ and β_1 for discrete distributions (hence, would be equal to zero for a continuous usage distribution). Similarly, the first term in (11) corresponds to the expected reduction because of opening a smaller quantity $y^* \leq \hat{x}$ and the second term to the difference between $F(y^*)$ and β_2 .

The value clearly depends on the cost parameters but also on the usage distribution. The proposition also provides a lower bound on the value in (12) and (13). The lower bound highlights the dependency of the value on the usage distribution, in particular, the mass probabilities of usage between the critical quantiles $[\beta_3, \beta_1]$ and $[\beta_2, \beta_3]$. To obtain a lower bound, in the proposition, we approximate these probabilities using their smallest and largest values between the respective quantiles. In particular, the terms $\lfloor (\beta_1 - \beta_3) / p_{\text{mode}_{\beta_3, \beta_1}} \rfloor$ and $\lfloor (\beta_3 - \beta_2) / p_{\text{mode}_{\beta_2, \beta_3}} \rfloor$, respectively, provide lower bounds for $x^* - \hat{x}$ and $\hat{x} - y^*$, that is, the optimal increase in the number of filled items and the optimal decrease in the number of opened items, when an open quantity is included on the SPC. As a result, the lower bound becomes tighter when the mass probabilities are close to that of a uniform distribution. For example, if usage probabilities are equal between the β_1 th and β_3 th quantiles, then $\lfloor (\beta_1 - \beta_3) / p_{\text{mode}_{\beta_3, \beta_1}} \rfloor = x^* - \hat{x}$.

The lower bound together with the exact expression in (10) and (11) provide insights on the properties of the usage distributions corresponding to a high value of including an open quantity. The first terms in (10) and (11) include sum of the integers between the quantile pairs (x^*, \hat{x}) and (\hat{x}, y^*) weighted by their respective probabilities. Because the sum of probabilities between each pair of quantiles is constant, the value is larger if the usage distribution allocates a higher probability mass to larger integers between the quantiles, with the value attaining its maximum if all mass is allocated to

the larger quantile value, that is, \hat{x} and y^* , in the first and second sums, respectively.

The lower bound also illustrates the dependency of the value on the cost ratios. Observe that, keeping $(u_1 + o_1)$ and $(o_2 - o_1 + u_2)$ fixed, the lower bound on ΔC increases as $\beta_1 - \beta_3$ and $\beta_3 - \beta_2$ increase, assuming that the largest and smallest probabilities between the quantiles remain unchanged. In other words, as the difference between the service levels for fill and open decisions increase, the value of including an open decision increases.

We are also interested in understanding the impact of the cost parameters on the magnitude of the value. In general, the dependency of the value on usage distribution makes it hard to establish structural results on the impact of the cost parameters. If we consider a simpler lower bound on the value, obtained by replacing $p_{\text{mode}_{\beta_3, \beta_1}}$, $p_{\text{mode}_{\beta_2, \beta_3}}$, and $p_{\min_{\beta_3, \beta_1}}$, $p_{\min_{\beta_2, \beta_3}}$, with their global counterparts, we can establish monotonicity properties on the dependency of the lower bound on cost parameters.

Corollary 3. *The value of open decision satisfies,*

$$\Delta C \geq \frac{(u_1 o_2 - u_1 o_1 - o_1 u_2) p_{\min}}{2(u_1 + u_2 + o_2) p_{\text{mode}}} \left[\frac{\beta_1 - \beta_2}{p_{\text{mode}}} - 4 \right], \quad (14)$$

where p_{mode} and p_{\min} correspond to the largest and smallest mass probability of usage, respectively. Furthermore, assuming that $(\beta_1 - \beta_2) / p_{\text{mode}} \geq 4$, where p_{mode} is the largest mass probability of usage, the lower bound in (14) is (i) nondecreasing in u_1 ; (ii) nonincreasing in o_1 , (iii) nonincreasing in u_2 , (iv) nondecreasing in o_2 , and (v) nondecreasing in $\beta_1 - \beta_2$.

The corollary states that the lower bound on the value of including an open decision is nonincreasing in the cost of opening delay u_2 and the cost of return o_1 and nondecreasing in the item price o_2 and shortage cost u_1 . The result is intuitive and expected to hold under more general assumptions. As the cost of returns increase, it becomes less cost saving to fill a larger quantity. Similarly, with a higher opening delay, cost savings from keeping items unopened decreases. Conversely, as the shortage and wastage costs increase, it becomes more valuable to fill a larger quantity and open less at the beginning of the surgery. We note that the assumption $(\beta_1 - \beta_2) / p_{\text{mode}} \geq 4$ is typically satisfied for practical cases where $(\beta_1 - \beta_2)$ is large enough or the largest mass probability of usage p_{mode} is sufficiently small. Furthermore, although the monotonicity results are based on a lower bound, numerical experiments suggest that they hold for the exact value as well; see Section EC.2 of the online supplement for examples. We further illustrate the value of including the open quantity using numerical experiments in Section 5.

4. Service Level–Based Formulation

The model presented in Section 3 requires specifying four cost parameters for each item. As we alluded to in Section 3.1, although the overage costs are easy to interpret, the underage costs correspond to the cost of introducing delays to the surgery and potentially affecting the clinical outcomes. As such, determining these costs could be hard in practice. A well-known approach for estimating the critical quantile in the classical newsvendor problem is to express it in terms of the ratio of underage to overage cost and elicit an estimate for this ratio from an expert. For our problem and focusing on the practically relevant case of $\beta_1 \geq \beta_2$, β_1 can be expressed in terms of the cost ratio u_1/o_1 , and hence an estimate of the ratio of shortage to return cost is sufficient for determining the optimal fill quantity. However, β_2 depends on three cost parameters and determining the optimal open quantity requires an estimate of the ratio $u_2/(o_2 - o_1)$, which could be harder to determine. Motivated by these difficulties, in this section we present an alternative cost-free formulation based on satisfying target service levels. The service levels are easy to interpret and can be elicited directly without having to deal with costs. We show that the solution of the alternative formulation is closely related to the original cost-based formulation presented in Section 3 and leads to identical solutions in the practical regime where $\beta_1 \geq \beta_2$.

As in Section 3, we begin by considering a procedure with N items. The proposed formulation is a two-stage optimization problem. In the first stage, the objective is to determine the open quantities $y_i, i = 1, \dots, N$ to minimize the total wastage, subject to satisfying item-specific service levels on the probability of not having an item open when needed:

$$\min_{y_1, \dots, y_N} \sum_{i=1}^N E[(y_i - D_i)^+] \quad (15)$$

$$\text{s.t. } P(D_i > y_i) \leq 1 - \gamma_i, \quad \forall i \in \{1, \dots, N\}, \quad (16)$$

$$y_i \geq 0, \quad \forall i \in \{1, \dots, N\}. \quad (17)$$

The service level constraints in (16) imply that the probability of not having a required item open in the surgery room must be smaller than $1 - \gamma_i$ for item i , where γ_i is an input determined by the decision maker depending on the criticality of the item.

Given the optimal solutions, denoted by $\tilde{y}_i, i = 1, \dots, N$, in the second stage, we find the fill quantities that minimize the total number of unused and unopened items that need to be returned while satisfying item-specific service levels on the probability of shortage, that is, not having the item in the surgery room

and requiring the fill quantity to be at least equal to the optimal open quantities from the first stage:

$$\min_{x_1, \dots, x_N} \sum_{i=1}^N E[(x_i - D_i \vee \tilde{y}_i)^+] \quad (18)$$

$$\text{s.t. } P(D_i > x_i) \leq 1 - \psi_i, \quad \forall i \in \{1, \dots, N\}, \quad (19)$$

$$x_i \geq \tilde{y}_i, \quad \forall i \in \{1, \dots, N\}. \quad (20)$$

Similar to the first-stage problem, Constraint (19) requires the probability of shortage for each item to be smaller than $1 - \psi_i$ for item i where ψ_i is an input determined by the decision maker depending on the criticality of the item.

It is easy to see that the problems in both stages are decomposable (separable) for each item. Again, dropping the index i from the decision variables and cost parameters, we have the following item-level formulation:

$$\min_x C_2(x, \tilde{y}) \equiv E[(x - D \vee \tilde{y})^+], \quad (21)$$

$$\text{s.t. } P(D > x) \leq 1 - \psi, \quad (22)$$

$$x \geq \tilde{y}, \quad (23)$$

where \tilde{y} is the optimal solution of the following problem:

$$\min_y C_1(y) \equiv E[(y - D)^+], \quad (24)$$

$$\text{s.t. } P(D > y) \leq 1 - \gamma, \quad (25)$$

$$y \geq 0. \quad (26)$$

The following proposition characterizes the optimal solution to the item-level, two-stage problem.

Proposition 3. *The optimal solution to Problems (21)–(26) is given by $\tilde{y} = F^{-1}(\gamma)$ and $\tilde{x} = F^{-1}(\gamma \vee \psi)$.*

Observe that the open decision \tilde{y} is simply the quantile of the usage distribution evaluated at the specified service level γ , whereas the fill decision, \tilde{x} , is determined based on the maximum of the two service levels γ and ψ . Therefore, if a higher service level is required for having the items in the OR than having them open at the beginning of the surgery, that is, $\psi \geq \gamma$, the optimal solution is given by $\tilde{x} = F^{-1}(\psi)$ and $\tilde{y} = F^{-1}(\gamma)$. This suggests a connection to the cost-based formulation of Section 3. The following result, which follows directly from Propositions 1 and 3, formalizes this connection.

Corollary 4. *Consider Problems (21)–(26) with service levels (ψ, γ) and Problems (2)–(4) with cost parameters (u_1, u_2) and (o_1, o_2) . Assume that $(\psi, \gamma) = (\beta_1, \beta_2)$, where β_1 and β_2 are the cost ratios defined in (6). If the cost parameters satisfy $\beta_1 \geq \beta_2$, then the optimal solutions of the two problems coincide, that is, $(\tilde{x}, \tilde{y}) = (x^*, y^*)$. Otherwise if $\beta_1 < \beta_2$, we have $\tilde{x} = \tilde{y} \geq x^* = y^*$.*

The result formalizes the service-level interpretation of the cost ratios discussed in Section 3 when $\beta_1 \geq \beta_2$.

It states that if we solve the service level-based problem with service levels set to β_1 and β_2 as defined in (6), the solution (\tilde{x}, \tilde{y}) is also optimal for the cost-based formulation. This is practically appealing because it implies that when a higher service level is required for the fill decision, one can bypass cost estimation and directly elicit the service levels. We note that in the less practically relevant case of $\beta_1 < \beta_2$, the solutions do not necessarily coincide, and the service-level based formulation results in equal fill and open quantities that are equal or larger than those of the cost-based formulation.

5. Numerical Experiments

In this section, we use numerical experiments to illustrate the proposed methods and results and investigate the potential benefit of design of SPCs through our proposed approach using a case study that uses real usage data from one of our partner hospitals.

5.1. Illustrative Examples

We conduct four set of experiments. First, we use an example to illustrate the optimal solution of the cost-based model and the cost imputation approach discussed in Section 3. Second, we use numerical examples to illustrate the benefits of introducing an “open” decision on the SPCs. Third, we examine the impact of changing the fill and open decisions on the four performance metrics introduced in Section 3. Finally, we compare the performance of the optimal solution of the cost-based model to a simple heuristic that appears to be consistent with current practice at our partner hospital.

5.1.1. Cost-Based Optimization and Imputing the Underage Costs.

Here we use two examples to illustrate the optimal solution and the cost imputation approach. We use the empirical usage distributions for two surgical items from knee arthroplasty revision and coronary artery bypass surgeries (see Section 5.2 for additional details of our usage data). The usage

cdf for both items is illustrated in Figure 3. Item prices are $o_2^h = \text{C}\$160$ for item A and $o_2^l = \text{C}\$12.24$ for item B.

For item A, we set the return cost to $o_1 = \text{C}\$1$ and vary the underage costs u_1, u_2 . Figure 4 illustrates the optimal solution for u_1, u_2 satisfying $u_1 \geq u_2$ and $u_1 \geq o_1$. First, note that the solution space is mainly determined by u_1 . Because the item is relatively expensive ($o_2^h = \text{C}\$160$) and there is a positive probability of zero usage, opening an item is only optimal if the cost of opening delay is sufficiently large. When we impute the underage costs implied by the current fill and open quantities ($x = 3, y = 3$) using Proposition 2, we find that the decision is optimal when $u_1 \in (1.14, 4]$ and $u_2 \in (181.71, 636]$. The imputed underage costs satisfy $u_2 > u_1$, implying that shortage in the surgery room is less costly than not having the item open. This either implies that it is extremely important to have the item open at the beginning of the surgery or there is a potential for significant cost saving through keeping at least a fraction of the items unopened.

For item B, which has a lower price of $o_2^l = \text{C}\$12.24$ and a minimum usage of two items, we similarly set the return cost to $o_1 = \text{C}\$1$. Figure 5 illustrates the optimal solution for u_1, u_2 satisfying $u_1 \geq u_2$ and $u_1 \geq o_1$. For this item, the SPC used by the hospital prescribes filling four items and opening two of them at the beginning of the surgery ($x = 4, y = 2$), which is optimal when $u_1 \in (0.789, 2.091]$ and $u_2 \in (0, 1.498]$. This indicates a small cost for opening delay and shortage. If a higher service level is required for the fill decision, that is, shortage cost (u_1) is higher, it would be optimal to, for example, fill five items and open two, which corresponds to a larger region in Figure 5.

5.1.2. Value of Including an Open Decision. We numerically evaluate the value of including an open decision on the SPCs. In the experiments, we fix o_1, β_2 , and β_3 , and vary β_1 . Specifically, we fix $o_1 = 1$ and $\beta_3 = 0.5$ and consider two scenarios for β_2 , namely $\beta_2 \in \{0.1, 0.3\}$. We then vary β_1 in $[\beta_3 + 0.05, 0.95]$ with 0.05 increments.

Figure 3. Usage Distributions for the Two Items Considered in the Illustrative Examples

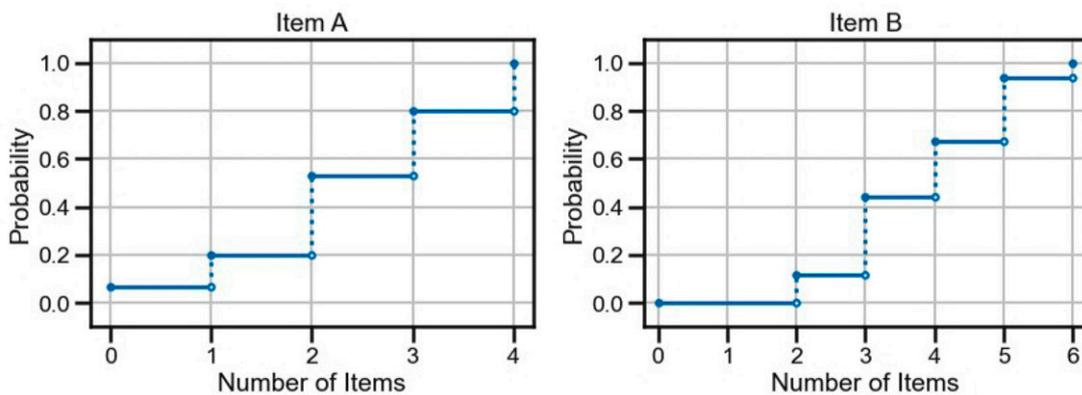
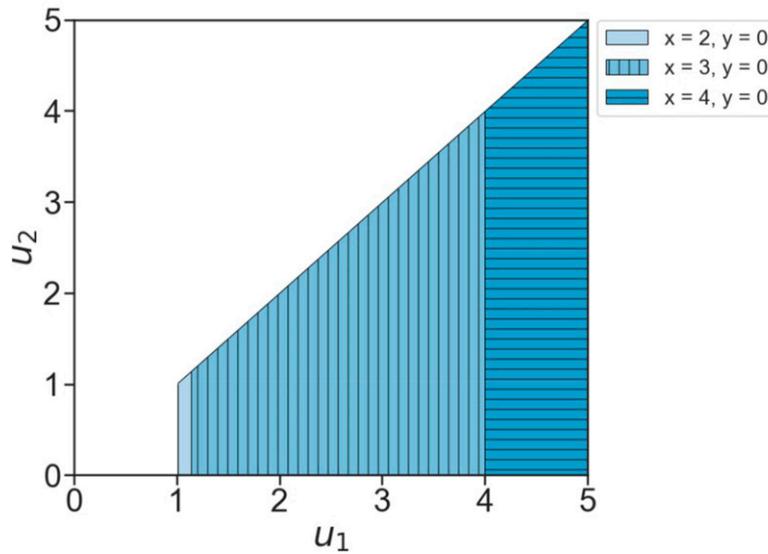


Figure 4. Optimal Decisions for Item A and for Different Underage Cost Values

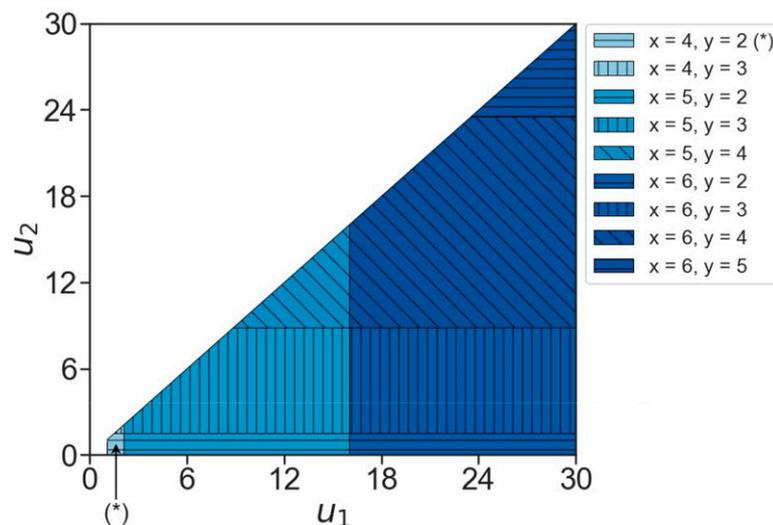


Note. The historical decision on the SPC is $x = 3, y = 3$, corresponding to $u_1 \in (1.14, 4]$ and $u_2 \in (181.71, 636]$, and is not displayed on the figure.

Once α_1, β_2 and β_3 are fixed, and then β_1 uniquely determines ΔC . To illustrate the impact of the usage distribution, we further consider three binomial(n, p) distributions with the number of trials n varying in $\{4, 6, 10\}$ and the probability of success p set such that the mean usage is equal to three for all cases; that is, p varies in $\{3/4, 1/2, 3/10\}$. We use a binomial distribution because for many items, it provides a good fit for the historical values observed in our usage data and has characteristics that make it suitable for modeling usage. Specifically, it captures the boundedness of the usage values and the range of variability observed in our usage data (see Section 5.2 for more details).

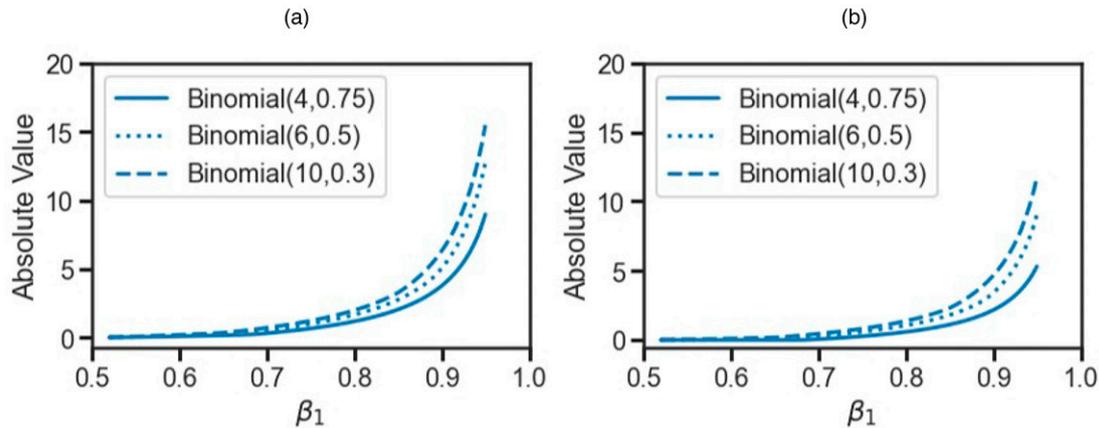
Figure 6 presents the results of the experiment. We can observe that, assuming α_1, β_2 , and β_3 are fixed, the value of including an open decision is increasing in β_1 . This also implies that the value is increasing in $\beta_1 - \beta_2$, which is consistent with the results established in Corollary 3 for the simple lower bound. We also observe that the value is convex increasing; that is, the marginal value is increasing in β_1 . This suggests that when a high service level on the availability of items in the OR is required, the value of including an open decision could be high. We further observe that the value is the highest for binomial(10,0.3) and lowest for binomial(4,0.75). This observation can be explained

Figure 5. Optimal Decisions for Item B and for Different Underage Cost Values



Notes. The historical decision on the SPC is $x = 4, y = 2$, which is marked with (*) on the figure.

Figure 6. Value of Including an Open Decision for Different Usage Distributions Plotted as a Function of β_1 by Fixing $\alpha_1, \beta_2,$ and β_3



Notes. (a) $\alpha_1 = 1, \beta_2 = 0.1, \beta_3 = 0.5$. (b) $\alpha_1 = 1, \beta_2 = 0.3, \beta_3 = 0.5$.

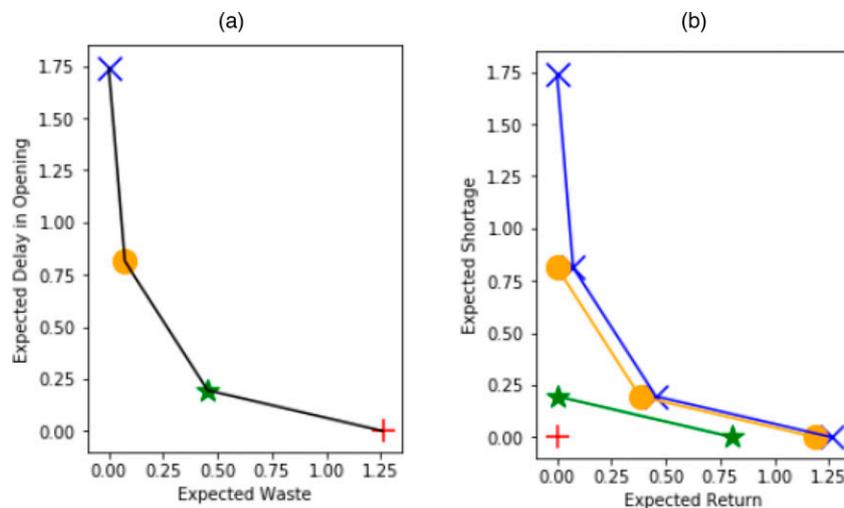
using the results of Section 3.2. The order holds because as $n \in \{4, 6, 10\}$ increases, the distribution allocates more probability mass to larger values between the critical quantiles.

5.1.3. On the Shortage-Return and Wastage-Opening Delay Tradeoffs. We use a numerical example to illustrate the tradeoff between the four performance metrics introduced in Section 3. The expected opening delay ($E[(D - y)^+]$) and waste ($E[(y - D)^+]$) depend only on the open decision y . Hence, we can illustrate the wastage-opening delay tradeoff as a function of y . In contrast, the expected return ($E[(x - y \vee D)^+]$) depends on both the fill x and the open decision y . Therefore, we investigate the shortage-return tradeoff by fixing the

open decision y and varying the fill decision x . We refer to this tradeoff as the first-stage tradeoff.

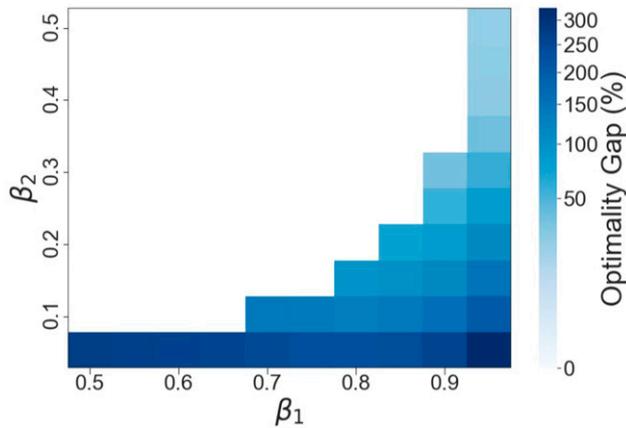
Figure 7 visualizes these two tradeoffs assuming that the item usage follows a binomial(n, p) distribution with $n = 3$ and $p = 0.59$. First consider the wastage-opening delay tradeoff in Figure 7(a). Each point on the figure corresponds to a different open decision $y \in \{0, 1, 2, 3\}$. For instance, the point marked with \times corresponds to $y = 0$, which does not lead to any waste but leads to the maximum expected delay in opening. In contrast, the point marked with $+$ corresponds to $y = 3$, which leads to the maximum expected waste but does not lead to any expected delay in opening. Next, consider the shortage-return tradeoff in Figure 7(b). Because shortage is a function of both fill x and open

Figure 7. Wastage-Opening Delay (Left) and Shortage-Return (Right) Tradeoffs



Notes. Each point on the left plot corresponds to an open decision. Each curve on the right figure corresponds to the shortage-return tradeoff for a fixed open decision on the left plot marked with the same symbol.

Figure 8. Relative Optimality Gap of the Heuristic with Respect to the Optimal Solution in the Practically Relevant Parameter Regime and Assuming Usage Follows $D \sim \text{Binomial}(3, 0.59)$, $o_1 = 1$ and $o_2 = 20$



y decisions, we include a separate tradeoff curve for each fixed open decision $y \in \{0, 1, 2, 3\}$. For instance, for the open decision $y = 0$ marked with \times , there is a corresponding curve in Figure 7(b) where possible solutions $x \in \{0, 1, 2, 3\}$ are also marked with \times . Both tradeoff curves are convex, suggesting that the expected shortage and waste can substantially be decreased in return for a small increase in expected return and opening delay.

5.1.4. Performance of a Simple Heuristic. Finally, here we compare the performance of the optimal fill and open quantities to that of a simple heuristic solution that we frequently observe in the data. The heuristic is to set both the fill and open quantities to the mode of the usage distribution. We observe this solution for 58% of the items in our data set. To examine the suboptimality of this simple solution, we assume that the item usage follows a binomial distribution with parameters $(3, 0.59)$, and set the price (o_2) to C\$20 and the return cost (o_1) to C\$1. We vary $\beta_1 \in [0.05, 0.95]$

and $\beta_2 \in [0.05, 0.95]$ in 0.05 increments. We focus on the practical parameter regime where $\beta_1 \geq \beta_2$ and eliminate the cases where $u_1 < u_2$ or $u_1 < o_1$. Figure 8 provides the optimality gap of the heuristic over these practically relevant instances. We observe that the optimality gap could be as high as 327% with an average suboptimality gap of 148%. Furthermore, we observe that the actual cost reduction could also be significant despite the relatively low price of the item and small variability in the usage distribution.

5.2. Case Study

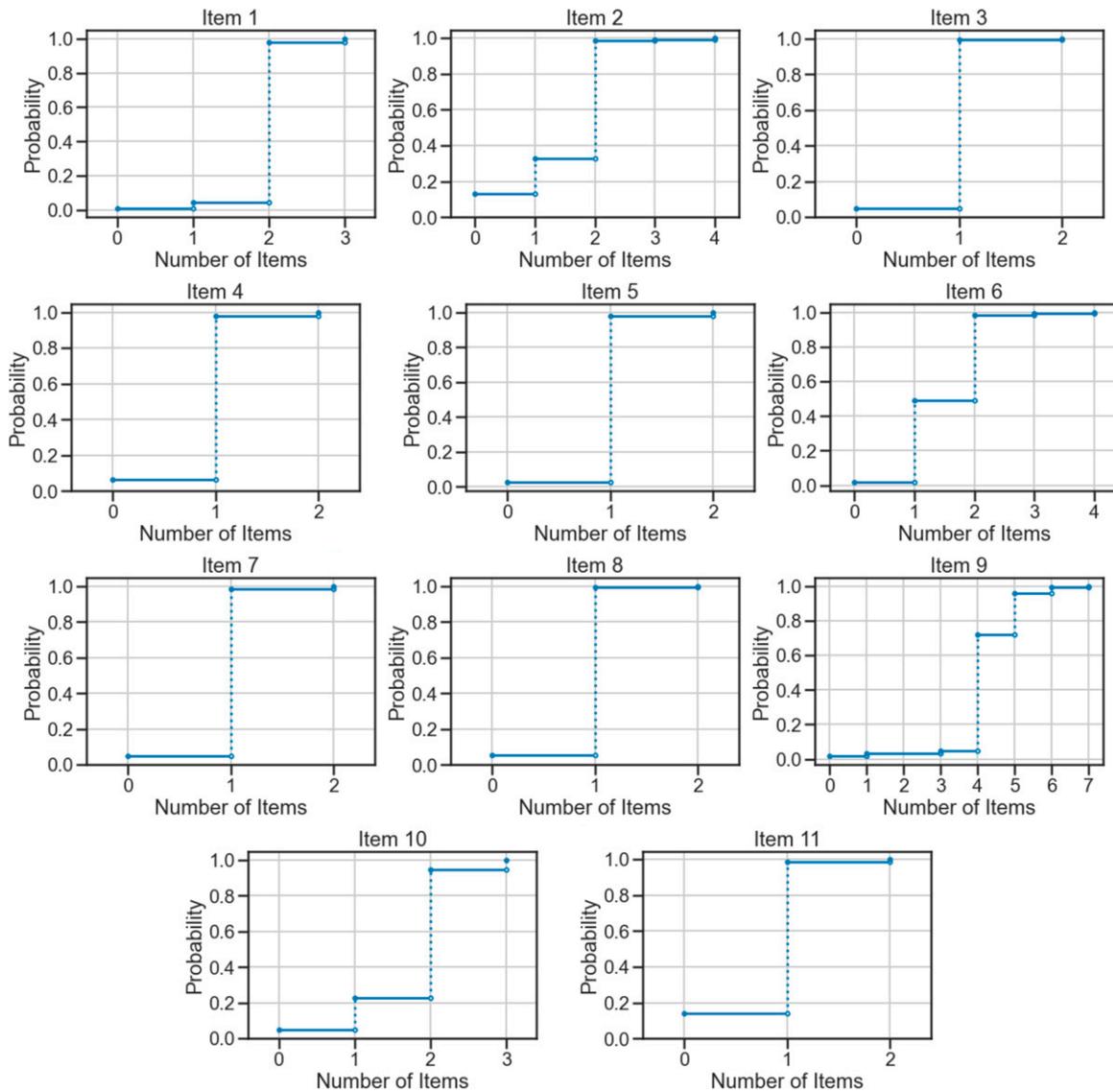
In this section, we present the results of a case study based on usage data of a single-procedure knee arthroplasty surgery performed regularly at our partner hospital. The main goal of the case study is to quantify and illustrate the potential benefits of redesigning SPCs based on our proposed approach and historical usage data.

The usage data we utilize in the case study was manually collected by OR nurses during August 2016 to March 2017 as part of a quality improvement program at the hospital and includes usage quantities from 127 historical cases. The preference card for this procedure contains 30 disposable items, 11 of which show variability in usage in the data. Table 1 provides the price and usage statistics for these items. The coefficient of variation (CV) is the ratio of standard deviation to mean usage and is a measure of variability in usage across different surgeries. For this surgery, we observe that the CVs are smaller than one and vary between 0.05 and 0.77, although for other surgeries, we observe higher variabilities as well. Table 1 also provides the historical open and fill quantities on the preference card for these 11 items that were used to pick the items. We note that the actual items picked for the surgery do not always exactly match the values on the SPC and may be modified before the surgery based on the knowledge of the head nurse. As we do not observe the actual number of picked items, we assume that the instructions on the SPCs were followed exactly.

Table 1. Historical Fill/Open Quantities, Prices, Usage Statistics, and Performance (per Surgery) for the Items Considered in the Case Study

Item ID	Fill	Open	Price (C\$)	Mean usage	CV	Expected shortage	Expected return	Expected wastage	Expected opening delay
1	2	2	2.32	1.97	0.05	0.024	0.000	0.055	0.024
2	2	2	80.00	1.56	0.37	0.024	0.000	0.465	0.024
3	1	1	66.95	0.96	0.06	0.008	0.000	0.047	0.008
4	1	1	38.63	0.96	0.09	0.024	0.000	0.063	0.024
5	1	1	20.24	1.08	0.77	0.024	0.000	0.024	0.024
6	2	1	5.72	1.52	0.23	0.024	0.468	0.016	0.535
7	1	1	8.36	0.97	0.06	0.016	0.000	0.047	0.016
8	1	1	78.11	0.95	0.06	0.007	0.000	0.055	0.008
9	4	4	2.28	4.20	0.19	0.328	0.000	0.126	0.331
10	2	2	3.35	1.78	0.21	0.055	0.000	0.195	0.125
11	1	1	29.99	0.87	0.16	0.016	0.000	0.142	0.016

Figure 9. Empirical Usage Distributions for the Items Considered in the Case Study



Finally, the table provides the average shortage, return, wastage, and opening delay per surgery computed based on the empirical usage distributions. Figure 9 presents the empirical usage distributions for all items.

From Table 1, we can observe that the historical fill and open decisions on the preference card tend to minimize returns and opening delays, leading to high wastage cost and considerable shortages in the surgery room. This suggests a potential in reducing wastage by opening less items and reducing shortages by filling more items at the beginning of the surgery.

We set the parameters of our model in the experiments as follows. For usage distribution, we directly use the empirical distribution observed in the data. This is equivalent to using a sample average approximation.

Although this approach does not lead to the optimal solution with respect to the unknown “true” usage distribution, performance bounds on the probability of optimality of the solution can be computed for a given sample size (Levi et al. 2007). For cost parameters, we set o_2 to the price reported in Table 1 for each item and vary $o_1 \in \{0.5, 1, 2\}$ to capture different return costs. The maximum return cost is set to two so that it is never optimal to waste an item in the OR, even for the cheapest item. For each value of o_1 , we then vary $\beta_1 \in [0.05, 0.95]$ and $\beta_2 \in [0.05, 0.95]$ in 0.05 increments. Fixing o_1 , o_2 , and β_1, β_2 uniquely determines u_1, u_2 . As shown in Section 4, when $\beta_1 \geq \beta_2$, optimizing costs is equivalent to minimizing waste and returns subject to service levels set to β_1 and β_2 . We focus on this practical parameter regime in our case study. Moreover, we eliminate the

Table 2. Shortage Quantities and Wastage Costs Obtained Using Historical and Optimal Decisions Across All Experiments

Item ID	Open/fill	Shortage			Waste		
		Min	Max	Mean	Min	Max	Mean
1	Historical	3.00	3.00	3.00	16.24	16.24	16.24
	Optimal	3.00	3.00	3.00	16.24	16.24	16.24
2	Historical	3.00	3.00	3.00	4,720.00	4,720.00	4,720.00
	Optimal	3.00	3.00	3.00	0.00	1,360.00	354.78
3	Historical	1.00	1.00	1.00	401.70	401.70	401.70
	Optimal	1.00	1.00	1.00	401.70	401.70	401.70
4	Historical	3.00	3.00	3.00	309.04	309.04	309.04
	Optimal	3.00	3.00	3.00	0.00	309.04	193.91
5	Historical	3.00	3.00	3.00	60.72	60.72	60.72
	Optimal	3.00	3.00	3.00	60.72	60.72	60.72
6	Historical	3.00	3.00	3.00	11.44	11.44	11.44
	Optimal	3.00	3.00	3.00	11.44	371.80	80.80
7	Historical	2.00	2.00	2.00	50.16	50.16	50.16
	Optimal	2.00	2.00	2.00	50.16	50.16	50.16
8	Historical	1.00	1.00	1.00	546.77	546.77	546.77
	Optimal	1.00	1.00	1.00	0.00	546.76	285.27
9	Historical	42.00	42.00	42.00	36.48	36.48	36.48
	Optimal	6.00	42.00	19.34	36.41	246.24	52.26
10	Historical	7.00	7.00	7.00	117.25	117.25	117.25
	Optimal	0.00	7.00	5.86	20.10	522.60	82.74
11	Historical	2.00	2.00	2.00	539.82	539.82	539.82
	Optimal	2.00	2.00	2.00	0.00	539.82	253.04
Total	Historical	70.00	70.00	70.00	6,809.62	6,809.62	6,809.62
	Optimal	27.00	70.00	46.19	596.84	4,425.09	1,831.62
Per surgery	Historical	0.55	0.55	0.55	53.62	53.62	53.62
	Optimal	0.21	0.55	0.36	4.70	34.84	14.42
Relative difference		-62%	0%	-34%	-91%	-35%	-73%

Note. The Min, Max, and Mean values, respectively, correspond to the minimum, maximum, and average values across all cost parameters considered in the experiments.

instances where the corresponding cost parameters do not satisfy the assumptions $u_1 \geq u_2$ or $u_1 \geq o_1$. In doing so, we aim to evaluate the potential benefits of using optimal fill and open quantities for a range of practically relevant cost parameters. Our case study includes a total of 1,904 instances.

Tables 2 and 3 summarize the results of the experiments. For each item, as well as in aggregate, Table 2 reports the minimum, maximum, and average shortage quantities and wastage costs corresponding to the optimal and historical fill/open decisions across all experiments. Table 3 presents the minimum, maximum, and average cost reduction obtained under the optimal fill and open decisions compared with (1) using the historical fill/open quantities specified on the SPC and (2) using optimal fill quantities under the restriction that all items are opened at the beginning of the surgery, that is, equal fill and open quantities.

We observe that the total cost can be reduced as much as 68.2% and on average by 43.5%, depending on the service levels required for shortage and opening delay. In particular, if opening delays can be tolerated, the wastage cost can be reduced as much as 91%, which corresponds to C\$6,212.78 over the 127 cases considered. Similarly, shortage can be reduced by as

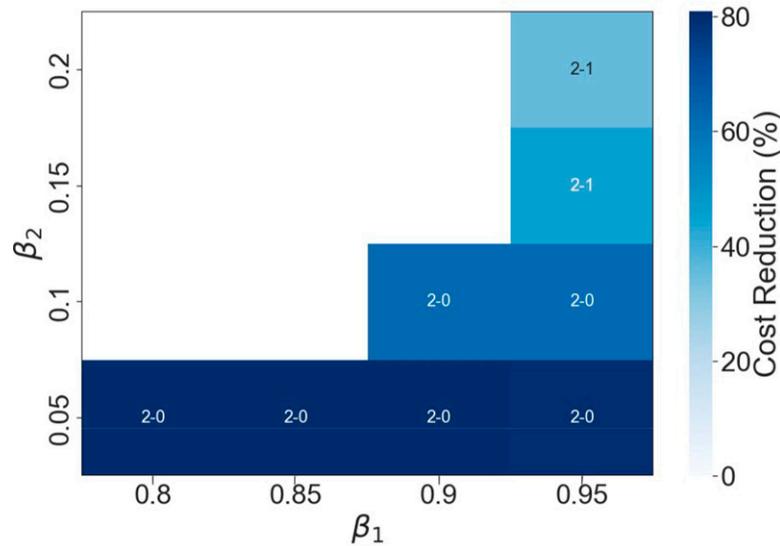
much as 62%, which corresponds to 43 items across all cases. There are two contributing factors to these observations: filling inadequately and opening excessively. As evident from Table 1, the historical fill/open decisions lead to returns for only one of the items. Our results suggest that by filling more items, the shortages in the surgery room (which are arguably costlier than returns) can be significantly reduced. In addition, by not opening all items at the beginning of the surgery, the wastage cost can be significantly reduced in return for incurring some opening delays

Table 3. Cost Reduction Under the Optimal Fill/Open Quantities Compared with the Historical Ones and the Value of Including an Open Decision

	Cost reduction			Value of open decision		
	Min	Max	Mean	Min	Max	Mean
Total	401.14	7,704.93	3,429.11	401.14	4,811.77	1,940.19
Per surgery	3.16	60.67	27.00	3.16	37.89	15.28
Relative reduction	5.6%	68.2%	43.5%	5.6%	63.7%	24.6%

Note. The Min, Max, and Mean values, respectively, correspond to the minimum, maximum, and average values across all cost parameters considered in the experiments.

Figure 10. Optimal Fill and Open Quantities for Item 2 for Varying Cost Ratios and the Corresponding Cost Savings



Note. The reported values correspond to optimal fill and open quantities (fill-open) and the colors represent the magnitude of relative cost reduction (%) compared with the historical decision (2-2).

during the surgery. More specifically, including an open decision can provide as much as 63.7% and on average 24.6% reduction in the total cost compared with using optimal but equal fill and open decisions. Consistent with our observations in Section 3.2, the value of including an open decision is high when a high fill service level is required in the surgery room, especially for the more expensive items.

To further illustrate the potential benefits, we next discuss the results for two individual items more closely. We fix the return cost to $o_1 = 1$ in the following experiments and consider the total cost savings that would have been realized over the 127 historical cases if the hospital followed “optimal” fill and open quantities.

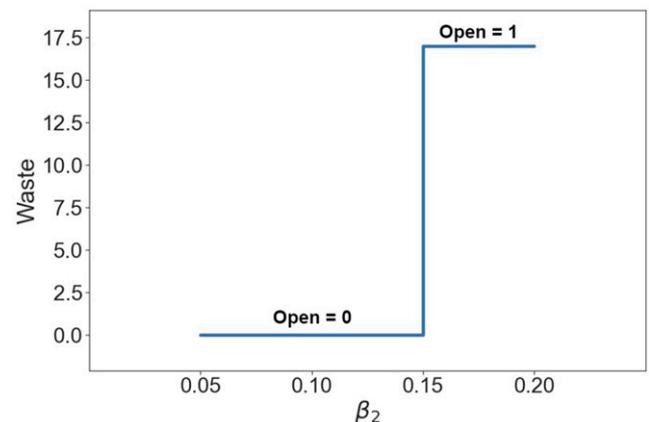
First, we consider item 2, which is relatively expensive ($o_2 = \text{C}\$80$) with historical fill and open decisions both set to two. Figure 10 presents the total cost reduction of using the optimal fill and open quantities under different β_1 and β_2 ratios, and Figure 11 presents the total wastage quantity of using the optimal open quantity under different β_2 ratios over the 127 cases. The historical decision (2,2) is optimal when $\beta_2 > 0.33$, which corresponds to $u_2 > 116.7$. If opening delays can be tolerated for this item, Figure 11 implies that wastage of 59 items, with total value of $\text{C}\$4,720$, could have been eliminated by keeping the items unopened at the beginning of the surgery. Alternatively, by only opening one item, wastage would have been reduced by 42 items, saving a total of $\text{C}\$3,360$.

Next, we consider item 9, which is a less expensive item ($o_2 = \text{C}\$2.28$) with historical fill and open decisions both set to four. Figures 12 and 13, respectively, present the total cost reduction and number of returns

under the optimal fill and open quantity for varying β_1 and β_2 ratios. Figure 14 presents the shortage values under the optimal fill/open quantities for different β_1 values. The current fill/open decision (4,4) is optimal when $\beta_1 \leq 0.76$ and $\beta_2 \leq 0.76$. As Figure 14 illustrates, this decision corresponds to a historical shortage of 42 items over the 127 cases. Considering a ratio $\beta_1 > 0.76$, that is, assuming a larger relative shortage to return cost, the optimal solution would be to fill five items but only open four of them at the beginning of the surgery. This would have reduced the shortage by 36 items and eliminated the potential delays to the surgery because of shortage.

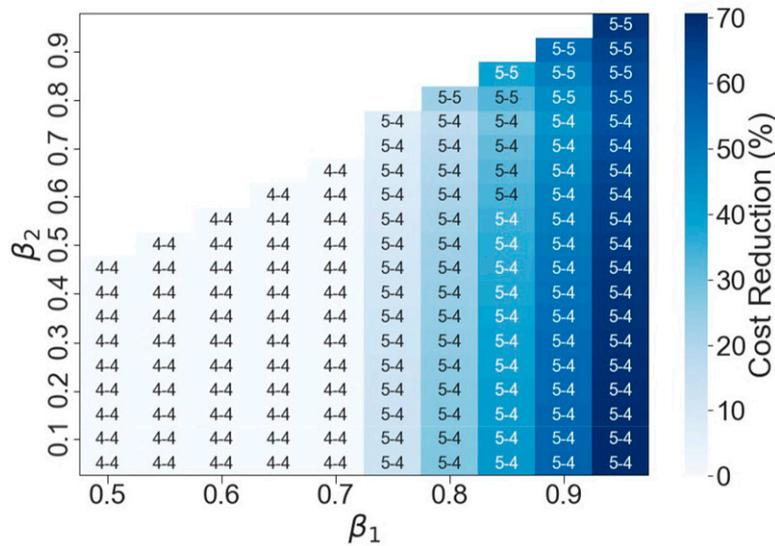
In this case study, we focus on a single procedure conducted 127 times throughout the year by a single

Figure 11. Wastage Quantity for Item 2 Under Optimal Open Decisions for Different β_2 Values



Note. The historical open decision for the item is 2.

Figure 12. Optimal Fill and Open Quantities for Item 9 for Varying Cost Ratios and the Corresponding Cost Savings



Note. The reported values correspond to optimal fill and open quantities (fill-open), and the colors represent the magnitude of relative cost reduction (%) compared with the historical decision (4-4).

surgeon. We find that it is possible to save a maximum of C\$60.67 and on average C\$27.00 per surgery. Our partner hospital performed on average 26,316 surgeries per year over the past five years. Extrapolating these savings to all surgeries suggests a potential for large savings of maximum C\$1,596,592 and on average C\$710,532 per year.

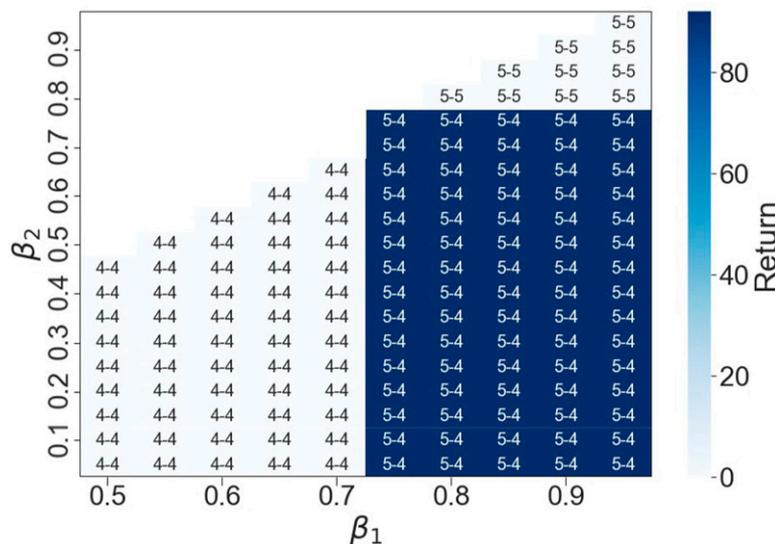
6. Discussion and Future Work

6.1. Summary of Main Results

In this work, we propose a newsvendor-based solution for design of SPCs. The approach allows for direct

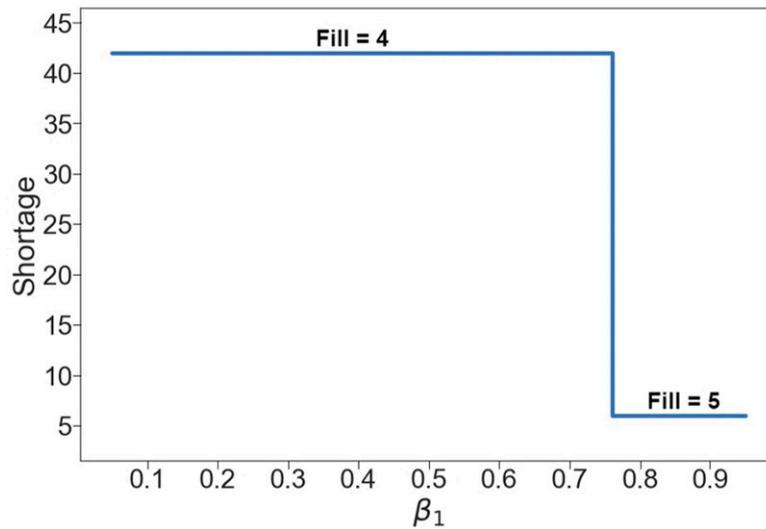
use of historical usage data to determine the fill and open quantities that minimize the wastage of items in the OR and returns from the OR while maintaining required service levels on the probability of shortage and delayed openings. The structure of the resulting solution is consistent with the current practice of including an open quantity on the SPC in addition to the fill quantity. Furthermore, it provides insights on the value of including an open decision. The fill quantity should be set at a higher quantile of the usage distribution compared with the open decision. This implies bringing a “larger” number of

Figure 13. Return Quantities for Item 9 Under the Optimal Decision for Different Cost Ratios



Note. The reported values correspond to optimal fill and open quantities (fill-open), and the colors represent the corresponding return quantity.

Figure 14. Shortage Quantity for Item 9 Under Optimal Fill Decisions for Different β_1 Values



Note. The historical fill decision for the item is four.

items to the OR but only opening a fraction of them at the beginning of the surgery. The results of our case study using real usage data suggest a potential for significant reduction of waste and shortage in the ORs by adjusting the fill and open quantities by leveraging the available usage data using our proposed approach. Considering the large volume of surgeries operated in hospitals (e.g., more than 26,000 per year in our partner hospital) and the number and value of the disposable supplies, our results suggest an opportunity for significant reduction of wastage and reducing the number of unnecessary returns and OR runs.

6.2. Limitations and Future Work

Here, we assume that the total cost is a linear combination of expected shortage, returns, waste, and delayed openings. In particular, the incurred cost linearly increases with the number of occurrences for each metric. This is a reasonable assumption for the waste and returns but is likely an approximation of reality for the shortage and delayed openings. A fixed setup cost may also be required as items can be brought up to the surgery room or be opened at the same time. Nevertheless, this assumption allows us to decompose the problem into an item-level formulation and uncover the simple newsvendor-form solution. We believe the newsvendor connection is useful as it renders the vast literature on this topic readily applicable. In particular, patient-level features that are correlated with the usage of the items in the OR can be used to design personalized SPCs using the Big Data newsvendor framework of Ban and Rudin (2019).

Extensions of our simple model to include more realistic cost structures can be considered in future

work. Including setup costs is straightforward and leads to a stochastic program that can be solved, for example, using sample average approximation. Including nonlinear costs or accounting for the costs shared among concurrent surgeries would be more challenging extensions. From a practical perspective, revising the SPCs using the approach proposed in this work and conducting a prospective study for empirical evaluation of the benefits could be invaluable in convincing medical practitioners to adopt the approach.

Acknowledgments

The authors thank the anonymous referees, associate editor, and department editor for constructive comments and suggestions that helped improve the paper. The authors are grateful to their collaborators at London Health Sciences for valuable discussions and help in accessing and understanding the data.

Appendix. Proof of Lemma 2

First, consider the case with $\beta_1 \geq \beta_2$ and note that it implies

$$\frac{u_1}{u_1 + o_1} \geq \frac{u_2}{o_2 - o_1 + u_2}. \quad (\text{A.1})$$

Because by assumption $o_2 > o_1$ and hence $o_2 - o_1 + u_2 > 0$, (A.1) implies

$$(u_1 o_2 - u_1 o_1 + u_1 u_2) \geq (u_2 u_1 + u_2 o_1) \Rightarrow u_1 o_2 - u_1 o_1 - u_2 o_1 \geq 0. \quad (\text{A.2})$$

Using the definition of β_1 and β_3 , we have,

$$\beta_1 - \beta_3 = \frac{u_1}{u_1 + o_1} - \frac{u_1 + u_2}{u_1 + u_2 + o_2} = \frac{u_1 o_2 - u_1 o_1 - u_2 o_1}{(u_1 + o_1)(u_1 + u_2 + o_2)}, \quad (\text{A.3})$$

which using (A.2) implies that $\beta_1 \geq \beta_3$. Similarly, using the definition of β_3 , we have

$$\begin{aligned}\beta_3 - \beta_2 &= \frac{u_1 + u_2}{u_1 + u_2 + o_2} - \frac{u_2}{o_2 - o_1 + u_2} \\ &= \frac{u_1 o_2 - u_1 o_1 - u_2 o_1}{(o_2 - o_1 + u_2)(u_1 + u_2 + o_2)}.\end{aligned}\quad (\text{A.4})$$

Again, using (A.2) and because $o_2 - o_1 + u_2 > 0$, we can conclude that $\beta_3 \geq \beta_2$. This proves the first claim, that is, $\beta_1 \geq \beta_3 \geq \beta_2$. Next, consider the case with $\beta_2 > \beta_1$. In this case and in contrast to (A.2), we have $u_1 o_2 - u_1 o_1 - u_2 o_1 < 0$. Therefore, both numerators in (A.3) and (A.4) become negative, and hence we can conclude that $\beta_2 > \beta_3 > \beta_1$. \square

References

- Ahmadi E, Masel DT, Metcalf AY, Schuller K (2019) Inventory management of surgical supplies and sterile instruments in hospitals: A literature review. *Health Systems (Basingstoke)* 8(2):134–151.
- Ban G-Y, Rudin C (2019) The big data newsvendor: Practical insights from machine learning. *Oper. Res.* 67(1):90–108.
- Ben-Zvi N (2014) $(OR)^2$: operations research applied to operating room supply chain. PhD thesis, Massachusetts Institute of Technology, Cambridge.
- Cardoen B, Beliën J, Vanhoucke M (2015) On the design of custom packs: Grouping of medical disposable items for surgeries. *Internat. J. Production Res.* 53(24):7343–7359.
- Carr S, Lovejoy W (2000) The inverse newsvendor problem: Choosing an optimal demand portfolio for capacitated resources. *Management Sci.* 46(7):912–927.
- Chasseigne V, Leguelinel-Blache G, Nguyen T, de Tayrac R, Prudhomme M, Kinowski J, Costa P (2018) Assessing the costs of disposable and reusable supplies wasted during surgeries. *Internat. J. Surgery* 53:18–23.
- Diamant A, Milner J, Queresby F, Xu B (2018) Inventory management of reusable surgical supplies. *Health Care Management Sci.* 21(3):439–459.
- Dobson G, Seidmann A, Tilson V, Froix A (2015) Configuring surgical instrument trays to reduce costs. *III Trans. Healthcare Systems Engrg.* 5(4):225–237.
- Eiferman D, Bhakta A, Khan S (2015) Implementation of a shared-savings program for surgical supplies decreases inventory cost. *Surgery* 158(4):996–1002.
- Green LV, Savin S, Savva N (2013) “Nursevendor problem”: Personnel staffing in the presence of endogenous absenteeism. *Management Sci.* 59(10):2237–2256.
- Harrison JM, Zeevi A (2005) A method for staffing large call centers based on stochastic fluid models. *Manufacturing Service Oper. Management* 7(1):20–36.
- Harvey LF, Smith KA, Curlin H (2017) Physician engagement in improving operative supply chain efficiency through review of surgeon preference cards. *J. Minimal Invasive Gynecology* 24(7):1116–1120.
- Huntley JS, Howard JJ, Simpson J, Sigalet DL (2018) Updating the surgical preference list. *Cureus* 10(7):e2997.
- Levi R, Roundy RO, Shmoys DB (2007) Provably near-optimal sampling-based policies for stochastic inventory control models. *Math. Oper. Res.* 32(4):821–839.
- Olivares M, Terwiesch C, Cassorla L (2008) Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Sci.* 54(1):41–55.
- Perakis G, Roels G (2008) Regret in the newsvendor model with partial information. *Oper. Res.* 56(1):188–203.
- Petrucci NC, Dada M (1999) Pricing and the newsvendor problem: A review with extensions. *Oper. Res.* 47(2):183–194.
- Qin Y, Wang R, Vakharia AJ, Chen Y, Seref MM (2011) The newsvendor problem: Review and directions for future research. *Eur. J. Oper. Res.* 213(2):361–374.
- Reymondon F, Pellet B, Marcon E (2008) Optimization of hospital sterilization costs proposing new grouping choices of medical devices into packages. *Internat. J. Production Econom.* 112(1):326–335.
- Simon KL, Frelich MJ, Gould JC (2018) Picking apart surgical pick lists: Reducing variation to decrease surgical costs. *Amer. J. Surgery* 215(1):19–22.
- van de Klundert J, Muls P, Schadd M (2008) Optimizing sterilization logistics in hospitals. *Health Care Management Sci.* 11(1):23–33.
- Williams J (2007) Dollars and sense engaging physicians in supply-cost control: Physicians could play a vital role in large-scale supply cost savings initiatives—but some hospitals struggle to include them in these efforts. Learn how three hospitals worked with their physicians to cut supply costs. *Healthcare Financial Management* 61(4):62–69.
- Wormer BA, Augenstein VA, Carpenter CL, Burton PV, Yokeley WT, Prabhu AS, Harris B, et al (2013) The green operating room: Simple changes to reduce cost and our carbon footprint. *Amer. Surgery* 79(7):666–671.
- Zhu Z, Zhang J, Ye Y (2013) Newsvendor optimization with limited distribution information. *Optim. Methods Software* 28(3):640–667.
- Zygourakis CC, Yoon S, Valencia V, Boscardin C, Moriates C, Gonzales R, Lawton MT (2017) Operating room waste: Disposable supply utilization in neurosurgical procedures. *J. Neurosurgery* 126(2):620–625.