

Data-driven Distributionally Robust Surgery Planning in Flexible Operating Rooms Over a Wasserstein Ambiguity

Karmel S. Shehadeh^{a,*}

^a*Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA*

Abstract

We study elective surgery planning in flexible operating rooms (ORs) where emergency patients are accommodated in the existing elective surgery schedule. Specifically, elective surgeries can be scheduled weeks or months in advance. In contrast, an emergency surgery arrives randomly and must be performed on the day of arrival. Probability distributions of the actual durations of elective and emergency surgeries are unknown, and only a possibly small set of historical realizations may be available. To address distributional uncertainty, we first construct an ambiguity set that encompasses all possible distributions of surgery durations within a 1-Wasserstein distance from the empirical distribution. We then define a *distributionally robust surgery assignment* (DSA) problem to determine optimal elective surgery assignment decisions to available surgical blocks in multiple ORs, considering the capacity needed for emergency cases. The objective is to minimize the total cost consisting of the fixed cost related to scheduling or rejecting elective surgery plus the maximum expected cost associated with OR overtime and idle time over all distributions defined in the ambiguity set. Using the DSA model's structural properties, we derive an equivalent mixed-integer linear programming (MILP) reformulation that can be implemented and solved efficiently using off-the-shelf optimization software. In addition, we extend the proposed model to determine the number of ORs needed to serve the two competing surgery classes and derive a MILP reformulation of this extension. We conduct extensive numerical experiments based on real-world surgery data, demonstrating our proposed model's computational efficiency and superior out-of-sample operational performance over two state-of-the-art approaches. In addition, we derive insights into surgery scheduling in flexible ORs.

Keywords: Surgery scheduling, operating rooms, distributionally robust optimization, Wasserstein metric, mixed-integer programming.

*Corresponding author.

Email address: kas720[at]lehigh.edu, karmelshehadeh[at]gmail.com (Karmel S. Shehadeh)

1. Introduction

Operating Rooms (OR) generates about 40–70% and 20–40% of hospitals’ revenues and operating costs, respectively (Bovim et al., 2020; Jackson, 2002; Li et al., 2016; Viapiano and Ward, 2000). As an essential area for cost management, planning and scheduling OR activities have received intense research attention (Cardoen et al., 2010; Hof et al., 2017; May et al., 2011; Samudra et al., 2016; Shehadeh and Padman, 2022; Zhu et al., 2019). In this paper, we focus on elective surgery planning in flexible operating rooms where emergency patients are accommodated in the existing elective surgery schedule. Elective cases can be scheduled weeks or months in advance. In contrast, the arrival of emergency surgeries is random and they must be performed on the day of arrival. The probability distributions of the durations of elective and emergency cases are unknown, and only small data on their durations is available. The goal is to construct a plan that specifies the assignments of a subset of elective cases from a waiting list to available surgical blocks in multiple ORs. The plan’s quality is a function of costs related to performing (scheduling) or delaying elective surgeries and costs related to OR overtime and idle time.

The concept of flexible ORs differs from dedicated ORs. In the latter type, one or more ORs are solely dedicated to accommodating emergency surgeries, and elective surgery can only be scheduled and performed in any ORs dedicated to its type. Therefore, we do not consider emergency surgeries when planning for elective surgeries in dedicated ORs. In contrast, in flexible ORs, the capacity of the ORs is shared among the two competing surgery classes. Thus, when planning for elective surgery in flexible ORs, we need to account for emergency surgery. Note that hospitals have either flexible or dedicated ORs. However, most of the existing elective surgery scheduling approaches focus on elective surgery scheduling in dedicated ORs (see Section 2).

Scheduling surgeries in ORs is a complex task primarily due to the significant variability of surgery duration and limited OR capacity. To measure this variability, we use three-years worth of actual surgery durations with respect to six different surgical specialties, namely General Cardiology (CARD), Gastroenterology (GASTRO), Gynecology (GYN), Medicine (MED), Orthopedics (ORTH), and Urology (URO) (from Mannino et al. (2012) and Mannino et al. (2010)). Figure 1 presents the empirical and fitted distributions of actual surgery durations data for each specialty. This figure clearly illustrates significant variability in durations within and across surgery types. Furthermore, there is a wide range of possible probability distributions for modeling the variability (uncertainty) in surgery durations of each type, suggesting distributional ambiguity (i.e., uncertainty of probability distribution). Such uncertainty and ambiguity are hard to predict and model in advance when elective surgery is scheduled and the OR schedule is constructed.

Ignoring uncertainty and ambiguity may lead to devastating consequences, most notably unpredictable OR utilization, overtime, idle time, and can lead to surgery cancellation and thus sub-optimal quality of care (Carello and Lanzarone, 2014; Denton et al., 2010; Hof et al., 2017;

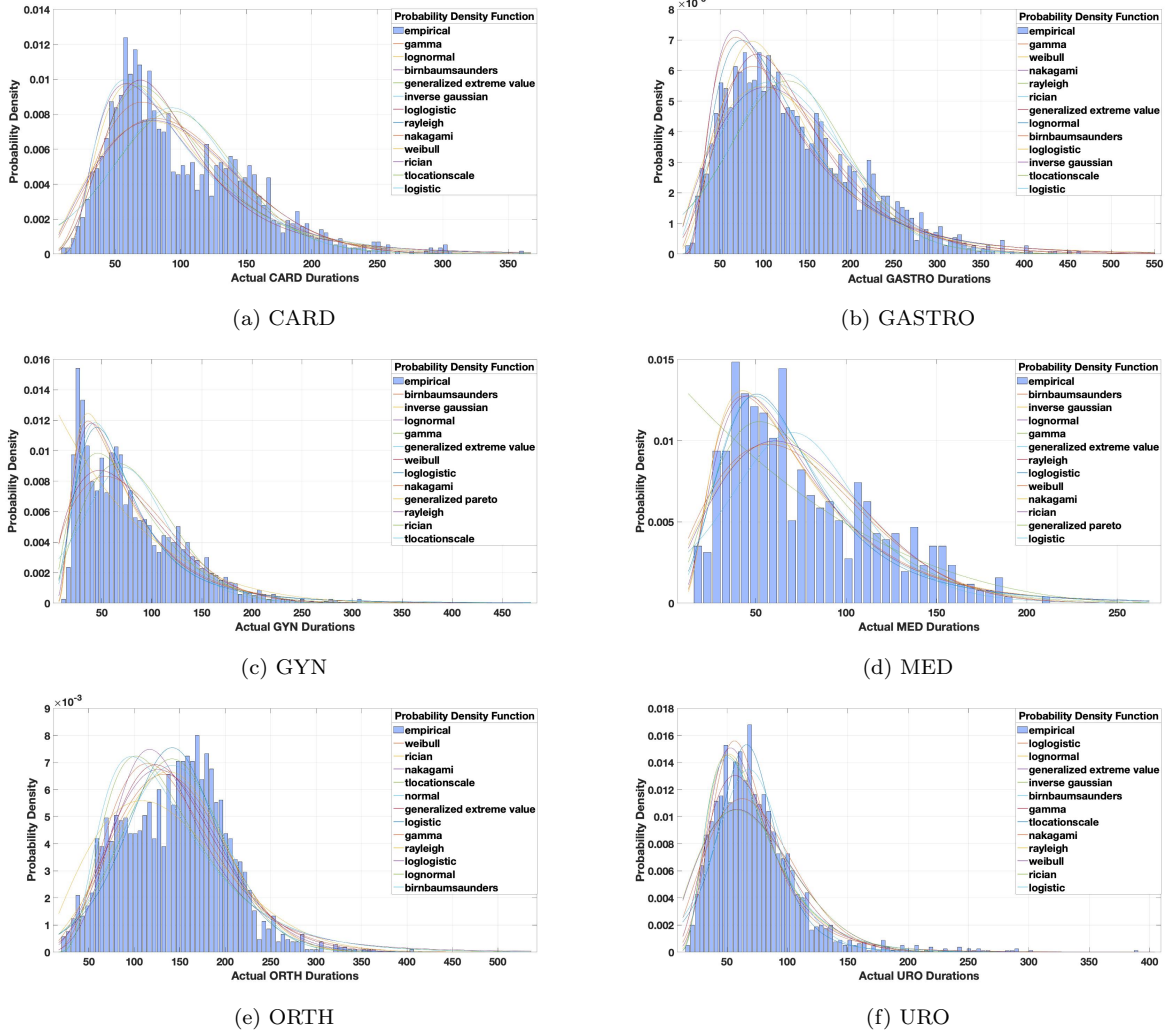


Figure 1: The empirical and fitted probability distributions for actual surgery duration.

May et al., 2011; Shehadeh et al., 2019; Wang et al., 2014, 2019; Xiao and Yoogalingam, 2021). To model uncertainty, most of the existing OR scheduling approaches assume that the exact distribution of surgery duration is known (often lognormal) and employ stochastic programming (SP) with sample average approximation (SAA) (see, e.g., Denton et al. (2007), Lamiri et al. (2008a), Shylo et al. (2013), Batun et al. (2011)). The SAA approaches assume that the hospital manager is risk-neutral and evaluates the overtime and idle time costs via the sample average, in which the approximation accuracy improves with the increase of sample size, but the computation becomes cumbersome as well (Birge and Louveaux, 2011; Shapiro et al., 2014).

SP remains the state-of-the-art approach to model uncertainty in healthcare scheduling and other application domains. However, the applicability of the SP approach is limited to the case in which we know the distribution of surgery durations of each type. In practice, however, it is unlikely that decision-makers have access to a sufficient amount of high-quality data to estimate

their probability distributions accurately (Viapiano and Ward, 2000; Keyvanshokoo et al., 2020; Shehadeh and Padman, 2022; Wang et al., 2019). Even when hospitals collect data on surgery durations, this data may not be readily available due to privacy issues.

In view of possible distributional ambiguity, if we solve a model with a data sample from a biased distribution (as in the SP approach), then the resulting (biased) optimal scheduling decisions may have a disappointing out-of-sample operational performance (e.g., excessive overtime) under the true probability distribution when the schedule is implemented in practice (Mohajerin Esfahani and Kuhn, 2018).

Alternatively, we can design a so-called *ambiguity set* (i.e., a family) of all distributions that possess certain partial information about the durations or are close enough to an empirical distribution. Then, we can formulate a distributionally robust optimization (DRO) problem that determines the optimal elective surgery assigning decisions that minimize the sum of patient-related costs plus the maximum expected cost associated with overtime and idle time of ORs over all distributions defined in the ambiguity set. Note that in DRO, the optimization is based on the worst-case distribution within the ambiguity set, i.e., the distribution is a decision variable (Rahimian and Mehrotra, 2019).

DRO has received significant attention recently in many application domains due to the following three primary benefits. First, DRO alleviates the unrealistic assumption of the decision-maker’s complete knowledge of distributions. Second, DRO approaches acknowledge the presence of distributional uncertainty. Thus, depending on the ambiguity set used, DRO solutions can anticipate the possibility of disappointing out-of-sample consequences. Third, several studies have proposed techniques to derive DRO models for real-life optimization problems that are more computationally tractable than their SP counterparts, and we have seen many successful applications of DRO to OR and surgery scheduling in recent years (see, e.g., Bansal et al. (2021); Keyvanshokoo et al. (2020); Jiang et al. (2017); Mak et al. (2014); Shehadeh et al. (2020); Shehadeh and Padman (2021); Wang et al. (2019) to name a few).

The computational tractability of a DRO model depends on the ambiguity set used (Rahimi and Gandomi, 2020). There are various techniques to construct the ambiguity set, and it could be based on moment information (Delage and Ye, 2010) or statistical measures such as the Wasserstein distance (Mohajerin Esfahani and Kuhn, 2018). Most of the early DRO approaches employ moment ambiguity sets, consisting of all distributions sharing certain moments, e.g., the first moment and support. Mean-support ambiguity sets often lead to computationally tractable reformulations (see, e.g., Jiang et al. (2017), Shehadeh et al. (2020), and Wang et al. (2019)). However, asymptotic properties cannot be guaranteed because moment-based ambiguity sets only incorporate descriptive statistics (i.e., moments information). New studies have shifted toward distance-based DRO approaches that define the ambiguity set using a distance metric (e.g., Wasserstein distance) to describe the deviation from a reference (often empirical) distribution. The main advantage of these approaches is that they enable observed (possibly small-size) data to be incorporated directly and

effectively in the optimization problem.

Despite the potential advantages of DRO, there are no moment-based, distance-based, or any other DRO models for elective surgery scheduling in flexible ORs. This inspires us to study a data-driven DRO approach that can use some partial information on the true underlying distribution of surgery duration from a small set of historical realizations and investigate the properties and performance of such an approach for this specific problem. In particular, we aim to investigate the value of the distance-based DRO approach in modeling uncertainty in the specific OR scheduling problem that we consider in this paper.

1.1. Contributions

In this paper, we address the distributional ambiguity of surgery durations using a data-driven distance-based DRO approach. First, we construct an ambiguity set of probability distributions, incorporating 1-Wasserstein distance and the empirical support set of surgery duration. We then define a data-driven *distributionally robust surgery assignment* (DSA) problem, which aim to determine optimal elective surgery assignment decisions to available surgical blocks over a planning horizon to minimize the sum of patient-related penalty costs (i.e., cost of performing or delaying elective surgery) plus the worst-case (i.e., maximum) expected cost associated with overtime and idle time of the ORs. We take the expectation overall distributions residing in the data-driven 1-Wasserstein ambiguity set. We summarize our main contributions as follows.

- To the best of our knowledge, and according to our literature review in Section 2, our paper is the first to propose a data-driven DRO model (denoted as W-DRO) with a 1-Wasserstein ambiguity set for elective surgery scheduling in flexible ORs. In addition, we derive a DRO model (denoted as M-DRO) based on a mean-range ambiguity set and compare the operational performance of the two DRO models, demonstrating where significant performance improvements could be gained with the proposed W-DRO model.
- Using DSA structural properties, we derive an equivalent mixed-integer linear programming (MILP) reformulation of our proposed min-max DRO model that can be implemented and efficiently solved using off-the-shelf optimization software, not requiring customized algorithmic development or parameter tuning, thereby enabling practitioners to use our MILP model. As argued by [Shehadeh et al. \(2019\)](#), such implementability *is necessary for an optimization-based decision support tool to gain wide adoption in OR and other healthcare systems that do not have ongoing access to support staff with optimization expertise*.
- To demonstrate the value of a DRO approach for DSA, we conduct an extensive numerical experiment using real-world surgery data. Specifically, our results demonstrate that our proposed W-DRO model (1) yields robust decisions with both small and large data size; (2) enjoys both asymptotic consistency and finite-data guarantees, (3) have a good out-of-sample

performance under perfect and misspecified distributional information, even when a small data set is used in the optimization; (4) is less conservative than the mean-support based DRO model; and (5) is computationally efficient with solution times sufficient for real-world implementation. In addition, we compare the flexible and dedicated ORs policies and show the pros and cons of each in terms of overtime, OR utilization, and access (measured by the number of scheduled surgeries under each policy).

- Finally, we extend our model to include the decisions of how many OR blocks to open. We drive an equivalent MILP of this extension and conduct an experiment demonstrating the superior performance of this extension over the classical SP approach.

1.2. Structure of the paper

The remainder of the paper is structured as follows. In Section 2, we review the relevant literature. In Section 3, we detail our problem setting. In Section 4, we present our SP model. In Section 5, we present and analyze our proposed DRO model. In Section 6, we extend our proposed DRO model to include the decisions of how many OR blocks to open. In Section 7, we present our numerical experiments and corresponding insights. Finally, we draw conclusions and discuss future directions in Section 8.

2. Relevant Literature

In this section, we review the literature most relevant to our problem. Specifically, we focus on studies that proposed stochastic optimization approaches for OR planning or surgery scheduling, with two competing classes: elective and emergency. For comprehensive surveys of OR and surgery planning, we refer to [Cardoen et al. \(2010\)](#), [Gartner and Padman \(2019\)](#), [Hof et al. \(2017\)](#), [May et al. \(2011\)](#), [Rahimi and Gandomi \(2020\)](#), [Samudra et al. \(2016\)](#), [Shehadeh and Padman \(2022\)](#), and [Zhu et al. \(2019\)](#). We refer to [Pinedo \(2016\)](#) for a detailed survey of a wide range of scheduling problems, including their theory, algorithms, and applications. We also recognize that there is prior literature that employs simulation to study elective surgery duration and emergency arrivals (see, e.g., [Wang et al. \(2016\)](#) and references therein), similar objectives to us in deterministic models (see, e.g., [Fei et al. \(2009\)](#) and references therein). Our problem is somewhat similar to Newsvendor problem, in which one must find a trade-off between reserving too much versus too little OR time for emergency surgeries (see, e.g., [Choi \(2012\)](#) and [Nagarajan and Shechter \(2014\)](#)). In addition to solving the optimization model directly, some studies propose algorithmic and decomposition approaches, including logic-based Benders decomposition approaches (see, e.g., the recent work of [Guo et al. \(2021\)](#) and references therein). We refer to the surveys mentioned above and references therein for other tactical and strategic decisions. For example, [Anjomshoa et al. \(2018\)](#) propose an approach for tactical planning and patient selection for elective surgeries motivated by a case study in a local hospital in Melbourne.

Most of the existing stochastic optimization approaches assume that elective and emergency surgeries have dedicated ORs and focus on elective surgery scheduling (see, e.g., [Anjomshoa et al. \(2018\)](#); [Batun et al. \(2011\)](#); [Denton et al. \(2007\)](#); [Deng et al. \(2019\)](#); [Guo et al. \(2021\)](#); [Gul et al. \(2011\)](#); [Neyshabouri and Berg \(2017\)](#); [Shehadeh et al. \(2019\)](#); [Zhu et al. \(2019\)](#), and the references therein). Dedicating ORs to emergency cases has the apparent advantage of having OR space readily available for emergency cases, which can be handled promptly without impacting the scheduled elective surgeries. However, the drawback is the low utilization of costly OR resources when emergency cases do not materialize, or emergency arrivals are lower than expected ([Xiao and Yoogalingam, 2021](#)). Herein, we focus on flexible OR where emergency patients are accommodated in the existing elective surgery schedule. However, our DRO model can be used for elective surgery scheduling in ORs where emergency cases have dedicated blocks. In addition, we focus on two-stage models with recourse. Recourse models result when some decisions must be fixed before information relevant to uncertainty is available (e.g., in our problem surgery assignment to OR), while some decisions can be delayed until this information is available (e.g., actual schedule on the day of service and the associated overtime and idle time of the ORs). Two-stage models with recourse have been widely employed to model and solve stochastic (healthcare) scheduling problems ([Ahmadi-Javid et al., 2017](#); [Shehadeh and Padman, 2022](#)).

[Gerchak et al. \(1996\)](#), [Lamiri et al. \(2008b\)](#), [Lamiri et al. \(2008a\)](#) proposed SP models for the elective surgery planning problem in flexible OR shared between elective and emergency surgeries. The planning problem in [Lamiri et al. \(2008b\)](#) is closely related to our work, and consists of assigning elective patients to different ORs over a planning horizon in order to minimize elective-patient-related costs (i.e., cost of performing or delaying surgery to the next planning horizon) and the expected operating rooms’ utilization (overtime and idle time) costs. The expectation in [Lamiri et al. \(2008b\)](#)’s model is taken with respect to the assumed known distribution of surgery duration. To solve their SP, [Lamiri et al. \(2008b\)](#) proposed a column generation approach, which can produce near-optimal solutions in a short computation time for problems with 12 operating rooms. In contrast, [Lamiri et al. \(2008a\)](#) employed Monte Carlo Optimization to solve a sample-average approximation of their SP.

These SP models assume that the hospital manager is risk-neutral and can fully characterize the probability distribution of surgery duration. In practice, it is implausible that hospital managers and other decision-makers have sufficient high-quality data to infer the distribution of random parameters, especially in healthcare ([Macario, 2010](#); [Viapiano and Ward, 2000](#); [Wang et al., 2019](#)). Moreover, the distribution of random parameters is often ambiguous ([Mak et al., 2014](#); [Qi, 2017](#); [Wang et al., 2019](#)). As pointed out by [Wang et al. \(2019\)](#), various decisions theory, empirical, and neural studies show that decision makers tend to be ambiguity-averse ([Halevy, 2007](#); [Hsu et al., 2005](#)). This is especially the case in healthcare and OR practice due to the high cost of care and indirect costs resulting from adverse outcomes such as overtime, leading to surgery cancellation and

OR staff fatigue and, consequently, sub-optimal care (Carello and Lanzarone, 2014; Denton et al., 2010; Hof et al., 2017; May et al., 2011; Shehadeh et al., 2019; Wang et al., 2019, 2014; Xiao and Yoogalingam, 2021).

Various robust approaches have been proposed to model the risk-averse nature of decision-makers and uncertain parameters based on partial information of the distribution. The classical robust optimization (RO) approach model the uncertain parameters based on an uncertainty set of possible outcomes with some structure (e.g., ellipsoid or polyhedron, see, Bertsimas and Sim (2004), Ben-Tal et al. (2015) Soyster (1973), Rowse (2015), Neyshabouri and Berg (2017)). In RO, optimization is based on the worst-case scenario within the uncertainty set. By optimizing based on the worst-case scenario, classical RO approaches may yield over-conservative solutions and sub-optimal decisions for the other more likely to be observed scenarios (Chen et al., 2020; Delage and Saif, 2021; Wang et al., 2019).

In contrast to RO, the DRO approach assumes that the distribution of random parameters resides in a so-called ambiguity set, i.e., a family of all possible distributions that share some common properties of random parameters. Accordingly, optimization is based on the distributions residing within the ambiguity set, i.e., the probability distribution of a random parameter is a decision variable in DRO. Most early DRO studies employ the moment ambiguity set, which consists of all distributions sharing particular moments (e.g., the first and second moments). Notably, Wang et al. (2019) is the first (and only study) to employ DRO for elective surgery scheduling and surgery block allocation. Using an ambiguity set based on mean, mean absolute deviation, and support of surgery duration, Wang et al. (2019) propose a DRO model that determines the number of ORs to open and assign the surgeries in a daily listing to the open ORs. Wang et al. (2019)' DRO model aims to minimize OR opening costs and expected OR overtime cost. Wang et al. (2019) did not consider emergency surgery and ignore idle time costs, which is an important metric of OR efficiency and utilization (Giroto et al., 2010; Jebali and Diabat, 2015; Liu et al., 2019; Shehadeh and Padman, 2021).

Moment-support-based DRO approaches often lead to tractable reformulations, but they known to have weak convergence properties. New studies have shifted toward distance-based DRO approaches that construct ambiguity sets in the vicinity of a reference (empirical) distribution. The distance-based DRO models' main advantage is that they enable possibly small-size data to be incorporated directly and effectively in the optimization problem. In this paper, we construct a 1-Wasserstein distance-based ambiguity set of all probability distributions of surgery durations. Various authors have shown that 1-Wasserstein ambiguity admits tractable reformulation in many real-world applications (see, e.g., Jiang et al. (2019) Saif and Delage (2021)). Then, we propose a DRO model for elective surgery planning in flexible ORs. The objective is to minimize the sum of patient-related penalty costs (i.e., cost of performing or delaying elective surgery) and the expectation of OR overtime and idle time cost. We take the expectation over all distributions residing in

the data-driven 1-Wasserstein ambiguity set. We derive an equivalent MILP reformulation of our min-max model. Our MILP can be implemented and solved directly with standard optimization software packages, not requiring customized algorithmic development or tuning, thereby enabling practitioners to use our MILP.

We use our MILP and real-world surgery data to conduct extensive computational experiments, demonstrating our approach’s advantages. Finally, we extend our model to include the decision of how many OR blocks to open, derive an equivalent MILP of this extension, and present an experiment demonstrating the performance of this extension. To the best of our knowledge, and according to recent survey papers on OR and surgery scheduling (Zhu et al., 2019), our paper is the first to propose and analyze DRO approaches for this specific surgery and OR planning problem.

3. Problem Description

Let us now start to introduce our problem. We consider a hospital with B available surgical blocks (or ORs) over an arbitrary planning horizon of T days (e.g., a week). Each surgical block $b \in [B] := \{1 \dots, B\}$ is assigned to a particular OR, has a pre-allocated length of time \mathcal{L}_b , and is dedicated to only one type of surgical specialty. For any $b \in [B]$, the time length \mathcal{L}_b is typically long enough so that multiple surgeries can be performed during that time. Note that there can be multiple blocks for the same specialty during a cycle of the OR schedule. The OR capacity is shared among two competing surgery classes: a known number I of elective surgeries that are to be planned in advance, and random emergency surgeries that must be performed on the day of arrival.

Each elective surgery $i \in [I]$ has a surgery type and can be assigned to any of the blocks dedicated to the corresponding surgery type during the planning horizon. Associated with each elective case, there is a cost for performing (scheduling) and for rejecting the surgery (i.e., cost of delaying surgery to the next planning horizon). Let’s assume that patients who are not assigned to any surgery block during the current planning horizon are assigned to a dummy block $b' \notin [B]$. Furthermore, let $c_{i,b}$ and $c_{i,b'}$ represent the costs of performing and postponing surgery $i \in [I]$, respectively. Then, it is reasonable to assume that $c_{i,b'} > c_{i,b}$, for all $b \in [B]$ (a common assumption in the elective surgery scheduling literature, see, e.g., Jebali and Diabat (2015), Min and Yih (2010), Shehadeh and Padman (2021), and Zhang et al. (2019)). Moreover, the cost of assigning a surgery to a block is block-dependent to take into account the surgery’s waiting time on the list and clinical priority, among other potential factors (Jebali and Diabat, 2015; Lamiri et al., 2008b,a). However, determining this cost is out of the scope of this paper.

Elective surgery duration (d_i , $i \in [I]$) are random and depend on surgery type. As defined in Lamiri et al. (2008b) and Lamiri et al. (2008a), the total random surgery duration or capacity, e_b , needed for emergency cases in each surgical block $b \in [B]$ is random. The probability distributions

of these random parameters are ambiguous (unknown).

Given a waiting list of I elective surgeries and their types, we are interested in determining a plan that specifies the number (or subset) of elective surgeries to schedule in each surgery block (equivalently, surgery assignments to the available surgery blocks). The plan's quality is a function of costs related to performing (scheduling) or not performing (rejecting) elective surgeries and costs related to OR overtime and total idle time. Overtime occurs when surgeries assigned to block b are not completed within $[0, \mathcal{L}_b]$. Idle time occurs when surgeries assigned to block b are completed before \mathcal{L}_b . Note that overtime and idle time are random (second-stage) performance metrics. That is, they are observed after the realization of random parameters while the planned schedule is being executed. In contrast, costs related to scheduling or not scheduling an elective surgery are fixed planning (first-stage) costs incurred when the OR schedule is constructed. We make the following assumption based on prior studies:

- A1. The set of elective patients, their surgery types, and priorities are known with certainty (a standard assumption in the elective surgery scheduling literature, see, e.g., [Jebali and Diabat \(2015\)](#), [Lamiri et al. \(2008b\)](#), [Lamiri et al. \(2008a\)](#), [Min and Yih \(2010\)](#), and [Neyshabouri and Berg \(2017\)](#)).
- A2. The planning horizon T is an integer multiple of the surgery schedule cycle length.
- A3. Surgical blocks and their assignments to operating rooms during a cycle of surgery schedule are previously determined ([Lamiri et al., 2008b,a](#); [Neyshabouri and Berg, 2017](#)).
- A4. For ease of modeling, we assume that patients who are not assigned to any surgery block during the current planning horizon are assigned to a dummy block b' ([Lamiri et al., 2008b,a](#)).
- A5. Patients of the same type have identical probability distributions of surgery duration (a common assumption in prior studies).
- A6. We assume that we know the lower $(\underline{d}, \underline{e})$ and upper (\bar{d}, \bar{e}) bounds on \mathbf{d} and \mathbf{e} (this is a mild and realistic assumption that holds in healthcare scheduling and mimic the OR practice). Mathematically, we consider support $\mathcal{S} = \mathcal{S}^d \times \mathcal{S}^e$, where \mathcal{S}^d and \mathcal{S}^e are respectively the supports of random parameters d and e , and defined as follows:

$$\mathcal{S}^d := \left\{ \mathbf{d} \geq 0 : \underline{d}_i \leq d_i \leq \bar{d}_i, \forall i \in [I] \right\}, \quad (1)$$

$$\mathcal{S}^e := \left\{ \mathbf{e} \geq 0 : \underline{e}_b \leq e_b \leq \bar{e}_b, \forall b \in [B] \right\}. \quad (2)$$

For modeling convenience, prior literature suggests that one can avoid excessive overtime by assigning a high weight (c_b^o) on overtime in the objective and/or setting $\mathcal{L}_b = (b - \Delta)$ with $\Delta > 0$. Our model can accommodate such constraints.

Additional Notation: For $a, b \in \mathbb{Z}$, we define $[a] := \{1, 2, \dots, a\}$ and $[a, b]_{\mathbb{Z}} := \{c \in \mathbb{Z} : a \leq c \leq b\}$, i.e., $[a, b]$ represent the set of running integer indices $\{a, a+1, a+2, \dots, b\}$. The abbreviations

Table 1: Notation.

Parameters and sets	
I	number, or set, of surgery
B	number, or set, of OR blocks
$c_{i,b}$	cost of assigning surgery i to block b
c_b^o	unit overtime cost for each block b
c_b^g	unit idle cost for each block b
d_i	duration of surgery i
\underline{d}/\bar{d}_i	lower/upper bounds of d_i
e	capacity needed for emergency cases
$\underline{e}_i/\bar{e}_i$	lower/upper bounds of e_i
\mathcal{L}_b	capacity or planned length of block b
First-stage decision variables	
$y_{i,b}$	$\begin{cases} 1, & \text{surgery } i \text{ is assigned to OR block } b \\ 0, & \text{otherwise.} \end{cases}$
Second-stage decision variables	
o_b	continuous decision variable capturing overtime in block b
g_b	continuous decision variable capturing idle time in block b

“w.l.o.g.” and “w.l.o.o.” respectively represent “without loss of generality” and “without loss of optimality.” We use boldface notation to denote vectors, e.g., $\mathbf{d} := [d_1, d_2, \dots, d_N]^\top$. A complete listing of the parameters and decision variables can be found in Table 1.

4. Stochastic Programming Model

In this section, we present a two-stage SP formulation of the problem that assumes that the probability distributions of surgery durations are known. First, let us introduce the variables and constraints defining the first-stage of this SP model. For each $i \in [I]$ and $b \in [B] \cup \{b'\}$, we define a binary decision variable $y_{i,b}$ that equals 1 if we assign elective surgery i to block b , and is zero otherwise. We define feasible region \mathcal{Y} of \mathbf{y} in (3) such that each elective surgery is assigned to one surgical block with the obvious convention that a surgery assigned to the dummy block b' (i.e., $y_{i,b'} = 1$) is postponed to the next planning period (i.e., rejected in the current planning period).

$$\mathcal{Y} := \left\{ \mathbf{y} : \sum_{b \in [B] \cup \{b'\}} y_{i,b} = 1, \forall i \in [I], y_{i,b} \in \{0, 1\}, \forall i \in [I], b \in [B] \cup \{b'\} \right\}. \quad (3)$$

Let us now introduce the parameters and variables defining our second-stage problem. For each $i \in [I]$, we let d_i represents the duration of elective surgery i . For all $b \in [B]$, we define e_b as the duration of emergency surgery in block b (i.e., capacity needed for emergency cases as defined in Lamiri et al. (2008b, 2009)). For all $b \in [B]$, we define nonnegative continuous decision variables o_b and g_b respectively to represent block b overtime and idle time. For all $b \in [B]$, we define c_b^o and c_b^g

as the nonnegative unit penalty costs of overtime and idle time, respectively. Finally, we define \mathbb{P}_ξ as the probability distribution of $\xi = [\mathbf{d}, \mathbf{e}]^\top$, and $\mathbb{E}_{\mathbb{P}_\xi}$ as the expectation under distribution \mathbb{P}_ξ .

The SP model can now be stated as follows (Lamiri et al., 2008a, 2009):

$$(\text{SP}) \quad \hat{Z} = \min_{\mathbf{y} \in \mathcal{Y}} \left\{ \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \mathbb{E}_{\mathbb{P}_\xi} [f(\mathbf{y}, \xi)] \right\}, \quad (4)$$

where for a feasible $\mathbf{y} \in \mathcal{Y}$ and a joint realization of uncertain parameters $\xi = [\mathbf{d}, \mathbf{e}]^\top$, we compute the random cost associated with overtime and idle time using the following linear program (LP).

$$f(\mathbf{y}, \xi) := \min_{\mathbf{o}, \mathbf{g}} \sum_{b \in B} (c_b^o o_b + c_b^g g_b) \quad (5a)$$

$$\text{s.t. } o_b - g_b = \sum_{i \in I} d_i y_{i,b} + e_b - \mathcal{L}_b, \quad \forall b \in [B], \quad (5b)$$

$$(o_b, g_b) \geq 0, \quad \forall b \in [B]. \quad (5c)$$

Formulation (4) aim to find first-stage scheduling decisions $\mathbf{y} \in \mathcal{Y}$ that minimize the first-stage cost (i.e., costs related to performing/scheduling or not performing/rejecting elective surgeries) plus the expected cost associated with overtime and idle time. Constraint (5b) yield either block overtime or idle time based on the durations of surgeries preformed in this block. Finally, constraints (5c) specify feasible ranges of the second-stage decision variables. It is easy to verify that the second-stage (recourse) formulation (5) is feasible for any feasible first-stage decisions. Thus, we have a relatively complete recourse.

5. DRO model over Wasserstein ambiguity set (W-DRO)

In this section, we present our proposed DRO models for the DSA that does not assume that the probability distributions of surgery durations are known. Specifically, we consider the case that the joint distribution \mathbb{P}_ξ of $\xi := [\mathbf{d}, \mathbf{e}]^\top$ may be observed via a possibly small finite set $\{\hat{\xi}^1, \dots, \hat{\xi}^N\}$ of N samples, which may come from the limited historical realizations or a reference empirical distribution. Accordingly, we construct an ambiguity set based on 1-Wasserstein distance (i.e., we use ℓ_1 -norm in the definition of Wasserstein metric), which allows us to derive a tractable model. Then, we formulate our DRO model (denoted as W-DRO) using this ambiguity set.

First, let us define the 1-Wasserstein distance. Suppose that random vectors ξ_1 and ξ_2 follow \mathbb{Q}_1 and \mathbb{Q}_2 , respectively, where distributions \mathbb{Q}_1 and \mathbb{Q}_2 are defined over the common support \mathcal{S} . The 1-Wasserstein distance $\text{dist}(\mathbb{Q}_1, \mathbb{Q}_2)$ between \mathbb{Q}_1 and \mathbb{Q}_2 is the cost of an optimal transportation plan for moving the probability mas in one so it becomes identical to the other, where the cost of moving from ξ_1 to ξ_2 equals to the norm $\|\xi_1 - \xi_2\|$. Mathematically,

$$\text{dist}(\mathbb{Q}_1, \mathbb{Q}_2) := \inf_{\Pi \in \mathcal{P}(\mathbb{Q}_1, \mathbb{Q}_2)} \left\{ \int_{\mathcal{S}} \|\xi_1 - \xi_2\| \Pi(d\xi_1, d\xi_2) \mid \begin{array}{l} \Pi \text{ is a joint distribution of } \xi_1 \text{ and } \xi_2 \\ \text{with marginals } \mathbb{Q}_1 \text{ and } \mathbb{Q}_1, \text{ respectively} \end{array} \right\} \quad (6)$$

where $\mathcal{P}(\mathbb{Q}_1, \mathbb{Q}_2)$ is the set of all joint distributions of (ξ_1, ξ_2) supported on \mathcal{S} with marginals \mathbb{Q}_1 and \mathbb{Q}_2 . Accordingly, we construct the following 1-Wasserstein ambiguity set:

$$\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon) = \left\{ \mathbb{Q}_\xi \in \mathcal{P}(\mathcal{S}) \mid \text{dist}(\mathbb{Q}_\xi, \hat{\mathbb{P}}_\xi^N) \leq \epsilon \right\}, \quad (7)$$

where $\mathcal{P}(\mathcal{S})$ is the set of all joint probability distributions supported on \mathcal{S} , $\hat{\mathbb{P}}_\xi^N = \frac{1}{N} \sum_{n=1}^N \delta_{\xi^n}$ is the empirical distribution of ξ based on N i.i.d samples, and $\epsilon > 0$ is the radius of the ambiguity set. The set $\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)$ represent a Wasserstein ball of radius ϵ centered at the empirical distribution $\hat{\mathbb{P}}_\xi^N$. Note that we do not make any assumptions on elective and emergency surgeries durations (i.e., ξ). Thus, they can be dependent, correlated, or independent.

Using the ambiguity set $\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)$, we formulate our W-DRO model for DSA as follows:

$$(\text{W-DRO}) \quad \hat{Z}(N, \epsilon) = \min_{\mathbf{y} \in \mathcal{Y}} \left\{ \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \sup_{\mathbb{P}_\xi \in \mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)} \mathbb{E}_{\mathbb{P}_\xi} [f(\mathbf{y}, \xi)] \right\}. \quad (8)$$

Formulation (5) finds first-stage decisions $\mathbf{y} \in \mathcal{Y}$ that minimize the first-stage cost plus the maximum expectation of the second-stage cost, where the expectation is taken over all distributions residing in $\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)$.

We often seek asymptotic consistency in data-driven applications such as W-DRO. In particular, one expects that as the sample size N increases to infinity, the optimal value of problem (5) $\hat{Z}(N, \epsilon)$ converges to Z^* (the optimal value of the SP model in (4) with perfect knowledge of \mathbb{P}_ξ). Accordingly, as N increases, an optimal solution \mathbf{y} to W-DRO tends to the optimal solution of problem (4). Additionally, if $\hat{Z}(N, \epsilon) > Z^*$ almost surely, then W-DRO provides a safe upper bound guarantee on the expected total cost with any finite data size N . Assumption A6 indicates that the support set \mathcal{S} is non-empty, convex, and compact. As such, we can make use of existing theory in establishing the asymptomatic consistency and finite sample guarantee of our W-DRO model. In particular, given assumption A6, lemma 1 in [Jiang et al. \(2019\)](#) and Theorem 2 in [Fournier and Guillin \(2015\)](#) assure that the ℓ_p -Wasserstein ambiguity set incorporates the true distribution \mathbb{P}_ξ of ξ with high confidence. Theorem 1 in [Jiang et al. \(2019\)](#) and Theorem 3.6 in [Mohajerin Esfahani and Kuhn \(2018\)](#) assure asymptotic consistency, and Theorem 2 in [Jiang et al. \(2019\)](#) and Theorem 3.5 in [Mohajerin Esfahani and Kuhn \(2018\)](#) assure finite-data guarantee of W-DRO. We refer to these papers for detailed proofs and discussion on these results. For completeness, below, we provided the adapted Lemma 1, Theorem 1, and Theorem 2 and provide customized proofs to fit our problem in [Appendix B–Appendix D](#).

Lemma 1. (Adapted from [Jiang et al. \(2019\)](#) and Theorem 2 in [Fournier and Guillin \(2015\)](#)). *Suppose that Assumption A6 holds. Then there exist none-negative constants c_1 and c_2 such that, for all $N \geq 1$ and $\beta \in (0, \min\{1, c_1\})$,*

$$\mathbb{P}^N \{ \text{dist}(\mathbb{P}_\xi, \hat{\mathbb{P}}_\xi^N) \leq \epsilon_N(\beta) \} \geq 1 - \beta,$$

where \mathbb{P}_ξ^N represents the product measure of N copies of \mathbb{P}_ξ and $\epsilon_N(\beta) = \left\lceil \frac{\log(c_1\beta^{-1})}{c_2N} \right\rceil^{\frac{1}{\max\{3p,n\}}}$ (see [Appendix B](#) for a proof).

As detailed in [Jiang et al. \(2019\)](#), Lemma 1 assures that the p -Wasserstein ambiguity (1-Wasserstein ambiguity in our problem) contains the true distribution \mathbb{P}_ξ with high confidence.

Theorem 1. (Asymptotic consistency, adapted from [Jiang et al. \(2019\)](#) and Theorem 3.6 of [Mohajerin Esfahani and Kuhn \(2018\)](#)). *Suppose that Assumption A6 holds. Consider a sequence of confidence levels $\{\beta_N\}_{N \in \mathbb{R}}$ such that $\sum_{N=1}^\infty \beta_N < \infty$ and $\lim_{N \rightarrow \infty} \epsilon_N(\beta_N) = 0$, and let $(\hat{y}(N, \epsilon_N(\beta_N)))$ represents an optimal solution to W-DRO with the ambiguity set $\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon_N(\beta_N))$. Then, \mathbb{P}_ξ^∞ -almost surely we have $\hat{Z}(N, \epsilon_N(\beta_N)) \rightarrow Z^*$ as $N \rightarrow \infty$. In addition, any accumulation points of $\{\hat{y}(N, \epsilon_N(\beta_N))\}_{N \in \mathbb{N}}$ is an optimal solution of (4) \mathbb{P}_ξ^∞ -almost surely (see [Appendix C](#) for a proof).*

Theorem 2. (Finite-data guarantee, adapted from [Jiang et al. \(2019\)](#) and Theorem 3.5 in [Mohajerin Esfahani and Kuhn \(2018\)](#)). *For any $\beta \in (0, 1)$, let $\hat{y}(N, \epsilon_N(\beta_N))$, represent an optimal solution of W-DRO with ambiguity set $\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon_N(\beta_N))$. Then,*

$$\mathbb{P}_\xi^N \{ \mathbb{E}_{\mathbb{P}_\xi} [f(\hat{y}(N, \epsilon_N(\beta_N), \xi))] \leq \hat{Z}(N, \epsilon_N(\beta_N)) \} \geq 1 - \beta.$$

(see [Appendix D](#) for a proof).

5.1. Reformulation of the W-DRO model

Recall that $f(\cdot)$ is defined by a minimization problem. Thus, in (5), we have an inner max-min problem. As such, it is not straightforward to solve the W-DRO model in (5). In this section, we derive an equivalent reformulation of the W-DRO model that is solvable. First, in Proposition 1, we present an equivalent dual formulation of the inner maximization problem in (5) (see [Appendix A](#) for a detailed proof).

Proposition 1. *For fixed $\mathbf{y} \in \mathcal{Y}$, problem $\sup_{\mathbb{P}_\xi \in \mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)} \mathbb{E}_{\mathbb{P}_\xi} [f(\mathbf{y}, \xi)]$ in (5) is equivalent to*

$$\inf_{\rho} \left\{ \epsilon \rho + \frac{1}{N} \sum_{n=1}^N \sup_{\xi \in \mathcal{S}} \{ f(\mathbf{y}, \xi) - \rho \|\xi - \hat{\xi}^n\| \} \right\} \quad (9a)$$

$$s.t. \ \rho \geq 0. \quad (9b)$$

Again, formulation (9) is not directly solvable using standard techniques. Fortunately, given that the supports of \mathbf{d} and \mathbf{e} are rectangular and finite (Assumption A6), we next show that we can reformulate each of the inner maximization problems in (9) as a linear program (LP) for fixed $\rho \geq 0$ and $\mathbf{y} \in \mathcal{Y}$. First, we observe that for each $\mathbf{y} \in \mathcal{Y}$, the recourse problem can be decomposed by each block, i.e., $f(\mathbf{y}, \xi) = \sum_{b \in B} f_b(y, \xi)$, where for each $b \in B$:

$$f_b(y, \xi) := \min_{o_b, g_b} \left(c_b^o o_b + c_b^g o_g \right) \quad (10a)$$

$$\text{s.t. } o_b - g_b = \sum_{i \in I} d_i y_{i,b} + e_b - \mathcal{L}_b, \quad (10b)$$

$$(o_b, g_b) \geq 0. \quad (10c)$$

Let β_b be the dual variables associated with constraint (10c). The dual of the LP in (10) is as follows

$$f_b(\mathbf{y}, \xi) := \max_{\beta} \left(\sum_{i \in I} d_i y_{i,b} + e_b - \mathcal{L}_b \right) \beta_b \quad (11a)$$

$$\text{s.t. } \mathcal{B} := \{-c_b^g \leq \beta_b \leq c_b^o\}. \quad (11b)$$

Next, for fixed $\mathbf{y} \in \mathcal{Y}$ and $\rho \geq 0$, we define $Q^n(\rho, \mathbf{y}) = \sup_{\xi \in \mathcal{S}} \{f(\mathbf{y}, \xi) - \rho|\xi - \hat{\xi}^n|\}$, for all $n = 1, \dots, N$. Given the dual formulation of $f_b(y, \xi)$ with feasible region $\mathcal{B} := \{-c_b^g \leq \beta_b \leq c_b^o\}$, we write the problem of computing $Q_b^n(\rho, \mathbf{y})$ as follows for each $n \in [N]$ and $b \in [B]$.

$$Q_b^n(\rho, \mathbf{y}) = \sup_{(\mathbf{d}, \mathbf{e}) \in \mathcal{S}} \sup_{\beta \in \mathcal{B}} \left\{ \left(\sum_{i \in I} d_i y_{i,b} + e_b - \mathcal{L}_b \right) \beta_b - \rho \sum_{i \in I} |d_i - \hat{d}_i^n| y_{i,b} - \rho |e_b - \hat{e}_b^n| \right\} \quad (12a)$$

$$\equiv \max_{(\mathbf{d}, \mathbf{e}) \in \mathcal{S}} \max_{\beta \in \mathcal{B}} \left\{ \sum_{i \in I} [d_i \beta_b - \rho |d_i - \hat{d}_i^n|] y_{i,b} + [e_b \beta_b - \rho |e_b - \hat{e}_b^n|] - \mathcal{L}_b \beta_b \right\}. \quad (12b)$$

It is easy to verify that for fixed $\mathbf{y} \in \mathcal{Y}$, $\mathbf{d} \in [\underline{d}, \bar{d}]$, $\mathbf{e} \in [\underline{e}, \bar{e}]$, function $\max_{\beta \in \mathcal{B}} \{(\sum_{i \in I} d_i y_{i,b} + e_b - \mathcal{L}_b) \beta_b\}$ in (12) is convex in variable β_b . Hence, problem $\max_{\beta_b \in \mathcal{B}} \{\cdot\}$ is a convex maximization problem. It follows from the fundamental convex analysis that there exists an optimal solution β_b^* to problem $\max_{\beta \in \mathcal{B}} \{\cdot\}$ at one of the extreme points $\hat{\beta}_b$ of polyhedron $\mathcal{B} := \{-c_b^g \leq \beta_b \leq c_b^o\}$. In any extreme point, constraint $-c_b^g \leq \hat{\beta}_b \leq c_b^o$ is binding at either the lower bound or upper bound. Thus, given that $\mathbf{d} \in [\underline{d}, \bar{d}]$ and $\mathbf{e} \in [\underline{e}, \bar{e}]$, it is easy to verify that the optimal objective value of problem (12) is equivalent to

$$\eta_b^{n*} = \max \left\{ \begin{array}{l} \sum_{i \in I} [\bar{d}_i c_b^o - \rho |\bar{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b c_b^o - \rho |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b c_b^o \\ \sum_{i \in I} [\underline{d}_i c_b^o - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b c_b^o - \rho |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b c_b^o \\ \sum_{i \in I} [\bar{d}_i (-c_b^g) - \rho |\bar{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b (-c_b^g) - \rho |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b (-c_b^g) \\ \sum_{i \in I} [\underline{d}_i (-c_b^g) - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b (-c_b^g) - \rho |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b (-c_b^g) \\ \sum_{i \in I} [\underline{d}_i c_b^o - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b c_b^o - \rho |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b c_b^o \\ \sum_{i \in I} [\underline{d}_i c_b^o - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b c_b^o - \rho |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b c_b^o \\ \sum_{i \in I} [\underline{d}_i (-c_b^g) - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b (-c_b^g) - \rho |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b (-c_b^g) \\ \sum_{i \in I} [\underline{d}_i (-c_b^g) - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b (-c_b^g) - \rho |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b (-c_b^g) \\ \sum_{i \in I} \hat{d}_i^n c_b^o y_{i,b} + \hat{e}_b^n c_b^o - \mathcal{L}_b c_b^o \\ \sum_{i \in I} \hat{d}_i^n (-c_b^g) y_{i,b} + \hat{e}_b^n (-c_b^g) - \mathcal{L}_b (-c_b^g) \end{array} \right\}. \quad (13)$$

Accordingly, we can formulate problem (12) as the following linear program for a fixed $\mathbf{y} \in \mathcal{Y}$

$$Q_b^n(\rho, \mathbf{y}) = \min \eta_b^n \quad (14a)$$

$$\eta_b^n \geq \sum_{i \in I} [\bar{d}_i c_b^\circ - \rho |\bar{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b c_b^\circ - \rho |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b c_b^\circ, \quad (14b)$$

$$\eta_b^n \geq \sum_{i \in I} [\underline{d}_i c_b^\circ - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b c_b^\circ - \rho |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b c_b^\circ, \quad (14c)$$

$$\eta_b^n \geq \sum_{i \in I} [\bar{d}_i(-c_b^g) - \rho |\bar{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b(-c_b^g) - \rho |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b(-c_b^g), \quad (14d)$$

$$\eta_b^n \geq \sum_{i \in I} [\underline{d}_i(-c_b^g) - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b(-c_b^g) - \rho |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b(-c_b^g), \quad (14e)$$

$$\eta_b^n \geq \sum_{i \in I} [\underline{d}_i c_b^\circ - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b c_b^\circ - \rho |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b c_b^\circ, \quad (14f)$$

$$\eta_b^n \geq \sum_{i \in I} [\bar{d}_i c_b^\circ - \rho |\bar{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b c_b^\circ - \rho |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b c_b^\circ, \quad (14g)$$

$$\eta_b^n \geq \sum_{i \in I} [\underline{d}_i(-c_b^g) - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b(-c_b^g) - \rho |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b(-c_b^g), \quad (14h)$$

$$\eta_b^n \geq \sum_{i \in I} [\bar{d}_i(-c_b^g) - \rho |\bar{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b(-c_b^g) - \rho |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b(-c_b^g), \quad (14i)$$

$$\eta_b^n \geq \sum_{i \in I} \hat{d}_i^n c_b^\circ y_{i,b} + \hat{e}_b^n c_b^\circ - \mathcal{L}_b c_b^\circ, \quad (14j)$$

$$\eta_b^n \geq \sum_{i \in I} \hat{d}_i^n(-c_b^g) y_{i,b} + \hat{e}_b^n(-c_b^g) - \mathcal{L}_b(-c_b^g). \quad (14k)$$

Summing over $b \in [B]$ and $n \in [N]$ and combining $Q_b^n(\rho, \mathbf{y})$ in the form of (14) with the outer minimization problem in (9), we derive the following equivalent reformulation of (9).

$$\inf_{\rho \geq 0, \boldsymbol{\eta}} \left\{ \epsilon^p \rho + \frac{1}{N} \sum_{n=1}^N \eta_b^n \right\} \quad (15a)$$

$$\text{s.t. } (14b) - (14k). \quad (15b)$$

Combining the inner problem in the form of (15) with the outer minimization problem in (8), we derive the following equivalent mixed-integer non-linear programming (MINLP) reformulation of the DSA model in (8).

$$\hat{Z}(N, \epsilon) = \min \left\{ \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \epsilon \rho + \frac{1}{N} \sum_{n=1}^N \eta_b^n \right\} \quad (16a)$$

$$\text{s.t. } \mathbf{y} \in \mathcal{Y}, \rho \geq 0, (14b) - (14i). \quad (16b)$$

Note that constraints (14b)-(14i) contains the interaction terms $\rho y_{i,b}$ with binary variables $y_{i,b}$ and the non-negative continuous variable ρ , for all $i \in [I]$ and $b \in [B]$. To linearize, we define

variables $\pi_{i,b} = \rho y_{i,b}$, for all $i \in [I]$ and $b \in [B]$, and introduce the following inequalities for variables $\pi_{i,b}$

$$\pi_{i,b} \geq 0, \quad \pi_{i,b} \geq \bar{\rho}(y_{i,b} - 1) + \rho, \quad (17a)$$

$$\pi_{i,b} \leq \bar{\rho}y_{i,b}, \quad \pi_{i,b} \leq \rho. \quad (17b)$$

Accordingly, formulation (16) is equivalent to the following MILP:

$$\hat{Z}(N, \epsilon) = \min \left\{ \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \epsilon \rho + \frac{1}{N} \sum_{n=1}^N \eta_b^n \right\} \quad (18a)$$

$$\text{s.t. } \mathbf{y} \in \mathcal{Y}, \quad \rho \geq 0, \quad (14b) - (14i), (17a) - (17b). \quad (18b)$$

6. Extension: Block Allocation and Surgery Scheduling in Flexible ORs

In this section, we extend our W-DRO model to determine which OR to open and then assign surgery to open OR. Specifically, we consider a Surgery Block Allocation (SBA) problem, which, as defined in Wang et al. (2019), aims to determine the number of surgical blocks (i.e., ORs) to open and the number of elective surgeries to schedule in each open block. The fixed cost of opening OR or block $b \in [B]$ is c_b^f , which incorporates staffing and equipment costs. The objective is to minimize the total cost consisting of the fixed cost of opening ORs or surgical blocks, fixed cost related to scheduling or rejecting elective surgery, and the expected cost associated with OR overtime and idle time. Next, we propose a distributionally robust SBA (DSBA) model with 1-Wasserstein ambiguity (denoted as W-DSBA) for this problem.

Let us introduce additional variables defining our W-DSBA model. For all $b \in [B]$, we define a binary decision variable x_b , which equal 1 if block b is open and 0 otherwise. We define feasible region \mathcal{Y}' of (\mathbf{x}, \mathbf{y}) in (19) such that each elective surgery is assigned to one open surgical block or postponed to the next planning period.

$$\mathcal{Y}' := \left\{ (\mathbf{x}, \mathbf{y}) : \begin{array}{l} y_{i,b} \leq x_b, \quad \forall b \in [B], \\ \sum_{b \in [B] \cup \{b'\}} y_{i,b} = 1, \quad \forall i \in [I], \\ x_b \in \{0, 1\}, \quad \forall b \in [B], \\ y_{i,b} \in \{0, 1\}, \quad \forall i \in [I], b \in [B] \cup \{b'\} \end{array} \right\}. \quad (19)$$

Using the ambiguity set $\mathcal{F}_\xi(\hat{\mathbb{P}}_\xi^N, \epsilon)$ defined in (7), we formulate the following DRO model for the DSBA problem.

$$(\text{W-DSBA}) \quad \hat{Z}'(N, \epsilon) = \min_{(\mathbf{x}, \mathbf{y}) \in \mathcal{Y}'} \left\{ \sum_{b \in B} c_b^f x_b + \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \sup_{\mathbb{P}_\xi \in \mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)} \mathbb{E}_{\mathbb{P}_\xi}[f(\mathbf{x}, \mathbf{y}, \xi)] \right\}, \quad (20a)$$

where for a feasible $(\mathbf{x}, \mathbf{y}) \in \mathcal{Y}'$ and a realization of ξ :

$$f(\mathbf{x}, \mathbf{y}, \xi) := \min_{\mathbf{o}, \mathbf{g}} \sum_{b \in B} (c_b^o o_b + c_b^g o_g) \quad (21a)$$

$$\text{s.t. } o_b - g_b = \sum_{i \in I} d_i y_{i,b} + e_b x_b - \mathcal{L}_b x_b, \quad \forall b \in [B] \quad (21b)$$

$$(o_b, g_b) \geq 0, \quad \forall b \in [B]. \quad (21c)$$

Formulation (20) finds first-stage decisions $(\mathbf{x}, \mathbf{y}) \in \mathcal{Y}'$ that minimize the sum of (1) fixed cost of opening ORs (first term); (2) fixed cost of scheduling or postponing elective surgeries (second term); and (3) maximum expected cost of OR overtime and idle time over all distributions defined in the ambiguity set $\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon)$. We remark that formulation (20) is the first DRO formulation with Wasserstein ambiguity for the DSBA problem in flexible ORs. Wang et al. (2019) proposed a moment-based DRO model for surgery block allocation in dedicated OR. The objective is to determine the ORs to open and assign the surgeries in a daily listing to the ORs, minimizing the weighted sum of OR opening costs and expected overtime costs. However, Wang et al. (2019)'s model cannot be used for flexible ORs because it does not account for emergency surgeries.

Using the same reformulation techniques in Section 5.1, we derive the following equivalent reformulation of the W-DSBA model in (20)

$$\hat{Z}'(N, \epsilon) = \min \left\{ \sum_{b \in V} c_b^f x_b + \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \epsilon \rho + \frac{1}{N} \sum_{n=1}^N \eta_b^n \right\} \quad (22a)$$

$$\text{s.t. } (\mathbf{x}, \mathbf{y}) \in \mathcal{Y}', \quad \rho \geq 0, \quad (22b)$$

$$\eta_b^n \geq \sum_{i \in I} [\bar{d}_i c_b^o - \rho |\bar{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b x_b c_b^o - \rho x_b |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b x_b c_b^o, \quad (22c)$$

$$\eta_b^n \geq \sum_{i \in I} [\underline{d}_i c_b^o - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b x_b c_b^o - \rho x_b |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b x_b c_b^o, \quad (22d)$$

$$\eta_b^n \geq \sum_{i \in I} [\bar{d}_i (-c_b^g) - \rho |\bar{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b x_b (-c_b^g) - \rho x_b |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b x_b (-c_b^g), \quad (22e)$$

$$\eta_b^n \geq \sum_{i \in I} [\underline{d}_i (-c_b^g) - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b x_b (-c_b^g) - \rho x_b |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b x_b (-c_b^g), \quad (22f)$$

$$\eta_b^n \geq \sum_{i \in I} [\underline{d}_i c_b^o - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b x_b c_b^o - \rho x_b |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b x_b c_b^o, \quad (22g)$$

$$\eta_b^n \geq \sum_{i \in I} [\underline{d}_i c_b^o - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b x_b c_b^o - \rho x_b |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b x_b c_b^o, \quad (22h)$$

$$\eta_b^n \geq \sum_{i \in I} [\bar{d}_i (-c_b^g) - \rho |\bar{d}_i - \hat{d}_i^n|] y_{i,b} + [\bar{e}_b x_b (-c_b^g) - \rho x_b |\bar{e}_b - \hat{e}_b^n|] - \mathcal{L}_b x_b (-c_b^g), \quad (22i)$$

$$\eta_b^n \geq \sum_{i \in I} [\underline{d}_i (-c_b^g) - \rho |\underline{d}_i - \hat{d}_i^n|] y_{i,b} + [\underline{e}_b x_b (-c_b^g) - \rho x_b |\underline{e}_b - \hat{e}_b^n|] - \mathcal{L}_b x_b (-c_b^g), \quad (22j)$$

$$\eta_b^n \geq \sum_{i \in I} \hat{d}_i^n c_b^o y_{i,b} + \hat{e}_b^n x_b c_b^o - \mathcal{L}_b x_b c_b^o, \quad (22k)$$

$$\eta_b^n \geq \sum_{i \in I} \hat{d}_i^n(-c_b^g)y_{i,b} + \hat{e}_b^n x_b(-c_b^g) - \mathcal{L}_b x_b(-c_b^g). \quad (22l)$$

Note that constraints (22c)-(22j) contains the interaction terms $\rho y_{i,b}$ and ρx_b with binary variables $y_{i,b}$ and x_b and the nonnegative continuous variable ρ , for all $i \in [I]$ and $b \in [B]$. To linearize, we use variables $\pi_{i,b} = \rho y_{i,b}$, for all $i \in [I]$ and $b \in [B]$, and introduce inequalities (17a)-(17b) for variables $\pi_{i,b}$. We also define variables $\tau_b = \rho x_b$, and introduce the following inequalities for variables τ_b , for all $b \in [B]$.

$$\tau_b \geq 0, \quad \tau_b \geq \bar{\rho}(x_b - 1) + \rho, \quad (23a)$$

$$\tau_b \leq \bar{\rho}x_b, \quad \tau_p \leq \rho. \quad (23b)$$

Accordingly, formulation (22) is equivalent to the following MILP:

$$\hat{Z}(N, \epsilon) = \min \left\{ \sum_{b \in V} c_b^f x_b + \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \epsilon \rho + \frac{1}{N} \sum_{n=1}^N \eta_b^n \right\} \quad (24a)$$

$$\text{s.t. } (y, x) \in \mathcal{Y}, \rho \geq 0, (22c) - (22l), (17a) - (17b), (23a) - (23b). \quad (24b)$$

7. Computational Experiments

In this section, we use publicly available actual surgery duration data to construct several instances of DSA and compare the performance of three proposed models: our 1-Wasserstein-based DRO model (W-DRO), moment-based DRO model (M-DRO), and a sample average approximation (SAA) of the SP model. In M-DRO, we construct an ambiguity set based on the mean and support of surgery duration (see Appendix E for the formulation). The SAA model solves model (4) with \mathbb{P}_ξ replaced by an empirical distribution based on N samples of random parameters (see Appendix F for the formulation). In Section 7.1, we describe the set of problem instances that we constructed and discuss other aspects of our experimental setup. In Section 7.2, we examine the effect of the radius ϵ on the performance of the W-DRO model. In Section 7.3, we compare the out-of-sample performance of the models under unseen data from the in-sample distribution (i.e., the distribution we used in the optimization step). In Section 7.4, we compare the out-of-sample performance of the models under unseen data from a distribution different than the one we used in the optimization. In Section 7.5, we analyze the W-DRO and SAA models solution times. In Section 7.6, we compare the flexible OR and dedicated OR policies. Finally, in Section 7.7, we present an experiment comparing the performance the proposed W-DASBA model for the surgical block allocation problem in Section 6 and its SP counterpart.

7.1. Description of the experiments

Our computational study is based on anonymized real-world surgery data presented by Mannino et al. (2012) and Mannino et al. (2010). This data set involves three-years (10,390 observations)

Table 2: Statistics for surgery duration (in minutes) based on surgery type. Notation: percent% is the percentage of patients needing a specific type of surgery, μ^d and σ^d are respectively the empirical mean and standard deviations of d .

Surgery type	percent%	μ^d	σ^d
CARD	14.01	99	53
GASTRO	17.79	132	76
GYN	27.81	78	52
MED	4.41	75	72
ORTH	17.81	142	58
URO	17.98	72	38

Table 3: Block schedule.

OR room	Monday	Tuesday	Wednesday	Thursday	Friday
1	GASTRO	GASTRO	GASTRO		
2			GASTRO	GASTRO	GASTRO
3	CARD		CARD		CARD
4	ORTH	ORTH		ORTH	ORTH
5		ORTH	MED		
6	GYN	GYN	GYN	GYN	
7		GYN	GYN	GYN	GYN
8	URO	URO		URO	URO
9	CARD		URO		CARD
10	URO		ORTH		

worth of daily surgery data that belong to six different surgical specialties, namely General Cardiology (CARD), Gastroenterology (GASTRO), Gynecology (GYN), Medicine (MED), Orthopedics (ORTH), and Urology (URO). Table 2 presents the mean and standard deviation of elective surgery duration based on surgery type. These values were computed from publicly available data that is referenced in [Mannino et al. \(2010\)](#) and [Mannino et al. \(2012\)](#).

We assume that we have 10 available ORs and $B = 32$ surgical blocks (the largest benchmark surgical suite and block schedule in the literature; see [Min and Yih \(2010\)](#) and [Shehadeh and Padman \(2021\)](#)). Table 3 presents the weekly assignments of surgical blocks to ORs. Each block $b \in [B]$ is $\mathcal{L}_b = 8$ -hours long, and an elective surgery can be assigned to any of the blocks allocated to the corresponding surgery type during the planning horizon. It is worth to mention that the distribution of surgery blocks through OR rooms and the week is often influenced by surgeons' and surgical teams' schedules, and the setup of the operating rooms, among other factors.

We consider two different cost structures for the objective function: (1) Cost1: $c^o = \$26/\text{min}$, $c^s = c^o/1.5$, and (2) Cost2: $c^o = \$26/\text{min}$, $c^s = 0$. An overtime cost of \$26 per minute is based on the work of [Min and Yih \(2010\)](#), [Shehadeh and Padman \(2021\)](#), and [Stodd et al. \(1998\)](#). We fix the ratio c^o/c^s to 1.5 as in prior surgery scheduling studies ([Jebali and Diabat, 2015](#); [Liu et al., 2019](#); [Shehadeh et al., 2019](#)).

We follow the same procedure in the literature to generate patient costs $c_{i,b}$ and $c_{i,b'}$ (with surgery of the same type having common values of these parameters). We compute the lower and upper bound of surgery duration from the data. [Lamiri et al. \(2008b\)](#) and [Lamiri et al. \(2008a\)](#) assumed that the daily capacity needed for emergency cases to be exponentially distributed with a mean of 3 and 2 hours, respectively, and the mean of the elective case to be 2 hours. This indi-

cates that, on average, the capacity required for emergency surgery in each block is approximately equivalent to the capacity needed for 1 scheduled elective surgery. We apply this logic to construct the parameters needed for emergency surgeries of each type.

We implemented and solved the three models using the AMPL modeling language and use CPLEX (version 12.6.2) as the solver with its default settings. We performed all experiments on a MacBook Pro with an Intel Core i7 processor, 2.6 GHz CPU, and 16 GB (2667 MHz DDR4) of memory.

7.2. Effect of ϵ in W-DRO model

The Wasserstein ball's radius ϵ is an input parameter to the W-DRO model, and a larger radius implies that we seek a more distributionally robust solutions. In this section, we investigate the impact of ϵ on the out-of-sample performance of the W-DRO's optimal solution $\hat{\mathbf{y}}(\epsilon, N)$ (i.e., the objective value obtained by simulating $\hat{\mathbf{y}}(\epsilon, N)$ under unseen data), with respect to the data size N .

We evaluate the out-of-sample performance for $\epsilon \in \{0.1, 0.5, 1, \dots, 100\}$ as follows. First, for each ϵ , we randomly sample 30 data sets of size $N \in \{5, 10, 50, 100\}$ from the empirical distribution of surgery duration. Second, we solve the W-DRO in (18) for each of the generated data sets and each ϵ . Third, we fix the first stage variables to $\hat{\mathbf{y}}(\epsilon, N)$, then re-optimize the second-stage of the SP using 10,000 out-of-sample (unseen) data. This is to compute the corresponding out-of-sample overtime and idle time. For illustrative and brevity purposes, in this experiment, we focus on an instance of $I = 60$ surgeries with Cost 1. We observe similar results for other instances (see [Appendix G](#) for results with $I = 80$).

Figure 2 illustrates the mean values (over the 30 independent replications) of the out-of-sample cost (i.e., first-stage cost plus the out-of-sample second-stage cost) as a function of ϵ . From this figure, we first observe that irrespective of N , the out-of-sample cost first improves with ϵ and then increases (or stabilize) after some (critical) value of ϵ . Such pattern is often observed in the literature (see, e.g., [Mohajerin Esfahani and Kuhn, 2018](#) and [Jiang et al., 2019](#)), and indicate that there exists a Wasserstein radius ϵ^* such that the corresponding optimal distributionally robust solutions have the best out-of-sample performance.

Second, we observe that ϵ^* decreases with the increase in N . Intuitively, a small sample does not have sufficient distributional information, and thus a larger ϵ produces distributionally robust solutions that better hedge against ambiguity. In contrast, with a larger sample, we may have more information from the data, and thus we can make a less conservative decision using a smaller ϵ .

In practice, it is unlikely that OR managers and other decision-makers have the expertise or the time to conduct the above exercise (or other iterative procedures) to obtain a suitable choice of ϵ for each surgery list and parameter settings. In addition, we often have a small data set, which is not enough for the optimization and simulation tasks (validation). Therefore, in the next sections, we

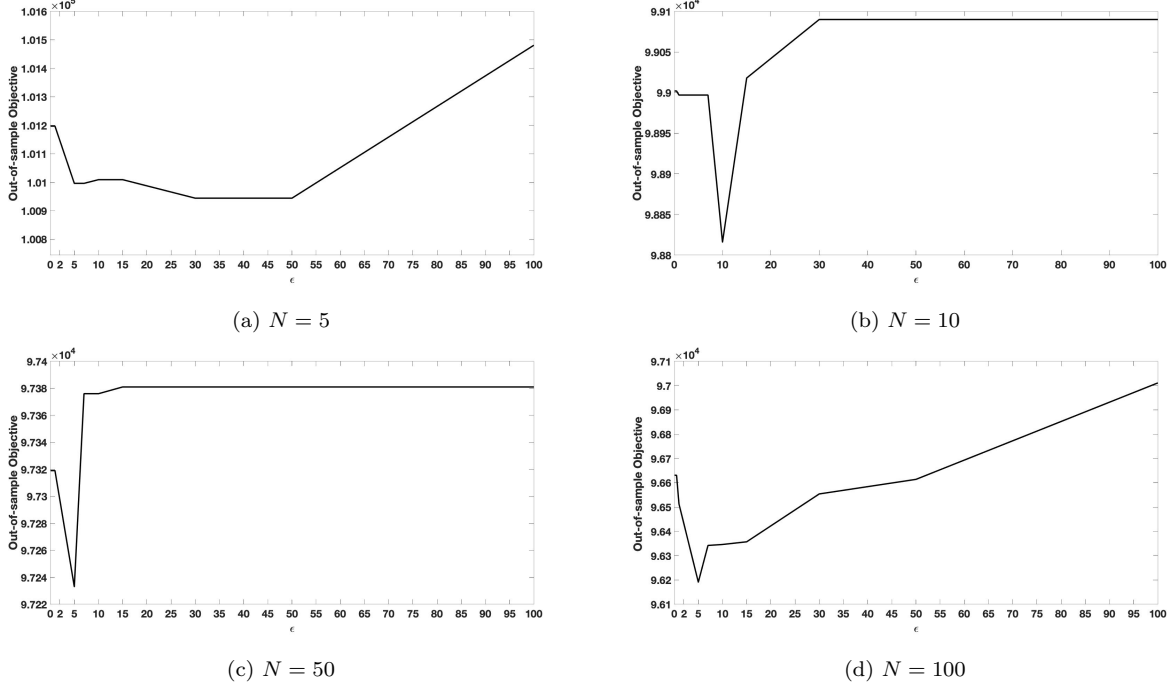


Figure 2: Out-of-sample performance as a function of the Wasserstein radius

derive insights with 0.1 (small), 10 (relatively average), and 100 (relatively large), which captures different extents of the robustness of the W-DRO model. This approach mimics, to a certain extent, decision-makers' thinking who do not have optimization expertise.

7.3. Solutions quality under the in-sample distribution

In this section, we compare the out-of-sample performance of the W-DRO, M-DRO, and SAA models under unseen data sample from the in-sample data distribution (i.e., perfect information case). We test the out-of-sample performance (i.e., the objective value obtained by simulating the optimal solution of a model under a larger unseen data) as follows. First, we sample data sets of sizes $N \in \{5, 10, \dots, 100\}$ from the empirical distribution. Second, using each set of N scenarios, we solve an instance of W-DRO, an instance M-DRO, and an instance of SAA. Third, we fix the optimal first-stage decisions \mathbf{y} yielded by each model in the SP model. Then, we solve the second-stage recourse problem in (5) using \mathbf{y} and $N' = 10,000$ out-of-sample (unseen) data to compute the corresponding out-of-sample cost (i.e., first-stage fixed cost plus simulated second-stage cost). We perform these steps 20 times using 20 independent samples of N scenarios. For the W-DRO model, we use $\epsilon = 0.1$ (small), $\epsilon = 10$ (relatively average), and $\epsilon = 100$ (relatively large), which captures different extents of the robustness of the W-DRO model. For illustrative purposes and brevity, we use an instance of $I = 60$ and 80 surgeries under Cost1. We observe similar results for other instances.

Figures 3 and 4 present the mean (line) and the (shaded) area between the 20% and 80%

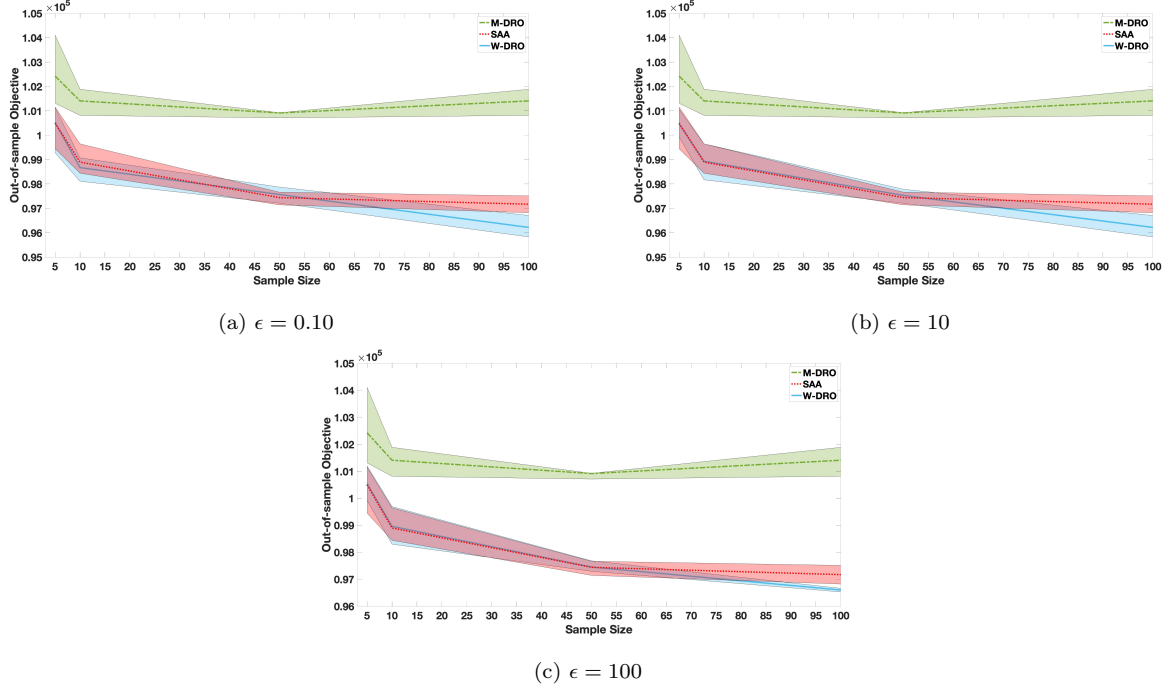


Figure 3: Out-of-sample performance under perfect information, $I = 60$

quantiles of the out-of-sample cost as a function of N and ϵ for $I = 60$ and $I = 80$, respectively. We compute these values over the 20 independent simulation runs of each optimal solution and N . Note that we do not know the optimal value Z^* of the stochastic optimization problem because we do not know the true (unknown) distribution. However, we know that Z^* should be less than or equal to the bounds yielded by all models. Therefore, for the sake of illustration and convenience of making comparisons, one can assume that Z^* is the horizontal axis in each sub-figure.

We observe the following from figures 3–4. First, the out-of-sample cost of the M-DRO model are significantly larger than those of the W-DRO and SAA model across all values of N and ϵ . Second, as N increases (i.e., more distributional information becomes available), the out-of-sample costs of the W-DRO model decrease and converges to Z^* , which is consistent with Theorem 1 (demonstrating that the W-DRO model enjoys the asymptotic consistency property). In contrast, the out-of-sample cost of the M-DRO model is relatively stable. These results are consistent with the theoretical results in Theorem 1 (see Appendix C) demonstrating the asymptotic consistency property of the W-DRO approach. The M-DRO model depends on the mean and support of surgery duration (i.e., descriptive statistics), and thus one cannot guarantee asymptotic consistency.

Third, we observe that W-DRO slightly outperforms SAA, especially when N is small. *This demonstrates that the W-DRO model is capable to effectively learn distributional information even from a very small amount of data (e.g., 5 and 10 scenarios).* Thus, the proposed W-DRO model is advantageous in ORs with scarce data on surgery duration. Finally, we observe a lower out-of-

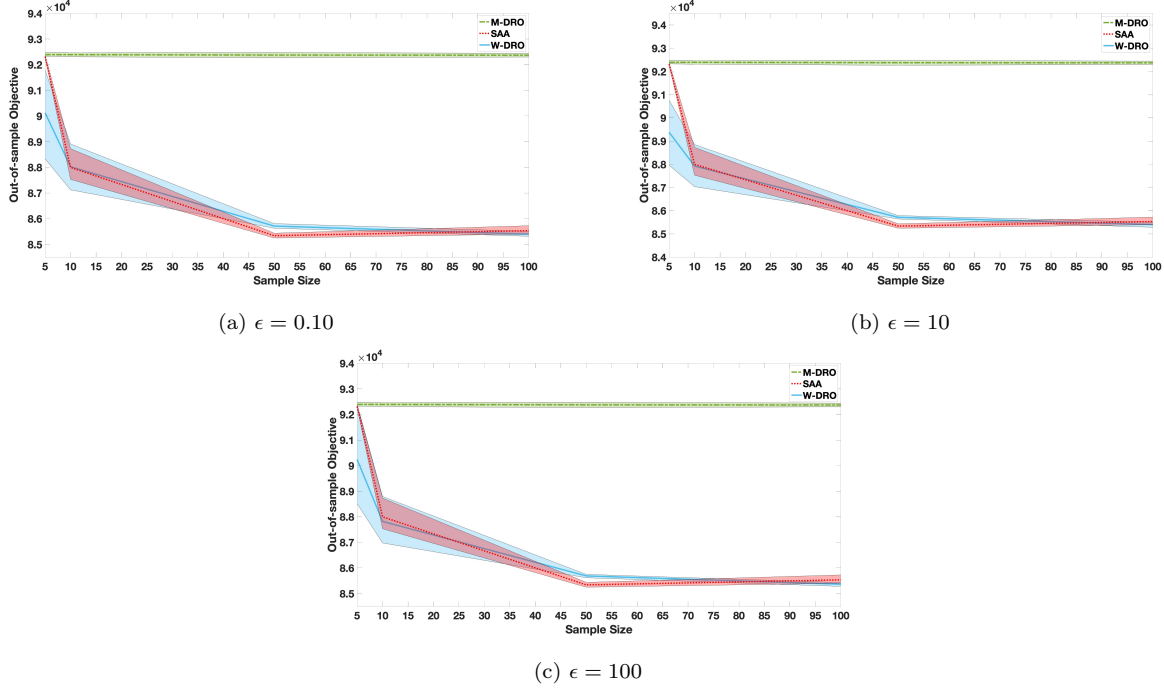


Figure 4: Out-of-sample performance under perfect information, $I = 80$

sample cost under higher ϵ , which reflects the fact that one can achieve a stronger probabilistic guarantee by increasing ϵ in W-DRO.

Finally, we attribute the higher out-of-sample cost of the M-DRO to the following. The W-DRO model schedule a larger number of elective surgeries than the M-DRO model to hedge against the cost of rejecting surgeries and excessive OR idle time. In contrast, the M-DRO model is more conservative in the sense that it attempts to hedge against observing extremely long surgery durations and overtime by scheduling fewer surgeries. However, in practice, we may not observe extremely long durations. Accordingly, the M-DRO schedule will result in a higher out-of-sample cost associated with larger idle time and rejecting a larger number of surgeries than the W-DRO schedule. For example, consider the instance of $I = 80$ surgeries with $\epsilon = 0.1$, $N = 50$, and Cost1. The W-DRO and M-DRO models schedule 74 and 57 surgery, respectively. The associated average (surgery cost, overtime, idle time) with the W-DRO and M-DRO scheduling decisions are (1196, 27 minutes, 108) and (1638, 26 minutes, 201 minutes), respectively. Note that a lower idle time indicates better utilization of the expensive OR resources. Moreover, scheduling more surgeries indicates better access to surgical care. Thus, a decision-maker may prefer the W-DRO solutions.

7.4. Solutions quality under misspecified distributional information

Let us now analyze the out-of-sample performance of the models under the case where the true distribution is different than the one used in the optimization. That is, we investigate the performance of optimal solutions of the models using N' data sample generated from a different distribution than

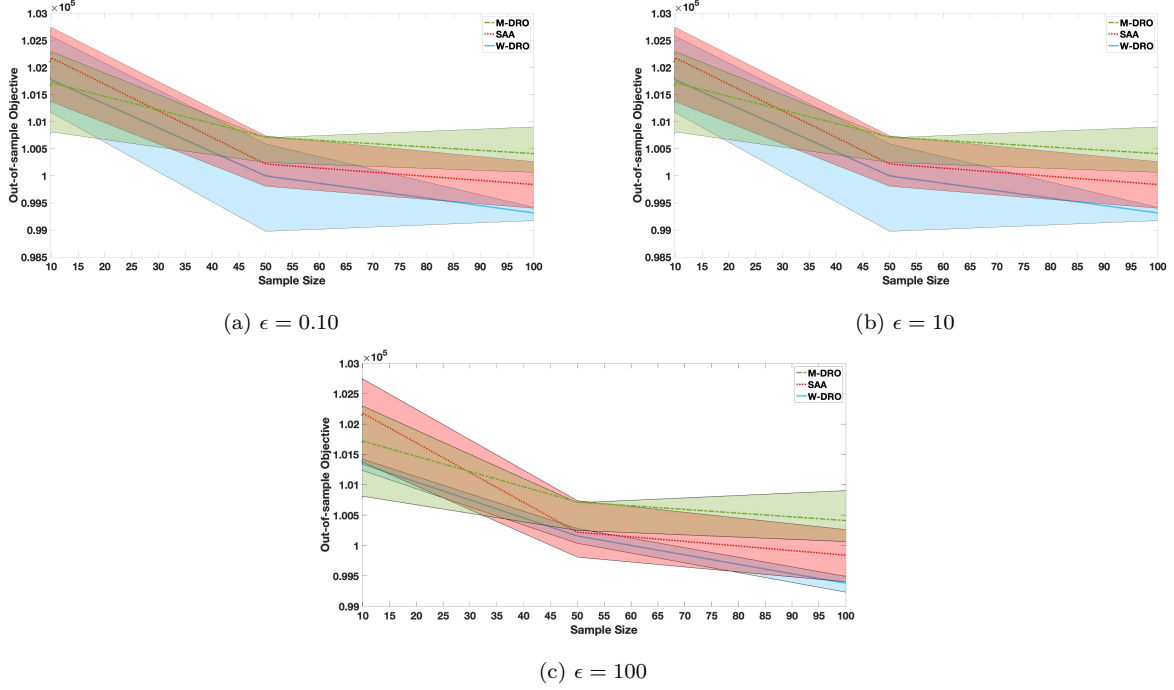


Figure 5: Out-of-sample performance under LogN, $I = 60$

the empirical distribution we used in the optimization. Specifically, we generate the $N' = 10,000$ out-of-sample data from a Lognormal (LogN) distribution with the same mean, variance, and support of the in-sample distribution. For brevity, in Figures 5, we present the out-of-sample cost under logN for an instance with $I = 60$.

Notably, both W-DRO and M-DRO yield lower out-of-sample costs on average and across all quantiles than the SP model when N is small (i.e., limited distributional information). In fact, the M-DRO model yields a slightly lower out-of-sample cost than the W-DRO model when N is small. Second, when N is larger, the W-DRO and SAA models yield lower costs than the M-DRO model. Third, the W-DRO model outperforms the SAA model in most instances, especially when N is small. These results indicate that the DRO approach is effective in an environment where the distribution of random parameters quickly changes, and there is small data or information on such variability. Moreover, these results emphasize the value of modeling uncertainty and distributional ambiguity of surgery duration.

7.5. CPU Time

In this section, we analyze the solution times of the W-DRO and SAA models. We take advantage of the fact that our problem is separable by block for some instances and thus is decomposable.

We analyze solution times of the models (i.e., total CPU time needed to construct the OR schedule) with three sizes of surgeries list $I \in \{60, 80, 100\}$ under Cost1 ($c^o = \$26/\text{min}$, $c^g = c^o/1.5$), and Cost2 ($c^o = \$26/\text{min}$, $c^g = 0$). Recall that the W-DRO and SAA models are scenario based.

Table 4: Solution times in seconds

Cost1															
Model	I	N	Min	Avg	Max	I	N	Min	Avg	Max	I	N	Min	Avg	Max
W-DRO	60	5	0.2	0.4	0.7	80	5	1	1	2	100	5	0.4	1	2
SAA			0.1	0.1	0.2			0.1	0.1	0.1			0.2	0.3	0.5
W-DRO		10	0.4	0.5	0.8		10	1	2	3		10	1	2	6
SAA			0.10	0.13	0.2			1	1	2			1	1	2
W-DRO		50	1	3	9		50	3	6	21		50	4	11	25
SAA			0.13	0.2	1			1	1	3			1	1	3
W-DRO		100	3	5	10		100	6	10	29		100	6	14	63
SAA			0.3	1	3			1	2	3			1	2	4
W-DRO		500	20	29	58		500	89	164	267		500	105	133	174
SAA			1	2	3			2	4	11			5	6	13
Cost2															
Model	I	N	Min	Avg	Max	I	N	Min	Avg	Max	I	N	Min	Avg	Max
W-DRO	60	5	0.3	0.4	1	80	5	0.2	0.3	0.4	100	5	0.4	0.5	0.6
SAA			0.1	0.2	0.2			0.1	0.1	0.1			0.2	0.3	0.4
W-DRO		10	0.4	0.5	1		10	0.25	0.30	0.4		10	0.4	0.6	1.1
SAA			0.1	0.2	0.2			0.1	0.2	0.3			0.2	0.3	1
W-DRO		50	0.5	0.6	1		50	0.5	0.6	0.7		50	1	1.4	1.8
SAA			0.2	0.2	0.3			0.2	0.3	0.5			1	1	1
W-DRO		100	0.6	0.7	1		100	0.9	1.1	1.5		100	3	4	6
SAA			0.3	0.3	0.4			0.3	0.4	0.5			1	1	1
W-DRO		500	1.3	1.5	2.6		500	5	9	15		500	16	27	63
SAA			1	1	1.1			2	2	3			4	5	7

Thus, solution times of these models depend on the sample size, N . Therefore, we next analyze solution times of these two models under $N \in \{5, 10, 50, 100, 500\}$. For each N , I , and cost structure, we generate 10 random instances and solve each using the W-DRO and SAA models. Table 4 presents the minimum (Min), Average (Avg), and maximum (Max) solution times (in seconds) of these instances.

We make the following observations from Table 4. First, solution times increase with I and N . Second, both models solve all instances fairly quickly under Cost2. Third, solution times are longer under Cost1, which makes sense as this cost structure require more variables (for idle time) and includes more criteria in the objective to optimize. Fourth, under Cost2, solution times of the W-DRO model are approximately similar to the SAA solution times when N is small (e.g., 5 or 10) and longer than those of the SAA model when N is large (e.g., 100 or 500). However, when $N = 500$, solution times of the W-DRO are all less than 5 minutes. Moreover, we have demonstrated that our W-DRO approach is effective in ORs with scares (small) data.

We attribute the difference in solution times between the SAA and W-DRO models to their respective sizes in terms of the number of variables and constraints. The W-DRO model has more variables and constraints than the SAA model. As pointed out by (Artigues et al., 2015; Catanzaro et al., 2015; Fortz et al., 2017; Jünger et al., 2009; Keha et al., 2009; Klotz and Newman, 2013; Morales-España et al., 2016; Shehadeh et al., 2019), an increase in model size often suggests an increase in solution time for the LP relaxation and, thus, the model's overall solution time.

Table 5: Optimal number of scheduled surgery under the flexible and dedicated ORs policies.

Policy	Base Emergency Rate		Double Emergency Rate	
	Cost1	Cost2	Cost1	Cost2
Flexible	79	76	76	72
Dedicated	79	79	79	79

Finally, it is worth mentioning that we were able to solve all instances using the M-DRO model fairly quickly. This makes sense, because this model is deterministic, i.e., does not depend on N .

7.6. Flexible versus dedicated ORs

As mentioned earlier, some hospitals dedicate one or more operating rooms to emergency cases. Dedicating some ORs to emergency cases may have the advantage of having some ORs readily available to handle them promptly without impacting the scheduled elective surgeries. However, this policy may results in a low utilization of costly OR resources (Xiao and Yoogalingam, 2021). In this section, we compare the flexible and dedicated OR policies. In the latter, we solve our W-DRO model assuming that emergency surgeries have some ORs dedicated for them. We compare the optimal number of scheduled surgeries and performance with an instance of $I = 80$ surgeries and with the base and twice (double) the base rate of emergency cases described in Section 7.1.

Table 5 presents the optimal number of scheduled surgery yielded by the flexible and dedicated policies. From this table, we first observe that the W-DRO model always schedules more elective surgeries under the dedicated policy than under the flexible policy. In fact, under the dedicated policy, the model always schedules 79/80 surgeries irrespective of the cost structure and rate of emergency cases. In contrast, we schedule fewer surgeries under the flexible policy, especially with Cost2 (which ignores the idle time cost) and a higher rate of emergency cases. These results make sense because by allocating an OR for emergency cases, we can accommodate more elective surgeries, i.e., we can use the capacity of the remaining ORs solely for elective surgeries.

Next, we analyze the simulation performance of these optimal schedules with $N' = 10,000$ scenarios from the empirical distribution. Table 6 presents the mean and quantiles of overtime (total OT hours per week) and OR utilization ($UT\% = \frac{\text{available OR time} - \text{idle time}}{\text{available OR time}} \times 100\%$, where a larger UT% indicate less idle time and higher utilization rate). From this table, we observe that there is no clear winner. While the dedicated policy yields significantly lower overtime, it yields a significantly lower utilization than the flexible policy on average and at all quantiles. This makes sense because the flexible policy uses some OR capacity for emergency surgeries (which must be performed), thus mitigating the risk of OR idle time. In contrast, under the dedicated policy, we perform emergency surgery in different ORs than those dedicated to elective surgery. This could thus mitigate the risk of overtime in ORs dedicated to elective surgery but potentially lead to idle time due to the unused capacity of the ORs dedicated to emergency cases.

These results show that the dedicated policy may improve access by scheduling a larger number of surgery. It may also lead to less overtime but higher idle time than the flexible policy. However,

Table 6: Simulation performance of optimal schedule of the flexible and dedicated policies. Notation: OT is overtime and UT=

Base Emergency Rate							
Cost1				Cost2			
Metric	Policy	OT	UT%	Metric	Policy	OT	UT%
Mean	Flexible	2.7	63	Mean	Flexible	1.3	61
	Dedicated	0.7	44		Dedicated	0.8	44
75-q	Flexible	3.8	61	75-q	Flexible	2.0	61
	Dedicated	1.0	43		Dedicated	1.4	44
95-q	Flexible	5.9	59	95-q	Flexible	3.7	59
	Dedicated	2.5	41		Dedicated	2.9	43
Double Emergency Rate							
Cost1				Cost2			
Metric	Policy	OT	UT%	Metric	Policy	OT	UT%
Mean	Flexible	4.6	69	Mean	Flexible	1.8	66
	Dedicated	0.7	44		Dedicated	0.7	44
75-q	Flexible	5.9	68	75-q	Flexible	2.6	64
	Dedicated	1.0	43		Dedicated	1.0	43
95-q	Flexible	8.4	66	95-q	Flexible	4.3	62
	Dedicated	2.5	41		Dedicated	2.5	41

as mentioned in prior literature, dedicating (or opening an OR) for emergency surgery is costly (because the OR is an expensive resource) and may not be a feasible option (e.g., financial constraints, etc).

7.7. Block allocation example

In this section, we compare the performance of our proposed W-DSBA model for the surgical block allocation (SBA) problem (presented in Section 6) and its SP counterpart. For simplicity, we use SP-SBA to denote the SP model.

For illustrative purposes, we use the data related to Gastro (described in Section 7.1) and parameter settings from prior literature. Specifically, we consider $B = 10$ and $B = 5$ surgical blocks or ORs, and regular time length of each block is 8 hours. By Min and Yih (2010) and Zhang et al. (2019), $c_b^f = 800$, $c_b^o = 780$, and $c_b^g = 0$. We set the cost of scheduling or rejecting an elective surgery as described in Section 7.1. Finally, we consider a list of $I = 25$ Gastro surgeries.

Let us first compare the performance of these models under the perfect distributional information case described in Section 7. That is under the case when N (optimization scenarios) and N' (simulation scenarios) are sampled from the empirical distribution. Figure 6 presents the mean (line) and the (shaded) area between the 20% and 80% quantiles of the out-of-sample cost (over 30 optimization and simulation replications). The out-of-sample cost of DRO model is computed for the best ϵ determined as described in Section 7.2.

We observe the following from Figure 6. Notably, the W-DSBA mode yields substantially lower out-of-sample cost than the SP-SBA model under all values of B and N . This is partly because the W-DSBA model schedules fewer surgeries and opens more blocks than the SP-SBA model to hedge against ambiguity and overtime. For example, when $B = 10$, the W-DSBA model opens

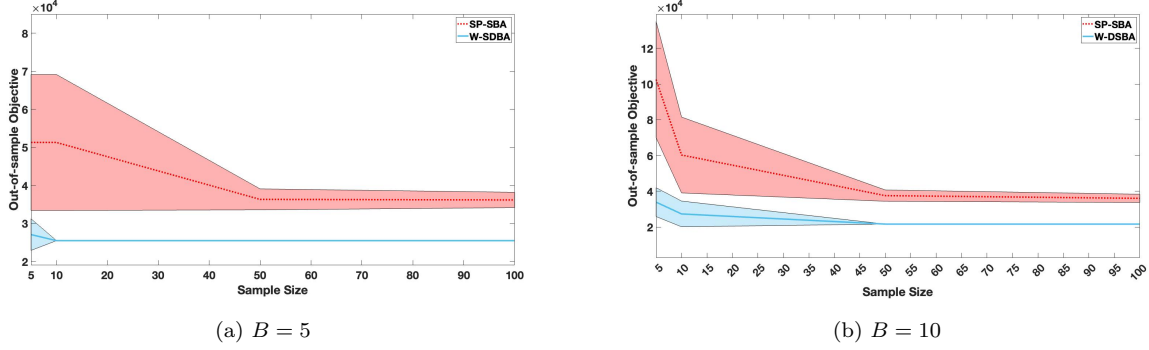


Figure 6: Out-of-sample performance for SBA

9–10 blocks and schedules 22–25 surgeries, while the SP-SBA model opens 8 blocks and schedules 24–25 surgeries. When $B = 5$, the DRO model opens all blocks and schedules 11–15 surgeries, while the SP-SBA model opens all blocks and schedules 13–16 surgeries. In addition, the W-DSBA decisions result in less overtime, which improve surgical team satisfaction, indirectly mitigate the risk of canceling surgeries due to overtime, and reduces OR cost. For example, when $B = 10$ and $N = 5$, the SP and DRO models’ average OR overtime is 190 and 58 minutes, respectively. Second, we observe that the out-of-sample cost of the W-DSBA model decrease and converges as N increases. These results are consistent with the theoretical results in Theorem 1 (see [Appendix C](#)), demonstrating the that W-DSBA model enjoys the asymptotic consistency property.

Next, we analyze the out-of-sample performance under the case when the data we use in the optimization is biased. Specifically, we investigate the out-of-sample performance of the optimal solution yielded by the SP-SBA and W-DSBA models using N' data sample from a LogN distribution defined on the same mean and support of the in-sample data. Figure 7 presents the out-of-sample cost under LogN. It is clear that the optimal solutions of the W-DSBA maintains a robust performance and yields a significantly lower out-of-sample cost than the optimal solutions of the SP-SBA model, even when small data set is used in the optimization. These results again confirm that our DRO approach is effective in environments with limited or no data on random parameters.

8. Conclusion

In this paper, we proposed DRO models for elective surgery planning in flexible ORs under random surgery duration. Specifically, we proposed a W-DRO model that determines optimal elective surgery assigning decisions to available surgical blocks in multiple ORs over a planning horizon that minimizes the sum of patient-related costs and the worst-case expected cost associated with overtime and idle time of ORs over all distributions residing in the 1-Wasserstein ambiguity set. We derived an equivalent MILP reformulation of our min-max W-DRO model. In addition, we extended

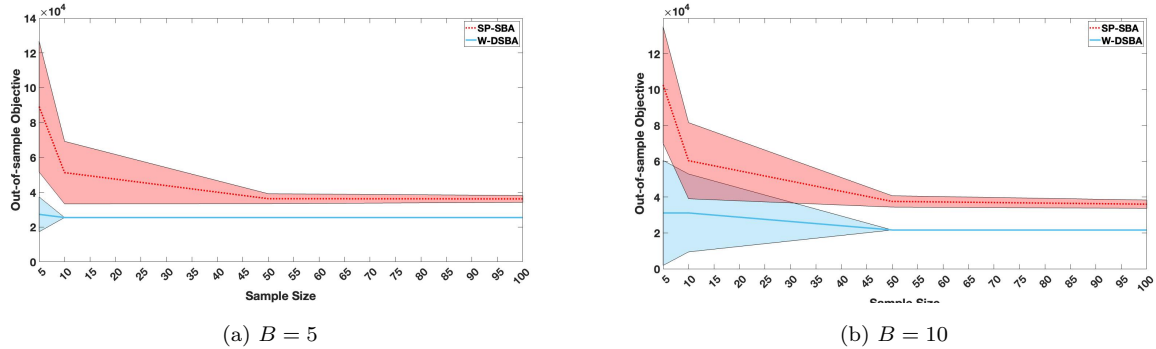


Figure 7: Out-of-sample performance for SBA misspecified distribution (LogN)

our W-DRO model to determine the number of ORs to open and then surgery assignments to open ORs. We derived an equivalent MILP of this extension.

Using real-world surgery data, we conduct extensive numerical experiments comparing our approach with the state-of-the-art approaches, namely the SP approach and a DRO approach based on mean-support ambiguity (M-DRO). Our results demonstrate that our W-DRO approach (1) enjoys both asymptotic consistency and finite-data guarantees, (2) yield robust decisions that have a good out-of-sample performance, especially under limited data and quickly varying distributions of uncertain surgery duration, and (3) is computationally efficient with solution times sufficient for real-world implementation. In addition, we investigated the pros and cons of the flexible and dedicated ORs policies. In particular, our results indicate that we schedule more elective surgeries under the dedicated policy. However, the dedicated policy yield longer overtime and lower utilization (when emergency cases do not materialize). In contrast, under the flexible policy, we schedule fewer elective surgeries and observe less overtime and better utilization of OR time. Finally, we show that our proposed W-DABA model for the surgical block allocation problem has superior out-of-sample performance over the SP model for this problem.

Overall, our results demonstrate that our proposed DRO approach can effectively learn distributional information even from a small amount of data (e.g., 5 and 10 scenarios). Thus, the proposed approach is advantageous in ORs with scarce data on surgery duration.

We suggest the following areas for future research. First, we want to extend our approach by incorporating other multi-modal sources of uncertainty. Second, our model can be considered as the first step toward building data-driven and robust OR and surgery planning. We aim to extend our model and build more comprehensive OR and surgery planning models, which consider all relevant organizational and technical constraints and various sources of uncertainties. Finally, we aim to incorporate other metrics such as surgery waiting time and model the probability of canceling surgery and unexpected ORs or surgical teams' unavailability.

Acknowledgments

We want to thank all colleagues and practitioners who have contributed significantly to the related literature. We are grateful to the anonymous reviewers for their insightful comments and suggestions that allowed us to improve the paper. Dr. Karmel S. Shehadeh dedicates her effort in this paper to every little dreamer in the whole world who has a dream so big and so exciting. Believe in your dreams and do whatever it takes to achieve them—the best is yet to come for you.

References

References

- Ahmadi-Javid, A., Jalali, Z., Klassen, K. J., 2017. Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research* 258 (1), 3–34.
- Anjomshoa, H., Dumitrescu, I., Lustig, I., Smith, O. J., 2018. An exact approach for tactical planning and patient selection for elective surgeries. *European Journal of Operational Research* 268 (2), 728–739.
- Artigues, C., Koné, O., Lopez, P., Mongeau, M., 2015. Mixed-integer linear programming formulations. In: Schwindt, C., Zimmermann, J. (Eds.), *Handbook on Project Management and Scheduling* Vol. 1. Springer, pp. 17–41.
- Bansal, A., Berg, B., Huang, Y.-L., 2021. A distributionally robust optimization approach for coordinating clinical and surgical appointments. *IIE Transactions*, 1–83.
- Batun, S., Denton, B. T., Huschka, T. R., Schaefer, A. J., 2011. Operating room pooling and parallel surgery processing under uncertainty. *INFORMS journal on Computing* 23 (2), 220–237.
- Ben-Tal, A., Den Hertog, D., Vial, J.-P., 2015. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical programming* 149 (1-2), 265–299.
- Bertsimas, D., Popescu, I., 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* 15 (3), 780–804.
- Bertsimas, D., Sim, M., 2004. The price of robustness. *Operations research* 52 (1), 35–53.
- Birge, J. R., Louveaux, F., 2011. *Introduction to stochastic programming*. Springer Science & Business Media.
- Bovim, T. R., Christiansen, M., Gullhav, A. N., Range, T. M., Hellemo, L., 2020. Stochastic master surgery scheduling. *European Journal of Operational Research*.
- Cardoen, B., Demeulemeester, E., Beliën, J., 2010. Operating room planning and scheduling: A literature review. *European journal of operational research* 201 (3), 921–932.

- Carello, G., Lanzarone, E., 2014. A cardinality-constrained robust model for the assignment problem in home care services. *European Journal of Operational Research* 236 (2), 748–762.
- Catanzaro, D., Gouveia, L., Labbé, M., 2015. Improved integer linear programming formulations for the job sequencing and tool switching problem. *European Journal of Operational Research* 244 (3), 766–777.
- Chen, Z., Sim, M., Xiong, P., 2020. Robust stochastic optimization made easy with rsome. *Management Science*.
- Choi, T.-M., 2012. Handbook of Newsvendor problems: Models, extensions and applications. Vol. 176. Springer.
- Delage, E., Saif, A., 2021. The value of randomized solutions in mixed-integer distributionally robust optimization problems. *INFORMS*.
- Delage, E., Ye, Y., 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research* 58 (3), 595–612.
- Deng, Y., Shen, S., Denton, B., 2019. Chance-constrained surgery planning under conditions of limited and ambiguous data. *INFORMS Journal on Computing* 31 (3), 559–575.
- Denton, B., Viapiano, J., Vogl, A., 2007. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health care management science* 10 (1), 13–24.
- Denton, B. T., Miller, A. J., Balasubramanian, H. J., Huschka, T. R., 2010. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations research* 58 (4-part-1), 802–816.
- Fei, H., Chu, C., Meskens, N., 2009. Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operations Research* 166 (1), 91–108.
- Fortz, B., Oliveira, O., Requejo, C., 2017. Compact mixed integer linear programming models to the minimum weighted tree reconstruction problem. *European Journal of Operational Research* 256 (1), 242–251.
- Fournier, N., Guillin, A., 2015. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields* 162 (3-4), 707–738.
- Gartner, D., Padman, R., 2019. Flexible hospital-wide elective patient scheduling. *Journal of the Operational Research Society*, 1–15.
- Gerchak, Y., Gupta, D., Henig, M., 1996. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science* 42 (3), 321–334.

- Giroto, J. A., Koltz, P. F., Drugas, G., 2010. Optimizing your operating room: or, why large, traditional hospitals don't work. *International Journal of Surgery* 8 (5), 359–367.
- Gul, S., Denton, B. T., Fowler, J. W., Huschka, T., 2011. Bi-criteria scheduling of surgical services for an outpatient procedure center. *Production and Operations management* 20 (3), 406–417.
- Guo, C., Bodur, M., Aleman, D. M., Urbach, D. R., 2021. Logic-based benders decomposition and binary decision diagram based approaches for stochastic distributed operating room scheduling. *INFORMS Journal on Computing*.
- Halevy, Y., 2007. Ellsberg revisited: An experimental study. *Econometrica* 75 (2), 503–536.
- Hof, S., Fügner, A., Schoenfelder, J., Brunner, J. O., 2017. Case mix planning in hospitals: a review and future agenda. *Health care management science* 20 (2), 207–220.
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., Camerer, C. F., 2005. Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310 (5754), 1680–1683.
- Jackson, R. L., 2002. The business of surgery. *Health management technology* 23 (7), 20–22.
- Jebali, A., Diabat, A., 2015. A stochastic model for operating room planning under capacity constraints. *International Journal of Production Research* 53 (24), 7252–7270.
- Jiang, R., Ryu, M., Xu, G., 2019. Data-driven distributionally robust appointment scheduling over wasserstein balls. *arXiv preprint arXiv:1907.03219*.
- Jiang, R., Shen, S., Zhang, Y., 2017. Integer programming approaches for appointment scheduling with random no-shows and service durations. *Operations Research* 65 (6), 1638–1656.
- Jünger, M., Liebling, T. M., Naddef, D., Nemhauser, G. L., Pulleyblank, W. R., Reinelt, G., Rinaldi, G., Wolsey, L. A., 2009. 50 years of integer programming 1958-2008: From the early years to the state-of-the-art. Springer Science & Business Media.
- Keha, A. B., Khowala, K., Fowler, J. W., 2009. Mixed integer programming formulations for single machine scheduling problems. *Computers & Industrial Engineering* 56 (1), 357–367.
- Keyvanshokoh, E., Kazemian, P., Fattahi, M., Van Oyen, M. P., 2020. Coordinated and priority-based surgical care: An integrated distributionally robust stochastic optimization approach. *Production and Operations Management*.
- Klotz, E., Newman, A. M., 2013. Practical guidelines for solving difficult mixed integer linear programs. *Surveys in Operations Research and Management Science* 18 (1), 18–32.

- Lamiri, M., Grimaud, F., Xie, X., 2009. Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics* 120 (2), 400–410.
- Lamiri, M., Xie, X., Dolgui, A., Grimaud, F., 2008a. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research* 185 (3), 1026–1037.
- Lamiri, M., Xie, X., Zhang, S., 2008b. Column generation approach to operating theater planning with elective and emergency patients. *Iie Transactions* 40 (9), 838–852.
- Li, F., Gupta, D., Potthoff, S., 2016. Improving operating room schedules. *Health care management science* 19 (3), 261–278.
- Liu, N., Truong, V.-A., Wang, X., Anderson, B. R., 2019. Integrated scheduling and capacity planning with considerations for patients’ length-of-stays. *Production and Operations Management*.
- Macario, A., 2010. Is it possible to predict how long a surgery will last? *medscape*.
- Mak, H.-Y., Rong, Y., Zhang, J., 2014. Appointment scheduling with limited distributional information. *Management Science* 61 (2), 316–334.
- Mannino, C., Nilssen, E. J., Nordlander, T. E., 2010. Sintef ict: Mss-adjusts surgery data. [WWW] Available from: <https://www.sintef.no/Projectweb/Health-care-optimization/Testbed/>.
- Mannino, C., Nilssen, E. J., Nordlander, T. E., 2012. A pattern based, robust approach to cyclic master surgery scheduling. *Journal of Scheduling* 15 (5), 553–563.
- May, J. H., Spangler, W. E., Strum, D. P., Vargas, L. G., 2011. The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management* 20 (3), 392–405.
- Min, D., Yih, Y., 2010. Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research* 206 (3), 642–652.
- Mohajerin Esfahani, P., Kuhn, D., 2018. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171 (1-2), 115–166.
- Morales-España, G., Correa-Posada, C. M., Ramos, A., 2016. Tight and compact mip formulation of configuration-based combined-cycle units. *IEEE Transactions on Power Systems* 31 (2), 1350–1359.

- Nagarajan, M., Shechter, S., 2014. Prospect theory and the newsvendor problem. *Management Science* 60 (4), 1057–1062.
- Neyshabouri, S., Berg, B. P., 2017. Two-stage robust optimization approach to elective surgery and downstream capacity planning. *European Journal of Operational Research* 260 (1), 21–40.
- Pinedo, M. L., 2016. *Scheduling: theory, algorithms, and systems*. Springer.
- Qi, J., 2017. Mitigating delays and unfairness in appointment systems. *Management Science* 63 (2), 566–583.
- Rahimi, I., Gandomi, A. H., 2020. A comprehensive review and analysis of operating room and surgery scheduling. *Archives of Computational Methods in Engineering*, 1–22.
- Rahimian, H., Mehrotra, S., 2019. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*.
- Rowse, E. L., 2015. Robust optimisation of operating theatre schedules. Ph.D. thesis, Cardiff University.
- Saif, A., Delage, E., 2021. Data-driven distributionally robust capacitated facility location problem. *European Journal of Operational Research* 291 (3), 995–1007.
- Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., Rademakers, F. E., 2016. Scheduling operating rooms: achievements, challenges and pitfalls. *Journal of Scheduling* 19 (5), 493–525.
- Shapiro, A., Dentcheva, D., Ruszczyński, A., 2014. *Lectures on stochastic programming: modeling and theory*, 2nd Edition. SIAM.
- Shehadeh, K. S., Cohn, A. E., Epelman, M. A., 2019. Analysis of models for the stochastic outpatient procedure scheduling problem. *European Journal of Operational Research* 279 (3), 721–731.
- Shehadeh, K. S., Cohn, A. E., Jiang, R., 2020. A distributionally robust optimization approach for outpatient colonoscopy scheduling. *European Journal of Operational Research* 283 (2), 549–561.
- Shehadeh, K. S., Padman, R., 2021. A distributionally robust optimization approach for stochastic elective surgery scheduling with limited intensive care unit capacity. *European Journal of Operational Research* 290 (3), 901–913.
- Shehadeh, K. S., Padman, R., 2022. Stochastic optimization approaches for elective surgery scheduling with downstream capacity constraints: Models, challenges, and opportunities. *Computers & Operations Research* 137, 105523.

- Shylo, O. V., Prokopyev, O. A., Schaefer, A. J., 2013. Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS Journal on Computing* 25 (4), 682–692.
- Soyster, A. L., 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research* 21 (5), 1154–1157.
- Stodd, K., Ortiz, A., Tenzer, I., et al., 1998. Operating room benchmarking: The kaiser permanente experience. *The Permanente Journal* 2 (4), 5–16.
- Viapiano, J., Ward, D. S., 2000. Operating room utilization: the need for data. *International anesthesiology clinics* 38 (4), 127–140.
- Wang, S., Roshanaei, V., Aleman, D., Urbach, D., 2016. A discrete event simulation evaluation of distributed operating room scheduling. *IIE Transactions on Healthcare Systems Engineering* 6 (4), 236–245.
- Wang, Y., Tang, J., Fung, R. Y., 2014. A column-generation-based heuristic algorithm for solving operating theater planning problem under stochastic demand and surgery cancellation risk. *International Journal of Production Economics* 158, 28–36.
- Wang, Y., Zhang, Y., Tang, J., 2019. A distributionally robust optimization approach for surgery block allocation. *European Journal of Operational Research* 273 (2), 740–753.
- Xiao, Y., Yoogalingam, R., 2021. Reserved capacity policies for operating room scheduling. *Operations Management Research*, 1–16.
- Zhang, J., Dridi, M., El Moudni, A., 2019. A two-level optimization model for elective surgery scheduling with downstream capacity constraints. *European Journal of Operational Research* 276 (2), 602–613.
- Zhu, S., Fan, W., Yang, S., Pei, J., Pardalos, P. M., 2019. Operating room planning and surgical case scheduling: a review of literature. *Journal of Combinatorial Optimization* 37 (3), 757–805.

Appendix A. Proof of Proposition 1

Proof. Recall that $\hat{\mathbb{P}}_\xi^N = \frac{1}{N} \sum_{n=1}^N \delta_{\hat{\xi}^n}$. The definition of Wasserstein distance indicates that there exist a joint distribution Π of $(\xi, \hat{\xi})$ such that $\mathbb{E}_\Pi[\|\xi - \hat{\xi}\|] \leq \epsilon$. In other words, for any $\mathbb{P}_\xi \in \mathcal{P}(\mathcal{S})$, we can rewrite any joint distribution $\Pi \in \mathcal{P}(\mathbb{P}_\xi, \hat{\mathbb{P}}_\xi^N)$ by the conditional distribution of ξ given $\hat{\xi} = \hat{\xi}^n$ for $n = 1, \dots, N$, denoted as \mathbb{Q}_ξ^n . That is, $\Pi = \frac{1}{N} \sum_{n=1}^N (\mathbb{Q}_\xi^n \times \delta_{\hat{\xi}^n})$. Notice that if we find one joint distribution $\Pi \in \mathcal{P}(\mathbb{P}_\xi, \hat{\mathbb{P}}_\xi^N)$ such that $\int \|\xi - \hat{\xi}\| d\Pi \leq \epsilon$, then $\text{dist}(\mathbb{P}_\xi, \hat{\mathbb{P}}_\xi^N) \leq \epsilon$. Hence, we can drop the infimum operator in Wasserstein distance and arrive at the following equivalent problem.

$$\sup_{\mathbb{Q}_\xi^n \in \mathcal{P}(\mathcal{S}), n \in [N]} \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{S}} f(\mathbf{y}, \xi) d\mathbb{Q}_\xi^n \quad (\text{A.1a})$$

$$\text{s.t.} \quad \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{S}} \|\xi - \hat{\xi}^n\| d\mathbb{Q}_\xi^n \leq \epsilon \quad (\text{A.1b})$$

Using a standard strong duality argument and letting $\rho \geq 0$ be the dual multiplier, we can reformulate problem (A.1) by its dual, i.e.

$$\begin{aligned} & \inf_{\rho \geq 0} \sup_{\mathbb{Q}_\xi^n \in \mathcal{P}(\mathcal{S}), n \in [N]} \left\{ \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{S}} f(\mathbf{y}, \xi) d\mathbb{Q}_\xi^n + \rho \left[\epsilon - \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{S}} \|\xi - \hat{\xi}^n\| d\mathbb{Q}_\xi^n \right] \right\} \\ &= \inf_{\rho \geq 0} \left\{ \epsilon \rho + \frac{1}{N} \sum_{n=1}^N \sup_{\mathbb{Q}_\xi^n \in \mathcal{P}(\mathcal{S})} \int_{\mathcal{S}} [f(\mathbf{y}, \xi) - \rho \|\xi - \hat{\xi}^n\|] d\mathbb{Q}_\xi^n \right\} \\ &= \inf_{\rho \geq 0} \left\{ \epsilon \rho + \frac{1}{N} \sum_{n=1}^N \sup_{\xi \in \mathcal{S}} \{f(\mathbf{y}, \xi) - \rho \|\xi - \hat{\xi}^n\|\} \right\}. \end{aligned} \quad (\text{A.2})$$

□

Appendix B. Proof of Lemma 1 Adapted from Jiang et al. (2019) and Theorem 2 in Fournier and Guillin (2015)

Lemma 1. (Adapted from Jiang et al. (2019) and Theorem 2 in Fournier and Guillin (2015)). Suppose that Assumption A6 holds. Then there exist none-negative constants c_1 and c_2 such that, for all $N \geq 1$ and $\beta \in (0, \min\{1, c_1\})$,

$$\mathbb{P}^N \left\{ \text{dist}(\mathbb{P}_\xi, \hat{\mathbb{P}}_\xi^N) \leq \epsilon_N(\beta) \right\} \geq 1 - \beta$$

where \mathbb{P}_ξ^n represents the product measure of N copies of \mathbb{P}_ξ and $\epsilon_N(\beta) = \left[\frac{\log(c_1 \beta^{-1})}{c_2 N} \right]^{\frac{1}{\max\{3p, n\}}}$

Proof. For completeness, in what follows, we provide the proof of Lemma 1 of Jiang et al. (2019) as detailed in their paper, which is adapted from Theorem 2 in Fournier and Guillin (2015).

First, given that the support is finite and compact (Assumption A6), then there exists $\alpha > p$ and $\gamma > 0$ such that $\mathbb{E}_{\mathbb{P}_\xi}[\exp\{\gamma \|\xi\|_p^\alpha\}] < \infty$. By Theorem 2 of Fournier and Guillin (2015), for any $N \geq 1$ and $\epsilon \in (0, \infty)$, there exists positive constants c and C depending only on p , n , α and γ such that

$$\mathbb{P}_\xi^N \left(\text{dist}(\mathbb{P}_\xi, \hat{\mathbb{P}}_\xi^N) \geq \epsilon^{1/p} \right) \leq a(N, \epsilon) \mathbb{1}_{(\epsilon \leq 1)} + b(N, \epsilon),$$

where

$$a(N, \epsilon) = C \begin{cases} \exp\{-cN\epsilon^2\} & \text{if } p > n/2 \\ \exp\{-cN(\epsilon/\log(2+1/\epsilon))^2\} & \text{if } p = n/2 \\ \exp\{-cN\epsilon^{n/p}\} & \text{if } p \in [1, n/2) \end{cases} \quad (\text{B.1})$$

and $b(N, \epsilon) = C \exp\{-cN\epsilon^{\alpha/p}\} \mathbb{1}_{(\epsilon > 1)}$, where n is the dimension of the random vector ξ . Second, to derive bounds on $a(N, \epsilon)$ and $b(N, \epsilon)$, we discuss the following two cases based on the value of ϵ

Case 1. $\epsilon \in (0, 1]$. Note that

$$\left[\frac{\epsilon}{\log(2+1/\epsilon)} \right]^2 \geq \frac{\epsilon^3}{[\log(3)]^2},$$

It follows that $\epsilon[\log(2+1/\epsilon)]^2 \leq [\log(3)]^2$. Thus, we can have the following upper bound on $a(N, \epsilon)$

$$a(N, \epsilon) \leq C \exp \left\{ -\frac{c}{[\log(3)]^2} N \epsilon^{\max\{3, n/p\}} \right\}.$$

Case 2. $\epsilon > 1$. Take $\alpha = \max\{3p, n\} > p$. It follows that

$$b(N, \epsilon) \leq C \exp \left\{ -cN\epsilon^{\max\{3, n/p\}} \right\}.$$

Summarizing the above two cases and letting $c_1 = C$ and $c_2 = c/[\log(3)]^2$, we derive

$$\mathbb{P}_\xi^N \left(\text{dist}(\mathbb{P}_\xi, \hat{\mathbb{P}}_\xi^N) \geq \epsilon^{1/p} \right) \leq c_1 \exp \left\{ -c_2 N \epsilon^{\max\{3, n/p\}} \right\},$$

By equating the right hand side of the last inequality to β and solving for $\epsilon^{1/p}$, we obtain $\epsilon = [(c_2 N)^{-1} \log(c_1 \beta^{-1})]^{-\max\{3, n/p\}^{-1}}$. Plugging this expression to the last inequality, we get the desired result. Note that in our ambiguity set $p = 1$. □

Appendix C. Proof of Asymptotic Consistency

Theorem 1. (Asymptotic consistency, adapted from [Jiang et al. \(2019\)](#) and Theorem 3.6 of [Mohajerin Esfahani and Kuhn \(2018\)](#)). *Suppose that Assumption A6 holds. Consider a sequence of confidence levels $\{\beta_N\}_{N \in \mathbb{R}}$ such that $\sum_{N=1}^{\infty} \beta_N < \infty$ and $\lim_{N \rightarrow \infty} \epsilon_N(\beta_N) = 0$, and let $(\hat{y}(N, \epsilon_N(\beta_N)))$ represents an optimal solution to W -DRO with the ambiguity set $\mathcal{F}(\hat{\mathbb{P}}_{\xi}^N, \epsilon_N(\beta_N))$. Then, $\mathbb{P}_{\xi}^{\infty}$ -almost surely we have $\hat{Z}(N, \epsilon_N(\beta_N)) \rightarrow Z^*$ as $N \rightarrow \infty$. In addition, any accumulation points of $\{\hat{y}(N, \epsilon_N(\beta_N))\}_{N \in \mathbb{N}}$ is an optimal solution of (4) $\mathbb{P}_{\xi}^{\infty}$ -almost surely*

Proof. Recall the dual of $f(y, \xi)$.

$$f(y, \xi) := \max_{\beta} \left\{ \sum_{b \in B} \left(\sum_{i \in I} d_i y_{i,b} + e_b - \mathcal{L}_b \right) \beta_b \right\}$$

$$\text{s.t. } \beta \in \mathcal{B} = \{-c_b^g \leq \beta_b \leq c_b^o\}$$

First, for fixed $y \in \mathcal{Y}$, this is an LP in β . Note that \mathcal{B} is bounded and \mathcal{Y} is a bounded set, which implies that $f(y, \xi)$ is finite and bounded. It follows that we have $|f(y, \xi)| \leq M(1 + \|\xi\|)$, where M is an upper bound on $f(\cdot)$. Second, we claim that $f(y, \xi)$ is continuous on \mathcal{Y} . For simplicity, write the objective function of the dual problem as $b(y, \xi)^{\top} \beta$, where the b -th entry of $b(y, \xi)$ is the coefficient associated to β_b . By fundamental theorem of LP, instead of maximizing over the entire \mathcal{B} , we can maximize over the set of finite extreme points of \mathcal{B} . Hence, $f(y, \xi)$ is the maximum of finite number of linear functions of y and ξ and the continuity follows. Indeed, we can show that for fixed $y \in \mathcal{Y}$, $f(y, \xi)$ is Lipschitz in ξ . Let $(\xi^1, \xi^2) \in \mathcal{S}$. Then,

$$\begin{aligned} f(y, \xi^1) - f(y, \xi^2) &= \max_{\beta \in \mathcal{B}} b(y, \xi^1)^{\top} \beta - \max_{\beta \in \mathcal{B}} b(y, \xi^2)^{\top} \beta \leq \max_{\beta \in \mathcal{B}} [b(y, \xi^1) - b(y, \xi^2)]^{\top} \beta \\ &\leq \max_{\beta \in \mathcal{B}} \|\beta\|_q \|b(y, \xi^1) - b(y, \xi^2)\|_p, \end{aligned} \tag{C.2}$$

where the last inequality follows from Holder's inequality. Notice that $\max_{\beta \in \mathcal{B}} \|\beta\|_q$ is bounded since \mathcal{B} is bounded. Since $p \geq 1$, then for any $b \in B$, $y \in \mathcal{Y}$ and $\xi^1, \xi^2 \in \mathcal{S}$

$$\begin{aligned} f_b(y, \xi^1) - f_b(y, \xi^2) &= \max_{\beta \in \mathcal{B}} \left\{ \left(\sum_{i \in I} d_i^1 y_{i,b} + e_b^1 - \mathcal{L}_b \right) \beta_b \right\} - \max_{\beta \in \mathcal{B}} \left\{ \left(\sum_{i \in I} d_i^2 y_{i,b} + e_b^2 - \mathcal{L}_b \right) \beta_b \right\} \\ &\leq \max_{\beta \in \mathcal{B}} \left\{ \left(\sum_{i \in I} d_i^1 y_{i,b} + e_b^1 - \mathcal{L}_b \right) \beta_b - \left(\sum_{i \in I} d_i^2 y_{i,b} + e_b^2 - \mathcal{L}_b \right) \beta_b \right\} \end{aligned}$$

$$\begin{aligned}
&= \max_{\beta \in \mathcal{B}} \beta_b [(\mathbf{d}^1 - \mathbf{d}^2) + (e_b^1 - e_b^2)] \\
&\leq \max_{\beta \in \mathcal{B}} \|\beta\|_q \|\boldsymbol{\xi}^1 - \boldsymbol{\xi}^2\|_p
\end{aligned} \tag{C.3}$$

$$= C \|\boldsymbol{\xi}^1 - \boldsymbol{\xi}^2\|_p \tag{C.4}$$

$$\tag{C.5}$$

where $C = \max_{\beta \in \mathcal{B}} \beta_b$ and the last inequality follows from Holder's inequality. This verifies that $f(y, \boldsymbol{\xi})$ is Lipschitz in $\boldsymbol{\xi}$. Finally, the result follows directly with $p = 1$ from Theorem 3.6 of [Mohajerin Esfahani and Kuhn \(2018\)](#), where the required conditions are shown. \square

Appendix D. Proof of Finite-data guarantee

Theorem 2. (Finite-data guarantee, adapted from [Jiang et al. \(2019\)](#) and Theorem 3.5 in [Mohajerin Esfahani and Kuhn \(2018\)](#)). *For any $\beta \in (0, 1)$, let $\hat{y}(N, \epsilon_N(\beta_N))$, represent an optimal solution of W-DRO with ambiguity set $\mathcal{F}(\hat{\mathbb{P}}_\xi^N, \epsilon_N(\beta_N))$. Then,*

$$\mathbb{P}_\xi^N \{ \mathbb{E}_{\mathbb{P}_\xi} [f(\hat{y}(N, \epsilon_N(\beta_N), \xi)] \leq \hat{Z}(N, \epsilon_N(\beta_N)) \} \geq 1 - \beta.$$

Proof. (adapted from [Jiang et al. \(2019\)](#)). By Assumption A6 on the support \mathcal{S} and Lemma 1 all conditions of Theorem 3.5 in [Mohajerin Esfahani and Kuhn \(2018\)](#) are satisfied. Therefore, the conclusions of Theorem 2 hold valid. \square

Appendix E. Moment Model (M-DRO)

In this section, we drive a DRO model for DSA based on the known mean and support (\mathcal{S}) of random parameters (we denote this model as M-DRO). The mean-support ambiguity set has been extensively employed in the healthcare scheduling literature for the following two primary reasons. First, from a clinical point of view and as reported in several studies, the mean μ , upper and lower bounds of surgery duration are intuitive statistics that can be easily approximated based on clinicians or OR manager experiences or from available data. Second, incorporating higher moments often undermines the DRO model's computational tractability, and therefore its applicability in practice. *We remark that there is no mean-support DRO model for the specific surgery planning in flexible ORs problem that we study in this paper (see Section 2).*

Let us introduce some additional notation defining our M-DRO model. First, we let μ^d and μ^e represent the mean values of \mathbf{d} and \mathbf{e} , respectively, and denote $\mu := \mathbb{E}_{\mathbb{P}}[\xi] = [\mu^d, \mu^e]^\top$ for notational brevity. Then, we consider the following mean-support ambiguity set $\mathcal{F}(\mathcal{S}, \mu)$:

$$\mathcal{F}(\mathcal{S}, \mu) := \left\{ \mathbb{P} \in \mathcal{P}(\mathcal{S}) \left| \begin{array}{l} \int_{\mathcal{S}} d\mathbb{P} = 1 \\ \mathbb{E}_{\mathbb{P}}[\xi] = \mu \end{array} \right. \right\}, \quad (\text{E.1})$$

where $\mathcal{P}(\mathcal{S})$ in $\mathcal{F}(\mathcal{S}, \mu)$ represents the set of probability distributions supported on \mathcal{S} with mean μ . Using the ambiguity set $\mathcal{F}(\mathcal{S}, \mu)$, we formulate M-DRO as the following min-max problem:

$$\min_{\mathbf{y} \in \mathcal{Y}} \left\{ \sum_{i \in I} \sum_{b \in [B] \cup \{b'\}} c_{i,b} y_{i,b} + \sup_{\mathbb{P} \in \mathcal{F}(\mathcal{S}, \mu)} \mathbb{E}_{\mathbb{P}}[f(\mathbf{y}, \xi)] \right\}. \quad (\text{E.2})$$

Next, we derive an equivalent reformulation of problem (E.2) that is solvable. We first consider the inner maximization problem $\sup_{\mathbb{P} \in \mathcal{F}(\mathcal{S}, \mu)} \mathbb{E}_{\mathbb{P}}[f(\mathbf{y}, \xi)]$ for a fixed first-stage decision $\mathbf{y} \in \mathcal{Y}$, where \mathbb{P} is the decision variable, i.e., we are choosing the distribution that maximizes $\mathbb{E}_{\mathbb{P}}[f(x, \xi)]$. For a fixed $\mathbf{y} \in \mathcal{Y}$, we can formulate this problem as the following linear functional optimization problem.

$$\max_{\mathbb{P} \geq 0} \int_{\mathcal{S}} f(\mathbf{y}, \xi) d\mathbb{P} \quad (\text{E.3a})$$

$$\text{s.t.} \quad \int_{\mathcal{S}} d_i d\mathbb{P} = \mu_i^d, \quad \forall i \in [I], \quad (\text{E.3b})$$

$$\int_{\mathcal{S}} e_b d\mathbb{P} = \mu_b^e, \quad \forall b \in [B], \quad (\text{E.3c})$$

$$\int_{\mathcal{S}} d\mathbb{P} = 1. \quad (\text{E.3d})$$

Letting $\boldsymbol{\rho} = [\rho_1, \dots, \rho_I]^\top$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_B]^\top$, and θ be the dual variable associated with constraints (E.3b), (E.3c), and (E.3d), respectively, we present problem (E.3) in its dual form:

$$\min_{(\boldsymbol{\rho}, \boldsymbol{\alpha}), \theta \in \mathbb{R}} \left\{ \sum_{i \in I} \mu_i^d \rho_i + \sum_{b \in B} \mu_b^e \alpha_b + \theta \right\} \quad (\text{E.4a})$$

$$\text{s.t. } \sum_{i \in I} d_i \rho_i + \sum_{b \in B} e_b \alpha_b + \theta \geq f(\mathbf{y}, \mathbf{d}, \mathbf{e}), \quad \forall (\mathbf{d}, \mathbf{e}) \in \mathcal{S}, \quad (\text{E.4b})$$

where $\boldsymbol{\rho}$, $\boldsymbol{\alpha}$, and θ are unrestricted in sign, and constraint (E.4b) is associated with the primal variable \mathbb{P} , and strong duality hold between (E.3) and (E.4) (see Bertsimas and Popescu (2005); Shapiro et al. (2014)).

Note that for fixed $(\boldsymbol{\rho}, \boldsymbol{\alpha}, \theta)$, constraints (E.4b) are equivalent to $\theta \geq \max_{(\mathbf{d}, \mathbf{e}) \in \mathcal{S}} \{f(\mathbf{y}, \mathbf{d}, \mathbf{e}) - (\sum_{i=1}^P d_i \rho_i + \sum_{b \in B} e_b \alpha_b)\}$. Since we are minimizing θ in (E.4), the dual formulation of (E.3) is equivalent to:

$$\min_{(\boldsymbol{\rho}, \boldsymbol{\alpha})} \left\{ \sum_{i \in I} \mu_i^d \rho_i + \sum_{b \in B} \mu_b^e \alpha_b + \max_{(\mathbf{d}, \mathbf{e}) \in \mathcal{S}} \{f(\mathbf{y}, \mathbf{d}, \mathbf{e}) - (\sum_{i \in I} d_i \rho_i + \sum_{b \in B} e_b \alpha_b)\} \right\} \quad (\text{E.5})$$

Using the dual formulation of $f(\mathbf{y}, \mathbf{d}, \mathbf{e})$ in (11) and the same techniques we used to derive (14), we derive the following equivalent reformulation of (E.5).

$$\min_{(\boldsymbol{\rho}, \boldsymbol{\alpha})} \left\{ \sum_{i \in I} \mu_i^d \rho_i y_{i,b} + \sum_{b \in B} \mu_b^e \alpha_b + \sum_{b \in B} \eta_b \right\} \quad (\text{E.6a})$$

$$\text{s.t. } \eta_b \geq \left[\sum_{i \in I} \bar{d}_i y_{i,b} + \bar{e}_b - \mathcal{L}_b \right] (c_b^o) - \left(\sum_{i \in I} \bar{d}_i \rho_i y_{i,b} + \bar{e}_b \alpha_b \right), \quad \forall b \in [B], \quad (\text{E.6b})$$

$$\eta_b \geq \left[\sum_{i \in I} \bar{d}_i y_{i,b} + \underline{e}_b - \mathcal{L}_b \right] (c_b^o) - \left(\sum_{i \in I} \bar{d}_i \rho_i y_{i,b} + \underline{e}_b \alpha_b \right), \quad \forall b \in [B], \quad (\text{E.6c})$$

$$\eta_b \geq \left[\sum_{i \in I} \bar{d}_i y_{i,b} + \bar{e}_b - \mathcal{L}_b \right] (-c_b^g) - \left(\sum_{i \in I} \bar{d}_i \rho_i y_{i,b} + \bar{e}_b \alpha_b \right), \quad \forall b \in [B], \quad (\text{E.6d})$$

$$\eta_b \geq \left[\sum_{i \in I} \bar{d}_i y_{i,b} + \underline{e}_b - \mathcal{L}_b \right] (-c_b^g) - \left(\sum_{i \in I} \bar{d}_i \rho_i y_{i,b} + \underline{e}_b \alpha_b \right), \quad \forall b \in [B], \quad (\text{E.6e})$$

$$\eta_b \geq \left[\sum_{i \in I} \underline{d}_i y_{i,b} + \bar{e}_b - \mathcal{L}_b \right] (c_b^o) - \left(\sum_{i \in I} \underline{d}_i \rho_i y_{i,b} + \bar{e}_b \alpha_b \right), \quad \forall b \in [B], \quad (\text{E.6f})$$

$$\eta_b \geq \left[\sum_{i \in I} \underline{d}_i y_{i,b} + \underline{e}_b - \mathcal{L}_b \right] (c_b^o) - \left(\sum_{i \in I} \underline{d}_i \rho_i y_{i,b} + \underline{e}_b \alpha_b \right), \quad \forall b \in [B], \quad (\text{E.6g})$$

$$\eta_b \geq \left[\sum_{i \in I} \underline{d}_i y_{i,b} + \bar{e}_b - \mathcal{L}_b \right] (-c_b^g) - \left(\sum_{i \in I} \underline{d}_i \rho_i y_{i,b} + \bar{e}_b \alpha_b \right), \quad \forall b \in [B], \quad (\text{E.6h})$$

$$\eta_b \geq \left[\sum_{i \in I} \underline{d}_i y_{i,b} + \underline{e}_b - \mathcal{L}_b \right] (-c_b^g) - \left(\sum_{i \in I} \underline{d}_i \rho_i y_{i,b} + \underline{e}_b \alpha_b \right), \quad \forall b \in [B]. \quad (\text{E.6i})$$

Combining the inner maximization problem in the form of (E.6) with the outer minimization problem in (E.2), we derive the following equivalent reformulation of the M-DRO model in (E.2) as follow

$$\min \left\{ \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \sum_{i \in I} \mu_i^d \rho_i y_{i,b} + \sum_{b \in B} \mu_b^e \alpha_b + \sum_{b \in B} \eta_b \right\} \quad (\text{E.7a})$$

$$\text{s.t. } \mathbf{y} \in \mathcal{Y}, \quad (\text{E.6b}) - (\text{E.6i}) \quad (\text{E.7b})$$

Note that the objective function (E.7a) and constraints (E.6b)–(E.6i) contains the interaction terms $\rho_i y_{i,b}$ with binary variables $y_{i,b}$ and continuous variables ρ_i . To linearize, we define $k_{i,b} = \rho_i y_{i,b}$ and introduce the following McCormick inequalities for variables $k_{i,b}$, for all $i \in [I]$ and $b \in [B]$

$$k_{i,b} \geq \underline{\rho}_i y_{i,b} \quad k_{i,b} \geq \bar{\rho}_i (y_{i,b} - 1) + \rho_i \quad (\text{E.8a})$$

$$k_{i,b} \leq \bar{\rho}_i y_{i,b} \quad k_{i,b} \leq \underline{\rho}_i (y_{i,b} - 1) + \rho_i \quad (\text{E.8b})$$

where $\underline{\rho}_i$ and $\bar{\rho}_i$ are respectively valid lower and upper bounds on variables ρ_i , for all $i \in [I]$. Accordingly, formulation (E.7) is equivalent to the following MILP:

$$\min \left\{ \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \sum_{i=1}^I \mu_i^d k_{i,b} + \sum_{b \in B} \mu_b^e \alpha_b + \sum_{b \in B} \eta_b \right\} \quad (\text{E.9a})$$

$$\text{s.t. } \mathbf{y} \in \mathcal{Y}, \quad (\text{E.6b}) - (\text{E.6i}), \quad (\text{E.8a}) - (\text{E.8b}) \quad (\text{E.9b})$$

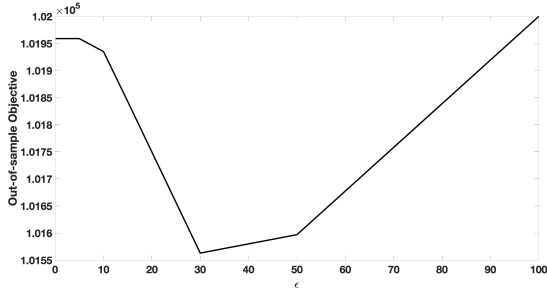
Appendix F. The Sample Average Approximation (SAA) Formulation

$$\min_{\mathbf{y} \in \mathcal{Y}} \left\{ \sum_{i \in I} \sum_{b \in B \cup \{b'\}} c_{i,b} y_{i,b} + \sum_{n=1}^N \frac{1}{N} \left[\sum_{b \in B} (c_b^o o_b^n + c_b^g o_g^n) \right] \right\} \quad (\text{F.1a})$$

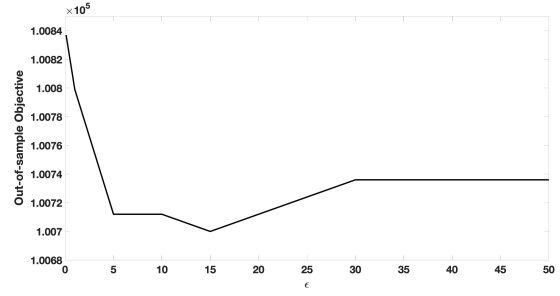
$$\text{s.t. } o_b^n - g_b^n = \sum_{i \in I} d_i^n y_{i,b} + e_b^n - \mathcal{L}_b, \quad \forall b \in B, \forall n \in [N], \quad (\text{F.1b})$$

$$(o_b^n, g_b^n) \geq 0, \quad \forall b \in B, \forall n \in [N]. \quad (\text{F.1c})$$

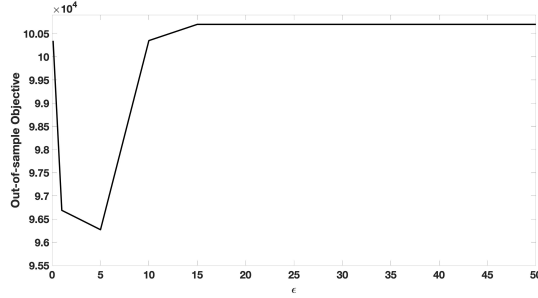
Appendix G. Effect of ϵ , $I = 80$ Surgeries



(a) $N = 5$

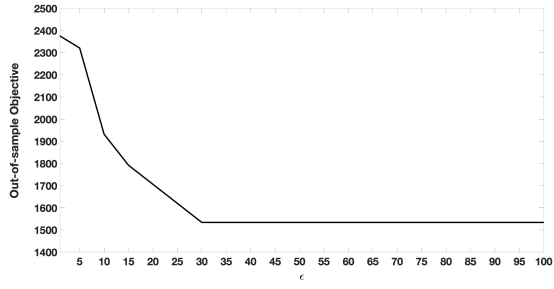


(b) $N = 10$

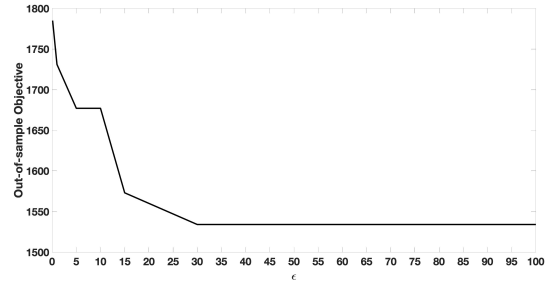


(c) $N = 50$

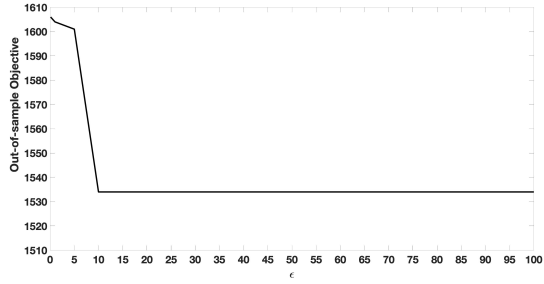
Figure G.8: Out-of-sample performance as a function of the Wasserstein radius for an instance of $I = 80$ under Cost1



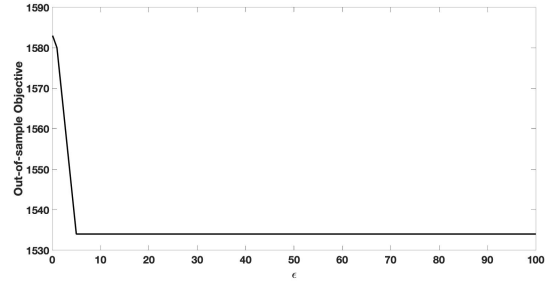
(a) $N = 5$



(b) $N = 10$



(c) $N = 50$



(d) $N = 100$

Figure G.9: Out-of-sample performance as a function of the Wasserstein radius for an instance of $I = 80$ under Cost2