Relief-based Anesthesiologist Scheduling with Stochastic Surgery Durations

Abhishrut Sinha* Ankit Bansal† Osman Ozaltin‡ Michael Russell§

Abstract

We present a two-stage stochastic programming model for scheduling anesthesiologists to operating rooms under uncertainty in surgery durations. The proposed model takes a relief order to balance anesthesiologists' workload as input and captures the trade-offs between anesthesiologist *relief times*, handoffs and under-staffing. To address the computational challenges of solving the proposed model, we derive supervalid equalities that exploit the structure of the second stage. Subsequently, we develop a tight monolithic reformulation whose size, in terms of variables and constraints, is independent of the number of scenarios. To further strengthen the formulation, we introduce valid inequalities for the first-stage problem and show that they describe convex hull of binary sets induced by subsets of the first-stage constraints. Our computational experiments, based on data from a tertiary medical center, show that the proposed reformulation significantly outperforms both the extensive form and the L-shaped algorithm. Compared with current practice, the proposed framework reduces handoffs and under-staffed periods while allowing anesthesiologists to be relieved earlier under uncertain surgery durations. Ultimately, the proposed approach supports data-driven workforce planning and policy design to align anesthesiologist schedules with institutional priorities to reduce handoffs and under-staffing.

Keywords: OR in health services; Anesthesiologist scheduling; Relief times; Two-stage stochastic program; Valid inequalities

*Department of Systems Science and Industrial Engineering, State University of New York, Binghamton, NY, United States

†Department of Systems Science and Industrial Engineering, State University of New York, Binghamton (*Corresponding author*)

‡Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, NC, United States

§Department of Anesthesiology, West Virginia University School of Medicine, WV, United States

1 Introduction

Anesthesiology is a critical component of surgical care, with anesthesiologists responsible for the safe administration of anesthesia, continuous monitoring of vital signs, and rapid intervention in response to perioperative complications (Menezes and Zahalka 2024). In the United States, millions of surgeries are performed each year, yet the healthcare system is experiencing a significant shortage of anesthesiologists, a problem that has intensified since the COVID-19 pandemic (Abouleish et al. 2024, Afonso et al. 2024). This shortage not only poses risks to patient health and safety when surgery schedules are disrupted due to the unavailability of anesthesiologists, but also drives up healthcare costs as hospitals become increasingly reliant on costly temporary anesthesia providers to meet demand (Abouleish et al. 2024, Menezes and Zahalka 2024).

This challenge is further exacerbated by a self-perpetuating cycle. Staffing shortages increase workload, which has been identified as a major contributor to anesthesiologist burnout (Afonso et al. 2024). In turn, anesthesiologists experiencing high levels of burnout are more likely to leave their positions, further deepening workforce shortages. Consequently, a vicious cycle emerges: shortages increase workload, which leads to burnout, and burnout accelerates attrition, thereby intensifying the shortage problem (Abouleish et al. 2024, Afonso et al. 2024, 2021).

An important step toward interrupting this cycle is to focus on retaining the anesthesiologists already in the workforce (Abouleish et al. 2024, Afonso et al. 2024). One approach to supporting this effort is to ensure an even workload among anesthesiologists. In this context, workload balance refers to distributing end-of-service times, also known as *relief times*, evenly among anesthesiologists across days. This prevents any anesthesiologist from being systematically burdened with disproportionate duties, thereby enhancing job satisfaction and supporting retention.

A practical mechanism for balancing workload through *relief times* is a rotation-based relief system (Wax and McCormick 2017, Bhutiani et al. 2018). In this setting, all anesthesiologists begin their duties at the start of the day, and their *relief times* are determined by a predefined relief order. This relief order specifies the sequence in which anesthesiologists are relieved, ensuring that no anesthesiologist is relieved before those preceding them in the sequence. For example, if the set of anesthesiologists is $\{1, 2, 3\}$ and the relief order for the day is $1 \to 2 \to 3$, then anesthesiologist 3 is not relieved before 2 and anesthesiologist 2 is not relieved before 1. The relief order is usually determined based on the *relief times* from the previous day and ensures that the earlier relief opportunity is distributed evenly among anesthesiologists, preventing any anesthesiologist from consistently experiencing extended work hours

(Wax and McCormick 2017).

For a given relief order, scheduling anesthesiologists involves two interrelated decisions: determining each anesthesiologist's *relief time* and assigning anesthesiologists to ORs for the periods prior to their relief. These decisions are shaped by the daily surgical schedule and uncertainty in surgery durations, and must balance competing operational considerations. For instance, a key consequence of these decisions is the occurrence of handoffs—transfers of responsibility for an OR from one anesthesiologist to another. Handoffs occur when an anesthesiologist assigned to an OR is relieved before the final surgery in that OR is completed, requiring another provider to take over. Such transitions can disrupt continuity of care, increase coordination effort, and are associated with adverse patient health outcomes (Boet et al. 2020, Saager et al. 2014). Delaying relief can reduce the frequency of handoffs, but increases anesthesiologist workload. Another critical operational consideration is under-staffing: if surgeries are ongoing but no anesthesiologist is available, additional coverage—often from on-call providers—must be arranged, leading to additional costs. Again, postponing relief can help mitigate under-staffing at the expense of a higher anesthesiologist workload. Overall, determining anesthesiologists' *relief times* and OR assignments requires carefully balancing the trade-offs among evenly distributing anesthesiologists' workload, minimizing handoffs, and preventing under-staffing.

To address these challenges, we propose a two-stage stochastic program for scheduling anesthesiologists to ORs, which we refer to as the Stochastic Anesthesiologist Scheduling with Handoffs (SASH). Given the daily allocation of surgeries to ORs and their scheduled start times, the proposed model determines, in the first stage, the *relief times* of anesthesiologists and assigns them to ORs for the periods prior to their relief. The second stage, given the first-stage decisions and a realization of surgery durations, evaluates the handoff and under-staffing costs. The model minimizes the combined cost of delaying anesthesiologists' *relief times*, along with the expected costs of handoffs and under-staffing. To the best of our knowledge, this is the first study to enforce a predetermined relief order to manage anesthesiologists' workload while explicitly modeling the trade-offs among delayed relief, handoffs, and under-staffing.

The proposed model, SASH, is a two-stage stochastic program with binary variables in the first stage and continuous variables in the second stage. To solve large-scale instances of our model, we derive supervalid equalities that exploit the structure of the second stage. Leveraging these equalities, we develop a strengthened monolithic reformulation with scenario-independent variables and constraints, that is, a formulation in which the number of variables and constraints does not grow with the number of scenarios. We also derive valid inequalities for the first stage and show that they characterize convex

hull of binary sets defined by subsets of the first-stage constraints. Computational results show that combining the monolithic reformulation with the proposed valid inequalities yields significant computational gains, outperforming both direct solution of the extensive form of `SASH` using mixed-integer programming solvers and the L-shaped algorithm (Birge and Louveaux 2011).

The remainder of this paper is organized as follows. Section 2 presents a review of the relevant literature. Section 3 presents the proposed anesthesiologist scheduling model and its monolithic reformulation. Section 4 details the derivation of valid inequalities. Section 5 presents computational experiments and managerial insights. Section 6 concludes the paper.

2 Literature Review

Section 2.1 reviews the literature on healthcare provider scheduling and Section 2.2 examines solution approaches for two-stage stochastic programs with continuous recourse. We summarize the main contributions of this study in Section 2.3.

2.1 Healthcare Provider Scheduling

Addressing workload has been a key consideration in healthcare scheduling problems, as workload is a primary driver of provider burnout and attrition (Afonso et al. 2021). Fügener et al. (2015) formulate a multicriteria optimization model for anesthesiologist task and workstation rostering to maximize coverage while penalizing violations of equity constraints. Sun et al. (2023) develop a two-step anesthesiologist scheduling framework under demand uncertainty that optimizes for equitable workload distribution using multiobjective mixed-integer programming. Rath et al. (2026) propose a data-driven optimization model for dynamically assigning anesthesiologists across multiple hospital locations, incorporating fairness constraints for on-call duties. However, these studies focus on shift-level scheduling, where workload is measured by the total number of hours over all assigned shifts. While valuable for strategic and tactical planning, such approaches don't consider operational dynamics affecting workload such as anesthesiologists' *relief time* at the end of a daily surgery schedule. In contrast, we focus on operational-level scheduling within a single day. For a given relief order of providers, we explicitly model their *relief times* under uncertainty in task durations, capturing a more granular notion of workload than is typically considered in the existing literature.

The existing literature on anesthesiologist scheduling in `OR`s (Breuer et al. 2020, Rath et al. 2017, Tsang et al. 2024) primarily focuses on integrated models that jointly determine surgery schedules and

anesthesiologist assignments. While these approaches account for uncertainty in surgery durations, they typically do not explicitly model handoffs between anesthesiologists, instead relying on overtime as the primary recourse to address uncertainty. In contrast, we study the allocation of anesthesiologists to ORs for a given set of surgery-to-OR assignments and surgery start times. Our model explicitly incorporates uncertainty in surgery durations and accounts for handoffs that occur when anesthesiologists are relieved at the end of their daily schedules, thereby capturing operational considerations largely overlooked in existing work.

A limited set of papers consider handoffs in healthcare scheduling and resource allocation. Kazemian et al. (2014) propose an integer programming model to create shift schedules for medical trainees with the aim of reducing patient handoffs. Similarly, Smalley et al. (2015) develop a mixed-integer programming approach for physician scheduling that reduces patient transfers among physicians. However, both studies are limited to deterministic, non-surgical settings. Sun et al. (2023) addresses anesthesiologist scheduling at a tactical level over a multi-week planning horizon. Their model focuses on designing shifts and allocating providers to them. Although handoffs are allowed in their framework, their impact on care quality and patient safety is not considered; thus, reducing handoffs is not a goal in their models. In contrast, we address anesthesiologist scheduling with a focus on reducing handoffs under uncertain surgery durations. Sinha et al. (2025) propose a two-stage stochastic model for assigning certified registered nurse anesthetists (CRNAs) to ORs with the goal of reducing handoffs. However, CRNA scheduling is fundamentally different from the anesthesiologist scheduling problem. In CRNA scheduling, each nurse is assigned to a shift with fixed start and end times, with relief occurring at the end of the shift. During the shift, each nurse is assigned to at most a single OR in each time period. On the other hand, in our setting, anesthesiologists' *relief times* and, hence, the length of their daily schedules are decisions restricted by a given relief order. Moreover, unlike CRNAs, each anesthesiologist may cover multiple ORs within the same time period.

2.2 Solution Methods for Two-stage Stochastic Programs with Continuous Recourse

The extensive form of two-stage stochastic integer programs can be solved using commercial mixed-integer programming solvers. However, this approach is computationally feasible only for instances with a small number of scenarios, as the size of the extensive form—both in terms of variables and constraints—increases with the number of scenarios. Consequently, the L-shaped algorithm is widely applied to problems with continuous recourse (Küçükyavuz and Sen 2017). This approach iteratively

incorporates scenario-specific Benders cuts, either through a classical iterative scheme or within a single branch-and-bound framework. However, the L-shaped algorithm requires solving a subproblem for each scenario, which can be computationally expensive when the number of scenarios is large. Moreover, its efficiency often depends on carefully selecting and strengthening Benders cuts to accelerate convergence (Glomb et al. 2026, Sinha et al. 2025, Aghaei et al. 2025). In this paper, we derive a reformulation of the SASH with scenario-independent variables and constraints whose size does not increase with the number of scenarios. This allows instances of SASH with a large number of scenarios to be solved efficiently using commercial mixed-integer programming solvers. Computational results show that the proposed reformulation significantly outperforms both the extensive form and the L-shaped algorithm.

2.3 Contributions

The main contributions of this paper include:

i. Formulating a two-stage stochastic program for scheduling anesthesiologists to ORs under uncertainty in surgery durations. The proposed model takes relief order to balance anesthesiologists' workload as input and captures the trade-offs between anesthesiologist *relief times*, handoffs and under-staffing. It is broadly applicable to relief order-based multi-period scheduling environments in which handoffs and under-staffing arise due to uncertainty in task durations, conditions common across many medical institutions and in several other industries (Patterson et al. 2004).

ii. Deriving supervalid equalities that exploit the structure of the second-stage problem and proposing a stronger monolithic reformulation with scenario-independent variables and constraints.

iii. Deriving valid inequalities for the first-stage problem and demonstrating that these valid inequalities define convex hull of subsets of first-stage constraints. Our computational experiments demonstrate that the proposed reformulation combined with these valid inequalities outperforms both the extensive form and the L-shaped algorithm.

iv. Demonstrating the application of SASH using surgery and anesthesiologist schedules from a tertiary medical hospital in the eastern United States. Our computational results show that SASH consistently produces schedules that result in fewer handoffs and under-staffed periods than real-life anesthesiologist schedules observed in our dataset. Furthermore, sensitivity analyses reveal

how anesthesiology departments can adjust model parameters to align anesthesiologist schedules with institutional priorities to reduce handoffs and under-staffed periods.

3 Model Formulation

We formulate a two-stage stochastic program for scheduling anesthesiologists to ORs. The proposed model captures the trade-offs between the *relief time* of anesthesiologists, handoffs, and under-staffing, while accounting for the uncertainty in surgery durations. Specifically, given the daily surgery schedules of ORs and a relief order of anesthesiologists to balance their workloads, the first stage determines the *relief times* of anesthesiologists and assigns them to ORs. The second stage calculates the expected costs of anesthesiologist handoffs and under-staffing. The model minimizes the cost of delaying anesthesiologists' *relief times*, along with the expected total cost of handoffs and under-staffing. We refer to this problem as Stochastic Anesthesiologist Scheduling with Handoffs (SASH).

Let I denote the set of anesthesiologists. Without loss of generality, we index the anesthesiologists by their relief order. For example, if $I = \{1, 2, 3\}$, then anesthesiologist 3 should not be relieved before anesthesiologist 2, and anesthesiologist 2 should not be relieved before anesthesiologist 1. The relief order for a given day aims to balance anesthesiologists' workload. This order is determined based on the previous day's *relief times* at our partner institution. It serves as an input parameter for our model; thus, any method for determining the relief order can be incorporated into the proposed framework. Let J be the set of ORs, T the set of time periods, and k the maximum number of ORs that an anesthesiologist can oversee at the same time. Let c_t^r denote the cost of relieving an anesthesiologist at the end of period t. This parameter is a composite cost that represents two distinct effects of extending an anesthesiologist's end-of-service time. The first is the direct staffing cost incurred by keeping the anesthesiologist in service until the end of period t. The second is an implicit cost reflecting the increased workload that comes with a later *relief time*, capturing effects such as burnout and reduced job satisfaction that need to be mitigated. The relative weights of these two components can be calibrated based on institutional priorities. Our model places no restriction on the functional form of c_t^r beyond it being non-negative and non-decreasing in t. Other non-negative cost parameters in our model include c^h, which denotes the cost of a handoff, and c^u, the per-period cost of anesthesiologist under-staffing in an OR. These costs also depend on institutional priorities, as with c_t^r. It is important to note that the relative values of these parameters with respect to one another determine the optimal solutions of our model, rather than their individual values. Therefore,

in real-world applications, they should be calibrated against each other to obtain practical, acceptable solutions.

Uncertainty in surgery durations is captured by a finite set of scenarios Ω, and p_ω is the probability of scenario $\omega \in \Omega$. Let $\bar{t}_j(\omega)$ be the last period of the last surgery in OR j under scenario ω and $\mathcal{T}_j(\omega) = \{1, \ldots, \bar{t}_j(\omega)\}$ denote the set of active periods of `OR` j under scenario ω.

In the first stage, binary variable $y_{i,j,t}$ takes the value 1 if anesthesiologist i is allocated to OR j in period t, and 0 otherwise. Furthermore, $x_{i,t}$ is a binary variable that equals 1 if anesthesiologist i is relieved at the end of period t, and 0 otherwise. In the second stage, $w_{j,t}(\omega)$ is a binary variable equal to 1 if there is a handoff in `OR` j at the end of period t under scenario ω, and 0 otherwise. Table 1 summarizes the notation. We use boldface lowercase letters to denote vectors and boldface uppercase letters to represent matrices. The two-stage stochastic program for `SASH` is given by:

Table 1: Description of Notations

Sets and Parameters	
I	set of anesthesiologists
J	set of `OR`s
T	set of periods
k	maximum number of `OR`s an anesthesiologist can cover in a time period
Ω	set of all scenarios
c_t^r	cost of relieving an anesthesiologist at the end of period t
c^h	cost of a handoff
c^u	per-period cost of anesthesiologist under-staffing in an `OR`
$\bar{t}_j(\omega)$	last period of the last surgery in `OR` j under scenario ω
$\mathcal{T}_j(\omega) \subseteq T$	set of active periods of `OR` j under scenario ω
Indices	
i	index of anesthesiologist, $i \in I$
j	index of `OR`, $j \in J$
t	index of period, $t \in T$
ω	index of scenario, $\omega \in \Omega$
First-stage decision variables	
$x_{i,t}$	1 if anesthesiologist i is relieved at the end of period t, 0 otherwise
$y_{i,j,t}$	1 if anesthesiologist i is allocated to `OR` j in period t, 0 otherwise
Second-stage decision variables	
$w_{j,t}(\omega)$	1 if a handoff occurs in `OR` j at the end of period t under scenario ω, 0 otherwise

$$(\texttt{SASH}) \quad \text{Min} \quad \sum_{i \in I} \sum_{t \in T} c_t^r x_{i,t} + \mathcal{Q}(\mathbf{x}, \mathbf{y}) \tag{1a}$$

subject to :

$$\sum_{i \in I} y_{i,j,t} \leq 1 \qquad \forall j \in J, t \in T, \tag{1b}$$

$$\sum_{t\in T} x_{i,t} = 1 \qquad \forall i \in I, \tag{1c}$$

$$\sum_{\tau=1}^{t} x_{i+1,\tau} \leq \sum_{\tau=1}^{t} x_{i,\tau} \qquad \forall i \in I \setminus \{|I|\}, t \in T, \tag{1d}$$

$$\sum_{j\in J} y_{i,j,t} \leq k \sum_{\tau=t}^{|T|} x_{i,\tau} \qquad \forall i \in I, t \in T, \tag{1e}$$

$$\mathbf{x} \in \mathbb{B}^{|I|\times|T|}, \ \ \mathbf{y} \in \mathbb{B}^{|I|\times|J|\times|T|}. \tag{1f}$$

Constraint (1b) ensures that an OR is allocated at most one anesthesiologist in each period while constraint (1c) ensures that each anesthesiologist is assigned to a *relief time*. Constraint (1d) enforces the relief order by ensuring that anesthesiologist $i+1$ is not relieved before anesthesiologist i for all $i \in I\setminus\{|I|\}$. Constraint (1e) ensures that anesthesiologist i is assigned to at most k ORs in the same period before being relieved. The objective function (1a) minimizes the cost of anesthesiologists' *relief times*, along with the expected total cost of handoffs and under-staffing. Specifically, $\mathcal{Q}(\mathbf{x},\mathbf{y}) \coloneqq \mathbb{E}_\omega Q(\omega,\mathbf{x},\mathbf{y}) = \sum_{\omega\in\Omega} p_\omega Q(\omega,\mathbf{x},\mathbf{y})$, where

$$(SS)\ \ Q(\omega,\mathbf{x},\mathbf{y}) = \text{Min} \sum_{j\in J} \sum_{t\in\mathcal{T}_j(\omega)\setminus\{\bar{t}_j(\omega)\}} c^h w_{j,t}(\omega) + \sum_{j\in J}\sum_{t\in\mathcal{T}_j(\omega)} c^u \left(1 - \sum_{i\in I} y_{i,j,t}\right) \tag{2a}$$

subject to:

$$w_{j,t}(\omega) \geq y_{i,j,t} - y_{i,j,t+1} \qquad \forall i \in I, j \in J, t \in \mathcal{T}_j(\omega)\setminus\{\bar{t}_j(\omega)\}, \tag{2b}$$

$$\mathbf{w}(\omega) \in \mathbb{R}_+^{\sum_{j\in J}(\bar{t}_j(\omega)-1)}. \tag{2c}$$

In the second stage, constraints (2b) account for handoffs in OR j at the end of period t. Specifically, if anesthesiologist i is assigned to OR j in period $t \in \mathcal{T}_j(\omega)\setminus\{\bar{t}_j(\omega)\}$ ($y_{i,j,t}=1$) but is not assigned to that OR in $t+1$ ($y_{i,j,t+1}=0$), then $w_{j,t}(\omega)=1$, indicating a handoff. The first term in (2a) captures the cost of handoffs while the second term represents the under-staffing cost across all ORs. The extensive form of SASH is given by:

$$(\texttt{ExF})\ \ \text{Min} \sum_{i\in I}\sum_{t\in T} c_t^r x_{i,t} + \sum_{\omega\in\Omega} p_\omega \left(\sum_{j\in J} \sum_{t\in\mathcal{T}_j(\omega)\setminus\{\bar{t}_j(\omega)\}} c^h w_{j,t}(\omega) + \sum_{j\in J}\sum_{t\in\mathcal{T}_j(\omega)} c^u \left(1 - \sum_{i\in I} y_{i,j,t}\right)\right)$$

subject to :

(1b) – (1f),

(2b) – (2c), $\forall\omega\in\Omega$.

This model can be solved using mixed-integer programming solvers. However, it includes a set of decision variables $w_{j,t}(\omega)$ and constraints (2b)–(2c) for each scenario. Therefore, as the number of scenarios increases, the size of ExF grows, making it computationally expensive to solve. To overcome this limitation, we exploit the structure of the second-stage problem to derive a stronger monolithic reformulation of ExF with scenario-independent variables and constraints. We establish the validity of the reformulation by means of a set of *supervalid* equalities, a concept we formalize next.

Definition 1. *An equality is supervalid for a mixed-integer program (MIP) if adding it to the MIP eliminates integer feasible solutions, but does not eliminate an optimal solution.*

For each $j \in J$ and $t \in T$, let $\mathcal{O}_{j,t} = \{\omega \in \Omega \mid t \in \mathcal{T}_j(\omega) \setminus \{\bar{t}_j(\omega)\}\}$ be the set of all scenarios for which a handoff in OR j at the end of period t is possible. Proposition 1 establishes that, at optimality, the values of second-stage variables $w_{j,t}(\omega)$ are identical across all scenarios in $\mathcal{O}_{j,t}$, yielding supervalid equalities for ExF.

Proposition 1. *For each $j \in J$ and $t \in T$, the equalities*

$$w_{j,t}(\omega_1) = w_{j,t}(\omega_2) \qquad \forall \omega_1, \omega_2 \in \mathcal{O}_{j,t} : \omega_1 \neq \omega_2 \tag{3}$$

are supervalid for ExF.

Proof. For any feasible first-stage solution $(\mathbf{x}, \mathbf{y})$, the optimal value of $w_{j,t}(\omega)$ in SS is determined by constraints (2b). Specifically, as $w_{j,t} \geq 0$ and $c^h \geq 0$, at optimality $w^*_{j,t}(\omega) = \underset{i \in I}{max} \, (y_{i,j,t} - y_{i,j,t+1})^+$, where $a^+ = max(a, 0)$. Now consider any two scenarios $\omega_1, \omega_2 \in \mathcal{O}_{j,t}$. By definition, t is an active period for OR j under both scenarios where a handoff may occur. As $w^*_{j,t}(\omega)$ depends only on the first-stage variables $y_{i,j,t}$ and $y_{i,j,t+1}$, we have $w^*_{j,t}(\omega_1) = \underset{i \in I}{max} \, (y_{i,j,t} - y_{i,j,t+1})^+ = w^*_{j,t}(\omega_2)$. □

The following example shows that (3) can eliminate integer feasible solutions to ExF, but not all optimal solutions.

Example 1. *Consider an instance of* ExF *with $I = \{1\}$, $J = \{1\}$, $T = \{1, 2, 3\}$, $\Omega = \{1, 2\}$, $k = 1$, $\bar{t}_1(1) = \bar{t}_1(2) = 3$. The solution, $x_{1,3} = y_{1,1,1} = y_{1,1,2} = y_{1,1,3} = w_{1,1}(1) = 1$ and $x_{1,1} = x_{1,2} = w_{1,1}(2) = w_{1,2}(1) = w_{1,2}(2) = 0$ is feasible to* ExF *but violates the supervalid equality* (3)*: $w_{1,1}(1) = w_{1,1}(2)$. Furthermore, the optimal solution with $w_{1,1}(1) = w_{1,1}(2) = 0$ satisfies this supervalid equality.*

Including supervalid equalities (3) in ExF is equivalent to replacing $w_{j,t}(\omega)$ for all $\omega \in \mathcal{O}_{j,t}$ by a single auxiliary variable $\eta_{j,t}$. This yields a stronger reformulation of ExF, which is formalized in Proposition 2.

Proposition 2. *By introducing auxiliary variables* $\boldsymbol{\eta}$, `ExF` *augmented with equalities* (3) *for all* $j \in J$ *and* $t \in T$ *can be reformulated as:*

$$
\begin{aligned}
(\texttt{ReF}) \quad \text{Min} \quad & \sum_{i\in I}\sum_{t\in T} c^r_t x_{i,t} + \sum_{\omega\in\Omega} p_\omega \left(\sum_{j\in J} \sum_{t\in\mathcal{T}_j(\omega)\setminus\{\bar{t}_j(\omega)\}} c^h \eta_{j,t} + \sum_{j\in J}\sum_{t\in\mathcal{T}_j(\omega)} c^u (1 - \sum_{i\in I} y_{i,j,t}) \right) \\
& \textit{subject to}: \\
& (1\text{b}) - (1\text{f}), \\
& \eta_{j,t} \geq y_{i,j,t} - y_{i,j,t+1} \qquad\qquad \forall i \in I, j \in J, t \in T \setminus \{|T|\} \\
& \boldsymbol{\eta} \in \mathbb{R}_+^{|J|\times(|T|-1)}.
\end{aligned}
$$

Proof. We show equivalence in both directions. Let $(\mathbf{x}^*, \mathbf{y}^*, [\mathbf{w}^*(\omega)]_{\omega\in\Omega})$ be a feasible solution of `ExF` augmented with equalities (3) for each $j \in J$ and $t \in T$. Due to the definition of (3), $w^*_{j,t}(\omega)$ takes the same value for all $\omega \in \mathcal{O}_{j,t}$. Define $\eta^*_{j,t} = w^*_{j,t}(\omega)$ for all $j \in J$, $t \in T \setminus \{|T|\}$ where $\mathcal{O}_{j,t} \neq \emptyset$. As constraints (2b) imply $w^*_{j,t}(\omega) \geq y^*_{i,j,t} - y^*_{i,j,t+1}$ for all $i \in I$, it follows that $\eta^*_{j,t} \geq y^*_{i,j,t} - y^*_{i,j,t+1}$ for all $i \in I, j \in J, t \in T \setminus \{|T|\}$ where $\mathcal{O}_{j,t} \neq \emptyset$. For $j \in J$, $t \in T \setminus \{|T|\}$ where $\mathcal{O}_{j,t} = \emptyset$, we set $\eta^*_{j,t} = \max_{i\in I} \left(y^*_{i,j,t} - y^*_{i,j,t+1} \right)^+$. $(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\eta}^*)$ satisfies all the constraints of `ReF`. By construction, $\sum_{\omega\in\Omega} p_\omega \sum_{j\in J} \sum_{t\in\mathcal{T}_j(\omega)\setminus\{\bar{t}_j(\omega)\}} c^h w^*_{j,t}(\omega) = \sum_{j\in J} \sum_{t\in T\setminus\{|T|\}} \sum_{\omega\in\mathcal{O}_{j,t}} p_\omega c^h w^*_{j,t}(\omega)$, and since $\eta^*_{j,t} = w^*_{j,t}(\omega)$ for all $j \in J, t \in T \setminus \{|T|\}, \omega \in \mathcal{O}_{j,t}$, the objective value of `ReF` evaluated at $(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\eta}^*)$ equals that of `ExF` augmented with (3) at $(\mathbf{x}^*, \mathbf{y}^*, [\mathbf{w}^*(\omega)]_{\omega\in\Omega})$.

Conversely, let $(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\eta}^*)$ be a feasible solution of `ReF`. For each $j \in J$, $t \in T \setminus \{|T|\}$, and $\omega \in \mathcal{O}_{j,t}$, define $w^*_{j,t}(\omega) = \eta^*_{j,t}$. Since $\eta^*_{j,t} \geq y^*_{i,j,t} - y^*_{i,j,t+1}$ for all $i \in I$, it follows that equalities (3) and constraints (2b) are satisfied by $w^*_{j,t}(\omega)$ for all $\omega \in \mathcal{O}_{j,t}$. As constraints (2b) are defined only for $\omega \in \mathcal{O}_{j,t}$ where $\mathcal{O}_{j,t} \neq \emptyset$, $(\mathbf{x}^*, \mathbf{y}^*, [\mathbf{w}^*(\omega)]_{\omega\in\Omega})$ satisfies all the constraints of `ExF` augmented with equalities (3). Furthermore, as $\sum_{\omega\in\Omega} p_\omega \sum_{j\in J} \sum_{t\in\mathcal{T}_j(\omega)\setminus\{\bar{t}_j(\omega)\}} c^h \eta^*_{j,t} = \sum_{j\in J} \sum_{t\in T\setminus\{|T|\}} \sum_{\omega\in\mathcal{O}_{j,t}} p_\omega c^h \eta^*_{j,t}$, it follows that the objective value of `ExF` augmented with equalities (3) at $(\mathbf{x}^*, \mathbf{y}^*, [\mathbf{w}^*(\omega)]_{\omega\in\Omega})$ equals that of `ReF` at $(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\eta}^*)$. □

Together, Propositions 1 and 2 establish that `ReF` is a stronger formulation than `ExF`. Specifically, the supervalid equalities (3) cut off feasible solutions to `ExF` but not all optimal solutions, while Proposition 2 shows that `ExF` augmented with these equalities is equivalent to `ReF`. Moreover, unlike `ExF`, the number of variables and constraints in `ReF` does not grow with the number of scenarios. In Section 5.1, we show that `ReF` yields substantially shorter solution times than `ExF`.

4 Valid Inequalities

In this section, we derive valid inequalities to strengthen the linear programming (LP) relaxation of ReF. Let $\mathcal{P} = \{(\mathbf{x}, \mathbf{y}) \mid (\mathbf{x}, \mathbf{y}) \text{ satisfies (1b)} - \text{(1f)}\}$ be the set of feasible first-stage solutions.

Proposition 3. *For each $i \in I$, $j \in J$, and $t \in T$, the inequality*

$$y_{i,j,t} \leq \sum_{\tau=t}^{|T|} x_{i,\tau} \tag{4}$$

is valid for $\mathcal{P}$.

Proof. From (1c) and (1f), it follows that in any feasible solution of $\mathcal{P}$, $\sum_{\tau=t}^{|T|} x_{i,\tau} \in \{0,1\}$. If $\sum_{\tau=t}^{|T|} x_{i,\tau} = 0$, then constraints (1e) and (1f) imply $y_{i,j,t} = 0$. In this case, inequality (4) reduces to $y_{i,j,t} \leq 0$, which is valid since $y_{i,j,t} \in \{0,1\}$. If $\sum_{\tau=t}^{|T|} x_{i,\tau} = 1$, then (4) reduces to $y_{i,j,t} \leq 1$. □

Inequalities (4) are valid because an anesthesiologist i can be assigned to OR j in period t only if it is relieved after t. Example 2 demonstrates that (4) cut off fractional solutions that are feasible to the LP relaxation of $\mathcal{P}$.

Example 2. *Consider $\mathcal{P}$ with $I = \{1\}$, $J = \{1,2,3\}$, $T = \{1,2\}$ and $k = 2$. The solution $x_{1,1} = 0.6$, $x_{1,2} = 0.4$, $y_{1,1,2} = 0.6$, $y_{1,2,2} = 0.2$, $y_{1,1,1} = y_{1,2,1} = y_{1,3,1} = y_{1,3,2} = 0$ is feasible to the LP relaxation of $\mathcal{P}$ but violates the valid inequality (4) for $i = 1, j = 1, t = 2$: $y_{1,1,2} \leq x_{1,2}$.*

Next, we demonstrate the strength of these valid inequalities by showing that including them gives convex hull of a subset of constraints in $\mathcal{P}$. For some $i \in I$, let $\mathcal{S}_i$ be the set defined by the following subset of constraints of $\mathcal{P}$:

$$\sum_{t \in T} x_{i,t} = 1, \tag{5a}$$

$$\sum_{j \in J} y_{i,j,t} \leq k \sum_{\tau=t}^{|T|} x_{i,\tau} \qquad \forall t \in T, \tag{5b}$$

$$x_{i,t}, y_{i,j,t} \in \{0,1\} \qquad \forall j \in J, t \in T. \tag{5c}$$

In the following proposition, we show that including valid inequalities (4) for $i \in I$ to the linear relaxation of $\mathcal{S}_i$ gives $\text{conv}(\mathcal{S}_i)$. The proof of this proposition is stated in Appendix A and follows a two-step approach. First, we use a disjunctive programming technique to derive a formulation for

$\text{conv}(\mathcal{S}_i)$ in an extended variable space. Then, we project this extended formulation in the space of $(\mathbf{x}, \mathbf{y})$ variables.

Proposition 4. $\text{conv}(\mathcal{S}_i)$ *is given by the following constraints:*

$$\sum_{t \in T} x_{i,t} = 1, \tag{6a}$$

$$\sum_{j \in J} y_{i,j,t} \leq k \sum_{\tau=t}^{|T|} x_{i,\tau} \qquad \forall t \in T, \tag{6b}$$

$$y_{i,j,t} \leq \sum_{\tau=t}^{|T|} x_{i,\tau} \qquad \forall j \in J, t \in T, \tag{6c}$$

$$x_{i,t}, y_{i,j,t} \geq 0 \qquad \forall j \in J, t \in T. \tag{6d}$$

While inequalities (4) describe the convex hull of $\mathcal{S}_i$ for each anesthesiologist i, they do not exploit the relief order constraints. In particular, constraints (1d) imply that if anesthesiologist i is relieved by period t, then every anesthesiologist preceding i in the relief order must also be relieved by period t. This observation enables us to strengthen inequalities (4) by aggregating the assignments of anesthesiologists preceding i in the relief order to `OR` j. In the following proposition, we show that inequalities (4) can be lifted to obtain a stronger valid inequality for $\mathcal{P}$.

Proposition 5. *For each $i \in I, j \in J$ and $t \in T$, the inequality*

$$\sum_{i'=1}^{i} y_{i',j,t} \leq \sum_{\tau=t}^{|T|} x_{i,\tau} \tag{7}$$

is valid for $\mathcal{P}$.

Proof. Suppose $(\mathbf{x}, \mathbf{y}) \in \mathcal{P}$ violates (7) for some $i \in I, j \in J, t \in T$, *i.e.*, $\sum_{i'=1}^{i} y_{i',j,t} > \sum_{\tau=t}^{|T|} x_{i,\tau}$. Since $\sum_{\tau=t}^{|T|} x_{i,\tau} \in \{0, 1\}$ and $\sum_{i \in I} y_{i,j,t} \leq 1$, this implies $\sum_{\tau=t}^{|T|} x_{i,\tau} = 0$ and $y_{i',j,t} = 1$ for some $i' \leq i$. From constraints (1c) and (1d), $\sum_{\tau=t}^{|T|} x_{i,\tau} = 0$ implies $\sum_{\tau=t}^{|T|} x_{i',\tau} = 0$ for all $i' \leq i$. This implies, due to constraint (1e), $y_{i',j,t} = 0$ for every $i' \leq i$, which is a contradiction. □

Inequality (7) strengthens (4) by ensuring that an anesthesiologist $i' \leq i$ can be assigned to `OR` j in period t only if anesthesiologist i is relieved in some period $\tau \geq t$. Otherwise, if anesthesiologist i is relieved before period t, then i' must also be relieved before t, and therefore cannot be assigned to any `OR` in that time period. Example 3 demonstrates that (7) cuts off fractional solutions that are feasible to the LP relaxation of $\mathcal{P}$.

Example 3. *Consider $\mathcal{P}$ with $I = \{1, 2\}$, $J = \{1\}$, $T = \{1, 2\}$, and $k = 2$. The solution $x_{1,1} = 0.6$, $x_{1,2} = 0.4$, $x_{2,1} = 0.5$, $x_{2,2} = 0.5$, $y_{1,1,2} = 0.5$, $y_{2,1,2} = 0.3$, $y_{1,1,1} = y_{2,1,1} = 0$ is feasible to the LP relaxation of $\mathcal{P}$ but violates the valid inequality (7) for $i = 2, j = 1, t = 2$: $y_{1,1,2} + y_{2,1,2} \leq x_{2,2}$.*

For some $j \in J$, let $\mathcal{R}_j$ be the set defined by the following set of constraints:

$$\sum_{i \in I} y_{i,j,t} \leq 1 \qquad \forall t \in T, \tag{8a}$$

$$\sum_{t \in T} x_{i,t} = 1 \qquad \forall i \in I, \tag{8b}$$

$$\sum_{\tau=1}^{t} x_{i+1,\tau} - \sum_{\tau=1}^{t} x_{i,\tau} \leq 0 \qquad \forall i \in I \setminus \{|I|\}, t \in T, \tag{8c}$$

$$y_{i,j,t} \leq \sum_{\tau=t}^{|T|} x_{i,\tau} \qquad \forall i \in I, t \in T, \tag{8d}$$

$$x_{i,t}, y_{i,j,t} \in \{0, 1\} \qquad \forall i \in I, t \in T. \tag{8e}$$

$\mathcal{R}_j$ describes a binary set induced by a subset of constraints in $\mathcal{P}$, along with the valid inequalities (4). In the following proposition, we show that lifting inequality (4) to (7) in the linear relaxation of $\mathcal{R}_j$ characterizes the convex hull of $\mathcal{R}_j$.

Proposition 6. $\text{conv}(\mathcal{R}_j)$ *is given by the following set of constraints:*

$$(\tilde{\mathcal{R}}_j) \quad (8a), (8b), (8c), \tag{9a}$$

$$\sum_{i'=1}^{i} y_{i',j,t} \leq \sum_{\tau=t}^{|T|} x_{i,\tau} \qquad \forall i \in I, t \in T \tag{9b}$$

$$0 \leq x_{i,t} \leq 1 \qquad \forall i \in I, t \in T, \tag{9c}$$

$$y_{i,j,t} \geq 0 \qquad \forall i \in I, t \in T. \tag{9d}$$

We state the proof of Proposition 6 in Appendix A. The proof relies on constructing a polyhedron $\tilde{\mathcal{R}}'_j$ whose constraint matrix is totally unimodular, and an integrality-preserving linear transformation $\mathcal{L}$ that maps $\tilde{\mathcal{R}}'_j$ to $\tilde{\mathcal{R}}_j$, a valid formulation for the binary set $\mathcal{R}_j$. The integrality of the extreme points of $\tilde{\mathcal{R}}_j$ then follows from the integrality of the extreme points of $\tilde{\mathcal{R}}'_j$, establishing that $\text{conv}(\mathcal{R}_j) = \tilde{\mathcal{R}}_j$. In Section 5.1, we show that incorporating valid inequalities (7) in `ReF` significantly improves its computational performance.

5 Numerical Experiments

Our computational experiments are based on data from a tertiary medical hospital in the eastern United States. The dataset comprises 14,883 surgeries across 24 specialties, spanning 353 working days. For each specialty, Table 6 in Appendix B provides the mean and standard deviation of the surgery durations, along with the percentage share of total surgeries for each specialty. We model the surgery duration for each specialty using a log-normal distribution (Neyshabouri and Berg 2017). We generate scenarios, Ω, by sampling the surgery durations from specialty-specific distributions, assuming uniform scenario probabilities. We demonstrate the application of the proposed model and solution methods using $|\Omega|= 500$ and 750 scenarios.

We consider a 14-hour planning horizon from 7 a.m. to 9 p.m., divided into equal 15-minute time intervals, yielding $|T|= 56$ periods. Each anesthesiologist can cover at most $k = 3$ ORs in each period. We experiment with two surgery schedules derived from our dataset, referred to as Schedules A and B. Table 7 in Appendix B reports the number of surgeries, open ORs, and available anesthesiologists for each schedule. More details about each scheduled surgery, including its specialty, planned start time, and the assigned OR, are provided in Tables 8–9. Schedules A and B are used in Sections 5.1-5.3 to conduct a detailed analysis of the proposed framework, while Section 5.4 reports results across 50 schedules from our dataset to assess more general performance.

As discussed in Section 3, c^r_t is a composite cost that reflects both the direct anesthesiologist staffing cost and an implicit penalty for the increased workload associated with a later *relief time*. In our experiments, we adopt a linear form, $c^r_t = tc^r$ with $c^r = 1000$, where c^r aggregates the per-period staffing and workload components. It is important to note that any non-decreasing function can be used to capture the relation between c^r_t and t to demonstrate the key trade-offs of interest. We refer to the quantity $\sum_{i\in I}\sum_{t\in T} c^r_t x_{i,t}$ as the *relief cost* throughout the rest of the computational study.

In contrast, estimating the costs of handoffs and under-staffing can be challenging because of their downstream impacts on clinic operations and patient health outcomes. We specify these cost parameters relative to c^r, informed by our medical collaborator's acceptable levels of expected handoffs and under-staffing in the resulting schedules. This expert-informed estimation approach is generalizable and supports the clinical feasibility of the proposed solutions. Subsequently, we set handoff cost (c^h) and per-period anesthesiologist under-staffing cost (c^u) using the parameters $\zeta_h = \frac{c^h}{c^r}$ and $\zeta_u = \frac{c^u}{c^r}$, respectively. We consider three values $\{5, 10, 15\}$ for ζ_h and ζ_u. These values reflect a practical middle ground while capturing meaningful trade-offs in scheduling decisions. Extreme values of c^h or c^u are

not computationally or practically interesting because they yield trivial solutions. In particular, if c^h or c^u are set too high—placing excessive weight on avoiding handoffs or under-staffing—then nearly all anesthesiologists are scheduled to be relieved at the end of the day. Conversely, if these parameters are set too low, most anesthesiologists are relieved early, leading to a large number of handoffs or under-staffed periods. We consider a full factorial experimental design over the values of ζ_h and ζ_u, resulting in 9 different cases. We generate three random instances for each of the two schedules, resulting in 6 instances per case. We perform all computations using Python 3.9.13 and Gurobi 11.0.2 on a machine with an 8-core Intel Xeon 2.60GHz processor and 64GB of RAM. Section 5.1 assesses the strength of the proposed valid inequalities and compares the performance of three different solution approaches. Section 5.2 offers managerial insights through a sensitivity analysis. Section 5.3 examines the value of incorporating uncertainty in surgery durations into anesthesiologist scheduling decisions. Finally, Section 5.4 compares the optimal solution of the proposed model with the anesthesiologist schedules observed in current practice at our partner institution.

5.1 Computational Performance Analysis

We first evaluate the effectiveness of the valid inequalities (7) by comparing `ReF` and `ReF-VI` that is augmented with the valid inequalities. We set a time limit of 3600 seconds and a target optimality gap of 1%. Table 2 demonstrates that `ReF-VI` yields a substantial reduction in runtime across all cases. We observe similar gains for $|\Omega|= 750$ in Appendix C.

Table 2: Average (max) runtime for `ReF-VI` and `ReF` with $|\Omega|= 500$ scenarios.

ζ_h–ζ_u	`ReF-VI`	`ReF`	Difference (%)
5—5	370 (1477)	1259 (2599)	78.23 (97.18)
5—10	353 (1459)	1016 (1803)	75.41 (94.92)
5—15	617 (3029)	842 (2063)	63.67 (94.05)
10—5	350 (933)	1001 (1887)	43.74 (95.98)
10—10	248 (754)	693 (1548)	62.40 (91.74)
10—15	423 (1552)	590 (1005)	46.29 (92.14)
15—5	257 (543)	1307 (2807)	63.01 (90.62)
15—10	257 (876)	941 (1696)	61.83 (96.74)
15—15	381 (994)	434 (655)	16.48 (52.19)

Next, we compare the performance of three solution approaches: (*i*) using Gurobi to solve `ReF-VI`, (*ii*) using Gurobi to solve `ExF-VI`: `Extensive Form` with valid inequalities (7), (*iii*) `LS-VI`: `L-shaped Algorithm` with valid inequalities (7). The details of `LS-VI` are presented in Appendix D. `ReF-VI` and `ExF-VI` return `0.01-optimal` solutions for all instances within 3600 seconds. In contrast, `LS-VI`

cannot achieve 1% optimality for any instance, with an average optimality gap of over 90% in each case. Table 3 shows that `ReF-VI` is significantly faster than `ExF-VI`. The performance difference becomes more pronounced as the number of scenarios increases. Consequently, we adopt `ReF-VI` for the remainder of the computational experiments. Additional computational results are presented in Tables 11 and 12 in Appendix C. All computational times are reported in seconds.

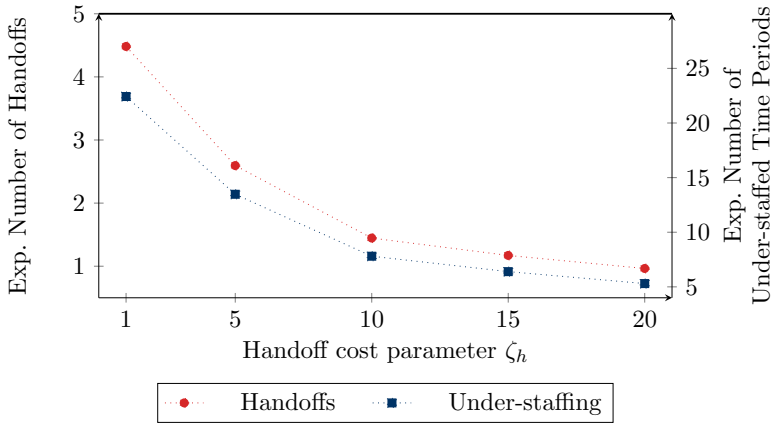
Table 3: Average (max) runtime for `ReF-VI` and `ExF-VI` with $|\Omega|= 500$ and 750.

	$\|\Omega\|= 500$		$\|\Omega\|= 750$	
ζ_h–ζ_u	`ReF-VI`	`ExF-VI`	`ReF-VI`	`ExF-VI`
5—5	370 (1477)	543 (801)	239 (676)	490 (869)
5—10	353 (1459)	535 (1056)	171 (395)	564 (1232)
5—15	617 (3029)	784 (2056)	224 (718)	754 (1955)
10—5	350 (933)	662 (788)	269 (422)	643 (1158)
10—10	248 (754)	680 (1238)	119 (335)	693 (967)
10—15	423 (1552)	675 (1398)	193 (498)	981 (1482)
15—5	257 (543)	462 (678)	172 (329)	442 (550)
15—10	257 (876)	626 (830)	183 (310)	613 (791)
15—15	381 (994)	875 (1460)	199 (421)	705 (971)

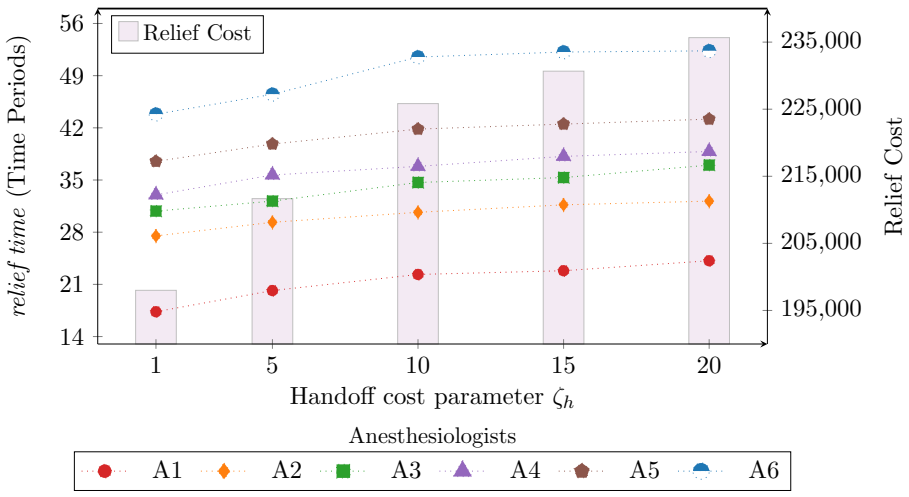
5.2 Sensitivity Analysis

We demonstrate how the expected number of handoffs, under-staffed time periods, and anesthesiologists' *relief times* vary with respect to handoff and under-staffing costs. To better understand the trade-offs, we consider five values $\{1, 5, 10, 15, 20\}$ for the handoff cost parameter ζ_h and the under-staffing cost parameter ζ_u. Results are averaged over six randomly generated instances for Schedules A and B with $|\Omega|= 500$.

In Figure 1a, both the average number of handoffs and the average number of under-staffed periods decrease as the handoff cost increases. This is because higher handoff costs lead to anesthesiologists being relieved later in the day, as shown in Figure 1b. Consequently, the anesthesiologist's relief cost rises. A similar trend is observed in Figures 2a and 2b, where increasing the under-staffing cost parameter ζ_u results in fewer handoffs and under-staffed periods. The results verify that `SASH` captures the trade-offs between delaying *relief times*, handoffs, and under-staffing as intended. A key takeaway from these results is that anesthesiologist schedules can be aligned with institutional priorities to reduce handoffs and under-staffing by adjusting model parameters. Furthermore, the required delay in anesthesiologist *relief times* to reduce the expected number of handoffs or under-staffed periods below a target can be analyzed using the proposed model. If major delays are required, the proposed model can inform expansion of the current anesthesiologist pool.

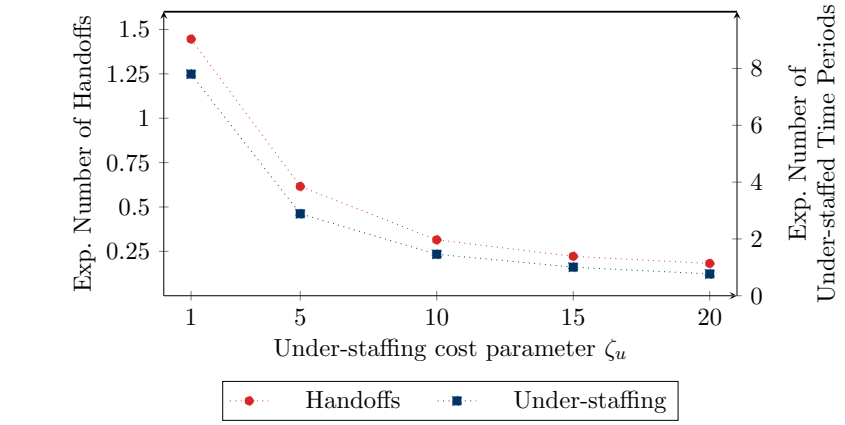


(a)

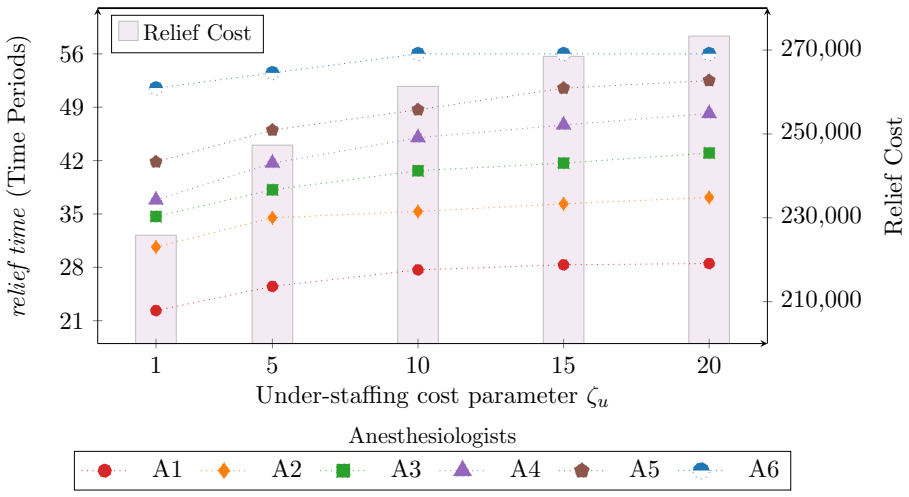


(b)

Figure 1: The trade-off between handoffs and *relief times* as a function of the handoff cost parameter ζ_h. Under-staffing cost parameter $\zeta_u = 1$.



(a)



(b)

Figure 2: The trade-off between handoffs and *relief times* as a function of the under-staffing cost parameter ζ_u. Handoff cost parameter $\zeta_h = 10$.

5.3 The Value of Considering Uncertainty in Surgery Durations

To demonstrate the value of modeling uncertainty in surgery durations, Table 4 compares the best solution obtained from `ReF-VI` (`Stochastic Solution`) with the `Mean Value Solution` (Birge and Louveaux 2011) for $|\Omega|= 500$ and 750. For each of the nine cases defined by ζ_h, $\zeta_u \in \{5, 10, 15\}$, we generate three random instances per schedule, resulting in a total of six instances per case.

For $|\Omega|= 500$, Column 2 of Table 4 reports the average Value of the Stochastic Solution (VSS), computed as $\frac{MV-UB}{MV} \times 100$, where UB and MV denote the objective value of the `Stochastic Solution` and `Mean Value Solution`, respectively. Columns 3–5 present the average percentage reduction in expected number of handoffs (`HO`), expected number of under-staffed periods (`US`), and anesthesiologist relief cost (`RF`), respectively, when using the `Stochastic Solution` relative to `Mean Value Solution`. Specifically, we report $\kappa_l = \frac{l_{MV}-l_{SS}}{l_{MV}} \times 100$ for $l = \texttt{HO}, \texttt{US}, \texttt{RF}$, where l_{MV} and l_{SS} are the values of l under `Mean Value Solution` and `Stochastic Solution`, respectively. Thus, $\kappa_l > 0$ indicates that the value of l under `Stochastic Solution` is lower than under `Mean Value Solution`. Analogous results for $|\Omega|= 750$ are presented in Columns 6-9. As shown in Column 2 of Table 4, incorporating uncertainty in surgery durations leads to substantial cost savings. This result is primarily driven by lower expected under-staffing and handoff costs. As seen in Column 5, `Stochastic Solution` tends to schedule anesthesiologists' *relief times* later in the day. This increases coverage throughout the day and reduces under-staffed periods and handoffs. The resulting cost savings from these reductions outweigh the higher anesthesiologist costs.

These observations remain the same for $|\Omega|= 750$. Moreover, the consistent magnitude of VSS across different combinations of handoff and under-staffing costs suggests that the benefits of modeling uncertainty are robust to changes in cost structure. These findings underscore the importance of accounting for uncertainty in operational settings with significant task-duration variability.

Table 4: The Value of `Stochastic Solution`.

	$\|\Omega\|= 500$				$\|\Omega\|= 750$			
ζ_h–ζ_u	VSS	$\kappa_{\texttt{HO}}$	$\kappa_{\texttt{US}}$	$\kappa_{\texttt{RF}}$	VSS	$\kappa_{\texttt{HO}}$	$\kappa_{\texttt{US}}$	$\kappa_{\texttt{RF}}$
5—5	38.89	90.17	91.43	-26.52	38.94	90.54	91.77	-26.49
5—10	56.29	95.18	95.95	-34.70	56.34	95.33	95.90	-33.94
5—15	65.99	96.49	97.14	-38.10	65.98	96.52	97.21	-37.75
10—5	43.21	92.07	92.83	-28.55	43.22	91.90	92.86	-28.17
10—10	58.55	95.95	96.37	-35.84	58.57	95.69	96.23	-34.88
10—15	67.33	97.14	97.49	-39.58	67.35	96.92	97.35	-38.34
15—5	46.94	93.26	93.67	-30.16	46.98	92.95	93.58	-29.49
15—10	60.59	96.26	96.44	-36.20	60.60	96.21	96.47	-35.75
15—15	68.58	97.16	97.52	-39.78	68.60	96.90	97.27	-38.18

5.4 Comparison with Real-Life Anesthesiologist Schedules

In this section, we compare the solution of our model (Opt) with the real-life OR anesthesiologist schedules implemented at our partner institution (CP). Specifically, we consider 50 daily surgical schedules observed in our dataset. For each day, we randomly generate three instances for each combination of ζ_h, $\zeta_u \in \{5, 10, 15\}$ case, resulting in a total of 150 instances per case. We compare the anesthesiologist relief cost (AC), expected handoff cost (HO), expected under-staffing cost (US), and overall objective function value (Obj) of the Opt and CP solutions. We report $\eta_l = \frac{\text{Cost}_l^{\texttt{CP}} - \text{Cost}_l^{\texttt{Opt}}}{\text{Cost}_l^{\texttt{CP}}} \times 100$ for $l = \texttt{AC}, \texttt{HO}, \texttt{US}, \texttt{Obj}$, where the higher η_l value implies better performance of the Opt solution. Table 5 presents the average (max) η_l values with $|\Omega| = 500$ scenarios.

Table 5: Average (max) %-improvement over 50 daily surgical schedules observed in the dataset.

ζ_h—ζ_u	$\eta_{\texttt{AC}}$	$\eta_{\texttt{HO}}$	$\eta_{\texttt{US}}$	$\eta_{\texttt{Obj}}$
5—5	22.50 (48.25)	98.42 (99.22)	97.01 (99.17)	76.80 (84.46)
5—10	17.96 (45.91)	99.17 (99.65)	98.57 (99.58)	83.89 (90.07)
5—15	15.36 (44.36)	99.45 (99.80)	99.11 (99.77)	87.64 (92.72)
10—5	21.39 (47.47)	98.69 (99.34)	97.45 (99.30)	80.54 (86.19)
10—10	17.26 (45.53)	99.29 (99.72)	98.72 (99.72)	85.77 (90.74)
10—15	14.92 (43.58)	99.51 (99.82)	99.18 (99.86)	88.77 (93.10)
15—5	20.38 (47.86)	98.88 (99.55)	97.80 (99.39)	83.19 (87.79)
15—10	16.62 (45.14)	99.37 (99.73)	98.86 (99.72)	87.24 (91.47)
15—15	14.53 (43.58)	99.56 (99.82)	99.24 (99.86)	89.70 (93.45)

The Opt solution substantially outperforms the CP for all cases. Table 5 reveals two main trends. First, average (max) $\eta_{\texttt{Obj}}$ increases as ζ_u increases for a fixed ζ_h, indicating that the Opt solution's advantage increases with under-staffing cost. Second, the consistently high values of $\eta_{\texttt{HO}}$ and $\eta_{\texttt{US}}$ across all cases indicate that the observed gains are robust and reflect systematic improvements in anesthesiologist schedules achievable with the proposed framework.

6 Conclusion

We propose a two-stage stochastic programming model to determine optimal schedules for anesthesiologists that balance delaying *relief times* with the number of handoffs and under-staffed periods, accounting for uncertainty in surgery durations. Given the daily surgery schedules in ORs and a relief order to balance workload among anesthesiologists, the proposed model determines the anesthesiologists' *relief times* and assignments to ORs. To address the computational challenges in solving SASH, we derive supervalid equalities and develop a strengthened monolithic reformulation whose size is independent of the number of scenarios. To further enhance computational performance, we introduce

valid inequalities for the first-stage problem and show that they describe the convex hull of binary sets defined by subsets of the first-stage constraints. These inequalities are generally applicable to multi-period scheduling problems with precedence constraints. Computational experiments demonstrate that the proposed reformulation, combined with the proposed valid inequalities, significantly outperforms both the extensive form and the L-shaped algorithm.

Our computational experiments employ a dataset from an academic medical center located in the eastern United States. We present a sensitivity analysis to explore trade-offs among anesthesiologist staffing levels, the number of handoffs, and under-staffed periods. We also compare the solution of our model with the institution's current anesthesiologist scheduling practices. The results demonstrate that the proposed framework can lead to substantial reductions in handoffs and under-staffed periods, while ensuring earlier *relief times* for anesthesiologists under uncertain surgery durations. Although the exact magnitude of improvements in the number of handoffs and under-staffed periods depends on the variability and distribution of surgery durations, the underlying modeling framework is broadly generalizable. It can be readily adapted to incorporate surgery duration distributions, staffing policies, and constraints at other institutions.

The proposed model offers valuable insights for anesthesiology departments across multiple levels of decision-making. At the operational level, it can assist in day-to-day scheduling decisions, such as assigning anesthesiologists to ORs and determining *relief times*. At the strategic level, the model can be applied to historical or projected surgery volumes (Hassanzadeh et al. 2022) to inform long-term planning. This includes assessing future staffing requirements, analyzing trends in anesthesiologist workload, and establishing baseline staffing benchmarks. For instance, using the model with projected surgery volumes for the upcoming year can help estimate the number of anesthesiologists needed to meet expected demand, identify ways to increase surgical capacity without expanding staff, and reduce instances of handoffs by ensuring adequate coverage. Ultimately, the model supports data-driven workforce planning and policy design, helping align anesthesiologist schedules with institutional priorities to reduce handoffs and avoid under-staffed periods.

In our computational experiments, we implement the proposed framework using specialty-specific log-normal distributions for surgery durations. Additional features such as surgeon experience, case complexity, procedure type, and time of day can be incorporated to improve the estimation of surgery duration (Kayis et al. 2012). Likewise, parameters related to handoff and under-staffing costs can be refined through empirical analysis of historical data. This enables the model to be tailored using institutional priorities and financial inputs, enhancing both its precision and adaptability across diverse

healthcare environments. For instance, handoff costs can be estimated by considering the expenses associated with systems implemented to maintain high-quality handoffs, as well as the hospital's costs related to adverse patient outcomes resulting from improper handoffs. Similarly, the cost of under-staffed time periods can be approximated by factoring in the cost of calling in on-call anesthesiologists and the additional expense of extending the *relief times* for on-duty anesthesiologists.

Future research will focus on integrating add-on surgeries into anesthesiologist scheduling. These surgeries, which are urgent and unplanned, are often added with little notice and can lead to significant delays in anesthesiologist *relief times*, increased handoffs, and under-staffed periods. Another promising extension of the proposed model involves integrating the scheduling of anesthesiologists and nurse anesthetists. Given their interdependence through the surgical schedule, a more advanced model could explicitly account for the coordination needed between their schedules and `OR` assignments.

References

Abouleish AE, Pomerantz P, Peterson MD, Cannesson M, Akeju O, Miller TR, Rathmell JP, Cole DJ (2024) Closing the chasm: Understanding and addressing the anesthesia workforce supply and demand imbalance. *Anesthesiology* 141(2):238–249.

Afonso AM, Cadwell JB, Staffa SJ, Sinskey JL, Vinson AE (2024) US attending anesthesiologist burnout in the post-pandemic era. *Anesthesiology* 140(1):38.

Afonso AM, Cadwell JB, Staffa SJ, Zurakowski D, Vinson AE (2021) Burnout rate and risk factors among anesthesiologists in the united states. *Anesthesiology* 134(5):683.

Aghaei S, Gómez A, Vayanos P (2025) Strong optimal classification trees. *Operations Research* 73(4):2223–2241.

Balas E (1998) Disjunctive programming: Properties of the convex hull of feasible points. *Discrete Applied Mathematics* 89(1-3):3–44.

Bhutiani M, Jablonski PM, Ehrenfeld JM, McEvoy MD, Fowler LC, Wanderer JP (2018) Decision support tool improves real and perceived anesthesiology resident relief equity. *Anesthesia & Analgesia* 127(2):513–519.

Birge JR, Louveaux F (2011) *Introduction to Stochastic Programming* (New York, NY: Springer Science & Business Media).

Boet S, Djokhdem H, Leir SA, Théberge I, Mansour F, Etherington C (2020) Association of intraoperative anaesthesia handovers with patient morbidity and mortality: a systematic review and meta-analysis. *British Journal of Anaesthesia* 125(4):605–613.

Breuer DJ, Lahrichi N, Clark DE, Benneyan JC (2020) Robust combined operating room planning and personnel scheduling under uncertainty. *Operations Research for Health Care* 27:100276.

Conforti M, Cornuéjols G, Zambelli G (2014) *Integer Programming*. Graduate Texts in Mathematics (Springer).

Fügener A, Brunner JO, Podtschaske A (2015) Duty and workstation rostering considering preferences and fairness: a case study at a department of anaesthesiology. *International Journal of Production Research* 53(24):7465–7487.

Glomb L, Liers F, Rösel F (2026) A novel pareto-optimal cut selection strategy for benders decomposition. *Mathematical Programming Computation* 18(1):211–257.

Hassanzadeh H, Boyle J, Khanna S, Biki B, Syed F (2022) Daily surgery caseload prediction: towards improving operating theatre efficiency. *BMC Medical Informatics and Decision Making* 22(1):151.

Kayis E, Wang H, Patel M, Gonzalez T, Jain S, Ramamurthi R, Santos C, Singhal S, Suermondt J, Sylvester K (2012) Improving prediction of surgery duration using operational and temporal factors. *AMIA Annual Symposium Proceedings*, volume 2012, 456.

Kazemian P, Dong Y, Rohleder TR, Helm JE, Van Oyen MP (2014) An IP-based healthcare provider shift design approach to minimize patient handoffs. *Health Care Management Science* 17:1–14.

Küçükyavuz S, Sen S (2017) An introduction to two-stage stochastic mixed-integer programming. *Leading Developments from INFORMS Communities*, 1–27 (INFORMS).

Menezes J, Zahalka C (2024) Anesthesiologist shortage in the united states: a call for action. *Journal of Medicine, Surgery, and Public Health* 2:100048.

Neyshabouri S, Berg BP (2017) Two-stage robust optimization approach to elective surgery and downstream capacity planning. *European Journal of Operational Research* 260(1):21–40.

Patterson ES, Roth EM, Woods DD, Chow R, Gomes JO (2004) Handoff strategies in settings with high consequences for failure: lessons for health care operations. *International Journal for Quality in Health Care* 16(2):125–132.

Rath S, Rajaram K, Hudson ME, Mahajan A (2026) Multilocation, dynamic staff planning for a healthcare system: Methodology and application. *Operations Research* 74(1):199–223.

Rath S, Rajaram K, Mahajan A (2017) Integrated anesthesiologist and room scheduling for surgeries: Methodology and application. *Operations Research* 65(6):1460–1478.

Saager L, Hesler BD, You J, Turan A, Mascha EJ, Sessler DI, Kurz A (2014) Intraoperative transitions of anesthesia care and postoperative adverse outcomes. *Anesthesiology* 121(4):695–706.

Sinha A, Bansal A, Ozaltin O, Russell M (2025) A two-stage stochastic programming approach for CRNA scheduling with handovers. Preprint, Optimization-Online, URL `https://optimization-online.org/wp-content/uploads/2025/02/Revision-Unmarked.pdf`.

Smalley HK, Keskinocak P, Vats A (2015) Physician scheduling for continuity: An application in pediatric intensive care. *Interfaces* 45(2):133–148.

Sun K, Sun M, Agrawal D, Dravenstott R, Rosinia F, Roy A (2023) Equitable anesthesiologist scheduling under demand uncertainty using multiobjective programming. *Production and Operations Management* 32(11):3699–3716.

Tsang MY, Shehadeh KS, Curtis FE, Hochman BR, Brentjens TE (2024) Stochastic optimization approaches for an operating room and anesthesiologist scheduling problem. *Operations Research* 73(3):1430–1458.

Vielma JP (2015) Mixed integer linear programming formulation techniques. *SIAM Review* 57(1):3–57.

Wax DB, McCormick PJ (2017) A real-time decision support system for anesthesiologist end-of-shift relief. *Anesthesia & Analgesia* 124(2):599–602.

Wolsey LA (2020) *Integer programming* (John Wiley & Sons).

Appendix A: Proofs of Propositions

Proof of Proposition 4: For the ease of presentation, we use x_t to represent $x_{i,t}$, and $y_{j,t}$ in place of $y_{i,j,t}$. We represent $\mathcal{S}_i$ as the union of disjoint sets V_t: $\mathcal{S}_i = \bigcup_{t\in T} V_t$ where

$$V_t = \begin{cases} x_t = 1, & \\ \sum_{j\in J} y_{j,\tau} \le k & \forall \tau \in \{1,\cdots,t\} \\ y_{j,\tau} = 0 & \forall j \in J, \tau \in \{t+1,\cdots,|T|\} \\ y_{j,\tau} \in \{0,1\} & \forall j \in J, \tau \in \{1,\cdots,t\}. \end{cases}$$

Let $\tilde{V}_t$ be the linear programming relaxation of V_t. As the constraint matrix of V_t is totally unimodular, $\text{conv}(V_t) = \tilde{V}_t$. Furthermore, as $\tilde{V}_t$ is bounded for all $t \in T$, $\text{conv}(\mathcal{S}_i) = \text{conv}(\bigcup_{t\in T} \tilde{V}_t)$ (Conforti et al. 2014). Through the application of disjunctive programming (Balas 1998, Conforti et al. 2014), $\text{conv}(\bigcup_{t\in T} \tilde{V}_t)$ is given by:

$$\sum_{j\in J} y^t_{j,\tau} \le k\, x_t \qquad \tau \in \{1,\cdots,t\}, t \in T \tag{10a}$$

$$0 \le y^t_{j,\tau} \le x_t \qquad \forall j \in J, \tau \in \{1,\cdots,t\}, t \in T \tag{10b}$$

$$y^t_{j,\tau} = 0 \qquad \forall \tau \in \{t+1,\cdots,|T|\}, t \in T \tag{10c}$$

$$\sum_{t=1}^{|T|} y^t_{j,\tau} = y_{j,\tau} \qquad j \in J, \tau \in T, \tag{10d}$$

$$x_t \ge 0 \qquad \forall t \in \{1,\cdots,|T|\}, \tag{10e}$$

$$\sum_{t=1}^{|T|} x_t = 1. \tag{10f}$$

Furthermore, let $\mathcal{T}_\tau \ \forall \tau \in T$ be defined by the following set of constraints:

$$y_{j,\tau} = \sum_{t=\tau}^{|T|} y_{j,\tau}^t \qquad \forall j \in J, \tag{11a}$$

$$\sum_{j \in J} y_{j,\tau}^t \leq k\, x_t \qquad \forall t \in \{\tau, \cdots, |T|\}, \tag{11b}$$

$$y_{j,\tau}^t \leq x_t \qquad \forall j \in J, t \in \{\tau, \cdots, |T|\}, \tag{11c}$$

$$y_{j,\tau}^t \geq 0 \qquad \forall j \in J, t \in \{\tau, \cdots, |T|\}. \tag{11d}$$

Using (10c), we get $\text{conv}(\mathcal{S}_i) = \text{conv}(\bigcup_{t \in T} \tilde{V}_t) = \cap_{\tau \in T} \mathcal{T}_\tau \cap (10\text{e}) \cap (10\text{f})$.

Next, we show that $\text{Proj}_{\left([y_{j,\tau}]_{j \in J},\, [x_t]_{t \in \{\tau,\ldots,|T|\}}\right)} \mathcal{T}_\tau = \dot{\mathcal{T}}_\tau$, where

$$\dot{\mathcal{T}}_\tau = \begin{cases} \displaystyle\sum_{j \in J} y_{j,\tau} \leq k \sum_{t=\tau}^{|T|} x_t, & \\ y_{j,\tau} \leq \displaystyle\sum_{t=\tau}^{|T|} x_t & \forall j \in J, \\ y_{j,\tau} \geq 0 & \forall j \in J, \\ x_t \geq 0 & \forall t \in \{\tau, \cdots, |T|\}, \end{cases} \tag{12}$$

thereby completing the proof. Here, the projection operator is defined as $\text{Proj}_x \mathcal{X} = \{\, x : \exists y \text{ such that } (x, y) \in \mathcal{X} \,\}$, and the notation $[x_\lambda]_{\lambda \in \Lambda}$ denotes the vector with components x_λ indexed by Λ.

Suppose that $\left([\bar{y}_{j,\tau}]_{j \in J},\ [\bar{x}_t]_{t \in \{\tau,\ldots,|T|\}},\ [\bar{y}_{j,\tau}^t]_{j \in J, t \in \{\tau,\ldots,|T|\}}\right) \in \mathcal{T}_\tau$. Summing (11a) over $j \in J$ and (11b) over $t \in \{\tau, \cdots, |T|\}$ gives $\sum_{j \in J} \bar{y}_{j,\tau} = \sum_{t=\tau}^{|T|} \sum_{j \in J} \bar{y}_{j,\tau}^t$ and $\sum_{t=\tau}^{|T|} \sum_{j \in J} \bar{y}_{j,\tau}^t \leq k \sum_{t=\tau}^{|T|} \bar{x}_t$, respectively. Thus, $\sum_{j \in J} \bar{y}_{j,\tau} \leq k \sum_{t=\tau}^{|T|} \bar{x}_t$. For any $j \in J$, summing (11c) over $t \in \{\tau, \cdots, |T|\}$ and applying (11a) yields $\bar{y}_{j,\tau} = \sum_{t=\tau}^{|T|} \bar{y}_{j,\tau}^t \leq \sum_{t=\tau}^{|T|} \bar{x}_t$. Finally, (11a) together with (11d) implies $\bar{y}_{j,\tau} \geq 0$ for all $j \in J$, and (11c) and (11d) imply $\bar{x}_t \geq 0 \ \forall t \in \{\tau, \cdots, |T|\}$. Hence, $\left([\bar{y}_{j,\tau}]_{j \in J},\ [\bar{x}_t]_{t \in \{\tau,\ldots,|T|\}}\right) \in \dot{\mathcal{T}}_\tau$. This implies $\text{Proj}_{\left([y_{j,\tau}]_{j \in J}, [x_t]_{t \in \{\tau,\ldots,|T|\}}\right)} \mathcal{T}_\tau \subseteq \ \dot{\mathcal{T}}_\tau$.

Next, suppose that $([\bar{y}_{j,\tau}]_{j \in J}, [\bar{x}_t]_{t \in \{\tau,\ldots,|T|\}}) \in \dot{\mathcal{T}}_\tau$ but $\nexists\, [y_{j,\tau}^t]_{j \in J, t \in \{\tau,\ldots,|T|\}}$ such that $([y_{j,\tau}]_{j \in J}$, $[\bar{x}_t]_{t \in \{\tau,\ldots,|T|\}}$, $[\bar{y}_{j,\tau}^t]_{j \in J, t \in \{\tau,\ldots,|T|\}}) \in \mathcal{T}_\tau$. This implies that the following polyhedron derived from the definition of $\mathcal{T}_\tau$ is empty:

$$(u_t^1) \qquad \sum_{j \in J} y_{j,\tau}^t \leq k\, \bar{x}_t \qquad \forall t \in \{\tau, \cdots, |T|\}, \tag{13a}$$

$$(u_j^2) \qquad \sum_{t=\tau}^{|T|} y_{j,\tau}^t = \bar{y}_{j,\tau} \qquad \forall j \in J, \tag{13b}$$

$$(u^3_{j,t}) \qquad y^t_{j,\tau} \le \bar{x}_t \qquad \forall j \in J, t \in \{\tau, \cdots, |T|\}, \tag{13c}$$

$$y^t_{j,\tau} \ge 0 \qquad \forall j \in J, t \in \{\tau, \cdots, |T|\}, \tag{13d}$$

and the linear program $(\mathrm{W}_\tau) := \{\max 0 \mid (13a) - (13d)\}$ is infeasible. Thus, its dual, stated as follows, must be either unbounded or infeasible:

$$\min \sum_{t=\tau}^{|T|} u^1_t k \cdot \bar{x}_t + \sum_{j \in J} u^2_j \bar{y}_{j,\tau} + \sum_{j \in J} \sum_{t=\tau} u^3_{j,t} \bar{x}_t \tag{14a}$$

subject to :

$$u^1_t + u^2_j + u^3_{j,t} \ge 0 \qquad \forall j \in J, t \in \{\tau, \ldots, |T|\}, \tag{14b}$$

$$u^1_t, u^3_{j,t} \ge 0 \qquad \forall j \in J, t \in \{\tau, \ldots, |T|\}. \tag{14c}$$

As $u^1_t = u^2_j = u^3_{j,t} = 0$, $\forall j \in J, t \in \{\tau, \ldots, |T|\}$ is feasible to (14), it must be unbounded. Let $\mathcal{C}_\tau$ be the polyhedral cone defined by (14b) and (14c). With slight abuse of notation, we represent each ray in $\mathcal{C}_\tau$ as $\left(\mathbf{u}^1, \mathbf{u}^2, \mathbf{U}^3\right)$, where $\mathbf{u}^1 \in \mathbb{R}^{|T|-\tau+1}$, $\mathbf{u}^2 \in \mathbb{R}^{|J|}$, and $\mathbf{U}^3 \in \mathbb{R}^{|J| \times (|T|-\tau+1)}$ but view $\left(\mathbf{u}^1, \mathbf{u}^2, \mathbf{U}^3\right)$ as a column vector with the last component converted into a vector by unpacking the columns of $\mathbf{U}^3$. Define $\mathbb{1}_j$ to be a vector of suitable dimension that is zero except for entry j that is equal to 1. Similarly, define $\mathbb{I}_{j,t}$ to be a matrix of appropriate dimensions that is zero except for entry (j,t) that is equal to 1. $\mathbb{1}$ is a vector of conformable dimensions with all elements equal to 1. We also use $\underline{0}$ to denote a vector whose components are all zero, and $\mathbb{0}$ to denote a matrix whose components are all zero. The matrix $\mathbb{1}_j \mathbb{1}^\intercal$ has ones in all entries of row j and zeros elsewhere. Now, consider the following rays in $\mathcal{C}_\tau$:

1. $r^1 = (\mathbb{1}, -\mathbb{1}, \mathbb{0})$
2. $r^2_j = (\underline{0}, -\mathbb{1}_j, \mathbb{1}_j \mathbb{1}^\intercal), \quad \forall j \in J$
3. $r^3_j = (\underline{0}, \mathbb{1}_j, \mathbb{0}), \quad \forall j \in J$
4. $r^4_t = (\mathbb{1}_t, \underline{0}, \mathbb{0}), \quad \forall t \in \{\tau, \ldots, |T|\}$
5. $r^5_{j,t} = (\underline{0}, \underline{0}, \mathbb{I}_{j,t}) \ , \quad \forall j \in J, t \in \{\tau, \ldots, |T|\}.$

Taking a conic combination of these rays with non-negative multipliers λ^1, λ^2_j, λ^3_j for $j \in J$, λ^4_t for $t \in \{\tau, \ldots, |T|\}$, and $\lambda^5_{j,t}$ for $j \in J, t \in \{\tau, \ldots, |T|\}$, we get:

$$u^1_t = \lambda^1 + \lambda^4_t \qquad \forall t \in \{\tau, \ldots, |T|\} \tag{15a}$$

$$u^2_j = -\lambda^1 - \lambda^2_j + \lambda^3_j \qquad \forall j \in J \tag{15b}$$

$$u^3_{j,t} = \lambda^2_j + \lambda^5_{j,t} \qquad \forall j \in J, t \in \{\tau, \dots, |T|\} \tag{15c}$$

$$\lambda^1, \lambda^2_j, \lambda^3_j, \lambda^4_t, \lambda^5_{j,t} \geq 0 \qquad \forall j \in J, t \in \{\tau, \dots, |T|\}. \tag{15d}$$

Applying Fourier-Motzkin elimination to eliminate λ^1, λ^2_j, λ^3_j for $j \in J$, λ^4_t for $t \in \{\tau, \dots, |T|\}$, and $\lambda^5_{j,t}$ for $j \in J, t \in \{\tau, \dots, |T|\}$ from (15), we get (14b) and (14c). Furthermore, the inequalities in $\dot{\mathcal{T}}_\tau$ ensure that the objective (14a) is non-negative at every ray defined above. This implies (14) is not unbounded and we have a contradiction. Thus, $\dot{\mathcal{T}}_\tau \subseteq \text{Proj}_{\left([y_{j,\tau}]_{j\in J},\, [x_t]_{t\in\{\tau,\dots,|T|\}}\right)} \mathcal{T}_\tau$. □

Proof of Proposition 6: We simplify notation by writing $y_{i,t}$ for $y_{i,j,t}$. Consider the polyhedron $\tilde{\mathcal{R}}'_j \subseteq \mathbb{R}^{2|I||T|}$, defined in variables $(\boldsymbol{\chi}, \boldsymbol{\upsilon})$ by the following constraints:

$$\upsilon_{|I|,t} \leq 1 \qquad \forall t \in T, \tag{16a}$$

$$\chi_{i,|T|} = 1 \qquad \forall i \in I, \tag{16b}$$

$$\chi_{i+1,t} - \chi_{i,t} \leq 0 \qquad \forall i \in I \setminus \{|I|\}, t \in T, \tag{16c}$$

$$\upsilon_{i,1} \leq 1 \qquad \forall i \in I \tag{16d}$$

$$\upsilon_{i,t} + \chi_{i,t-1} \leq 1 \qquad \forall i \in I, t \in T \setminus \{1\} \tag{16e}$$

$$0 \leq \chi_{i,1} \leq 1 \qquad \forall i \in I \tag{16f}$$

$$0 \leq \chi_{i,t} - \chi_{i,t-1} \leq 1, \qquad \forall i \in I, t \in T \setminus \{1\} \tag{16g}$$

$$\upsilon_{1,t} \geq 0 \qquad \forall t \in T, \tag{16h}$$

$$\upsilon_{i,t} - \upsilon_{i-1,t} \geq 0 \qquad \forall i \in I \setminus \{1\}, t \in T. \tag{16i}$$

Let $\mathcal{A}$ be the constraint matrix of (16). $\mathcal{A}$ contains entries only from $\{-1, 0, 1\}$ and each row has at most two nonzero elements. Thus, considering the column partition $(\tilde{N}_1, \tilde{N}_2)$ of $\mathcal{A}$, where $\tilde{N}_1 = \{\chi_{i,t} \mid \forall i \in I, t \in T\}$ and $\tilde{N}_2 = \{\upsilon_{i,t} \mid \forall i \in I, t \in T\}$, it follows that $\mathcal{A}$ is totally unimodular (Wolsey 2020).

Define the linear transformation $\mathcal{L} : \mathbb{R}^{2|I||T|} \to \mathbb{R}^{2|I||T|}$ by $\mathcal{L}(\boldsymbol{\chi}, \boldsymbol{\upsilon}) = (\mathbf{x}, \mathbf{y})$, where

$$x_{i,t} = \begin{cases} \chi_{i,1} & \forall i \in I, \\ \chi_{i,t} - \chi_{i,t-1} & \forall i \in I, t \in T \setminus \{1\}, \end{cases} \qquad y_{i,t} = \begin{cases} \upsilon_{1,t}, & \forall t \in T, \\ \upsilon_{i,t} - \upsilon_{i-1,t}, & \forall i \in I \setminus \{1\}, t \in T. \end{cases}$$

Each component of $\mathcal{L}(\chi, \upsilon)$ is a linear combination of the components of $(\boldsymbol{\chi}, \boldsymbol{\upsilon})$ with coefficients in $\{-1, 0, 1\}$. Hence, $\mathcal{L}$ is a linear transformation that preserves integrality. Furthermore, $\mathcal{L}(\tilde{\mathcal{R}}'_j) = \tilde{\mathcal{R}}_j$.

Let $\text{ext}(H)$ denote the set of all extreme points of a polyhedron H. Applying Proposition 3.10 of Vielma (2015) to the polyhedron $\tilde{\mathcal{R}}'_j$ using the linear transformation $\mathcal{L}$, we obtain $\text{ext}(\mathcal{L}(\tilde{\mathcal{R}}'_j)) \subseteq \mathcal{L}(\text{ext}(\tilde{\mathcal{R}}'_j))$, which simplifies to $\text{ext}(\tilde{\mathcal{R}}_j) \subseteq \mathcal{L}(\text{ext}(\tilde{\mathcal{R}}'_j))$. Since the constraint matrix for $\tilde{\mathcal{R}}'_j$ is totally unimodular, all the extreme points of $\tilde{\mathcal{R}}'_j$ are integral. Since $\mathcal{L}$ preserves integrality, every extreme point of $\tilde{\mathcal{R}}_j$ is integral. As $\tilde{\mathcal{R}}_j$ is a valid formulation for the binary set $\mathcal{R}_j$, we get $\text{conv}(\mathcal{R}_j) = \tilde{\mathcal{R}}_j$. □

Appendix B: Surgery Data

Table 6: Summary of 353-Day surgery data

Specialty	Duration average (min)	Duration standard deviation (min)	Case mix (%)
Ophthalmology	47	32	23.42
Urology	51	34	23.40
Orthopedics	98	51	14.59
Otolaryngology	109	77	13.05
Gynecology	87	53	5.22
Pediatric ENT	72	58	3.95
Pediatric Ophthalmology	65	25	3.82
Surgical Oncology	99	56	2.90
Plastics	88	58	1.84
Dentistry	131	39	1.77
Pain	61	35	1.53
Podiatry	104	69	1.48
General	67	39	1.16
Pediatric Urology	60	33	1.06
Gastroenterology	71	47	0.28
Pediatric Orthopedics	94	35	0.17
Neurology	44	8	0.14
Pediatric Neurology	42	25	0.07
Pediatric General	93	30	0.05
Oral Surgery	100	46	0.05
Gynecology Oncology	52	0	0.01
Interventional Radiology	52	4	0.01
Obstetrics	53	17	0.01
Pediatric Gastroenterology	50	1	0.01

Table 7: Size of Schedules A and B

	Schedule A	Schedule B
Number of surgeries	80	70
Number of ORs	18	18
Number of anesthesiologists	14	14

Table 8: Surgery Scheduling Details - Schedule A

Specialty		(OR, starting time period)
Type	No of Surgeries	
Dentistry	5	(8, 1), (8, 14), (17, 1), (17, 10), (17, 20)
Gynecology	7	(2, 1), (2, 12), (2, 17), (6, 1), (6, 17), (6, 24), (6, 31)
Ophthalmology	18	(7, 10), (7, 17), (13, 1), (13, 5), (13, 11), (13, 17), (13, 22), (13, 27), (13, 31), (16, 1), (16, 3), (16, 6), (16, 8), (16, 11), (16, 13), (16, 19), (16, 26), (16, 30)
Orthopedics	11	(5, 1), (5, 8), (5, 14), (5, 21), (11, 21), (14, 1), (14, 8), (14, 19), (14, 27), (15, 1), (15, 12)
Otolaryngology	8	(1, 1), (1, 9), (1, 16), (3, 16), (3, 21), (12, 1), (12, 9), (12, 23)
Pediatric Ophthalmology	2	(7, 1), (7, 4)
Plastics	7	(4, 1), (4, 5), (4, 12), (4, 17), (4, 21), (4, 25), (4, 38)
Pediatric ENT	9	(3, 1), (3, 4), (3, 7), (3, 12), (10, 1), (10, 5), (10, 13), (10, 22), (10, 29)
Surgical Oncology	3	(11, 5), (11, 9), (11, 15)
Urology	10	(7, 26), (9, 1), (9, 8), (9, 15), (9, 19), (9, 23), (9, 33), (18, 1), (18, 5), (18, 9)

Table 9: Surgery Scheduling Details - Schedule B

Specialty		(OR, starting time period)
Type	No of Surgeries	
Dentistry	4	(8, 1), (8, 10), (17, 1), (17, 20)
Gynecology	9	(6, 1), (6, 7), (14, 1), (14, 10), (14, 19), (14, 29), (15, 1), (15, 14), (15, 19)
Ophthalmology	6	(7, 1), (7, 8), (16, 1), (16, 10), (16, 17), (16, 22)
Orthopedics	9	(2, 1), (2, 4), (2, 12), (2, 17), (5, 1), (5, 8), (5, 16), (5, 22), (15, 29)
Otolaryngology	7	(1, 1), (1, 5), (1, 11), (1, 29), (10, 7), (10, 12), (12, 17)
Pediatric Ophthalmology	15	(4, 1), (4, 4), (4, 11), (4, 18), (4, 23), (4, 26), (4, 31), (4, 34), (13, 1), (13, 4), (13, 14), (13, 19), (13, 24), (13, 29), (13, 34)
Plastics	4	(11, 1), (11, 5), (11, 9), (11, 16)
Podiatry	1	(5, 28)
Pediatric ENT	6	(3, 1), (3, 6), (3, 9), (10, 1), (12, 1), (12, 6)
Urology	9	(9, 1), (9, 6), (9, 12), (9, 18), (9, 25), (18, 1), (18, 7), (18, 19), (18, 26)

Appendix C: Additional Experimental Results

Table 10: Average (max) runtime for `ReF-VI` and `ReF` with $|\Omega|= 750$ scenarios.

ζ_h–ζ_u	`ReF-VI`	`ReF`	Difference (%)
5—5	239 (676)	943 (1970)	79.04 (97.69)
5—10	171 (395)	987 (1791)	85.58 (92.17)
5—15	224 (718)	856 (1914)	78.33 (94.55)
10—5	269 (422)	1582 (2771)	71.45 (96.17)
10—10	119 (335)	814 (1272)	84.61 (97.31)
10—15	193 (498)	536 (949)	67.97 (90.62)
15—5	172 (329)	1797 (3107)	82.02 (95.02)
15—10	183 (310)	1206 (1903)	84.71 (88.77)
15—15	199 (421)	507 (968)	63.59 (90.79)

Table 11: Comparison of the three algorithms over average (max) optimality gaps (%) for $|\Omega|= 500$ and $|\Omega|= 750$.

	$\|\Omega\|= 500$			$\|\Omega\|= 750$		
ζ_h–ζ_u	`ReF-VI`	`ExF-VI`	`LS-VI`	`ReF-VI`	`ExF-VI`	`LS-VI`
5—5	0.93 (0.99)	0.91 (0.98)	91.94 (92.05)	0.91 (1.00)	0.90 (1.00)	92.01 (92.06)
5—10	0.94 (0.99)	0.85 (0.99)	94.28 (95.45)	0.83 (0.97)	0.84 (0.99)	91.92 (95.42)
5—15	0.83 (0.98)	0.88 (1.00)	92.48 (96.79)	0.83 (1.00)	0.93 (0.99)	92.75 (96.79)
10—5	0.83 (1.00)	0.90 (1.00)	92.70 (92.84)	0.95 (1.00)	0.88 (1.00)	92.92 (93.06)
10—10	0.92 (1.00)	0.89 (0.99)	95.31 (95.39)	0.89 (0.99)	0.91 (0.99)	94.80 (95.43)
10—15	0.88 (0.99)	0.95 (1.00)	96.70 (96.77)	0.81 (0.98)	0.86 (0.97)	94.72 (96.75)
15—5	0.85 (0.99)	0.84 (0.97)	92.93 (93.24)	0.93 (1.00)	0.80 (1.00)	93.11 (93.34)
15—10	0.90 (1.00)	0.96 (0.99)	95.51 (95.66)	0.90 (1.00)	0.90 (0.99)	97.85 (100.00)
15—15	0.89 (1.00)	0.89 (1.00)	96.68 (96.76)	0.79 (1.00)	0.85 (0.92)	98.17 (100.00)

Table 12: Comparison of the three algorithms over average (max) computational times (seconds) for $|\Omega|= 500$ and $|\Omega|= 750$.

	$\|\Omega\|= 500$			$\|\Omega\|= 750$		
ζ_h–ζ_u	`ReF-VI`	`ExF-VI`	`LS-VI`	`ReF-VI`	`ExF-VI`	`LS-VI`
5—5	370 (1477)	543 (801)	3602 (3610)	239 (676)	490 (869)	3601 (3602)
5—10	353 (1459)	535 (1056)	3610 (3657)	171 (395)	564 (1232)	3620 (3679)
5—15	617 (3029)	784 (2056)	3601 (3603)	224 (718)	754 (1955)	3628 (3683)
10—5	350 (933)	662 (788)	3608 (3645)	269 (422)	643 (1158)	3628 (3681)
10—10	248 (754)	680 (1238)	3600 (3601)	119 (335)	693 (967)	3637 (3682)
10—15	423 (1552)	675 (1398)	3608 (3630)	193 (498)	981 (1482)	3629 (3680)
15—5	257 (543)	462 (678)	3610 (3655)	172 (329)	442 (550)	3609 (3648)
15—10	257 (876)	626 (830)	3610 (3656)	183 (310)	613 (791)	3615 (3678)
15—15	381 (994)	875 (1460)	3610 (3657)	199 (421)	705 (971)	3612 (3663)

Appendix D: Details of the `L-shaped Algorithm`

We implement `L-shaped Algorithm` for `SASH` on a single `Branch-and-Bound` tree. We initialize the algorithm by solving the following master problem using the `Branch-and-Bound` algorithm:

$$(\mathcal{M}) \quad \text{Min} \quad \sum_{i\in I}\sum_{t\in T} c_t^r x_{i,t} + \sum_{\omega\in\Omega} p_\omega \theta_\omega \tag{17a}$$

subject to :

$$\sum_{i\in I} y_{i,j,t} \le 1 \qquad \forall j \in J, t \in T, \tag{17b}$$

$$\sum_{t\in T} x_{i,t} = 1 \qquad \forall i \in I, \tag{17c}$$

$$\sum_{\tau=1}^{t} x_{i+1,\tau} \le \sum_{\tau=1}^{t} x_{i,\tau} \qquad \forall i \in I \setminus \{|I|\}, t \in T, \tag{17d}$$

$$\sum_{j\in J} y_{i,j,t} \le k \sum_{\tau=t}^{|T|} x_{i,\tau} \qquad \forall i \in I, t \in T, \tag{17e}$$

$$\sum_{i'=1}^{i} y_{i',j,t} \le \sum_{\tau=t}^{|T|} x_{i,\tau} \qquad \forall i \in I, \forall j \in J, t \in T, \tag{17f}$$

$$\theta_\omega \ge 0 \qquad \forall \omega \in \Omega, \tag{17g}$$

$$\mathbf{x} \in \mathbb{B}^{|I|\times|T|}, \ \mathbf{y} \in \mathbb{B}^{|I|\times|J|\times|T|}. \tag{17h}$$

For each integer solution $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\boldsymbol{\theta}})$ encountered within any node of the `Branch-and-Bound` tree, the second-stage problem for each scenario $\omega \in \Omega$ is given by:

$$Q(\omega, \bar{\mathbf{x}}, \bar{\mathbf{y}}) = \text{Min} \sum_{j\in J} \sum_{t\in\mathcal{T}_j(\omega)\setminus\{\bar{t}_j(\omega)\}} c^h w_{j,t}(\omega) + \sum_{j\in J}\sum_{t\in\mathcal{T}_j(\omega)} c^u s_{jt}(\omega) \tag{18a}$$

subject to:

$$(\mu^1_{j,t}) \qquad s_{jt}(\omega) \ge 1 - \sum_{i\in I} y_{i,j,t} \qquad \forall j \in J, t \in \mathcal{T}_j(\omega), \tag{18b}$$

$$(\mu^2_{i,j,t}) \qquad w_{j,t}(\omega) \ge y_{i,j,t} - y_{i,j,t+1} \qquad \forall i \in I, j \in J, t \in \mathcal{T}_j(\omega) \setminus \{\bar{t}_j(\omega)\}, \tag{18c}$$

$$\mathbf{s}(\omega), \mathbf{w}(\omega) \ge 0. \tag{18d}$$

The optimal dual solution, indicated in parentheses before the constraints, admit closed-form expressions. In particular, for all $j \in J$ and $t \in \{1, \ldots, \bar{t}_j(\omega)\}$: $\mu^1_{j,t} = c^u$. Moreover, for all $i \in I$, $j \in J$, and

$t \in \{1, \ldots, \bar{t}_j(\omega) - 1\}$: $\mu^2_{i,j,t} = \begin{cases} c^h, & \text{if } \bar{y}_{i,j,t} - \bar{y}_{i,j,t+1} > 0 \\ 0, & \text{otherwise} \end{cases}$. For each $\omega \in \Omega$, if $Q(\omega, \bar{\mathbf{x}}, \bar{\mathbf{y}}) > \bar{\theta}_\omega$, we add the following optimality cut using *callback* to exclude the solution $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\boldsymbol{\theta}})$:

$$\theta_\omega \geq \sum_{\substack{j \in J, \\ t \in \mathcal{T}_j(\omega)}} \mu^1_{j,t}(1 - \sum_{i \in I} y_{i,j,t}) + \sum_{\substack{i \in I, j \in J, \\ t \in \mathcal{T}_j(\omega) \setminus \{\bar{t}_j(\omega)\}}} \mu^2_{i,j,t}(y_{i,j,t} - y_{i,j,t+1}). \tag{19}$$