Collaborative Operating Room Planning and Scheduling

Article *in* INFORMS Journal on Computing · August 2017

DOI: 10.1287/ijoc.2017.0745

Vahid Roshanaei
University of Toronto

Dionne M Aleman
University of Toronto

INFORMS Journal on Computing

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Collaborative Operating Room Planning and Scheduling

Vahid Roshanaei, Curtiss Luong, Dionne M. Aleman, David R. Urbach

INFORMS JOURNAL ON COMPUTING
Vol. 29, No. 3, Summer 2017, pp. 558–580
ISSN 1091-9856 (print), ISSN 1526-5528 (online)
http://pubsonline.informs.org/journal/ijoc/

Collaborative Operating Room Planning and Scheduling

Vahid Roshanaei,[a] Curtiss Luong,[a] Dionne M. Aleman,[a,b,c] David R. Urbach[b,c]

[a] Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada; [b] Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario M5S 3E3, Canada; [c] Techna Institute at University Health Network, Toronto, Ontario M5G 1P5, Canada

Contact: vroshana@mie.utoronto.ca (VR); curtiss@mie.utoronto.ca (CL); aleman@mie.utoronto.ca (DMA); david.urbach@uhn.on.ca (DRU)

Received: February 3, 2015
Revised: April 1, 2016; December 4, 2016
Accepted: December 7, 2016
Published Online: July 11, 2017

http://dx.doi.org/10.1287/ijoc.2017.0745

Abstract. Operating rooms (ORs) play a substantial role in hospital profitability, and their optimal utilization is conducive to containing the cost of surgical service delivery, shortening surgical patient wait times, and increasing patient admissions. We extend the OR planning and scheduling problem from a single independent hospital to a coalition of multiple hospitals in a strategic network, where a pool of patients, surgeons, and ORs are collaboratively planned. To solve the resulting mixed-integer dual resource constrained model, we develop a novel logic-based Benders' decomposition approach that employs an allocation master problem, sequencing sub-problems for each hospital-day, and novel multistrategy Benders' feasibility and optimality cuts. We investigate various patient-to-surgeon allocation flexibilities, as well as the impact of surgeon schedule tightness. Using real data obtained from the General Surgery Departments of the University Health Network (UHN) hospitals, consisting of Toronto General Hospital, Toronto Western Hospital, and Princess Margret Cancer Centre in Toronto, Ontario, Canada (who already engage in some collaborative resource sharing), we find that on average, collaborative OR scheduling with traditional patient-to-surgeon allocation flexibility results in 6% cost-savings, while flexible patient-to-surgeon allocation flexibility increases cost-savings to 40%, and surgeon schedule tightness can impact costs by 15%. The collective impact of our collaboration and patient flexibility results in between 45% and 63% savings per surgery. We also use a game theoretic approach to fairly redistribute the payoff acquired from a coalition of hospitals and to empirically show coalitional stability among hospitals.

History: Accepted by Allen Holder, Area Editor for Applications in Biology, Medicine, and Health Care.
Supplemental Material: Data and the online supplement are available at https://doi.org/10.1287/ijoc.2017.0745.

Keywords: healthcare • operating room • open scheduling strategy • mixed-integer linear programming • logic-based Benders' decomposition • coalition • patient-to-surgeon allocation flexibility • surgeon schedule tightness • game theory • convex games • Shapley value

1. Introduction

Operating rooms (ORs) account for 9%–10% of hospitals' annual budget and 42% of hospitals' profit (Clinical Advisory Board 2001), while they are utilized at only 68% of their nominal capacities (Health Care Financial Management Association 2005). Since many OR resources can be considered fixed expenses, improving throughput by just one additional procedure per day per OR can generate anywhere from $4 to $7 million in additional annual revenue for an average-sized organization (Health Care Financial Management Association 2005).

Collaborative scheduling in manufacturing companies has recently garnered considerable attention (Hooker 2007; Naderi and Ruiz 2010, 2014; Naderi and Azab 2014). However, in a healthcare context, ORs are not usually incorporated into network-wide re-engineering efforts, thereby limiting the positive impact of those efforts (Health Care Financial Management Association 2005). Collaborative OR planning and scheduling (CORPS) among hospitals improves resource use by enabling network-based tactical (surgeon-to-OR allocation) and operational (surgery-to-OR allocation) OR planning, allowing for surgeon and surgery allocations that are optimal for the entire network. The increased flexibility resulting from collaborative resource pooling will potentially lead to fewer required ORs and surgeons, less required overtime, and higher patient admissions (Santibanez et al. 2007). OR pooling (Batun et al. 2011) and OR block sharing (Day et al. 2012) in a single hospital has also been proven to be a cost-effective way to improve the utilization of ORs and surgeons.

We extend the OR planning and scheduling problem from a single independent hospital to a coalition of hospitals in a network, where patients, surgeons, and ORs are collaboratively planned. CORPS is a centralized bi-horizon multihospital allocation-sequencing approach to OR scheduling, which integrates (1) master surgical scheduling (how much OR time to allocate to each surgeon within each specialty) at the tactical level; (2) advanced scheduling(which day and OR to operate on a patient); and (3) allocation scheduling (which order of operations to choose) at the operational level, under an open scheduling strategy (surgeons can change their ORs in each hospital-day). CORPS aggregates the wait list of all surgeons and OR capacities of all hospitals, and allows surgeons to move among hospitals to operate on their patients. In our specific scenario, CORPS is an integrated rolling planning and scheduling problem, carried out on a weekly basis for a particular surgical department to schedule a week of surgeries two months in the future.

Collaborative resource pooling among hospitals may cause an OR capacity surplus. The possible OR time surplus can be utilized to accommodate patients that were not anticipated to be operated on in the current planning horizon. Therefore, similar to single-hospital OR scheduling literature (Guinet and Chaabane 2003, Jebali et al. 2006, Fei et al. 2009, Marques et al. 2012, Hashemi Doulabi et al. 2016), CORPS considers scheduling of two sets of patients: patients who are scheduled for the current planning horizon (mandatory patients), and patients who are scheduled for the next planning horizon (optional patients). CORPS provides a reward for each optional patient who can be fit into the current horizon's schedule. This idea has been proven to be effective in single-hospital scheduling when a patient's preallocated surgeon has some vacancy in the current planning horizon (Guinet and Chaabane 2003, Jebali et al. 2006, Fei et al. 2009, Marques et al. 2012, Hashemi Doulabi et al. 2016).

Collaborative scheduling can be further enriched by incorporating a new decisional aspect that we call discretionary patient-to-surgeon allocation flexibility. We investigate the impact of three different types of patient-to-surgeon allocation flexibilities: dedicated (Fei et al. 2009, Marques et al. 2012, Hashemi Doulabi et al. 2016), semiflexible (novel), and flexible (novel). In dedicated patient-to-surgeon allocation, each patient is assigned to exactly one surgeon a priori, the advantage of which is the certainty in the day on which a surgery is executed; however, rigid a priori assignments decrease potential utilization gains. In semiflexible allocation, patients from the current planning horizon are assigned to exactly one surgeon a priori; however, patients from the next planning horizon are willing to be operated on in the current planning horizon by any eligible surgeons. The advantage of having flexible optional patients is to increase the likelihood of surgeon and OR utilization, while providing the comfort of surgeon certainty for mandatory patients. In flexible patient-to-surgeon allocation, patients both from the current and next planning horizon are willing to be operated on by any eligible surgeon. The advantage of flexibility for all patients is an ability to achieve maximum utilization; however, in reality, it is impossible to consider all patients as flexible due to patients' distinctive perceptions of surgeons' capabilities.

ORs also have room for improvements in scheduling-related issues, including turnover rates (Health Care Financial Management Association 2005). The average patient-out-to-patient-in time is 31.5 minutes, but best practice is 15 minutes; therefore, effective preparation and cleaning times are essential to improve turnover (Health Care Financial Management Association 2005). Some studies in OR scheduling have considered patients' preparation, surgical, and cleaning times toward a surgeon's availability time (Jebali et al. 2006, Fei et al. 2009, Denton et al. 2010), while others have just considered the surgical time of each patient toward surgeon availability time (Pham and Klinkert 2008, Batun et al. 2011), allowing a surgeon to possibly circumvent cleaning and preparation times of a surgery. The first approach is conservative as it admits fewer patients given the availability times of surgeons, while the second one is more aggressive because more patients can be scheduled for surgery. Hence, we incorporate a new coefficient into our model to determine the best trade-off between the two approaches. The coefficient determines the tightness of a surgeon's schedule among his/her surgeries. We empirically show how fine-tuning this coefficient increases OR and surgeon throughput and consequently results in better utilization of OR capacities.

We additionally examine surgical sequencing for efficiency gains. The incorporation of surgical case sequencing into already-hard surgery-to-multiple-OR allocation makes the problem even more computationally challenging even in the presence of small numbers of patients (Hashemi Doulabi et al. 2016). The complexity of this problem is further intensified with more patients and patient-to-surgeon allocation flexibility in a collaborative environment. An easy yet sub-optimal workaround for this problem is to ignore the feasible sequencing among surgeries and solve the problem under an open scheduling strategy (Fei et al. 2008, 2009) and hope that the open scheduling strategy will provide an optimal solution with respect to sequencing constraints (Fei et al. 2009). However, there is no guarantee that such an approach will lead to a feasibly sequenceable solution, much less an optimal solution.

The scheduling procedure that we use to allocate patients to ORs is based on an open scheduling strategy, the main aim of which is to accommodate as many surgical cases as possible (Fei et al. 2009). Due to the potential increase in the number of patients that can be scheduled (Fei et al. 2009), open scheduling has recently garnered significant attention (Guinet and Chaabane 2003, Jebali et al. 2006, Pham and Klinkert 2008, Fei et al. 2009, Batun et al. 2011, Marques et al. 2012, Hashemi Doulabi et al. 2016). Open scheduling is equally popular in both private (Batun et al. 2011) and publicly-funded hospitals (Marques et al. 2012, Vijayakumar et al. 2013), and UHN currently uses a combination of block and open scheduling to allocate OR times to surgeons in each day, so a move to a fully open schedule is not unrealistic. The actual number of patients that can be scheduled in ORs depends also on the availability of beds in downstream units of ORs (Jebali et al. 2006), which is mostly relaxed in OR scheduling optimization due to computational complexity (Fei et al. 2009, Denton et al. 2010, Batun et al. 2011, Marques et al. 2012, Hashemi Doulabi et al. 2016). An open scheduling strategy might generate infeasible schedules with respect to sequencing constraints, and thus we employ a two-stage allocation-sequencing approach similar to existing methods (Jebali et al. 2006, Fei et al. 2010) in which patients are allocated to ORs in the first stage and sequenced in the second stage. Jebali et al. (2006) proposed two strategies for the sequencing of surgical cases in the second stage: (1) pure sequencing in which the optimal precedence and starting time of surgeries are determined given the patient-to-OR allocation decision in the first-stage, and (2) sequencing with possible patient-to-OR reallocation. The second approach is more likely to ensure a feasible schedule with respect to the first stage constraints, while the first approach is more computationally tractable. Fei et al. (2009) solved an OR planning problem using a column-generation heuristic to determine a date of surgery for each patient, and then used a hybrid genetic algorithm to solve a daily surgical case sequencing as a two-stage hybrid flow-shop problem. Hashemi Doulabi et al. (2016) resolved the issue of feedback mechanism between the allocation and sequencing stage by developing a branch-and-price-and-cut approach, which was effective in solving small-to-medium instances of OR scheduling and reached 12% optimality gap when scheduling 100 patients in five hours CPU time. However, neither study considered downstream resource availability for tractability, which is a common simplification in OR scheduling (Fei et al. 2009, Denton et al. 2010, Batun et al. 2011, Hashemi Doulabi et al. 2016).

Differently from the column-generation-based feedback mechanism proposed by Hashemi Doulabi et al. (2016), we develop a novel logic-based Benders' decomposition (LBBD) algorithm to optimize both resource allocation and sequencing in CORPS. LBBD (Hooker and Ottosson 2003) has been proven to be faster than existing mathematical models for scheduling (Hooker 2007, Tran et al. 2016) and location-allocation (Fazel-Zarandi and Beck 2012) problems by multiple orders of magnitude. Our LBBD consists of a global nonhomogeneous dual resource constrained (DRC) allocation master problem (MP) based on a novel mixed-integer open scheduling strategy and a novel local DRC allocation-sequencing subproblem (SP). The purpose of our LBBD is to remedy the inherent infeasibilities in the open scheduling strategy and to compute more accurate cost values for the MP in the presence of sequencing constraints. Therefore, the sequencing SPs are used to refine the MP allocations and guide the MP toward optimality.

Systems are known as DRC when capacity constraints stem from both machines (ORs) and human operators (surgeons). More specifically, a system can be defined as DRC when operators (surgeons) are the constraining resource who can transfer across various workstations (ORs) as required (Xu et al. 2011). Generally, DRC scheduling problems consist of resource allocation and operation sequencing problems; the former is to allocate the required resources (ORs and surgeons) to each operation, and the latter is to determine the operational sequence of all jobs (patients) on each resource. CORPS is a DRC OR scheduling problem, stemming from the limited availability time of surgeons and ORs. DRC scheduling has garnered significant attention in parallel machine (Ruiz-Torres and Centeno 2007, Chaudhry and Drake 2009) and job shop (ElMaraghy et al. 1999, 2000; Lei and Guo 2014) environments, and is mostly solved with heuristics, including genetic algorithms (ElMaraghy et al. 1999, 2000; Chaudhry and Drake 2009), variable neighborhood search (Lei and Guo 2014), and customized approaches (Ruiz-Torres and Centeno 2007). Many of the previous studies on DRC systems only consider homogeneous workers (surgeons are nonhomogeneous), and mathematical modeling has not received much attention in DRC environments (Xu et al. 2011).

The problems of deterministic single-objective surgery sequencing (Jebali et al. 2006), deterministic bi-objective surgery allocation (Fei et al. 2009), and two-stage stochastic multiobjective surgery allocation-sequencing (Batun et al. 2011) DRC OR mixed-integer scheduling models have already been studied. To the best of our knowledge and according to three recent OR scheduling review papers (Guerriero and Guido 2011, Cardoen et al. 2010, Samudra et al. 2016), there are no DRC models that consider patient-to-surgeon allocation flexibility. Therefore, our DRC OR mixed-integer scheduling approach that considers nonhomogeneity of OR equipment and surgeon skills with patient-to-surgeon allocation flexibility is novel. Our DRC sequencing SP model can be used by both manufacturing companies and hospitals, where the execution of a job requires the availability of both the eligible operator and the machine.

We use the notion of Shapley values (Shapley 1953, Leng and Parlar 2009) to fairly redistribute the acquired payoff (cost-savings) among collaborating hospitals. We also use the notion of game convexity (Shapley 1971, Leng and Parlar 2009) to empirically determine the most stable coalition for each hospital in the network. We show that the grand coalition among hospitals becomes convex and stable as we increase flexibilities and problem size. Few game theoretic approaches with healthcare applications have been studied (Westhoff et al. 2012, Agee and Zane 2013). Westhoff et al. (2012) investigated Nash equilibrium and Prisoner's dilemma on four hypothetical cooperative and competitive public health case studies, and concluded that public health agencies should move toward more collaborative models of service delivery to enhance efficient and effective service delivery. Agee and Zane (2013) used game theory to show that a cooperative pricing framework between health insurers and hospitals benefits both parties and they could lower costs and increase profits. To the best of our knowledge, no study has used Shapley value and game convexity in a healthcare environment.

We consider CORPS in the University Health Network (UHN) hospitals, consisting of three hospitals in Toronto, Ontario, Canada: Toronto General Hospital (TGH), Toronto Western Hospital (TWH), and Princess Margaret Cancer Centre (PMCC). UHN hospitals already engage in wait list sharing and resource pooling in a nonsystematic way, and are culturally receptive to sharing resources. In our study, CORPS is a cooperative game among the UHN hospitals because they operate under a universal healthcare system in which hospitals aim to minimize their surgical delivery costs, reduce wait times, and increase patient throughput given existing resources. Using real data, we make the following contributions. We investigate collaborative DRC scheduling with flexible patient-to-surgeon allocation (which has not been explored in the literature), and develop an exact LBBD optimization algorithm for general collaborative DRC scheduling problems and prove the validity of the Benders' feasibility and optimality cuts. We further propose a novel approach to assess the fairness of collaborative schedules using game theory techniques. To test our methods, we examine OR scheduling in a collaborative environment, which has not been investigated in the healthcare literature, and we quantify the value of various patient-to-surgeon flexibility schemes, surgeon schedule tightness, and hospital collaboration. We also develop a novel scheme for generating correlated random preparation and surgical times inspired by real UHN data, which can be applied to a wide variety of OR scheduling problems.

2. Logic-Based Benders' Decomposition Approach

CORPS is an integrated rolling planning and scheduling problem, carried out on a weekly basis and includes a set of patients ($p \in \mathcal{P}$), surgeons ($s \in \mathcal{S}$), hospitals ($h \in \mathcal{H}$), days ($d \in \mathcal{D}$), and ORs in each hospital ($r \in \mathcal{R}_h$). Through a primary care physician, each patient selects a preferred list of one or more surgeons (Ω_p) to perform his or her operation. The execution of a surgery hinges on the availability of dual resources: eligible surgeons (Ω_p) and eligible ORs ($\mathcal{Q}_{ph}$). Since surgeons can freely move among collaborating hospitals in different days of a week, CORPS allocates surgeons to hospital-days ($z_{shd} \in \{0,1\}$). Further, it determines whether an OR should be opened in each hospital-day ($y_{hdr} \in \{0,1\}$) and how much overtime per opened OR should be used ($v_{hdr} \geq 0$). Given daily OR regular time (B_{hdr}) and possible overtime (v_{hdr}) in each hospital-day, surgeon availability time in each day (A_{sd}), and the due date of each patient (θ_p), CORPS allocates each patient to a surgeon, hospital, day, and an OR ($x_{pshdr} \in \{0,1\}$) before his or her due-date and ensures that the required resources (i.e., eligible surgeons and ORs), are feasibly allocated to each hospital-day. All the mandatory patients ($\theta_p \leq |\mathcal{D}|$) are allocated to a surgeon-hospital-day-OR and if OR and surgeon time surpluses exist, CORPS may schedule optional patients ($\theta_p > |\mathcal{D}|$). CORPS additionally ensures feasible sequencing. CORPS minimizes the cost of opening ORs, using surgeons, and incurring OR overtime, and maximizes the utility of optional patients (U_p) if operated on in the current planning horizon. CORPS notation is shown in Table 1.

While CORPS can be formulated as a single mathematical model and solved via optimization solvers, the integrated model explodes with the number of hospitals, ORs, days, and surgeons in the planning horizon. As empirical evidence, the branch-and-price-and-cut of Hashemi Doulabi et al. (2016) solved problems with 80 and 100 patients to 9.92% and 12.01% optimal gaps, respectively, in five hours computational time, but did not include multiple hospitals and patient flexibility. CORPS with only 40 patients as a single mixed-integer optimization was unable to achieve an optimality gap of 20% within 10 hours of CPU time.

Therefore, we use an LBBD approach to solve CORPS to optimality. The classical Benders' approach cannot handle CORPS due to its mixed-integer progamming (MIP) SPs rather than linear programming SPs. Unlike classical Benders' decomposition (Benders 1962), LBBD imposes no structural restrictions (e.g., linearity) on the different components of the decomposition (Tran et al. 2016); therefore, LBBD can accommodate MPs and SPs of various mathematical natures. Like classical Benders' decomposition, LBBD decomposes a problem into an MP (consisting of primary variables) and one or more SPs (consisting of secondary variables), and connects them via Benders' feasibility or optimality cuts. Unlike classical Benders' decomposition, which relies on the solution of its SP dual to derive a Benders' cut, LBBD provides no standard scheme for generating Benders' cuts and thus cuts must be devised and proven valid for each problem class uniquely. Hooker and Ottosson (2003) maintain that the notion of duality must be definable for any SPs, not just for linear ones. Thus, a generalized dual can

Table 1. CORPS Notation

Indices:	
p, s, h, d, r	Index for patients, surgeons, hospitals, days, and ORs, respectively
Sets:	
$\mathcal{P}$	Set of patients ($p = 1, \ldots, \lvert\mathcal{P}\rvert$)
$\mathcal{S}$	Set of surgeons ($s = 1, \ldots, \lvert\mathcal{S}\rvert$)
$\mathcal{H}$	Set of hospitals ($h = 1, \ldots, \lvert\mathcal{H}\rvert$)
$\mathcal{D}$	Set of days in the planning horizon ($d = 1, \ldots, \lvert\mathcal{D}\rvert$)
$\mathcal{R}_h$	Set of ORs in each hospital, ($r = 1, \ldots, \lvert\mathcal{R}_h\rvert$)
$\mathcal{P}_{hdr}$	Set of patients that can be operated on in room r of hospital h with $\theta_p \geq d$
Λ_{sd}	Set of patients belonging to surgeon s and $\theta_p \geq d$
Ω_p	Set of preferred surgeons for patient p
Δ_s	Set of days in which surgeon s operates
$\mathcal{Q}_{ph}$	Set of ORs at each hospital h eligible for patient p
Parameters:	
K_{hdr}, C_{hdr}	Fixed and variable costs of OR r in hospital h on day d, respectively
L_{shd}	Fixed cost of surgeon s operating in hospital h on day d
B_{hdr}	Regular time of each OR r in hospital h on day d
T_{ps}	Total time of preparation, surgery, and cleaning time of patient p operated on by surgeon s
F_p	Preparation time of patient p
G_p	OR turnover time after patient p, including OR cleaning time
E_{ps}	Time required for the surgical procedure of patient p operated on by surgeon s, $E_{ps} = T_{ps} - (G_p + F_p)$
A_{sd}	Available time of surgeon s on day d
U_p	Reward assigned to patient p from next horizon if operated on in the current horizon
θ_p	Due date of patient p
V_{hdr}	Maximum allowable amount of overtime
α	Surgeon schedule tightness coefficient
Variables:	
x_{pshdr}	1 if patient p is operated by surgeon s in hospital h on day d in room r, 0 otherwise
y_{hdr}	1 if room r of hospital h on day d is opened, 0 otherwise
z_{shd}	1 if surgeon s is in hospital h on day d, 0 otherwise
v_{hdr}	Continuous variable for overtime of OR r in hospital h on day d

be defined as an inference dual, which is responsible for inferring the tightest possible bound from the SP constraint set. The SP inference dual solution is used as the tightest bound on the optimal MP value, and the solution of the dual inference is incorporated into the LBBD cuts to guide the master search toward feasibility or optimality by providing a feedback on the quality of the previous MP solution. The LBBD iterates until the MP solution is feasible and equal to the SP solution.

The LBBD for CORPS (Figure 1) has a novel MIP allocation MP, novel multiple single-objective MIP sequencing SPs, and novel multistrategy Benders' feasibility and optimality cuts with associated proofs. CORPS is decomposed into hospital-day SPs rather than hospital SPs because there is no resource-sharing after resources are allocated to each hospital-day, leading to fully independent SPs. The MIP/MIP-based LBBD decomposition has been proven effective for scheduling problems (Harjunkoski and Grossmann 2002). The MIP allocation MP allocates patients and surgeons to hospital-days and decides which ORs to open. Then, the SPs allocate patient-surgeons to ORs and sequence them. The MP minimizes the fixed costs of opening ORs, the variable costs of OR overtime, the cost of surgeons, and the negative of the rewards assigned to optional patients if operated on in the current planning horizon. The MIP surgical case scheduling SP minimizes OR overtime in the presence of sequencing constraints. The MP derives a lower bound on the amount of overtime for each SP, as it cannot precisely determine the actual amount of OR overtime because it does not include sequencing constraints. Given the MP solution at each iteration, the SPs perform the feasible sequencing, and return feasibility and optimality cuts to the MP if the MP solution results in infeasible or suboptimal SPs, respectively. LBBD converges to optimality when the optimal value of MP overtime for each SP ($\hat{v}_{hdr}^{(i)}$ at iteration i) is equal to the value of overtimes of SPs ($\bar{v}_{hdr}^{(i)}$). Thus, the linking variable between the MP and SPs is the continuous variables v_{hdr} (Figure 2). Additionally, our LBBD incorporates an SP feasibility check before solving SPs, reducing computational burden by only solving SPs when all SPs are feasible (Harjunkoski and Grossmann 2002).

For computational tractability, CORPS is solved based on the following common assumptions: (1) ORs are not functionally identical (Jebali et al. 2006, Roland et al. 2010); (2) surgeons' capabilities are not homogeneous; (3) the allocation of anesthetists and nurses-to-ORs have been decided a priori (Denton et al. 2010, Batun et al. 2011, Marques et al. 2012, Hashemi Doulabi et al. 2016); (4) durations of surgical cases are deterministic (Jebali et al. 2006, Marques et al. 2012, Hashemi Doulabi et al. 2016); and (5) the number of beds in post-anesthesia and intensive care units is well-sized and does not cause any cancellation (Denton et al. 2010, Batun et al. 2011, Marques et al. 2012, Hashemi Doulabi et al. 2016). The efficiency gained in unused ORs can be used to accommodate emergency patients or to mitigate uncertainties in surgery durations.

2.1. Allocation Master Problem

The LBBD MP is a MIP allocation problem, which relaxes the sequencing of surgical operations and allocates patients and surgeons to open hospital-day-ORs. If the MP is infeasible, the optimization process is terminated. Alternatively, if the MP is feasible, the primary variables produced by the MP are then used as inputs in the SPs. If the MP solution results in infeasible or suboptimal SPs, we develop Benders' cuts and

Figure 1. LBBD Algorithm

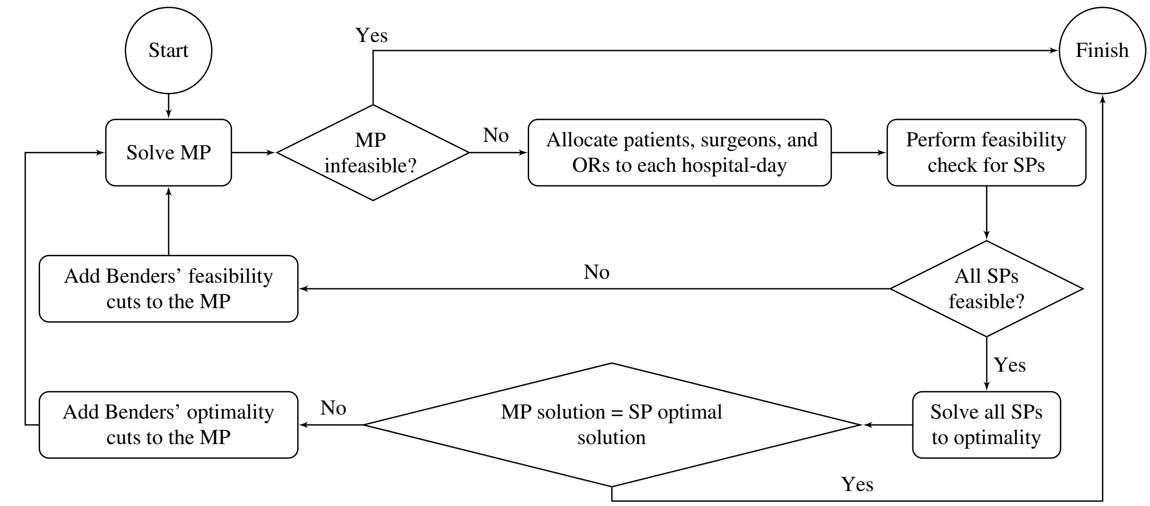
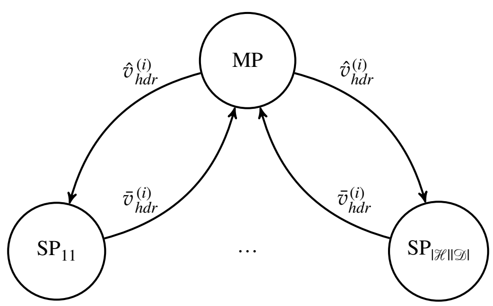


Figure 2. The Linking Variable Between the MP and SPs



Note. An MP solution is considered optimal if $\hat{v}^{(i)}_{hdr} = \bar{v}^{(i)}_{hdr}$.

incorporate them into the MP. We then re-solve the MP with the Benders' cuts.

The LBBD MP minimizes the cost of opening ORs, using surgeons, and incurring OR overtime, and maximizes the utility of optional patients if operated on in the current planning horizon. The MIP MP is as follows:

$$\min \Bigg\{ \sum_{h\in\mathscr{H}} \sum_{d\in\mathscr{D}} \sum_{r\in\mathscr{R}_h} K_{hdr} y_{hdr} + \sum_{s\in\mathscr{S}} \sum_{h\in\mathscr{H}} \sum_{d\in\mathscr{D}} L_{shd} z_{shd} + \sum_{h\in\mathscr{H}} \sum_{d\in\mathscr{D}} \sum_{r\in\mathscr{R}_h} C_{hdr} v_{hdr} - \sum_{p\in\mathscr{P}|\theta_p>|\mathscr{D}|} U_p \sum_{s\in\Omega_p} \sum_{h\in\mathscr{H}} \sum_{d\in\mathscr{D}} \sum_{r\in\mathscr{Q}_{ph}} x_{pshdr} \Bigg\} \quad \text{(MP)}$$

$$\text{s.t.} \quad \sum_{s\in\Omega_p} \sum_{h\in\mathscr{H}} \sum_{d\le\theta_p} \sum_{r\in\mathscr{Q}_{ph}} x_{pshdr} = 1 \quad \forall\, p\in\mathscr{P} \mid \theta_p \le |\mathscr{D}| \quad (1)$$

$$\sum_{s\in\Omega_p} \sum_{h\in\mathscr{H}} \sum_{d\in\mathscr{D}} \sum_{r\in\mathscr{Q}_{ph}} x_{pshdr} \le 1 \quad \forall\, p\in\mathscr{P} \mid \theta_p > |\mathscr{D}| \quad (2)$$

$$\sum_{h\in\mathscr{H}} z_{shd} \le 1 \quad \forall\, s\in\mathscr{S};\ d\in\Delta_s \quad (3)$$

$$x_{pshdr} \le z_{shd} \quad \forall\, p\in\mathscr{P};\ s\in\Omega_p;\ h\in\mathscr{H};\ d\in\Delta_s;\ r\in\mathscr{Q}_{ph} \quad (4)$$

$$x_{pshdr} \le y_{hdr} \quad \forall\, p\in\mathscr{P};\ s\in\Omega_p;\ h\in\mathscr{H};\ d\in\Delta_s;\ r\in\mathscr{Q}_{ph} \quad (5)$$

$$\sum_{p\in\mathscr{P}_{hdr}} \sum_{s\in\Omega_p} T_{ps} x_{pshdr} \le B_{hdr} y_{hdr} + v_{hdr} \quad \forall\, h\in\mathscr{H};\ d\in\Delta_s;\ r\in\mathscr{R}_h \quad (6)$$

$$\sum_{p\in\Lambda_{sd}} \sum_{r\in\mathscr{Q}_{ph}} (\alpha E_{ps} + (1-\alpha) T_{ps}) x_{pshdr} \le A_{sd} z_{shd} \quad \forall\, s\in\mathscr{S};\ h\in\mathscr{H};\ d\in\Delta_s \quad (7)$$

$$V_{hdr} \ge v_{hdr} \ge 0 \quad \forall\, h\in\mathscr{H};\ d\in\mathscr{D};\ r\in\mathscr{R}_h \quad (8)$$

$$x_{pshdr}, y_{hdr}, z_{shd} \in \{0,1\} \quad \forall\, p\in\mathscr{P};\ s\in\Omega_p;\ h\in\mathscr{H};\ d\in\Delta_s;\ r\in\mathscr{Q}_{ph}.$$

Constraint (1) ensures that mandatory patients are scheduled, while constraint (2) states that some of the optional patients may be operated on in this horizon. Constraint (3) ensures that each surgeon works at most in one hospital in each day. Constraints (4) and (5) ensure that a patient is feasibly assigned to a hospital-day if and only if the eligible surgeon is available and the eligible OR is open. Constraint (7) enforces that the daily workload of a surgeon will not exceed his/her available time. The α values range between $[0,1]$, where 0 and 1 indicate loosest and tightest surgeon schedules, respectively. Constraint (8) sets an upper bound on the amount of overtime that can be used for each OR. Constraint (6) limits the surgical workload of an OR to the allotted regular time and overtime. Dedicated, semiflexible, and flexible patient-to-surgeon allocations are differentiated in the input parameter Ω_p (dedicated: $|\Omega_p| = 1$, $\forall\, p\in\mathscr{P}$; semiflexible: $|\Omega_p| = 1$, $\forall\, p\in\mathscr{P} \mid \theta_p \le |\mathscr{D}|$ and $|\Omega_p| > 1$, $\forall\, p\in\mathscr{P} \mid \theta_p > |\mathscr{D}|$; flexible: $|\Omega_p| = |\mathscr{S}|$, $\forall\, p\in\mathscr{P}$).

The MP solution in iteration i is denoted $\hat{y}^{(i)}_{hdr}$, $\hat{z}^{(i)}_{shd}$, $\hat{x}^{(i)}_{pshdr}$, and $\hat{v}^{(i)}_{hdr}$. The sets of ORs, surgeons, and patients that the MP assigns to hospital h on day d are denoted $\hat{\mathcal{R}}^{(i)}_{hd}$, $\hat{\mathcal{S}}^{(i)}_{hd}$, and $\hat{\mathcal{P}}^{(i)}_{hd}$, respectively. These variable and set solutions for hospital-day hd are used as inputs to the SP for hospital-day hd. We use two stopping criteria for the MP: (1) a specific optimality gap and (2) maximum time per iteration, which is 90% of the total allowable CPU time. The first stopping criterion provides a solution to SPs when the MP reaches the predetermined gap, whereas the second criterion ensures that at least one MP incumbent solution (with a known gap) is given to the SPs for sequence feasibility verification, if the MP does not reach the predetermined gap. If the MP incumbent solution results in suboptimal and optimal SPs, the global optimality gap is computed as follows:

$$\text{Optimality gap} = \frac{(z^{\text{MP}} - z^{\hat{V}} + z^{\bar{V}}) - z^{\text{LP}}}{z^{\text{LP}}},$$

where z^{LP} is the objective function value of the MP linear programming relaxation, z^{MP} is the MP incumbent objective function value, $z^{\hat{V}}$ is the overtime cost computed by the MP, and $z^{\bar{V}}$ is the sum of overtime costs in all SPs.

2.2. Subproblems

Surgical case sequencing SPs are created based on the MP outputs to optimize the order of operations. We create an SP for each hospital-day to which at least one patient, surgeon, and OR has been assigned. The primary variables inherited from the MP (x_{psr}) are reoptimized in each SP in the presence of a new set of constraints, ensuring that no two operations (i.e., patients, denoted p and k) belonging to a surgeon ($\pi_{pks} \in \{0,1\}$) or an OR ($\eta_{pkr} \in \{0,1\}$) overlap. Despite varying degrees of total surgery time reduction via $\alpha \in \{0,1\}$, surgeries are considered for sequencing in their entirety, including actual preparation, surgical, and cleaning times. The SP notation is given in Table 2.

The SP is a MIP model, minimizing the cost of incurring overtime for open ORs in each hospital-day. The SP determines the optimal starting and finishing time of each surgeon. The optimal solution of the SP at iteration i ($\bar{v}^{(i)}_{hdr}$) may differ from that of the MP ($\hat{v}^{(i)}_{hdr}$) due to the reoptimization of allocation decisions to obtain feasible sequencing among operations. Before optimality, $\hat{v}^{(i)}_{hdr} < \bar{v}^{(i)}_{hdr}$, and at optimality, $\hat{v}^{(i)}_{hdr} = \bar{v}^{(i)}_{hdr}$. The SP is as follows:

$$\text{minimize} \quad \bar{v}^{(i)}_{hdr} = \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}} c_r v_r \cancel{v_r} \qquad \text{(SP)}$$

Table 2. SP Notation

Sets:	
$\mathcal{Q}_p$	Set of qualified ORs for patient p
$\mathcal{Q}_{pk}$	Set of qualified shared ORs between patient p and patient k
Ω_p	Set of qualified surgeons for patient p
Ω_{pk}	Set of shared surgeons between patient p and patient k
Parameters:	
A_s	Available time of surgeon s
B_r	Available time of OR r
M	A large positive number
Binary variables:	
x_{psr}	1 if patient p is operated by surgeon s in OR r, and 0 otherwise
η_{pkr}	1 if patient p is operated after patient k in OR r, and 0 otherwise
π_{pks}	1 if patient p is operated after patient k on surgeon s list, and 0 otherwise
Continuous variables:	
f_p	Finishing time of surgical case p
c_r	Completion time of OR r
v_r	Overtime of OR r
i_s	Starting time of surgeon s
e_s	Ending time of surgeon s

$$\text{s.t.} \quad \sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}\cap\Omega_p} \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_p} x_{psr} = 1 \quad \forall\, p\in\hat{\mathcal{P}}^{(i)}_{hd} \tag{9}$$

$$f_p \geq F_p + \sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}\cap\Omega_p} \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_p} E_{ps} x_{psr} \quad \forall\, p\in\hat{\mathcal{P}}^{(i)}_{hd} \tag{10}$$

$$\begin{aligned} f_p \geq{}& f_k + G_k + F_p + \sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}\cap\Omega_p} E_{ps} x_{psr} \\ &- M\Bigg(3 - \eta_{pkr} - \sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}\cap\Omega_p} x_{psr} - \sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}\cap\Omega_k} x_{ksr}\Bigg) \\ &\forall\, p,k\in\hat{\mathcal{P}}^{(i)}_{hd} \mid p<k;\, r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_{pk} \end{aligned} \tag{11}$$

$$\begin{aligned} f_k \geq{}& f_p + G_p + F_k + \sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}\cap\Omega_k} E_{ks} x_{ksr} \\ &- M\Bigg(2 + \eta_{pkr} - \sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}\cap\Omega_p} x_{psr} - \sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}\cap\Omega_k} x_{ksr}\Bigg) \\ &\forall\, p,k\in\hat{\mathcal{P}}^{(i)}_{hd} \mid p<k;\, r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_{pk} \end{aligned} \tag{12}$$

$$\begin{aligned} f_p \geq{}& f_k + E_{ps} \\ &- M\Bigg(3 - \pi_{pks} - \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_p} x_{psr} - \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_k} x_{ksr}\Bigg) \\ &\forall\, p,k\in\hat{\mathcal{P}}^{(i)}_{hd} \mid p<k;\, r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_{pk} \end{aligned} \tag{13}$$

$$\begin{aligned} f_k \geq{}& f_p + E_{ks} \\ &- M\Bigg(2 + \pi_{pks} - \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_p} x_{psr} - \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_k} x_{ksr}\Bigg) \\ &\forall\, p,k\in\hat{\mathcal{P}}^{(i)}_{hd} \mid p<k;\, r\in\hat{\mathcal{R}}^{(i)}_{hd}\cap\mathcal{Q}_{pk} \end{aligned} \tag{14}$$

$$f_p + G_p - M\left(1 - \sum_{s\in\hat{\mathscr{S}}^{(i)}_{hd}\cap\Omega_p} x_{psr}\right) \le B_r + v_r \quad \forall\, p\in\hat{\mathscr{P}}^{(i)}_{hd};\, r\in\hat{\mathscr{R}}^{(i)}_{hd}\cap\mathscr{Q}_p \tag{15}$$

$$e_s \ge f_p - M\left(1 - \sum_{r\in\hat{\mathscr{R}}^{(i)}_{hd}\cap\mathscr{Q}_p} x_{psr}\right) \quad \forall\, p\in\hat{\mathscr{P}}^{(i)}_{hd}\cap\Lambda_s;\, s\in\hat{\mathscr{S}}^{(i)}_{hd}\cap\Omega_p \tag{16}$$

$$i_s \le f_p - E_{ps} + M\left(1 - \sum_{r\in\hat{\mathscr{R}}^{(i)}_{hd}\cap\mathscr{Q}_p} x_{psr}\right) \quad \forall\, p\in\hat{\mathscr{P}}^{(i)}_{hd}\cap\Lambda_s;\, s\in\hat{\mathscr{S}}^{(i)}_{hd}\cap\Omega_p \tag{17}$$

$$e_s - i_s \le A_s \quad \forall\, s\in\hat{\mathscr{S}}^{(i)}_{hd} \tag{18}$$

$$c_r \ge f_p + G_p - M\left(1 - \sum_{s\in\hat{\mathscr{S}}^{(i)}_{hd}\cap\Omega_p} x_{psr}\right) \quad \forall\, p\in\hat{\mathscr{P}}^{(i)}_{hd};\, r\in\hat{\mathscr{P}}^{(i)}_{hd}\cap\mathscr{Q}_p \tag{19}$$

$$0 \le v_r \le V_r \quad \forall\, r\in\hat{\mathscr{R}}^{(i)}_{hd} \tag{20}$$

$$v_r \ge c_r - B_r \quad \forall\, r\in\hat{\mathscr{R}}^{(i)}_{hd} \tag{21}$$

$$x_{psr}\in\{0,1\} \quad \forall\, p\in\hat{\mathscr{P}}^{(i)}_{hd};\, s\in\hat{\mathscr{S}}^{(i)}_{hd}\cap\Omega_p;\, r\in\hat{\mathscr{R}}^{(i)}_{hd}\cap\mathscr{Q}_p$$

$$\pi_{pkr}\in\{0,1\} \quad \forall\, p=1,\ldots,|\hat{\mathscr{P}}^{(i)}_{hd}|-1;\, p<k\le|\hat{\mathscr{P}}^{(i)}_{hd}|;\; r\in\hat{\mathscr{R}}^{(i)}_{hd}\cap\mathscr{Q}_{pk}$$

$$\eta_{pks}\in\{0,1\} \quad \forall\, p=1,\ldots,|\hat{\mathscr{P}}^{(i)}_{hd}|-1;\, p<k\le|\hat{\mathscr{P}}^{(i)}_{hd}|;\; s\in\hat{\mathscr{S}}^{(i)}_{hd}\cap\Omega_{pk}$$

$$f_p, e_s, i_s, c_r, v_r \ge 0 \quad \forall\, p\in\hat{\mathscr{P}}^{(i)}_{hd};\, s\in\hat{\mathscr{S}}^{(i)}_{hd};\, r\in\hat{\mathscr{R}}^{(i)}_{hd}.$$

Constraint (9) ensures that each patient is assigned to only one surgeon-OR. Constraint (10) computes the finishing time of each surgery. Constraints (11) and (12) are either-or constraints, which determine the precedence among operations assigned to OR r, while constraints (13) and (14) determine the precedence among the operations assigned to surgeon s. If two patients belong to one surgeon and they are assigned to two different ORs, the surgeon may not have to wait for the cleaning time of the previous operation. Constraints (13) and (14) also ensure that an operation cannot be simultaneously the successor and the predecessor of another operation within a surgeon's surgery list. Constraint (15) states that the allocated workload to an OR should not exceed its allotted regular and overtimes. Constraint (16) states that the ending time of a surgeon is after the finishing time of all surgical cases assigned to that surgeon, and constraint (17) states that the beginning time of a surgeon is before the starting time of all surgical cases assigned to that surgeon. Constraint (18) ensures that the surgical workload of a surgeon should not exceed the surgeon's daily allotted OR time. Constraint (19) computes the finishing time of each OR. Constraint (20) ensures that the workload of an OR does not exceed the maximum session length of ORs determined by the hospital. Constraint (21) computes the value of overtime for each OR. Big M parameters are used to enforce logical relationships between the scheduled time of patients that are scheduled on the same day, in the same hospital. Big M for each SP is equal to the maximum of the total OR time for that day, in the ORs in the hospital considered by that SP, plus one. Since patients cannot be scheduled further apart than the available time in a day, these values for Big M correctly enforce relationships between patients.

2.3. Feasibility Check

In each LBBD iteration, we may have up to 15 SPs ($|\mathscr{H}|\times|\mathscr{D}|$), depending on $|\mathscr{P}|$. Since sequencing SPs are hard to solve, we avoid solving them unless they are all feasible, determined by solving the model with a constant (zero) objective. All SPs are sequentially screened for feasibility and all the obtained feasibility cuts are developed and incorporated into the MP. Using a feasibility check allows for many more iterations in a shorter period of time as the feasibility problem is solved faster than the full SP; however, each individual cut may be weaker using this approach compared to directly solving SPs to optimality.

2.4. Benders' Cuts

The key component of LBBD is Benders' cut development. We develop novel multistrategy logic-based Benders' feasibility and optimality cuts for the CORPS problem. To the best of our knowledge, no Benders' cuts in the literature communicate more than two remedial strategies to the MP simultaneously, but our Benders' cuts communicate three or more remedial strategies to the MP simultaneously. Benders' cuts are inequality constraints that are added to the MP each time SPs are infeasible or suboptimal ($\hat{v}^{(i)}_{hdr} \neq \bar{v}^{(i)}_{hdr}$). Two types of Benders' cuts can be generated: feasibility and optimality. Benders' feasibility cuts are developed from those hospital-day SPs whose constraints have been violated given the MP solution, and Benders' optimality cuts are developed when $\hat{v}^{(i)}_{hdr} \neq \bar{v}^{(i)}_{hdr}$. The set of infeasible SPs is denoted $\bar{\mathscr{U}}^{(i)}$ and the set of SPs whose $\hat{v}^{(i)}_{hdr} \neq \bar{v}^{(i)}_{hdr}$ is denoted $\bar{\mathscr{J}}^{(i)}$. The set of eligible unused ORs in hospital h on day d is denoted by $\mathscr{R}'_{hd} = (\bigcup_{p\in\hat{\mathscr{P}}^{(i)}_{hd}} \mathscr{Q}_{ph}) \cap (\mathscr{R}_h \backslash \hat{\mathscr{R}}^{(i)}_{hd})$ and the set of eligible unused surgeons is denoted by $\mathscr{S}'_{hd} = (\bigcup_{p\in\hat{\mathscr{P}}^{(i)}_{hd}} \Omega_p) \cap (\mathscr{S} \backslash \hat{\mathscr{S}}^{(i)}_{hd})$.

If the SP is not feasibly sequenceable given $\hat{\mathcal{P}}_{hd}^{(i)}$, $\hat{\mathcal{R}}_{hd}^{(i)}$, and $\hat{\mathcal{S}}_{hd}^{(i)}$, the following Benders' feasibility cut is generated to remove the current infeasible solution:

$$|\hat{\mathcal{P}}_{hd}^{(i)}| - \sum_{p\in\hat{\mathcal{P}}_{hd}^{(i)}} \sum_{s\in\hat{\mathcal{S}}_{hd}^{(i)}} \sum_{r\in\hat{\mathcal{R}}_{hd}^{(i)}} x_{pshdr} + \sum_{r\in\mathcal{R}'_{hd}} y_{hdr} + \sum_{s\in\mathcal{S}'_{hd}} z_{shd} \geq 1 \quad \forall (h,d) \in \bar{\mathcal{U}}_d^{(i)}.$$

This cut is a valid cut (see the proof of Theorem 1), defined as a logical expression that has two properties (Chu and Xia 2004): (1) the cut must eliminate the current infeasible MP solution, and (2) the cut must not remove any globally feasible solution. Each Benders' feasibility cut communicates the following three feasibility remedial strategies to the MP: (1) open at least one more OR from $\mathcal{R}'_{hd}$; (2) use at least one more surgeon from $\mathcal{S}'_{hd}$; and/or (3) subtract at least one patient from $\hat{\mathcal{P}}_{hd}^{(i)}$. At least one of these strategies must be enforced by the MP to break the SP infeasibility.

Theorem 1. *The Benders' feasibility cut is valid.*

Proof. Let $\hat{\mathcal{P}}_{hd}^{(i)}$, $\hat{\mathcal{R}}_{hd}^{(i)}$, and $\hat{\mathcal{S}}_{hd}^{(i)}$ be the set of assigned patients, ORs, and surgeons to hospital h on day d in MP solution $(\hat{x}_{pshdr}^{(i)}, \hat{y}_{hdr}^{(i)}, \hat{z}_{shd}^{(i)})$ that is infeasible to the SP. By definition, these binary variables take value 1 in the current iteration i of the MP. Denoting $\mathcal{R}'_{hd}$ and $\mathcal{S}'_{hd}$ as the set of unused eligible ORs and surgeons, respectively, our Benders' feasibility cut is

$$|\hat{\mathcal{P}}_{hd}^{(i)}| - \sum_{p\in\hat{\mathcal{P}}_{hd}^{(i)}} \sum_{s\in\hat{\mathcal{S}}_{hd}^{(i)}} \sum_{r\in\hat{\mathcal{R}}_{hd}^{(i)}} x_{pshdr} + \sum_{r\in\mathcal{R}'_{hd}} y_{hdr} + \sum_{s\in\mathcal{S}'_{hd}} z_{shd} \geq 1 \quad \forall (h,d) \in \bar{\mathcal{U}}_d^{(i)}.$$

We first show that the cut rules out the current infeasible MP solution (Property 1). Instantiating this cut with the current infeasible MP solution, it is clear that the infeasible MP solution is cut off, because the combination of patients, the selection of ORs, and the selection of surgeons has not changed:

$$\overbrace{|\hat{\mathcal{P}}_{hd}^{(i)}| - \sum_{p\in\hat{\mathcal{P}}_{hd}^{(i)}} \sum_{s\in\hat{\mathcal{S}}_{hd}^{(i)}} \sum_{r\in\hat{\mathcal{R}}_{hd}^{(i)}} \hat{x}_{pshdr}^{(i)}}^{0} + \overbrace{\sum_{r\in\mathcal{R}'_{hd}} \hat{y}_{hdr}^{(i)}}^{0} + \overbrace{\sum_{s\in\mathcal{S}'_{hd}} \hat{z}_{shd}^{(i)}}^{0} = 0 < 1 \quad \forall (h,d) \in \bar{\mathcal{U}}_d^{(i)}.$$

Thus, Property 1 is satisfied.

We next show that our Benders' feasibility cut does not remove any globally feasible solutions in future iterations (Property 2). Consider a hypothetical future solution $(\tilde{x}_{pshdr}, \tilde{y}_{hdr}, \tilde{z}_{shd})$ consisting of a new set of patients $(\tilde{\mathcal{P}}_{hd})$, ORs $(\tilde{\mathcal{R}}_{hd})$, and surgeons $(\tilde{\mathcal{S}}_{hd})$, leading to a feasible or infeasible future solution.

We first assume that the future patient allocation variable $(\tilde{x}_{pshdr})$ selects a new combination of patients $(\tilde{\mathcal{P}}_{hd})$ and other variables maintain their previous values, that is, $\tilde{y}_{hdr} = \hat{y}_{hdr}^{(i)}$ and $\tilde{z}_{shd} = \hat{z}_{shd}^{(i)}$. The possible relationships between $\hat{\mathcal{P}}_{hd}^{(i)}$ and $\tilde{\mathcal{P}}_{hd}$ are shown in Figure 3. Given the same selection of ORs and surgeons, a future solution is infeasible if the future solution increases the current surgical load (Figure 3(a)); other scenarios are potentially feasible. By keeping the same selection of ORs and surgeons, the cut is reduced to

$$|\hat{\mathcal{P}}_{hd}^{(i)}| - \sum_{p\in\hat{\mathcal{P}}_{hd}^{(i)}} \sum_{s\in\hat{\mathcal{S}}_{hd}^{(i)}} \sum_{r\in\hat{\mathcal{R}}_{hd}^{(i)}} \tilde{x}_{pshdr} \geq 1 \quad \forall (h,d) \in \bar{\mathcal{U}}_d^{(i)},$$

which further decomposes to

$$\overbrace{|\hat{\mathcal{P}}_{hd}^{(i)} \cap \tilde{\mathcal{P}}_{hd}| - \sum_{p\in\hat{\mathcal{P}}_{hd}^{(i)}\cap\tilde{\mathcal{P}}_{hd}} \sum_{s\in\hat{\mathcal{S}}_{hd}^{(i)}} \sum_{r\in\hat{\mathcal{R}}_{hd}^{(i)}} \tilde{x}_{pshdr}}^{a}$$

$$+ \overbrace{|\hat{\mathcal{P}}_{hd}^{(i)} \setminus \tilde{\mathcal{P}}_{hd}| - \sum_{p\in\hat{\mathcal{P}}_{hd}^{(i)}\setminus\tilde{\mathcal{P}}_{hd}} \sum_{s\in\hat{\mathcal{S}}_{hd}^{(i)}} \sum_{r\in\hat{\mathcal{R}}_{hd}^{(i)}} \tilde{x}_{pshdr}}^{b} \geq 1 \quad \forall (h,d) \in \bar{\mathcal{U}}_d^{(i)}.$$

Note that a is always zero, and thus we only need to show that $b < 1$ for infeasible scenarios (Figure 3(a)), and $b \geq 1$ for feasible scenarios (Figures 3(b)–3(d)):

- $\tilde{\mathcal{P}}_{hd}$ adds some new patients to $\hat{\mathcal{P}}_{hd}^{(i)}$ (Figure 3(a)), yielding $b < 1$ because all patients in $\hat{\mathcal{P}}_{hd}^{(i)}$ are in $\tilde{\mathcal{P}}_{hd}$.
- $\tilde{\mathcal{P}}_{hd}$ subtracts at least one patient from $\hat{\mathcal{P}}_{hd}^{(i)}$ (Figure 3(b)), yielding $b \geq 1$ because there is at least one patient in $\hat{\mathcal{P}}_{hd}^{(i)}$ that does not exist in $\tilde{\mathcal{P}}_{hd}$.
- $\tilde{\mathcal{P}}_{hd}$ shares some patients with $\hat{\mathcal{P}}_{hd}^{(i)}$ (Figure 3(c)), yielding $b \geq 1$ by the same argument used for Figure 3(b).
- $\tilde{\mathcal{P}}_{hd}$ shares no patient with $\hat{\mathcal{P}}_{hd}^{(i)}$ (Figure 3(d)), yielding $b \geq 1$ by the same argument used for Figure 3(b).

Next, we similarly examine changing the combination of ORs in a future solution $(\tilde{y}_{hdr})$ given fixed patients and surgeons. Given the same selection of patients and surgeons, a future solution is infeasible if it decreases one or more ORs from the existing set of ORs (Figure 4(b)); other scenarios are potentially feasible. In this scenario, the Benders' feasibility cut is reduced to

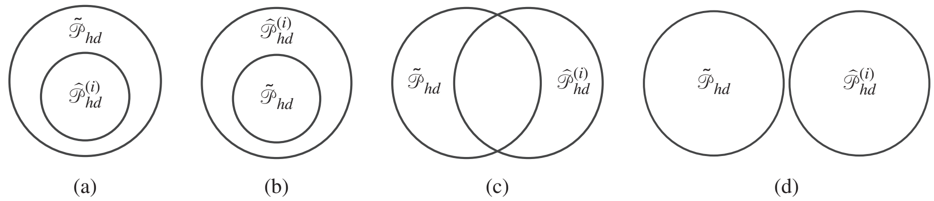
$$\sum_{r\in\mathcal{R}'_{hd}} \tilde{y}_{hdr} \geq 1 \quad \forall (h,d) \in \bar{\mathcal{U}}_d^{(i)},$$

which further decomposes to

$$\overbrace{\sum_{r\in\tilde{\mathcal{R}}_{hd}\cap\hat{\mathcal{R}}_{hd}^{(i)}} (1 - \tilde{y}_{hdr})}^{c} + \overbrace{\sum_{r\in\tilde{\mathcal{R}}_{hd}\setminus\hat{\mathcal{R}}_{hd}^{(i)}} \tilde{y}_{hdr}}^{d} \geq 1 \quad \forall (h,d) \in \bar{\mathcal{U}}_d^{(i)}.$$

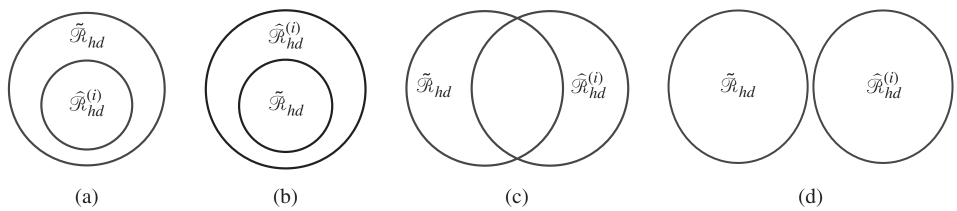
Note that c is always zero, and thus we only need to show that $d < 1$ for infeasible scenarios (Figure 4(b)),

Figure 3. Possible Relationships Between the Current MP Solution ($\hat{\mathscr{P}}^{(i)}_{hd}$) and a Hypothetical Future Solution ($\tilde{\mathscr{P}}_{hd}$) with Only a Change in Patients



Note. Only adding load in terms of more patients is prohibited.

Figure 4. Possible Relationships Between the Current MP Solution ($\hat{\mathscr{R}}^{(i)}_{hd}$) and a Hypothetical Future Solution ($\tilde{\mathscr{R}}_{hd}$) with Only a Change in ORs



Note. Only reducing resources in terms of fewer ORs is prohibited.

and $d \geq 1$ for feasible scenarios (Figures 4(a), 4(c), and 4(d)):

- $\tilde{\mathscr{R}}_{hd}$ adds some eligible ORs to $\hat{\mathscr{R}}^{(i)}_{hd}$ (Figure 4(a)), yielding $d \geq 1$ because there is at least one new OR in $\tilde{\mathscr{R}}_{hd}$ that does not exist in $\hat{\mathscr{R}}^{(i)}_{hd}$.
- $\tilde{\mathscr{R}}_{hd}$ subtracts at least one OR from $\hat{\mathscr{R}}^{(i)}_{hd}$ (Figure 4(b)), yielding $d = 0 < 1$ because all ORs in $\hat{\mathscr{R}}^{(i)}_{hd}$ are in $\tilde{\mathscr{R}}_{hd}$.
- $\tilde{\mathscr{R}}_{hd}$ shares some ORs with $\hat{\mathscr{R}}^{(i)}_{hd}$ (Figure 4(c)), yielding $d \geq 1$ by the same argument used for Figure 4(a).
- $\tilde{\mathscr{R}}_{hd}$ shares no ORs with $\hat{\mathscr{R}}^{(i)}_{hd}$ (Figure 4(d)), yielding $d \geq 1$ by the same argument used for Figure 4(a).

A procedure identical to the one for ORs is repeated for surgeons. Finally, we focus on future infeasible solutions stemming from combinatorial cases. The following combination of strategies leads to infeasible future solutions, which are removed by our Benders' feasibility cut:

1. $\hat{\mathscr{P}}^{(i)}_{hd} \subseteq \tilde{\mathscr{P}}_{hd}$, while $\tilde{\mathscr{R}}_{hd} = \hat{\mathscr{R}}^{(i)}_{hd}$, and $\tilde{\mathscr{S}}_{hd} = \hat{\mathscr{S}}^{(i)}_{hd}$
2. $\hat{\mathscr{P}}^{(i)}_{hd} \subseteq \tilde{\mathscr{P}}_{hd}$, while $\tilde{\mathscr{R}}_{hd} \subset \hat{\mathscr{R}}^{(i)}_{hd}$, and $\tilde{\mathscr{S}}_{hd} = \hat{\mathscr{S}}^{(i)}_{hd}$
3. $\hat{\mathscr{P}}^{(i)}_{hd} \subseteq \tilde{\mathscr{P}}_{hd}$, while $\tilde{\mathscr{R}}_{hd} = \hat{\mathscr{R}}^{(i)}_{hd}$, and $\tilde{\mathscr{S}}_{hd} \subset \hat{\mathscr{S}}^{(i)}_{hd}$
4. $\hat{\mathscr{P}}^{(i)}_{hd} \subseteq \tilde{\mathscr{P}}_{hd}$, while $\tilde{\mathscr{R}}_{hd} \subset \hat{\mathscr{R}}^{(i)}_{hd}$, and $\tilde{\mathscr{S}}_{hd} \subset \hat{\mathscr{S}}^{(i)}_{hd}$.

All these relationships are captured in the following expanded Benders' feasibility cut:

$$\overbrace{|\hat{\mathscr{P}}^{(i)}_{hd} \cap \tilde{\mathscr{P}}_{hd}| - \sum_{p \in \hat{\mathscr{P}}^{(i)}_{hd} \cap \tilde{\mathscr{P}}_{hd}} \sum_{s \in \hat{\mathscr{S}}^{(i)}_{hd}} \sum_{r \in \hat{\mathscr{R}}^{(i)}_{hd}} \tilde{x}_{pshdr}}^{0}$$
$$+ \overbrace{|\hat{\mathscr{P}}^{(i)}_{hd} \backslash \tilde{\mathscr{P}}_{hd}| - \sum_{p \in \hat{\mathscr{P}}^{(i)}_{hd} \backslash \tilde{\mathscr{P}}_{hd}} \sum_{s \in \hat{\mathscr{S}}^{(i)}_{hd}} \sum_{r \in \hat{\mathscr{R}}^{(i)}_{hd}} \tilde{x}_{pshdr}}^{0}$$
$$+ \overbrace{\sum_{r \in \tilde{\mathscr{R}}_{hd} \cap \hat{\mathscr{R}}^{(i)}_{hd}} (1 - \tilde{y}_{hdr})}^{0} + \overbrace{\sum_{r \in \tilde{\mathscr{R}}_{hd} \backslash \hat{\mathscr{R}}^{(i)}_{hd}} \tilde{y}_{hdr}}^{0} + \overbrace{\sum_{r \in \tilde{\mathscr{S}}_{hd} \cap \hat{\mathscr{S}}^{(i)}_{hd}} (1 - \tilde{z}_{shd})}^{0}$$
$$+ \overbrace{\sum_{r \in \tilde{\mathscr{S}}_{hd} \backslash \hat{\mathscr{S}}^{(i)}_{hd}} \tilde{z}_{shd}}^{0} = 0 < 1 \quad \forall (h, d) \in \bar{\mathscr{U}}^{(i)}_{d}. \tag{22}$$

Note that if the directionality of any of the subsets in scenarios (1)–(4) is reversed, then the corresponding term in Equation (22) will assume a value ≥ 1, and thus no feasible solutions are removed. □

The MP overtime is precisely calibrated by SPs in the presence of sequencing constraints. Each OR optimal overtime amount found by the SP ($\bar{v}^{(i)}_{hdr}$) is either equal to or greater than the optimal amount of overtime computed by the MP ($\hat{v}^{(i)}_{hdr}$) for the set of ORs whose SP overtime is not equal to the MP overtime ($\hat{v}^{(i)}_{hdr} \neq \bar{v}^{(i)}_{hdr}$), denoted by $\bar{\mathscr{F}}^{(i)}$. From the SPs, the following Benders' optimality cut is added to the MP:

$$\sum_{r \in \mathscr{R}_h} v_{hdr} \geq \sum_{r \in \hat{\mathscr{R}}^{(i)}_{hd}} \bar{v}^{(i)}_{hdr} \Bigg(1 - \Bigg(|\hat{\mathscr{P}}^{(i)}_{hd}| - \sum_{p \in \hat{\mathscr{P}}^{(i)}_{hd}} \sum_{s \in \hat{\mathscr{S}}^{(i)}_{hd}} \sum_{r_1 \in \hat{\mathscr{R}}^{(i)}_{hd}} x_{pshdr_1}$$
$$+ \sum_{r_1 \in \mathscr{R}'_{hd}} y_{hdr_1} + \sum_{s \in \mathscr{S}'_{hd}} z_{shd}\Bigg)\Bigg) \quad \forall (h, d) \in \bar{\mathscr{F}}^{(i)}_{d}.$$

Because of the optimal status of the SPs, this cut is called optimality cut. This cut is a valid Benders' optimality cut (see the proof of Theorem 2). The Benders' optimality cut removes the current MP solution by enforcing at least one of the following four strategies: (1) subtract at least one patient from the set of $\hat{\mathscr{P}}^{(i)}_{hd}$;

(2) open at least one more eligible OR; (3) use at least one more eligible surgeon; and/or (4) increase the amount of overtime (v_{hdr}) if none of the previous strategies is adopted by the MP. In the case where the SP is optimal, this cut reduces to

$$\sum_{r\in\mathcal{R}_h} v_{hdr} \geq \sum_{r\in\bar{\mathcal{R}}^{(i)}_{hd}} \bar{v}^{(i)}_{hdr} \quad \forall (h,d) \in \bar{\mathcal{J}}^{(i)}_d$$

and $\sum_{r\in\bar{\mathcal{R}}^{(i)}_{hd}} \bar{v}^{(i)}_{hdr}$ becomes a new lower bound for the MP previous allocation variables.

Theorem 2. *The Benders' optimality cut is valid.*

Proof. Let $\hat{\mathcal{P}}^{(i)}_{hd}$, $\hat{\mathcal{R}}^{(i)}_{hd}$, $\hat{\mathcal{S}}^{(i)}_{hd}$, and $\hat{v}^{(i)}_{hdr}$ be the set of assigned patients, ORs, surgeons, and overtime amount to hospital h on day d, respectively, from suboptimal MP solution $(\hat{x}^{(i)}_{pshdr}, \hat{y}^{(i)}_{hdr}, \hat{z}^{(i)}_{shd}, \hat{v}^{(i)}_{hdr})$. The following Benders' optimality cut is developed from those SPs whose $\hat{v}^{(i)}_{hdr} \neq \bar{v}^{(i)}_{hdr}$:

$$\sum_{r\in\mathcal{R}_h} v_{hdr} - \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}} \bar{v}^{(i)}_{hdr}\left(1-\left(|\hat{\mathcal{P}}^{(i)}_{hd}| - \sum_{p\in\hat{\mathcal{P}}^{(i)}_{hd}}\sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}}\sum_{r_1\in\hat{\mathcal{R}}^{(i)}_{hd}} x_{pshdr_1} + \sum_{r_1\in\mathcal{R}'_{hd}} y_{hdr_1} + \sum_{s\in\mathcal{S}'_{hd}} z_{shd}\right)\right) \geq 0 \quad \forall (h,d) \in \bar{\mathcal{J}}^{(i)}_d.$$

We first show that the cut removes the current suboptimal MP solution (Property 1). Instantiating this cut with the current suboptimal MP solution, it is clear that the suboptimal MP solution is ruled out, because the combination of patients, ORs, and surgeons has not changed:

$$\overbrace{\sum_{r\in\mathcal{R}_h} \hat{v}^{(i)}_{hdr} - \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}} \bar{v}^{(i)}_{hdr}}^{<0}\left(1-\left(\overbrace{|\hat{\mathcal{P}}^{(i)}_{hd}| - \sum_{p\in\hat{\mathcal{P}}^{(i)}_{hd}}\sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}}\sum_{r_1\in\hat{\mathcal{R}}^{(i)}_{hd}} \hat{x}^{(i)}_{pshdr_1}}^{0} + \overbrace{\sum_{r_1\in\mathcal{R}'_{hd}} \hat{y}^{(i)}_{hdr_1}}^{0} + \overbrace{\sum_{s\in\mathcal{S}'_{hd}} \hat{z}^{(i)}_{shd}}^{0}\right)\right) < 0 \quad \forall (h,d) \in \bar{\mathcal{J}}^{(i)}_d.$$

Thus, the Property 1 is satisfied,

We next show that no globally integer feasible solution is ruled out by our Benders' feasibility cut (Property 2). Similar to the proof of our Benders' feasibility cut, we fix all resources except overtime, and investigate whether or not our Benders' optimality cut removes a hypothetical future MP solution ($\tilde{v}_{hdr}$). By fixing other resources, our Benders' optimality cut is reduced to

$$\sum_{r\in\mathcal{R}_h} v_{hdr} - \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}} \bar{v}^{(i)}_{hdr} \geq 0 \quad \forall (h,d) \in \bar{\mathcal{J}}^{(i)}_d. \tag{23}$$

Consider hypothetical MP overtime values ($\tilde{v}_{hdr}$). The relational possibilities between $\tilde{v}_{hdr}$ and $\bar{v}^{(i)}_{hdr}$ can be of one of the following forms: (1) $\tilde{v}_{hdr} < \bar{v}^{(i)}_{hdr}$, in which Inequality (23) cuts off $\tilde{v}_{hdr}$; and (2) $\tilde{v}_{hdr} \geq \bar{v}^{(i)}_{hdr}$ and $\tilde{v}_{hdr}$ satisfies Inequality (23). The following combinatorial strategies lead to infeasible future solutions, which are removed by our Benders' optimality cut:

1. $\tilde{v}_{hdr} < \bar{v}^{(i)}_{hdr}$, $\hat{\mathcal{P}}^{(i)}_{hd} \subseteq \tilde{\mathcal{P}}_{hd}$, $\tilde{\mathcal{R}}_{hd} = \hat{\mathcal{R}}^{(i)}_{hd}$, and $\tilde{\mathcal{S}}_{hd} = \hat{\mathcal{S}}^{(i)}_{hd}$,
2. $\tilde{v}_{hdr} < \bar{v}^{(i)}_{hdr}$, $\hat{\mathcal{P}}^{(i)}_{hd} \subseteq \tilde{\mathcal{P}}_{hd}$, $\tilde{\mathcal{R}}_{hd} \subset \hat{\mathcal{R}}^{(i)}_{hd}$, and $\tilde{\mathcal{S}}_{hd} = \hat{\mathcal{S}}^{(i)}_{hd}$,
3. $\tilde{v}_{hdr} < \bar{v}^{(i)}_{hdr}$, $\hat{\mathcal{P}}^{(i)}_{hd} \subseteq \tilde{\mathcal{P}}_{hd}$, $\tilde{\mathcal{R}}_{hd} = \hat{\mathcal{R}}^{(i)}_{hd}$, and $\tilde{\mathcal{S}}_{hd} \subset \hat{\mathcal{S}}^{(i)}_{hd}$,
4. $\tilde{v}_{hdr} < \bar{v}^{(i)}_{hdr}$, $\hat{\mathcal{P}}^{(i)}_{hd} \subseteq \tilde{\mathcal{P}}_{hd}$, $\tilde{\mathcal{R}}_{hd} \subset \hat{\mathcal{R}}^{(i)}_{hd}$, and $\tilde{\mathcal{S}}_{hd} \subset \hat{\mathcal{S}}^{(i)}_{hd}$.

All these relationships are captured in the following expanded Benders' optimality cut, resulting in

$$\sum_{r\in\mathcal{R}_h} \tilde{v}_{hdr} - \sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}} \bar{v}^{(i)}_{hdr}\left(1-\left(\overbrace{|\hat{\mathcal{P}}^{(i)}_{hd}\cap\tilde{\mathcal{P}}_{hd}| - \sum_{p\in\hat{\mathcal{P}}^{(i)}_{hd}\cap\tilde{\mathcal{P}}_{hd}}\sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}}\sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}} \tilde{x}_{pshdr}}^{0} + \overbrace{|\hat{\mathcal{P}}^{(i)}_{hd}\setminus\tilde{\mathcal{P}}_{hd}| - \sum_{p\in\hat{\mathcal{P}}^{(i)}_{hd}\setminus\tilde{\mathcal{P}}_{hd}}\sum_{s\in\hat{\mathcal{S}}^{(i)}_{hd}}\sum_{r\in\hat{\mathcal{R}}^{(i)}_{hd}} \tilde{x}_{pshdr}}^{0} + \overbrace{\sum_{r\in\tilde{\mathcal{R}}_{hd}\cap\hat{\mathcal{R}}^{(i)}_{hd}} (1-\tilde{y}_{hdr})}^{0} + \overbrace{\sum_{r\in\tilde{\mathcal{R}}_{hd}\setminus\hat{\mathcal{R}}^{(i)}_{hd}} \tilde{y}_{hdr}}^{0} + \overbrace{\sum_{r\in\tilde{\mathcal{S}}_{hd}\cap\hat{\mathcal{S}}^{(i)}_{hd}} (1-\tilde{z}_{shd})}^{0} + \overbrace{\sum_{r\in\tilde{\mathcal{S}}_{hd}\setminus\hat{\mathcal{S}}^{(i)}_{hd}} \tilde{z}_{shd}}^{0}\right)\right) < 0 \quad \forall (h,d) \in \bar{\mathcal{J}}^{(i)}_d. \tag{24}$$

Note that if the directionality of any of the subsets in scenarios (1)–(4) is reversed, then the corresponding term in Inequality (24) will assume a value ≥ 1, and thus no feasible solution is removed. □

Because Properties 1 and 2 are satisfied in both Benders' feasibility and optimality cuts and $\mathcal{R}_h$, $\mathcal{S}$, and v_{hdr} have finite domains, the proposed LBBD results in a finite convergence to optimality.

3. Data

We obtained data including the pre-incision (preparation), incision (surgical time), and post-incision (cleaning) times of 7,500 emergency and elective patients, operated on during June 2011–June 2013 in UHN General Surgery Departments. After removing the data related to emergency patients, 5,573 elective patients remain, consisting of 1,750, 1,914, and 1,909 patients for TGH, TWH, and PMCC, respectively. Our data generation framework is based on these remaining elective patients. The complete data set is available as an online supplement.

3.1. Choice of Statistical Probability Distribution for Surgical Times

We fitted 25 parametric (including Pearson, log-normal, log-logistic, and inverse Gaussian) and non-parametric (kernel density estimation [KDE]; Sheater and Jones 1991, Wand and Jones 1995, Chan et al. 2016) distributions to the pre-incision and incision data

Table 3. KS Values for Best Fits of Distributions to UHN Pre-Incision and Incision Times

	Distribution	KS statistic	$a = 0.05$	$a = 0.01$
Pre-incision time	KDE	0.0551	0.0182	0.0218
	Loglogistic	0.0781	0.0182	0.0218
	Lognormal	0.1029	0.0182	0.0218
	Inv Gaussian	0.1160	0.0182	0.0218
	Gamma	0.1409	0.0182	0.0218
	Weibull	0.1591	0.0182	0.0218
Incision time	KDE	0.0424	0.0182	0.0218
	Pearson 5	0.0535	0.0182	0.0218
	Pearson 6	0.0573	0.0182	0.0218
	Inv Gaussian	0.0704	0.0182	0.0218
	Loglogistic	0.0728	0.0182	0.0218
	Lognormal	0.0760	0.0182	0.0218

Notes. Lower KS values indicate better fit. Critical thresholds are denoted a.

and found the best statistical distribution for each hospital as well as for the overall UHN data based on Kolmogorov-Smirnov (KS) tests (Table 3). Figure 5 illustrates the best parametric distribution and KDE fitted to each hospital's incision times.

KDE is the best fit for both pre-incision (unbounded support, bandwidths 1.4275) and incision (unbounded support, bandwidths 12.6802) data as it is able to handle irregular peaks and valleys in the data. However, in the data, there exists a correlation of 0.74 between the pre-incision and incision times, which must be preserved in the generated data. To generate correlated data, we first generate a two-dimensional Copula with a correlation of 0.74 between the two variables. Then, we calculate KDE-distributed cumulative density functions (CDFs) for both the pre-incision and incision time variables using the support and bandwidth parameters derived from the raw data. Finally, we transform the correlated uniformly distributed variables in the Copula into correlated KDE-distributed variables using the KDE-distributed CDF functions via the inverse-transform method (Nelsen 2006). The surgical time for each patient can vary by up to 20% depending on the surgeon, so we generate specific patient-surgeon surgical times using a normal distribution with the average equal to the KDE-generated time and the variance equal to 20% of that time. The correlated random

Figure 5. (Color online) Distributions Fitted to Incision Times

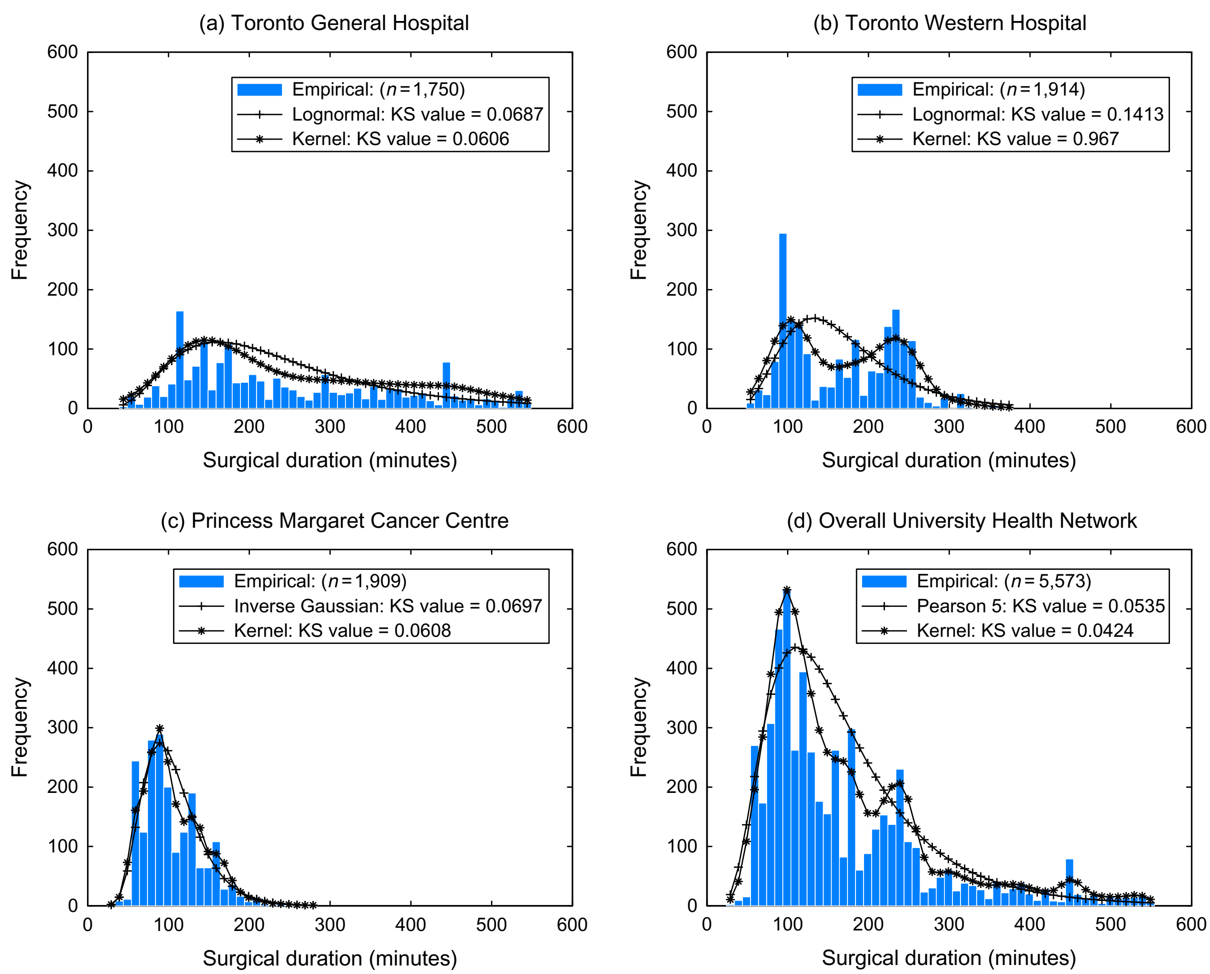
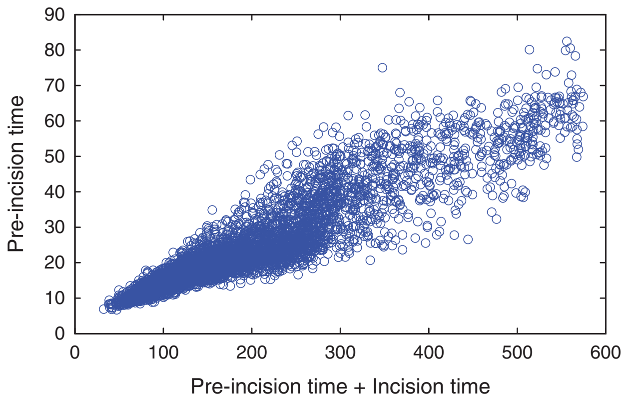


Figure 6. (Color online) Simulated Correlated Data for Pre-Incision and Incision Time Based on Copula Formula ($n = 5{,}573$)



numbers between pre-incision and incision time are shown in Figure 6.

There also exists a correlation between incision time and post-incision time in different hospitals with an average of 7, 10, and 15 minutes in PMCC, TWH, and TGH, respectively. Despite this correlation, the patient-out-to-patient-in time (turnaround time) in UHN is 25 minutes. The cleaning time for each patient is therefore considered constant at 25 minutes. If one is interested in generating correlated random data between three different data vectors (pre-incision, incision, and post-incision times), our KDE approach is also capable of handling this situation.

3.2. Cost Parameters

Cost parameters were provided by the Case Costing Department of UHN. From the data, per hour OR and surgeon fixed costs follow a uniform distribution with ranges $[400, 500]$ and $[500, 600]$, respectively. These surgeon costs are in contrast to literature (Batun et al. 2011) that considers highly variable ranges for a surgeon's per hour availability cost ([\$1065, \$5324]). The cost per hour for OR overtime is calculated as OR per hour fixed costs ×1.5. The average surgical cost of each patient without the cost of surgeons is 955CAD, and the average surgeon cost per surgery is 1650CAD. To approximate the benefit of optional patients (U_p) if operated on in the current planning horizon and to also be fair in giving priority to patients based on due dates, the following formula is used: $U_p = (2|\mathcal{D}|/\theta_p) \times (955 + 1{,}650)$, where $|\mathcal{D}|$ is the last day of the current planning horizon. This formula encourages earlier scheduling of optional patients whose due dates are sooner.

3.3. Initial Single- and Multihospital Feasible Schedules

To ensure that there is at least one existing feasible solution, we create a dummy surgical schedule by randomly assigning each mandatory patient to a specific day, OR, and surgeon, based on which an initial patient-to-surgeon assignment is extracted. We also create a baseline surgeon availability schedule by assigning a surgeon to work when he or she is performing a surgery in the dummy schedule. Because surgeons work full shifts, daily surgeon availability times are rounded up to four, eight, or 10 hours, and more availability is added in increments of four hours until a minimum of eight hours per surgeon per week is reached. The patient's surgery day in the dummy schedule is the patient's due date. We also randomly generate a single surgeon and OR assignment for each optional patient, but we do not account for these patients in surgeon availability time as they do not need to be scheduled in this planning horizon. We call this dummy schedule the dedicated single-hospital baseline schedule (DSHBS).

We aggregate the DSHBS of each hospital to create a dedicated multihospital baseline schedule, preserving the feasibility of each DSHBS as we just add extra patient-to-OR allocation eligibilities, due to additional OR availabilities in other hospitals. First, we hypothetically assume a full patient-to-surgeon allocation flexibility ($|\mathcal{P}| \times |\mathcal{S}|$) and generate surgical times for each patient-to-surgeon combination not already generated in the DSHBS. Since there is no patient-to-surgeon allocation flexibility in the DSHBS, we randomly add additional surgeon eligibilities until an average of three surgeons per patient is reached. Then, this flexible multihospital schedule is translated into a semiflexible multihospital schedule by setting $|\Omega_p| = 1$ for the mandatory patients ($\theta_p \le |\mathcal{D}|$). Note that this data can easily be disaggregated into single-, bi-, and tri-hospital data sets.

4. Results

The model was implemented in MATLAB 2015b (The Mathworks, Inc.) on a 4 Dual-Core AMD Opteron™ Processor 2.2 GHz in a CentOS 2.6 platform with 40 GB RAM. Gurobi Optimizer 6.00 (Gurobi Optimization, Inc.) is used to solve mixed-integer programs. We test values of α in $[0, 1]$ in increments of 0.2. For all hospitals collaborating together, trials are solved given 10 hours CPU time. All bi-hospital trials are solved to a 5% optimality gap and trials for individual hospitals are solved to a 0% optimality gap. For relaxed optimality gaps (5%), we solve the MP to that relaxed gap, and iterate LBBD until it converges, similar to Tran and Beck (2012).

The problem instances consist of 35–135 patients varying in size (small data sets are 35, 40, and 45 patients; medium are 70, 80, and 90; and large are 105, 120, and 135) and schedule density (sparse has an average of 3.5 surgeries/week/surgeon, normal has 4.0, and dense has 4.5). Each problem instance is generated five times. We consider two planning horizons, five days

Table 4. Average Computation Time (Seconds) With and Without the Feasibility Check Mechanism Over All Small Instances

	LBBD without feasibility check			LBBD with feasibility check		
$\lvert\mathscr{P}\rvert$	Total	[min, max]	[MP, SP]	Total	[min, max]	[MP, SP]
35	50.96	[0.17, **72.40**]	[50.34, 0.62]	**29.97**	[**0.14**, 120.49]	[**29.81**, **0.16**]
40	39.75	[**0.35**, 56.48]	[38.90, 0.85]	**30.17**	[0.56, **52.16**]	[**29.76**, **0.41**]
45	93.06	[68.37, **116.29**]	[62.96, 30.10]	**76.00**	[1.34, 200.22]	[**50.12**, **25.88**]

Note. Bold represents the best performance in each category.

each, and three equal-sized hospitals with three ORs each.

4.1. Value of Feasibility Check Before SP Optimization

To quantify the value of the feasibility check mechanism on the LBBD convergence, we apply LBBD with and without the feasibility check to small data sets with sparse, normal, and dense schedules. Table 4 shows that LBBD with the feasibility check consistently outperforms the LBBD without feasibility check, on average, though LBBD without the feasibility check may have better minimum and maximum times for some scenarios.

4.2. Value of Surgeon Schedule Tightness

Regardless of α value, most of the total CPU time is spent solving the MP (Figure 7). As expected, larger α values result in improved objective function values (Figure 8), but at the expense of worsened computation time (Figure 9) and feasibility (Figure 10) as the sequencing SPs become more difficult to solve, requiring more LBBD iterations. Schedule density predictably increases computation time (Figure 9) (though less than would be expected), but decreases costs (Figure 8). The large-sized problems are difficult to solve starting with $\alpha = 0.6$, though schedule density also impacts feasibility for medium-sized problems with starting with $\alpha = 0.8$. Increased number of iterations and Benders' cuts (Figure 10) indicate that the MP solution cannot be easily sequenced in the SPs and LBBD requires multiple iterations to find an MP solution that is feasibly sequenceable in the SPs. Thus, $\alpha = 0.6$ is chosen for all other experiments because it is a good trade-off among objective function value, CPU time, and feasibility.

4.3. Value of Collaboration, Surgeon Schedule Tightness, and Patient Flexibility

To quantify the value of collaboration, surgeon schedule tightness, and patient flexibility, we examine nine instances of single-hospital schedules, generated from

Figure 7. Total Average CPU Time (Minutes) Percentage Breakdown Between the MP and SPs for Different Values of α for Solved Trials with Dedicated Patient-to-Surgeon Allocation

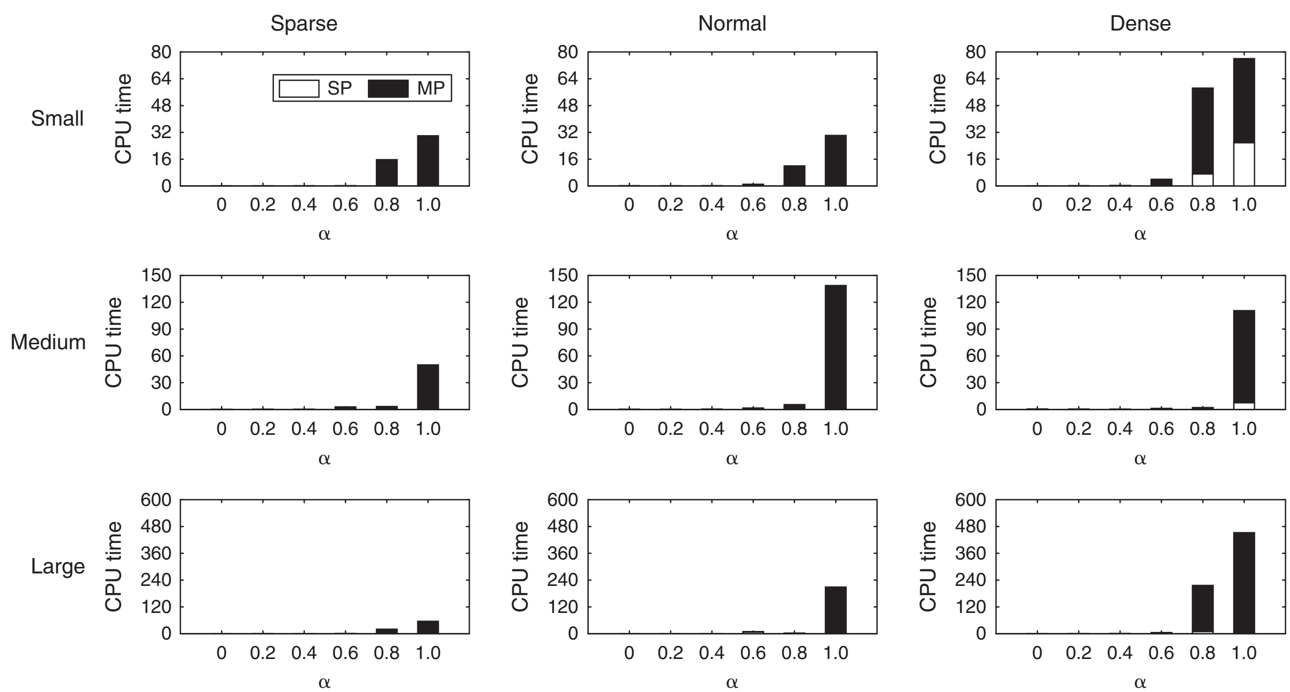
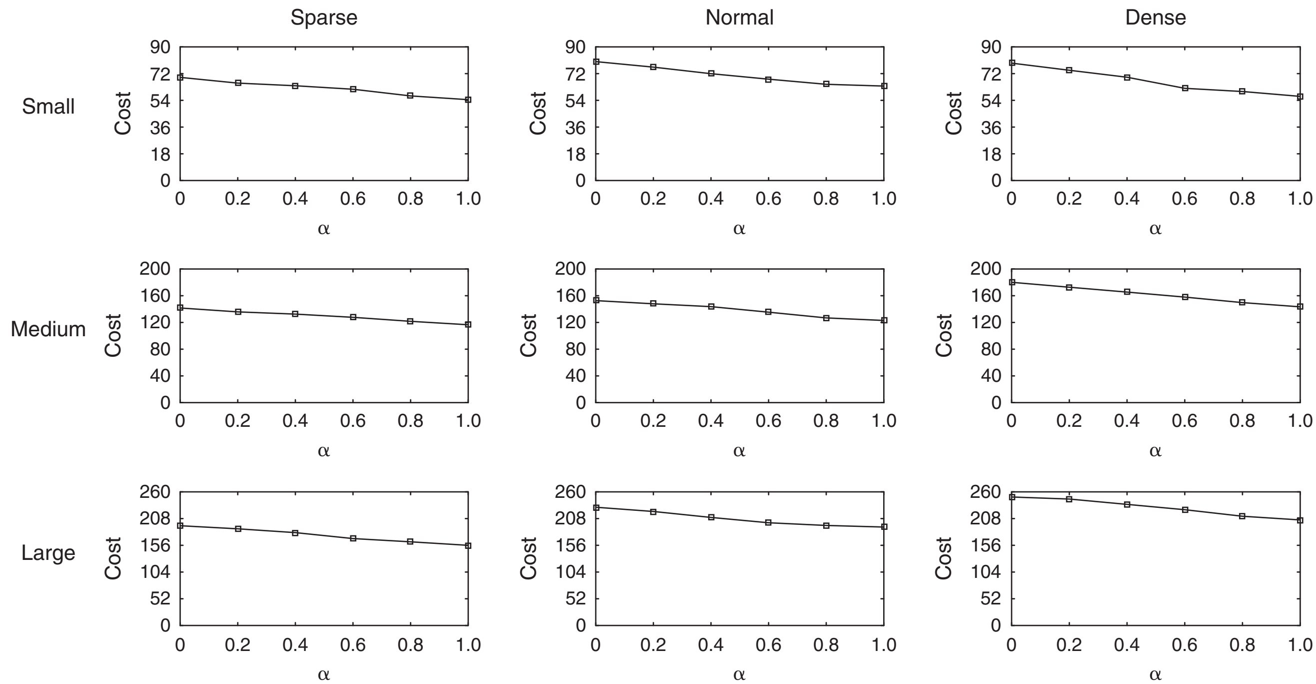


Figure 8. Objective Function Values (in Thousands) for Different Values of α for Dedicated Patient-to-Surgeon Allocation



the combination of small (20 patients, three surgeons), medium (30 patients, six surgeons), and large (40 patients, nine surgeons) cases. The aggregation of any of these single hospitals gives rise to bi- or tri-hospital data. Each problem instance is generated five times. The average number of binary variables, continuous variables, constraints, and sparsity of each trial is shown in Table 5.

We define the UHN optimal baseline cost as the sum of individual OR scheduling costs of each hospital with $\alpha = 0$. Collaboration with dedicated patients results in 6.08% cost-savings, despite the suboptimality of the

Figure 9. CPU Time (Minutes) for Different Values of α for Dedicated Patient-to-Surgeon Allocation

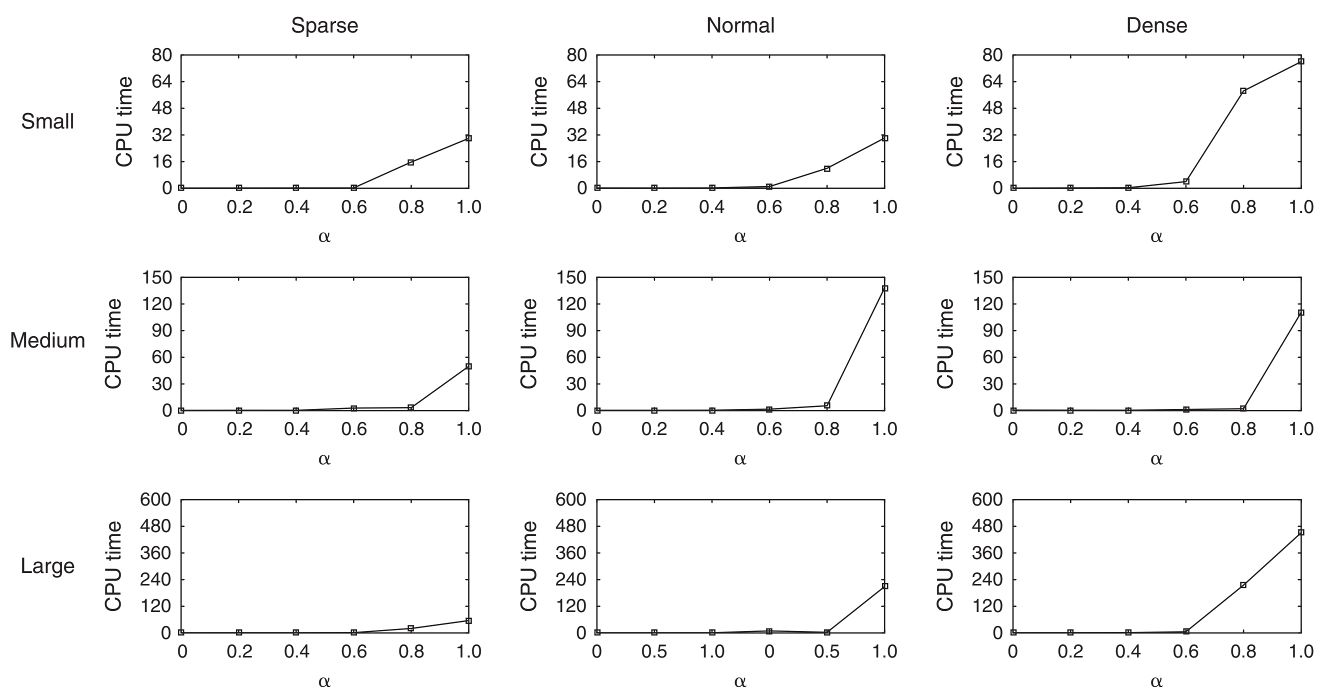


Figure 10. Average Number of Iterations, Feasibility, and Optimality Cuts for Different Values of α for Solved Trials with Dedicated Patient-to-Surgeon Allocation

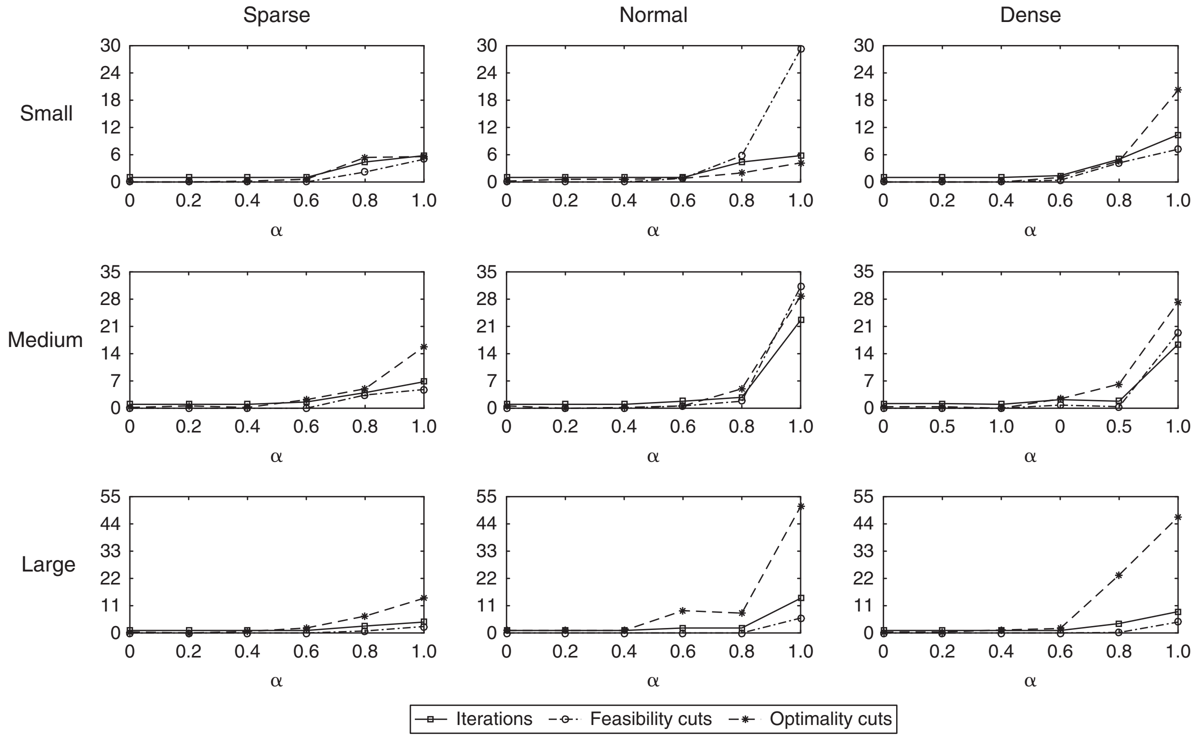


Table 5. Average Number of Constraints (Rows), Binary Variables (BV), Continuous Variables (CV), and Non Zeros (NZ)

		MP				SP			
	$\lvert\mathcal{P}\rvert$	Rows	BV	CV	NZ	Rows	BV	CV	NZ
Dedicated	60	3,649	2,368	45	16,885	97	19	415	50
	90	6,821	3,567	45	25,289	116	21	509	61
	120	9,197	5,896	45	39,434	156	25	696	85
Semiflexible	60	6,820	3,493	45	20,683	96	18	413	49
	90	13,103	6,670	45	50,417	137	22	612	71
	120	20,628	11,575	45	90,545	200	26	922	105
Flexible	60	8,058	4,112	45	30,219	107	18	465	53
	90	16,944	8,360	45	63,861	157	22	708	79
	120	28,795	18,568	45	126,361	228	26	1,070	116

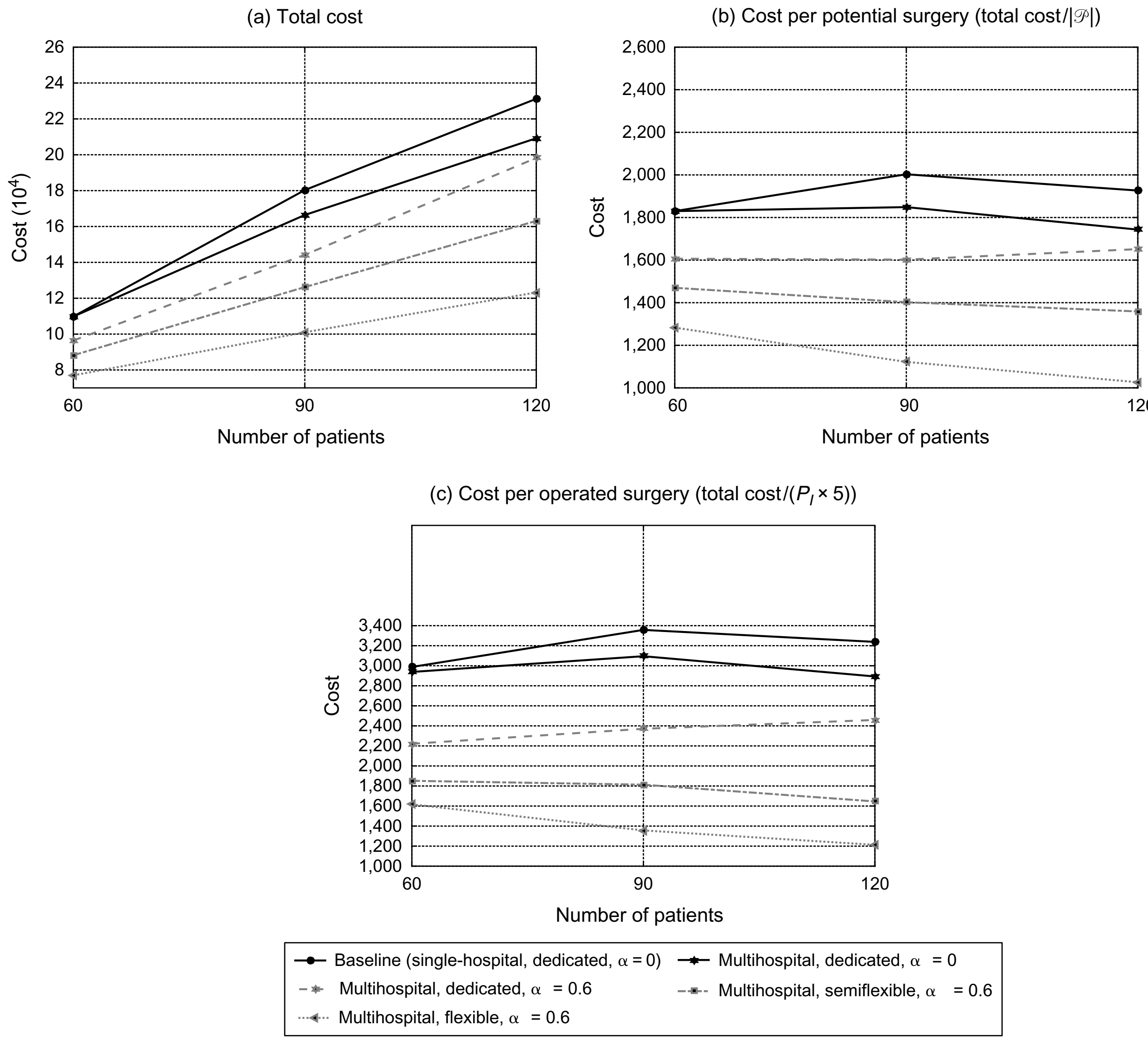
multihospital solutions, and adjusting surgeon schedule tightness further increases cost-savings to 15.50% (Figure 11(a)). The additional inclusion of semiflexible and flexible patients increases the average cost-savings to 26.37% and 39.89%, respectively.

The competitive advantage of CORPS with dedicated, semiflexible, and flexible patients increases with more patients (Figure 11(a)), though per-scheduled patient savings only significantly increase with semiflexible and flexible patients (Figure 11(b)). Because all patients in the schedule may not be operated on, we divide total cost by the number of operated patients to compute per-patient actual savings (Figure 11(c)). Comparing the cost per operated patient of the baseline with that of the multihospital with flexible patients with $\alpha = 0.6$ shows 45.00%, 59.57%, and 62.53% average cost-savings for 60, 90, and 120 patients, respectively.

4.4. Throughput Gain

The sources of cost-savings are quantified in Table 6, where indices I and C represent individual and collaborative scheduling, respectively. The surgeon (P/S) and OR (P/R) throughput are represented by S^* and R^*, respectively, where P, S, and R are the daily average number of operated patients, used surgeons, and opened ORs, respectively. In individual scheduling (12.00 patients), P is less than in collaborative scheduling (13.11 patients), amounting to 9.25% higher

Figure 11. Total Cost and Cost Per Patient as Scenario Flexibility Increases



patient admission for collaborative scheduling. However, this figure is not the only source of cost-savings; the decrease in the number of surgeons and ORs in collaborative scheduling should additionally be considered. Thus, we consider the relative percentage throughput gain (RPTG) of surgeons and ORs, calculated as RPTG $= 100\% \times (S_C^* - S_I^*)/S_I^*$ for surgeons and RPTG $= 100\% \times (R_C^* - R_I^*)/R_I^*$ for ORs.

Collaborative scheduling improves RPTG for ORs (16.02% on average) and surgeons (11.42% on average) in all scenarios (Table 7). These cost-savings can mostly be attributed to better use of surgeon and OR availability because surgeries can be assigned to the most time- and cost-efficient surgeons in the network. The patient-to-surgeon flexibility also causes more optional patients to be scheduled in the current planning horizon because some surgeons have unutilized availability that can be filled by a subset of optional patients.

4.5. Value of Shorter and Longer Planning Horizons

We investigate the impact of the number of days in the planning horizon on the CORPS daily objective function value, optimality gaps, and CPU times (Figure 12). As we increase the days in the planning horizon, CORPS daily costs decrease but CPU times and optimality gaps increase. Six days in the planning horizon seems to be the best length for the planning horizon, which goes against existing UHN practice (five days in the planning horizon). However, a five-day planning horizon also provides a good balance among costs, CPU time, and optimality. The different patient allocation schemes perform mostly as expected (more flexibility means more cost-savings, but at a higher computational cost), though maximum CPU time was reached with just six days in the horizon for flexible patients, and eight days for semiflexible patients,

Table 6. Daily Average Number of Operated Patients (P_I), Opened ORs (R_I), and Surgeons (S_I), and Surgeon (S^*) and ORs (R^*) Throughput

		Individual (0%)					Collaborative				
Allocation type	$\|\mathscr{P}\|$	P_I	S_I	OR_I	S_I^*	R_I^*	P_C	S_C	OR_C	S_C^*	R_C^*
Dedicated $\alpha = 0.0$	60	7.35	4.78	3.90	**1.54**	1.88	7.48	4.85	3.43	**1.54**	**2.18**
	90	10.74	6.86	5.42	1.57	1.98	10.75	6.83	4.85	**1.58**	**2.22**
	120	14.28	8.98	6.76	1.59	2.11	14.46	9.00	6.06	**1.61**	**2.39**
Average		10.79	6.87	5.36	1.57	1.99	10.90	6.89	4.78	**1.58**	**2.26**
Dedicated $\alpha = 0.6$	60	8.28	4.84	4.12	1.71	2.01	8.68	4.68	3.76	**1.85**	**2.31**
	90	12.08	6.36	5.68	1.90	2.13	12.16	6.28	5.08	**1.94**	**2.39**
	120	15.96	8.84	7.2	1.81	2.22	16.12	8.68	6.8	**1.86**	**2.37**
Average		12.11	6.68	5.67	1.81	2.12	12.32	6.55	5.21	**1.88**	**2.36**
Semi-flex $\alpha = 0.6$	60	8.44	4.92	4.16	1.72	2.03	9.52	4.96	4.00	**1.92**	**2.38**
	90	12.36	6.48	5.60	1.91	2.21	13.92	6.64	5.52	**2.10**	**2.52**
	120	16.88	8.84	7.40	1.91	2.28	19.56	9.2	7.64	**2.13**	**2.56**
Average		12.56	6.75	5.72	1.84	2.17	14.33	6.93	5.72	**2.05**	**2.49**
Flexible $\alpha = 0.6$	60	8.48	4.84	4.16	1.75	2.04	9.52	4.64	3.72	**2.05**	**2.56**
	90	12.52	8.44	5.60	1.48	2.24	14.88	6.32	5.36	**2.35**	**2.78**
	120	16.92	8.44	7.20	2.00	2.35	20.30	8.35	7.00	**2.43**	**2.90**
Average		12.64	7.24	5.65	1.75	2.21	14.9	6.44	5.36	**2.28**	**2.75**
Grand average		12.00	6.89	5.60	1.74	2.09	13.11	6.70	5.22	**1.95**	**2.46**

Note. Bold represents improved throughput for surgeons and ORs.

Table 7. Daily Average Surgeon and OR RPTG

	Individual		Collaborative	
Allocation type	OR	Surgeon	OR	Surgeon
Dedicated ($\alpha = 0.0$)	0.00	0.00	**13.57**	**0.006**
Dedicated ($\alpha = 0.6$)	0.00	0.00	**11.32**	**4.00**
Semi-flex ($\alpha = 0.6$)	0.00	0.00	**14.75**	**11.41**
Flexible ($\alpha = 0.6$)	0.00	0.00	**24.43**	**30.28**
Total average	0.00	0.00	**16.02**	**11.42**

Note. Bold represents improved throughput for surgeons and ORs.

which diminishes the cost-savings of those scenarios with larger horizon lengths.

5. Game Theoretic Implications

Table 8 shows that the suboptimal costs of collaborative scheduling are never worse than the optimal costs of individual hospitals, indicating that hospitals are better off collaborating than scheduling their own patients. We now use game theory to investigate whether each individual hospital in the network benefits from joining a coalition.

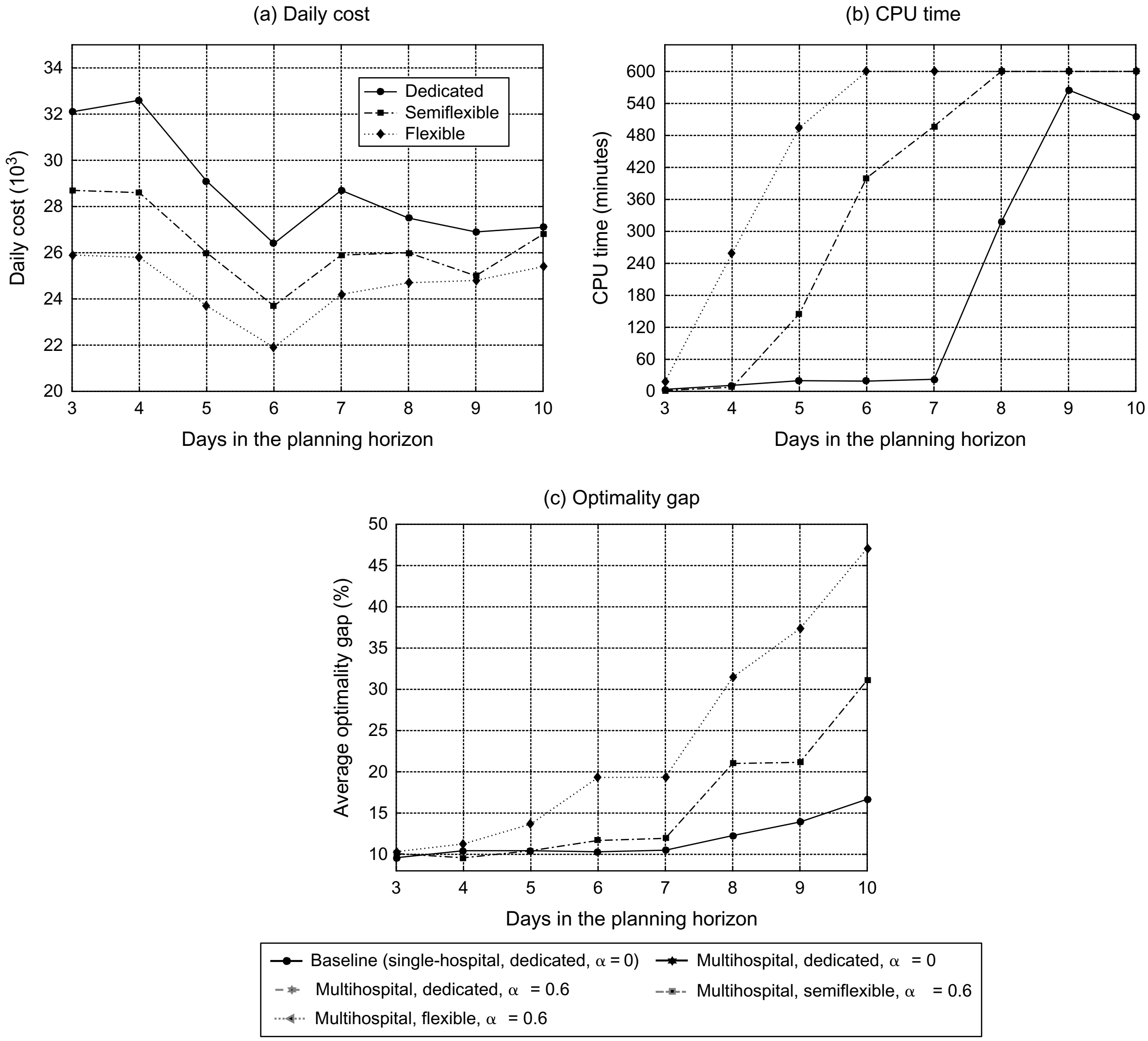
Multiperson games, first proposed by Neumann and Morgenstern (1944), assume that various subgroups of players might cooperate to form coalitions. CORPS is a cooperative game among UHN hospitals as they operate under a universal healthcare system, aiming to best serve their patients given existing resources. We therefore use the notions of Shapley value (Shapley 1953, Leng and Parlar 2009) and game convexity (Shapley 1971) to determine (1) how the acquired collaborative payoff has to be fairly redistributed among hospitals; and (2) which hospital coalitions are stable. The answers to these questions are equally important as a coalition may be stable but the payoff division scheme may not be fair. To answer these questions, we use problem instances with 20, 30, and 40 patients and three, six, and nine surgeons for individual hospitals, which are aggregated for bi- and tri-hospital instances.

Table 8. Comparison of Cost-Savings (in Tens of Thousands) of Suboptimal Collaborative Scheduling vs. Individual Scheduling Trials with 0% Optimality Gap with $\alpha = 0.6$

		Individual (0%)				
Allocation type	Size $\|\mathscr{P}\|, \|\mathscr{S}\|$	H_1	H_2	H_3	Total cost	Collaborative total cost
Dedicated ($\alpha = 0$)	60, 9	3.60	4.02	3.96	11.58	**10.98**$^{(1.00)}$
	90, 18	6.31	6.16	5.56	18.03	**16.64**$^{(2.11)}$
	120, 27	7.72	7.51	7.89	23.12	**20.92**$^{(2.84)}$
Dedicated ($\alpha = 0.6$)	60, 9	3.42	3.69	3.87	10.98	**9.64**$^{(4.90)}$
	90, 18	5.82	5.31	4.63	15.76	**14.42**$^{(4.62)}$
	120, 27	7.07	6.78	7.27	21.12	**19.82**$^{(5.32)}$
Semi-flex ($\alpha = 0.6$)	60, 9	3.39	3.62	3.80	10.81	**8.82**$^{(6.18)}$
	90, 18	5.67	5.09	4.44	15.21	**12.62**$^{(7.72)}$
	120, 27	6.67	6.41	6.76	19.85	**16.31**$^{(9.03)}$
Flexible ($\alpha = 0.6$)	60, 9	3.33	3.55	3.66	10.55	**7.70**$^{(9.95)}$
	90, 18	5.44	4.87	4.28	14.59	**10.10**$^{(16.02)}$
	120, 27	6.46	5.85	4.26	16.58	**12.32**$^{(15.77)}$

Notes. Superscripts are the optimality gaps of collaborative trials in 10 hours CPU time. Bold represents better total costs.

Figure 12. The Impact of Planning Horizon Length



5.1. Payoff Division

We use a game theoretic approach to determine the average marginal contribution of each hospital with respect to the total collaborative scheduling cost (grand coalition cost), and show how much each hospital benefited from the coalition. Real-valued payoffs associated with subcoalitions S and the grand coalition of $\mathscr{H}$ hospitals are denoted $v(S)$ and $v(\mathscr{H})$, respectively. Given a coalitional game $(\mathscr{H}, v)$, there is a unique payoff division $x(v) = \phi(\mathscr{H}, v)$ that divides the acquired payoff of the grand coalition among its members. The payoff division for each hospital $(\phi_h(\mathscr{H}, v))$ is computed via the Shapley value as follows (Shapley 1953):

$$\phi_h(|\mathscr{H}|, v) = \frac{1}{|\mathscr{H}|!} \sum_{S \subseteq |\mathscr{H}| \setminus \{h\}} |S|!(|\mathscr{H}| - |S| - 1)! \cdot [v(S \cup \{h\}) - v(S)].$$

The Shapley value captures the average marginal contributions of hospital h, averaging over all subcoalitions according to which the grand coalition could be built up from an empty coalition.

As expected, the acquired payoff of hospitals is increased as the number of patients and patient flexibility increases (Table 9).

5.2. Coalitional Stability

Convexity states that the grand coalition has the highest payoff and no member of a coalition has incentive to leave the coalition with the hope of forming a more cost-effective coalition (Shapley 1971). A reward-based cooperative game $G = (\mathscr{H}, v)$ is convex if $v(S \cup T) \geq v(S) + v(T) - v(S \cap T)$, $\forall\, S, T \subset \mathscr{H}$ (Shapley 1971). Since our game involves costs instead of rewards, we use the negative of the objective function value to show the convexity of our coalitional game. Each convex game

Table 9. Individual Scheduling Costs, Collaborative Shapley Values ($\phi_h(|\mathscr{H}|, v)$), and Collaborative Payoffs

Allocation type	$\|\mathscr{P}\|$		H_1	H_2	H_3	Total
Dedicated $\alpha = 0.6$	60	Individual cost	3.42	3.69	3.87	10.98
		Collaborative Shapley value	3.12	3.10	3.42	9.64
		Collaborative payoff	0.30	0.59	0.45	1.35
	90	Individual cost	5.82	5.31	4.63	15.76
		Collaborative Shapley value	5.11	5.09	4.22	14.42
		Collaborative payoff	0.71	0.22	0.41	1.34
	120	Individual cost	7.07	6.78	7.27	21.12
		Collaborative Shapley value	6.46	6.55	6.81	19.82
		Collaborative payoff	0.61	0.23	0.46	1.30
Semiflexible $\alpha = 0.6$	60	Individual cost	3.39	3.62	3.80	10.81
		Collaborative Shapley value	2.88	2.83	3.11	8.82
		Collaborative payoff	0.51	0.79	0.69	1.99
	90	Individual cost	5.67	5.09	4.44	15.21
		Collaborative Shapley value	4.45	4.52	3.65	12.62
		Collaborative payoff	1.22	0.57	0.79	2.60
	120	Individual cost	6.67	6.41	6.76	19.85
		Collaborative Shapley value	5.34	5.41	5.56	16.31
		Collaborative payoff	1.33	1.00	1.20	3.54
Flexible $\alpha = 0.6$	60	Individual cost	3.33	3.55	3.66	10.55
		Collaborative Shapley value	2.57	2.38	2.75	7.70
		Collaborative payoff	0.76	1.17	0.91	2.85
	90	Individual cost	5.44	4.87	4.28	14.59
		Collaborative Shapley value	3.59	3.77	2.74	10.10
		Collaborative payoff	1.85	1.10	1.54	4.49
	120	Individual cost	6.46	5.85	4.26	16.58
		Collaborative Shapley value	4.09	4.21	4.02	12.32
		Collaborative payoff	2.37	1.64	0.24	4.26

has a nonempty core and the core of an $\mathscr{H}$-hospital game is the set of feasible outcomes that cannot be improved upon by any subcoalitions of hospitals (Shapley 1971).

Table 10 shows the costs associated with the grand coalition and possible bi-hospital subcoalitions. The collaborative OR scheduling cost in a bi-hospital coalition between hospitals i and j is denoted H_{ij}. The individual OR scheduling cost of hospital i is denoted H_i, which is subtracted from the summation of hospital i's bi-hospital coalitions to determine whether or not the subcoalitions are stable (convex). For subcoalitions to be convex, the total cost of the grand coalition should not be greater than those of possible subcoalitions. The Shapley value is the unique symmetric payoff division procedure and is strongly monotonic (Young 1985, Theorem 2). Coalitional monotonicity corresponds to monotonic changes in the absolute value of the coalitions including hospital h. That is, the value of the coalitions including hospital h increase relative to the value of the coalitions excluding hospital h. Adding a new hospital to the existing coalition of hospitals may be monotonic (nonincreasing) if the optimal costs of the new coalition is readily obtainable; however, the optimal solution of the new coalition may be computationally prohibitive, leading to more cost-effective subcoalitions than the grand coalition. Thus, we hypothesize that some of dedicated, semiflexible, and flexible OR scheduling coalitions are nonconvex due to the optimality of individual OR scheduling test cases (0% optimality gap) compared to suboptimal collaborative scheduling with high optimality gaps (up to 16% optimality gap).

6. Discussion

CORPS novelly encompasses collaboration, surgeon schedule tightness, and patient flexibility. Our LBBD for CORPS obtains near-optimal solutions for dedicated patients (1%–3% optimality gaps), and computational performance is not significantly affected by the inclusion of new hospitals. As expected, the tighter surgeon schedules increase cost-savings, but slightly worsen CPU time. The most influential factor in CPU time and optimality gap is patient-to-surgeon allocation flexibility; however, patient flexibility also accounts for most of the cost-savings achieved by CORPS. We note that our investigation of surgeon schedule tightness is to understand how it impacts costs and solvability in the optimization planning process, and our selection of $\alpha = 0.6$ does not necessarily mean that a surgeon should be present for 40% of a patient's preparation and cleaning time in the OR; instead, it means that surgeons should be willing to move between ORs to avoid

Table 10. The Value of Grand Coalition vs. Subcoalitions (Bold Represents Convex and Stable Coalitions)

	Individual (0%)			Bi-hospital (5%)			Subcoalitions			Grand
	H_1	H_2	H_3	H_{12}	H_{13}	H_{23}	$(H_{12}+H_{23})-H_2$	$(H_{13}+H_{23})-H_3$	$(H_{12}+H_{13})-H_1$	H_{123}
Dedicated										
60	3.42	3.69	3.87	6.49	6.58	6.90	9.69	**9.61**	9.65	$9.64^{(4.90)}$
90	5.82	5.31	4.63	10.40	9.82	9.34	14.44	14.53	**14.41**	$14.42^{(4.62)}$
120	7.07	6.78	7.27	13.18	13.67	13.21	**19.61**	**19.61**	**19.63**	$19.82^{(5.32)}$
Semi-flex										
60	3.39	3.62	3.80	6.08	6.14	6.46	8.92	**8.80**	8.83	$8.82^{(6.18)}$
90	5.67	5.09	4.44	9.46	9.09	8.40	12.77	13.05	12.88	$\mathbf{12.62}^{(7.72)}$
120	6.67	6.41	6.76	11.52	11.44	11.86	16.96	16.53	**16.29**	$16.314^{(9.03)}$
Flexible										
60	3.33	3.55	3.66	5.43	5.47	6.06	7.93	7.86	**7.56**	$7.70^{(9.95)}$
90	5.44	4.87	4.28	8.31	7.75	6.84	10.28	10.31	10.60	$\mathbf{10.10}^{(16.02)}$
120	6.46	5.85	4.26	9.77	9.85	9.00	12.91	14.59	13.16	$\mathbf{12.32}^{(15.77)}$

Note. Individual, bi-hospital, and grand solutions have been solved to 0%, 5%, and varying optimality gaps shown in superscript.

some preparation and cleaning times. Despite a large optimality gap for flexible patients (16%), CORPS still achieves 40% cost-savings with respect to UHN optimal baseline costs. The cost-savings are also in agreement with previous studies finding that collaborative master surgical scheduling leads to at least 8%–9% ward bed savings (Santibanez et al. 2007) and OR scheduling under open scheduling leads to 21%–59% cost-savings (Batun et al. 2011).

We demonstrated that our novel LBBD is a powerful tool to optimize collaborative allocation and sequencing problems. In the absence of multiple hospitals, surgeon schedule tightness, patient flexibility, and surgical duration stochasticity, Hashemi Doulabi et al. (2016) solved an integrated single-hospital operating room planning (allocation) and scheduling (sequencing) for 80 and 100 patients to 9.92% and 12.01% optimality in 260 and 259 minutes, respectively, using a branch-and-price-and-cut algorithm. In contrast, our LBBD solved all CORPS trials with $\alpha = 0$ to 10% optimality in less than 90 seconds. This difference in computation time may be partially attributed to a difference in the optimization solver (IBM ILOG CPLEX Optimization Studio 12.4) and the computing platform (Computer Intel Xeon X5675 processors, 3.07 GHz, and a total of 12 cores) used in Hashemi Doulabi et al. (2016), but most of the CPU time difference is likely attributable to LBBD outperforming branch-and-price-and-cut.

We also used the notions of Shapley value and game convexity to determine that both UHN and each hospital in the network benefit from the coalition. On the other hand, we showed that the grand coalition is not consistently stable likely due to suboptimal solutions, but becomes more attractive as the number of patients and patient flexibility increases. It is noteworthy that including more hospitals in the coalition may not be cost-effective because the size dimensionality of the resulting scheduling problem may be computationally prohibitive, leading to better suboptimal solutions (lower optimality gap) for subcoalitions than for the grand coalition.

CORPS is hard to solve even with the simplification of deterministic surgery times (Jebali et al. 2006, Fei et al. 2009, Marques et al. 2012, Hashemi Doulabi et al. 2016; see review papers by Guerriero and Guido 2011, Cardoen et al. 2010). While this simplification may not be realistic, in the UHN data, 58% of the realized surgical times were less than the booked times, resulting in systematic OR under-utilization in the real schedules. We therefore anticipate that stochasticity will not dramatically diminish the cost-savings achievable with CORPS. However, for other hospitals where the realized incision time is mostly higher than booked times, projected cost-savings may be impacted.

While availability of downstream beds also impacts OR patient throughput (Jebali et al. 2006, Wang et al. 2016), most studies in the literature ignore this factor for tractability (Fei et al. 2009, Batun et al. 2011, Marques et al. 2012, Hashemi Doulabi et al. 2016, Roshanaei et al. 2017). However, very few surgeries were canceled due to downstream bed availability in UHN. Thus, we do not expect the cost-savings of CORPS to be significantly impacted by the omission of downstream beds.

Although the benefits of a centralized system are clear from a mathematical perspective, healthcare institutions are often resistant to quantitative planning approaches—collaborative or not—for a variety of reasons, including conservative attitudes toward change and a mistrust of mathematical methods to design healthcare delivery. In the case of our collaborator, UHN already engages in some collaborative surgical efforts, and though the collaborations are not systematic, it is a strong indication of receptiveness to a fully collaborative approach. Another collaborative example is Sunnybrook Health Sciences Centre

(Toronto, Ontario, Canada), whose orthopedic department engages in central wait list planning for some surgeons and some surgeries, which is also an example of patient-to-surgeon allocation flexibility. While this attitude toward collaboration and flexibility is likely more common in Canada (and other countries with universal healthcare) than in the United States, we note that in our example, each hospital was better off cost-wise collaborating than operating individually, contrary to the instinct to remain isolated in a zero-sum game. Though our model focuses on costs, not profits, we anticipate that a similar study of CORPS with a profit-motivated objective will find that collaboration is preferred to isolation for each hospital. In Canada, we anticipate that the two biggest challenges to implementing CORPS are (1) ensuring that surgeons are able to perform their desired number of specific types of surgeries each year, and (2) balancing workload between hospitals and surgeons. These concerns can be incorporated into the CORPS model, though they require significant modifications to the LBBD framework, and are planned for future investigation.

7. Conclusions

We studied the problem of collaborative operating room planning and scheduling in which human and physical resources are collaboratively planned amongst a coalition of hospitals in a strategic network. We incorporated various patient-to-surgeon allocation flexibilities and included a mechanism to empirically tune the tightness of surgeon schedules, leading to 40% cost-savings. Using a game theoretic approach, we provided compelling evidence for hospitals to share resources to gain individual benefits.

For future research, a branch-and-check algorithm can be developed for this problem. CORPS could be extended to incorporate workload balancing among surgeons, ORs, and hospitals, as well as to stochastic and robust variations. Additionally, different heuristic approaches can be incorporated to give the MP a stronger initial bound, expediting LBBD convergence to a more acceptable suboptimal solution. Collaborative case mix planning among hospitals to satisfy their surgical demands while maximizing their profit would be another interesting topic. It also would be interesting to formulate CORPS as a multilayered game theory problem, wherein hospitals cooperate as strategic players and surgeons compete over OR allocation as tactical players in a zero-sum game.

References

Agee MD, Zane G (2013) Lessons from game theory about healthcare system price inflation. *Appl. Health Econom. Health Policy* 11(1):45–51.

Batun S, Denton BT, Huschka TR, Schaefer AJ (2011) Operating room pooling and parallel surgery processing under uncertainty. *INFORMS J. Comput.* 23(2):220–237.

Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik* 4(1): 238–252.

Cardoen B, Demeulemeester E, Belien J (2010) Operating room planning and scheduling: A literature review. *Eur. J. Oper. Res.* 201(3):921–932.

Chan TCY, Demirtas D, Kwon RH (2016) Optimizing the deployment of public access defibrillators. *Management Sci.* 62(12):3617–3635.

Chaudhry IA, Drake PR (2009) Minimizing total tardiness for the machine scheduling and worker assignment problems in identical parallel machines using genetic algorithms. *Internat. J. Adv. Manufacturing Tech.* 42(5–6):581–594.

Chu Y, Xia Q (2004) Generating Benders cuts for a general class of integer programming problems. Régin J-C, Rueher M, eds. *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, Lecture Notes in Computer Science, Vol. 3011 (Springer, Berlin), 127–141.

Clinical Advisory Board (2001) Surgical services reform: Executive briefing for clinical leaders. Technical report, Advisory Board Company, Washington, DC.

Day R, Garfinkel R, Thompson S (2012) Integrated block sharing: A win-win strategy for hospitals and surgeons. *Manufacturing Service Oper. Management* 14(4):567–583.

Denton BT, Miller AJ, Balasubramanian HJ, Huschka TR (2010) Optimal allocation of surgery blocks to operating rooms under uncertainty. *Oper. Res.* 58(4, Pt. 1):802–816.

ElMaraghy H, Patel V, Abdallah IB (1999) Genetic algorithm based approach for scheduling of dual-resource constrained manufacturing systems. *CIRP Ann.—Manufacturing Tech.* 48(1):369–372.

ElMaraghy H, Patel V, Abdallah IB (2000) Scheduling of manufacturing systems under dual-resource constraints using genetic algorithms. *J. Manufacturing Systems* 19(3):186–201.

Fazel-Zarandi MM, Beck JC (2012) Using logic-based Benders decomposition to solve the capacity- and distance-constrained plant location problem. *INFORMS J. Comput.* 24(3):387–398.

Fei H, Chu C, Meskens N (2009) Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Ann. Oper. Res.* 166(1):91–108.

Fei H, Meskens N, Chu C (2010) A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Comput. Indust. Engrg.* 58(2):221–230.

Fei H, Chu C, Meskens N, Artiba A (2008) Solving surgical cases assignment problem by a branch-and-price approach. *Internat. J. Production Econom.* 112(1):96–108.

Guerriero F, Guido R (2011) Operational research in the management of the operating theatre: A survey. *Health Care Management Sci.* 14(1):89–114.

Guinet A, Chaabane S (2003) Operating theatre planning. *Internat. J. Production Econom.* 85(1):69–81.

Harjunkoski I, Grossmann IE (2002) Decomposition techniques for multistage scheduling problems using mixed-integer and constraint programming methods. *Comput. Chemical Engrg.* 26(11): 1533–1552.

Hashemi Doulabi SH, Rousseau L-M, Pesant G (2016) A constraint-programming-based branch-and-price-and-cut approach for operating room planning and scheduling. *INFORMS J. Comput.* 28(3):432–448.

Health Care Financial Management Association (2005) Achieving operating room efficiency through process integration. Technical report, Health Care Financial Management Association, Westchester, IL.

Hooker JN (2007) Planning and scheduling by logic-based Benders decomposition. *Oper. Res.* 55(3):588–602.

Hooker JN, Ottosson G (2003) Logic-based Benders decomposition. *Math. Programming* 96(1):33–60.

Jebali A, Alouane ABH, Ladet P (2006) Operating rooms scheduling. *Internat. J. Production Econom.* 99(1–2):52–62.

Lei D, Guo X (2014) Variable neighbourhood search for dual-resource constrained flexible job shop scheduling. *Internat. J. Production Res.* 52(9):2519–2529.

Leng M, Parlar M (2009) Allocation of cost savings in a three-level supply chain with demand information sharing: A cooperative-game approach. *Ann. Oper. Res.* 57(1):200–213.

Marques I, Captivo ME, Margarida VP (2012) An integer programming approach to elective surgery scheduling. *OR Spectrum* 34(2):407–427.

Naderi B, Azab A (2014) Modeling and heuristics for scheduling of distributed job shops. *Expert Systems Appl.* 41(17):7754–7763.

Naderi B, Ruiz R (2010) The distributed permutation flowshop scheduling problem. *Comput. Oper. Res.* 37(4):754–768.

Naderi B, Ruiz R (2014) A scatter search algorithm for the distributed permutation flowshop scheduling problem. *Eur. J. Oper. Res.* 239(2):323–334.

Nelsen RB (2006) *An Introduction to Copulas*, Springer Series in Statistics (Springer, New York).

Neumann JV, Morgenstern O (1944) *An Introduction to Copulas*, Springer Series in Statistics (Princeton University Press, Princeton, NJ).

Pham DN, Klinkert A (2008) Surgical case scheduling as a generalized job shop scheduling problem. *Eur. J. Oper. Res.* 185(3): 1011–1025.

Roland B, Di Martinelly C, Riane F, Pochet Y (2010) Scheduling an operating theatre under human resource constraints. *Comput. Indust. Engrg.* 58(2):212–220.

Roshanaei V, Luong C, Aleman D, Urbach D (2017) Propagating logic-based Benders' decomposition approaches for distributed operating room scheduling. *Eur. J. Oper. Res.* 257(2):439–455.

Ruiz-Torres AJ, Centeno G (2007) Scheduling with flexible resources in parallel workcenters to minimize maximum completion time. *Comput. Oper. Res.* 34(1):48–69.

Samudra M, Van Riet C, Demeulemeester E, Cardoen B, Vansteenkiste N, Rademakers FE (2016) Scheduling operating rooms: Achievements, challenges and pitfalls. *J. Scheduling* 19(5): 493–525.

Santibanez P, Begen M, Atkins D (2007) Surgical block scheduling in a system of hospitals: An application to resource and wait list management in a British Columbia health authority. *Health Care Management Sci.* 10(3):269–282.

Shapley LS (1953) A value for n-person games. *Ann. Math. Stud.* 28(2):307–317.

Shapley LS (1971) Cores of convex games. *Internat. J. Game Theory* 1(1):11–26.

Sheater SJ, Jones MC (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Res.* 53(3):683–690.

Tran TT, Beck JC (2012) Logic-based Benders decomposition for alternative resource scheduling with sequence dependent setups. De Raedt L, Dubois D, Bessiere C, eds. *ECAI'12 Proc. 20th Eur. Conf. Artificial Intelligence* (IOS Press, Amsterdam), 774–779.

Tran TT, Araujo A, Beck JC (2016) Decomposition methods for the parallel machine scheduling problem with setups. *INFORMS J. Comput.* 28(1):83–95.

Vijayakumar B, Parikh PJ, Scott R, Barnes A, Gallimore J (2013) A dual bin-packing approach to scheduling surgical cases at a publicly-funded hospital. *Eur. J. Oper. Res.* 224(3):583–591.

Wand MP, Jones MC (1995) *Kernel Smoothing* (Chaplan & Hall/CRC, Boca Raton, FL).

Wang S, Roshanaei V, Aleman D, Urbach D (2016) A discrete event simulation evaluation of distributed operating room scheduling. *IIE Trans. Healthcare Systems Engrg.* 6(4):236–245.

Westhoff WW, Cohen CF, Cooper EE, Corvin J, McDermott RJ (2012) Cooperation or competition: Does game theory have relevance for public health? *Amer. J. Health Ed.* 43(3):175–183.

Xu J, Xu X, Xie SQ (2011) Recent developments in dual resource constrained (DRC) system research. *Eur. J. Oper. Res.* 215(2): 309–318.

Young HP (1985) Monotonic solutions of cooperative games. *Internat. J. Game Theory* 14(2):65–72.