The Impact of Surgeon Daily Workload and its Implications for Operating Room Scheduling

Yiwen Shen
Information Systems, Business Statistics and Operations Management,
Hong Kong University of Science and Technology, yiwenshen@ust.hk

Carri W. Chan
Decision, Risk, and Operations, Columbia Business School, cwchan@columbia.edu

Fanyin Zheng
Decision, Risk, and Operations, Columbia Business School, fanyin.zheng@columbia.edu

Michael Argenziano
Department of Surgery, New York Presbyterian/Columbia University Irving Medical Center, ma66@cumc.columbia.edu

Paul Kurlansky
Department of Surgery, New York Presbyterian/Columbia University Irving Medical Center, pk2245@cumc.columbia.edu

Problem definition: In many service systems, individual server's workload can have a substantial impact on service time and quality. Such effects are particularly important in healthcare systems which often operate under resource and time constraints. In much of the literature, this has been primarily considered at the system level rather than the individual level. In this study, we investigate this relationship in the context of cardiac surgery, i.e., how surgery duration and patient outcomes are affected by individual surgeon's daily workload.

Methodology/results: Using a detailed data set of more than 5,600 cardiac operations in a large hospital, we quantify how individual surgeon's daily workload (e.g., the number of operations performed by the focal surgeon) affects surgery duration and patient outcomes. To handle the endogeneity of surgeon daily workload, we construct instrumental variables using the cardiac department's operational factors, including the block schedule of surgeons. We find that high daily workload for the focal surgeon is associated with longer incision times and worse patient outcomes. Specifically, the surgeon's higher daily workload leads to longer post-surgery length-of-stay in the ICU and hospital as well as higher likelihoods of reoperation and readmission for their patients. These results highlight the potential negative impact of high individual surgeon workload which may result in surgeon fatigue and operational constraints. We then develop a surgical scheduling model that incorporates the estimated impact of surgeon workload. We solve the model by mixed-integer quadratic programming and show that our proposed schedule can substantially reduce total incision time and post-surgery length-of-stay.

Managerial implications: Our results suggest that hospitals should take into account the effects of individual surgeon's daily workload when managing their ORs. Specifically, they can substantially improve patient flow and patient outcomes by smoothing individual surgeon's workload across days.

Key words: healthcare delivery, empirical operations management, behavioral operations, operating room scheduling, surgeon workload, quality of care

1. Introduction

The relationship between system workload and service performance has drawn increasing attention in the operations management community. Traditional operations management models generally assume service time is fixed and independent of system workload (e.g., Wolff 1989). However, a growing body of empirical research shows that the service time of human-involved systems can be endogenously impacted by the overall workload (e.g. Staats and Gino 2012 and Tan and Netessine 2014). Such effects are particularly important in healthcare, where resources are often constrained, and timely access to medical services can have a substantial impact on patient satisfaction and clinical outcomes. System workload has been shown to affect service time in different healthcare settings, such as intensive care units (ICUs) (Kc and Terwiesch 2012), patient transportation and cardiac surgery (Kc and Terwiesch 2009), as well as emergency departments (Kc 2014, Batt and Terwiesch 2016). While some studies found that service time can increase with workload levels (Green and Nguyen 2001, Tan and Netessine 2014), the opposite pattern, i.e., service time decreasing in workload levels, has also been observed (Kc and Terwiesch 2009, Chan et al. 2012). Partially reconciling the two patterns, some recent studies show that service time can react non-monotonically to workload levels. These studies have found an inverted-U shaped pattern, i.e., service time first increases and then decreases in the workload measures (see, e.g., Tan and Netessine 2014, Batt and Terwiesch 2016, and Berry Jaeker and Tucker 2017). Beyond service time, the effect of hospital's workload on the quality of care has also been investigated in both the operations management and medical community. There is empirical evidence that increased workload can lead to worse medical outcomes, such as higher mortality and readmission rates (e.g., Kc and Terwiesch 2009, Kc and Terwiesch 2012, and Needleman et al. 2011). Such negative effects have been explained by the mental strain of healthcare workers (Kuntz et al. 2015) and the delay in treatment received by patients (Chalfin et al. 2007).

In this paper, we empirically investigate the impact of workload on service time and quality in the context of cardiac surgery. Specifically, we focus on the daily workload of individual surgeons, e.g., the number of operations performed by the focal surgeon on a given day. In most of the existing literature in healthcare settings, workload is measured on a system level, typically as bed occupancy in different hospital units at the time of patient's admission (e.g., Kc and Terwiesch 2012, Kuntz et al. 2015, Kim et al. 2015, and Berry Jaeker and Tucker 2017). A possible reason is that workload data at the individual level is often more difficult to obtain. Different from these works, our study considers an alternative type of workload; namely we measure the workload for *individual* surgeons on each day. In the context of cardiac operations, surgeons often have high ownership of their patients, thus the workload at individual level would be a more relevant measure than that at the system level (e.g., aggregated across the entire cardiac department). In addition, the effect of workload at individual level may be subject to more behavioral and operational variations. To the best of our knowledge, we are the first to study the impact of individual surgeon's daily workload in the field of operations management. The work of Kc (2014) also uses operational data at the individual level. However, Kc (2014) studies how multitasking (caring for multiple patients at the same time) of ED physicians affects service time and outcomes, i.e., the impact of processor-sharing. This is a very different workflow from our work on the impact of surgeon's daily workload, where patients are served in sequence. Note that our work is in contrast to the literature on surgical volume over much longer time horizons. In these works, increased volume has been associated with improved outcomes due to more surgery experience (e.g. Kc and Staats 2012).

It is common for a cardiac surgeon to perform multiple operations a day. In our study hospital, the median surgeon daily workload is two operations, and the maximum is four operations. On average, each operation takes more than seven hours to complete. Although some parts of the operation can be done by other members of the medical team, performing multiple operations a day is a heavy physical and cognitive burden for the surgeon. With long working hours, surgeons can suffer from physical and mental fatigue, which may lead to worse medical outcomes (Janhofer et al. 2019). In addition, high surgeon workload may strain operational resources. For example, there may be less experienced staff for operations performed in the late evening, and the bed flow could be slower outside of the normal working hours if auxiliary resources are not available. In this study, our goal is to understand the effect of surgeon workload, which is important for hospitals to improve their surgical and system performance. Due to data limitations (e.g., lack of shift schedule of ancillary medical staff and nurses), we do not aim to fully differentiate the factors leading to such effect, e.g., surgeon fatigue versus operational constraints.

In this paper, we examine the impact of surgeons' daily workload using a data set of cardiac surgery from a large academic medical center. Our data comes from the Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database for our partner hospital and contains detailed information of more than 5,600 cardiac operations that are performed over a horizon of four years. We measure the impact of surgeon daily workload – at the individual level – on multiple outcomes. First, we examine how surgeon daily workload affects the surgery duration of each case, as measured by its incision time. This sheds light on the relationship between workload and service time in the context of cardiac surgery. Next, we analyze the effects of surgeon daily workload on the patient's post-surgery length-of-stay (LOS) in the ICU and in the hospital. The post-surgery LOS is important for the hospital as it affects the demand for downstream resources (e.g., ICU and ward beds) and overall throughput efficiency. Finally, we check the impact of surgeon workload on the likelihood of adverse post-operation events for their patients, including reoperation, readmission, and mortality. We also examine whether the treatment effect of surgeon workload is heterogeneous for different types of patients. For example, urgent and emergent patients are generally more severe than the elective patients, thus their surgical outcomes may be more sensitive to surgeon daily workload.

Our detailed data set allows us to control for a variety of demographic, risk, and operative factors that may also affect the surgical outcomes. However, we still face a major challenge in identifying the true effect of surgeon daily workload. That is, the surgeon daily workload is endogenous. This is because there are likely risk factors that are considered by the surgeons when they schedule their cases, but these factors are not observable in the data (e.g., patient's cognitive status). These unobservable factors will affect both the surgeon daily workload and the surgical outcomes, thus violating the exogeneity condition for identification. For example, a surgeon may schedule more cases by packing in low risk, "easy" cases. If these measures of low risk are unobservable in the data, this will generate a negative bias in the estimated causal effect of surgeon daily workload. We handle the endogeneity bias by utilizing an instrumental variable (IV) approach. The IV method has been widely used in healthcare operations management for patient admission decisions (e.g., Kc and Terwiesch 2011, Kc and Terwiesch 2012, and Kim et al. 2015). We now apply it in the context of cardiac surgery to control for the endogeneity in surgeon's daily workload.

A valid IV in our study should influence the surgical outcomes only via the surgeon daily workload. We construct two IVs by leveraging operational factors in cardiac surgery. The first IV is the number of cases performed by other cardiac surgeons on the same day. As many resources (e.g. staff, operating rooms, ICU beds, etc.) are shared by surgeons in the cardiac department, more operations performed by *other* surgeons tends to limit the daily workload of the focal surgeon. We then construct another IV using a second data set of novel operational data: the block schedule of surgeons. Specifically, we define the second IV as the number of days until the next scheduled block of the focal surgeon. This IV is based on the following surgeon behavior: the surgeon may "squeeze in" more cases if his or her next scheduled block is far away. We validate the two IVs empirically with our data and show they indeed significantly impact surgeon's daily workload, after a comprehensive set of demographic, risk, and operational factors (e.g., block status and cardiac patient census) are controlled. In addition, we find the two IVs are essential for correctly estimating the effect of surgeon daily workload.

We find that higher daily workload for a surgeon is associated with longer incision time of the operation and worse outcomes for the patients. Specifically, adding one more case to a surgeon's daily workload increases the incision time by 24 minutes for each case performed by the surgeon in the day. This is an 8% relative increase. As performing multiple operations on a single day is a demanding task, our result on incision time shows how workload affects service time when the workload level is already high, and is consistent with the second tipping point empirically observed in Berry Jaeker and Tucker (2017). Surgeon daily workload leads to longer post-surgery LOS of patients in both the ICU and the hospital: when the surgeon needs to do one more case in a day, the affected patients are expected to stay in the ICU (resp. hospital) for 1.07 (resp. 1.40) more days after their operations. In addition, we find higher daily workload increases the patient's likelihood of reoperation and readmission. These consistent results highlight the negative impacts of high surgeon daily workload.

We further show that there is substantial heterogeneity in the effect of daily workload for elective and non-elective patients. The non-elective patients refer to those categorized as urgent, emergent, and salvage status. We find the effect of surgeon daily workload on incision time is statistically significant for the elective patients, but not for the non-elective ones. On the contrary, the effects on post-surgery LOS (in both ICU and hospital), reoperation, and readmission are significant only for non-elective patients. One possible explanation for such heterogeneity in treatment effect is that the operations for non-elective patients are more time-constrained, thus their incision time is less impacted by surgeon's workload. On the other hand, the non-elective patients are generally more severe, thus their surgical outcomes (post-LOS, reoperation, and readmission) are more sensitive to surgeon workload.

Based on the empirical results, we develop a surgery scheduling model that incorporates the effect of surgeon daily workload. Operating rooms are expensive medical resources and generate up to a half of hospital's revenues (McDermott et al. 2017); accordingly, the literature on surgical scheduling is large (see, e.g., Keskinocak and Savva 2020). In most of the existing literature, the surgery duration and patient outcomes are assumed to be exogenous with deterministic or stochastic distributions. However, as shown by our study, they can endogenously depend on surgeon daily workload. We thus propose a scheduling model that accounts for such effect. We consider the intervention of changing the dates of the operations in order to mitigate the negative impact of high daily workload for a surgeon. This intervention does not require any expansion of the OR capacity, and, thus, is relatively easy to implement. In our model, the objective is to minimize the total expected incision time, post-LOS, or ICU time of the patients in the sample. We formulate and solve the model as a mixed-integer quadratic programming problem and show that the improvement from the model can indeed be impactful. Using our estimated effects, we find the new schedules from our model can reduce the total post-LOS and ICU time by up to 5.34% and 8.34% respectively, which are economically substantial for the hospital. This highlights the benefits of accounting for the impact of surgeon daily workload in surgery scheduling. Moreover, in contrast to much of the literature of workload in service systems, where the primary lever to address the issue is through procuring more resources, our work demonstrates the benefit of load balancing at the *individual* level.

In summary, we make the following contributions in this paper.

- **Impact of surgeon daily workload:** We empirically estimate the causal impact of individual surgeon daily workload on surgery duration and patient outcomes using a detailed data set of cardiac surgery. We find surgeon workload increases surgery duration and leads to worse patient outcomes (post-surgery LOS, reoperation, and readmission). The effects are highly heterogeneous for different patient outcomes and different types of patients. Our result provides a potential mechanism for hospitals to focus on in order to improve their surgery operations and patient outcomes.
- **Estimation methodology:** To address the endogeneity bias in surgeon daily workload, we propose two novel IVs using operational factors in the cardiac surgery department. The first IV is based on the resource sharing by surgeons in the department. The second IV leverages the surgeons' block schedule data to capture the "squeeze-in" behavior that affects the surgeon workload. We validate the two IVs empirically and show they are essential for estimating the effect of surgeon workload without bias.

- **Surgery scheduling:** Our findings suggest that surgeon daily workload can substantially affect the surgery duration and patient outcomes. However, such impacts are largely ignored in the previous literature. Thus, we develop a surgery scheduling model that incorporates the impacts of a surgeon's daily workload. Using the estimated effects, we show that accounting for such effects in surgery scheduling can substantially reduce the total incision time and patient's LOS.

The rest of the paper is organized as follows. The next section is a brief overview of related literature. Section 2 describes the data and clinical setting of our study. In Section 3, we develop the econometric framework and estimation methodology. Section 4 provides the main empirical findings. We discuss our surgery scheduling model in Section 5. Section 6 concludes the paper and discusses future directions. The online supplement includes variable definitions, model formulation details, and supplementary tables.

1.1. Literature Review

Our study is related to four streams of literature: (1) the effect of system workload on service rate and quality, (2) volume-outcome relationship, (3) the impact of surgeon fatigue, and (4) operating room scheduling.

While traditional models usually assume a constant and exogenous service rate, there is rich literature, both analytical and empirical, that focuses on the relationship between system workload and service rate. The dynamic queueing control literature has derived optimal service rates that balance the costs of acceleration and waiting time (e.g., George and Harrison 2001). In reality, such optimal policies are not always feasible for human workers. To examine how human workers actually behave under varying workload, various empirical research has been conducted using observational data in real-world settings, and the results are mixed. Kc and Terwiesch (2009) show that workers for patient transport and cardiac surgery increase their service rate under high workload. Kc and Terwiesch (2012) and Chan et al. (2012) find hospitals are likely to discharge patients early when ICU occupancy is high, i.e., decreasing the service time. The opposite direction of the impact is also observed empirically. For example, Green and Nguyen (2001) show patient's LOS can increase when patient load becomes higher. The seemingly opposite effects of workload can be partially reconciled by an inverted-U shape pattern between service time and workload. That is, the service time first increases and then decreases with the workload level. Empirical evidence for this inverted-U shape pattern is found using restaurant chain data in Tan and Netessine (2014), and in the healthcare setting in Batt and Terwiesch (2016) and Berry Jaeker and Tucker (2017). Different mechanisms have been proposed to explain the impact of workload on service time, such as server speedup (Kc and Terwiesch 2009, Tan and Netessine 2014), task reduction (Oliva and Sterman 2001, Kuntz et al. 2015), multitasking (Tan and Netessine 2014, Freeman et al. 2017), and server fatigue (Kuntz et al. 2015, Berry Jaeker and Tucker 2017). On the analytical side, Delasay et al. (2016) develop a state-dependent queueing model to capture the adaptive mechanisms leading to the nonlinear pattern between service rate and workload.

There is also a rich literature studying the effect of workload on servers' behavior and quality. Hopp et al. (2007) use an analytical queueing model to show increasing servers may worsen congestion when servers have discretion over task completion time. Freeman et al. (2017) find that gatekeeper-providers would alter their service configuration and referral decisions in response to their workload. In multiple healthcare settings, the quality of care is found to suffer under high workload, such as higher mortality and readmission rate (Kc and Terwiesch 2009, Kuntz et al. 2015, and Berry Jaeker and Tucker 2017), as well as longer LOS and higher likelihood of transfer-up (Kim et al. 2015). The positive linkage between hospital workload and mortality is also observed in the medical literature (e.g., Neuraz et al. 2015). Our study contributes to this line of literature by considering a novel type of workload in healthcare setting, i.e., number of operations performed in a day by the focal surgeon. We find a surgeon's high workload is associated with longer surgery duration and worse patient outcomes, providing consistent evidence for the negative impact of very high workload level. We also show such effects are highly heterogeneous across different patients and outcomes.

Next, our work is related to the literature on volume-outcome relationship in healthcare management. In the medical community, there is vast evidence supporting a positive relationship between a surgeon's (or a hospital's) volume and surgical outcomes (see, e.g., Bashir et al. 2017). The volume in these studies usually refers to the number of operations performed by the surgeon in a relatively long period (e.g., the past one year). The volume-outcome relationship has also drawn attention in the field of operations management. Research in different empirical settings has been conducted to investigate the driver and mechanism behind the relationship, e.g., learning and specialization. For example, Kc and Terwiesch (2011) show that after controlling for selective patient admissions, the benefit of specialization disappears at the hospital level, but it still exists at the operating unit level in terms of a shorter patient's LOS. Kc and Staats (2012) find a surgeon's focal experience has a greater impact on surgical outcome than related experience. Clark and Huckman (2012) identify the existence of complementarities resulting from cospecialization in focal and related segments. Staats and Gino (2012) disentangle the role of specialization and variety in improving worker's productivity. Recent work by Wang and Pourghannad (2020) shows the effects of surgical volume on surgery duration is heterogeneous across patients. Complementing this line of research, we investigate the impact of a surgeon's short-term volume, i.e., number of cases performed in a day, on surgery duration and surgical outcomes.

Our work also relates and contributes to the literature on surgeon fatigue. As the work of a surgeon is highly demanding both physically and mentally, the potential negative impact of surgeon fatigue has long been a focus of the medical community (see a recent survey in Janhofer et al. 2019). Under different medical settings, multiple studies have shown surgeon fatigue is associated with worse surgical outcomes (Shanafelt et al. 2010, Thomas et al. 2012). However, other studies have found no significant impact from surgeon fatigue on surgical outcomes (e.g., Govindarajan et al. 2015). In general, the medical literature does not have a clear conclusion on the relation between surgeon fatigue and worse patient outcomes. Our work sheds light on this important problem using a detailed empirical data set of cardiac surgery and rigorous econometric analysis. We acknowledge that there may be other factors (e.g., limited resources in the evening) that contribute and/or explain the effects of surgeon workload; however, surgeon fatigue is one plausible explanation for our findings. As an important difference from existing medical literature, which focuses primarily on correlations rather than trying to tease out causal effects, we use IVs to control for the endogeneity in surgeon daily workload and generate causal estimates. This accounts for the possibility that surgeons will schedule less severe cases when they know their workload is high. Ignoring such endogeneity may make it difficult, or even impossible, to identify the true effect.

We also contribute to the literature of operating room scheduling. Operating rooms are big cost centers and revenue generators of the hospital. The literature on operating room scheduling is huge. Some review of the current research, challenges, and future directions of this field can be found in May et al. (2011) and Samudra et al. (2016) among many others. Different objectives are considered in operating room scheduling, including minimizing costs, maximizing profit and utilization, as well as smoothing downstream census (e.g., Gupta 2007, Denton et al. 2010, Freeman et al. 2016, and Zenteno et al. 2016). Combinations of these objectives are also considered (e.g., Gul et al. 2011). In addition, staff planning in the operating room environment is also widely studied (Rath and Rajaram 2021). From a different aspect, Olivares et al. (2008) apply a structural estimation method to identify how the hospital actually balances the overage and underage costs when reserving operating room capacity. However, most of the existing literature assumes the surgery duration, either deterministic or stochastic, to be exogenous and independent of surgeon workload. To the best of our knowledge, we are the first to develop a scheduling model that incorporates the effects of surgeon daily workload. A recent example of endogenous surgery duration is the work by Wang and Pourghannad (2020), in which the surgery duration is affected by the attending surgeon's past volume. Our work differs from theirs in several important aspects. First, we focus on the impact of surgeon daily workload, and apply an IV method to address the endogeneity bias. Next, our model considers the assignment of operations to available days, which is different from their decisions of matching patients and surgeons. In addition, we examine clinical outcomes in addition to surgery duration, including patient post-LOS and ICU time.

2. Data and Clinical Setting

In this study, we use cardiac surgery data from a large academic hospital over the period of July 2015 to July 2019. The data comes from two sources. The first one is the cardiac surgery data collected from the Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database.[1] The STS data contains detailed information of patient demographics, risk factors, preoperative status, operative procedures and timelines, as well as postoperative events for all cardiac operations occurred in the sample period in our partner hospital. This comprehensive data set allows us to control for the severity of patients and complexity of operations when

[1] https://www.sts.org/registries-research-center/sts-national-database/adult-cardiac-surgery-database/data-collection

analyzing the impact of daily workload. The STS data contains basic patient demographics such as gender, age, and race. It also includes risk factors such as a patient's status for liver illness, lung disease, diabetes control, and renal failure. The preoperative section records whether the patient experienced heart failure, cardiogenic shock, or myocardial infarction (MI) before the operation. We also obtain from the STS data each patient's hospital admission date, surgery date, and discharge date. Thus, we can compute the patient's LOS before and after the operation. For each operation, we can determine its operating room (OR) time and incision time from the STS data using the timestamps of its OR entry and exit, as well as skin incision start and end. We also have the surgeon's identifier for each case, which enables us to control for surgeon-specific differences in outcomes (e.g., Wang and Pourghannad 2020). Finally, the STS data contains multiple outcomes of each operation, such as the time spent in ICU, reoperation, readmission, and mortality.

The second data source is the block schedule of surgeons provided by the cardiac department of our partner hospital. The hospital employs a block booking framework to schedule their cardiac surgery. Under the block booking, surgeons are assigned with fixed time slots (blocks) and dedicated resources (e.g., OR and staff) to perform their operations. The block schedule is decided in advance by the management board and adjusted infrequently, e.g., each quarter or twice a year. The block scheduling framework is widely accepted in the US as it is convenient for surgeons and hospitals to employ (Erdogan and Denton 2010). Each block in our data specifies the date, OR number, and the surgeon assigned – e.g., OR 1 is assigned to Surgeon A on October 1st, 2016. We note that in most cases in our study, each OR is assigned to only one surgeon for the entire day. That is, OR sharing by multiple surgeons is rare (only 3%). There are in total eight ORs for cardiac operations. However, some blocks of these ORs are assigned to other departments in the hospital (e.g., the pediatrics department). This is also documented in the block schedule data.

In principle, the block schedule data allows us to determine for each operation whether it happens in or out of the block schedule. Here "in block schedule" means the operation is performed in an OR that is assigned to the surgeon on the operation date. However, the data has two limitations. First, a significant proportion of the block schedule data is missing: out of the 48 months (resp. 5,604 cases) in our surgery data, we only have the block schedule data for 22 months (resp. 2,499 cases). Thus, we would need to impute the block schedule information for the missing periods. We provide the details of the imputation in Online Supplement S.2. We emphasize that the missing block data is due to the absence of administrative staff in the department (e.g., personal leave), thus it does not introduce any selection of patients or operations. Second, we do not have the location information (the OR number) for each operation in our data. Because of this limitation, we can only determine the block status on the surgeon-day level, i.e., whether a surgeon is assigned a block in one of the ORs on a specific day. We then use this block status for all the cases performed by the surgeon on that day. We verify with our partner hospital that this is a reasonable assumption.

2.1. Data Selection

In this section, we describe the data cleaning process and provide some summary statistics of the final sample in our study. We start from 5,604 cases from the STS data. We first drop 20 cases that are cancelled before or during the operation. We then drop 232 cases from seven "infrequent" surgeons in our sample. These surgeons performed a very small number of cases during the four years in the sample period. They are dropped for the following two reasons. First, these surgeons are more likely to only perform unusual procedures that require special expertise. Second, the small sample size of these surgeons does not allow us to effectively control for surgeon fixed effects. Thus, we focus on the cases from the other eight surgeons, each of which performed at least 200 cases in the sample period. This leaves us with a sample of 5,352 cases in total, which consists of 95.5% of the initial sample.

In the final full sample of our study, we have the block schedule information for 2,492 of the cases (46.5%). We refer to these cases as the block sample. As we explained above, we can determine the block status on the surgeon-day level. In the 1,744 surgeon-day pairs with block information, 1,343 (77%) of them are in block schedule, i.e., the surgeon has a block assignment on that day, while the remaining 401 (23%) pairs happen out of schedule. Given a surgeon operates in a day, on average, the surgeon performs 1.4 (resp. 1.2) cases a day if he or she is in (resp. out of) block schedule. In total, we have 2,010 cases classified as in block schedule, 482 as out of block schedule, and 2,860 cases of which we do not have block schedule information and use imputed block information in our analysis. We run robustness checks on the block sample for the cases with complete block schedule information only, as discussed in Section 4.3.

2.2. Patient Risk Factors

The STS data set provides a comprehensive set of risk factors of each patient, which allows us to control for the patient's severity and condition. Table 1 reports the summary statistics of patients' gender, age, and critical status for both the full sample and the block sample. Specifically, a patient is classified as critical if he or she experiences a cardiogenic shock or syncope before the operation. In Online Supplement S.1, we provide a detailed description of other factors in our econometric framework and their summary statistics.

Table 1 Summary Statistics of Patients for the Full Sample and Block Sample (Full Sample: N = 5,352, Block Sample: N = 2,492)

	Full Sample			Block Sample		
	Mean	Median	Std	Mean	Median	Std
Gender: Male	0.675	-	-	0.671	-	-
Age	64.73	66.00	12.56	65.06	66.00	12.33
Critical	0.103	-	-	0.102	-	-

The cardiac operations are divided into four main risk categories (surgery status) in increasing order of patient severity and urgency in need of operation: elective, urgent, emergent, and salvage. The elective cases are those operations that can be deferred without increased risk; the urgent cases are supposed to be performed during the same clinical stay to reduce further risk; the emergent and salvage cases refer to the situation that requires emergent operations with no delay upon the outbreak.[2] The surgery status has important implications on the surgical scheduling. While the hospital has relatively high flexibility in scheduling the elective cases, the schedules of urgent cases are more difficult to change, and the hospital has little control over the timing of emergent and salvage cases. In our data, a significant proportion of the operations are urgent or emergent cases, which consists of 53.5% of the full sample and 52.3% of the block sample. Indeed, the numbers of elective and urgent cases are almost the same (2,479 versus 2,488 for the full sample, and 1,184 versus 1,124 for the block sample). Moreover, the distribution of the surgery status are very similar for the full and block samples.

In addition to the surgery status, we also obtain the procedure information for each case from the STS data, i.e., which types of procedures are performed during the operation. To control for differences across different types of procedures, we classify the operations to different types as follows. First, there are eight standard types of the most commonly performed cardiac operations. For those cases, we directly use the classification provided by the STS data (e.g., coronary artery bypass graft, aortic valve replacement, etc.). For the non-standard procedures which do not fall in one of the eight types above, we define their surgery types based on the actual procedures performed. The detailed definitions of the non-standard types are provided in Online Supplement S.1. In total, we have 14 procedure types for the cases in our data. There are 3,420 cases (63.9%) that belong to one of the eight standard types, and 1,932 cases (36.1%) that belong to one of the six non-standard types.

We compute the pre-surgery LOS (pre-LOS) for each patient as the number of days between the hospital admission and the operation, and the post-surgery LOS (post-LOS) as that between the OR exit and hospital discharge. The OR time of each case is calculated as the time elapsed between its OR entry and OR exit. The OR time can be decomposed to three stages: pre-incision time, incision time, and post-incision time. The incision stage corresponds to the time between skin incision start and end, and the pre-incision (resp. post-incision) stage refers to the time before (resp. after) it. Different tasks are performed in the three stages. The pre-incision stage includes pre-operative tests, positioning the patient in OR, and anesthesia. The post-incision stage includes closing the incision and cleaning up. In cardiac operations, these tasks can be largely performed by medical staff or surgical fellows without the presence of the focal surgeon. On the other hand, the incision stage requires relatively high level of participation of the surgeon. Thus, the incision time is a more accurate measure for a surgeon's working time than the total OR time.

We present the summary statistics for the pre-incision measures including pre-LOS and pre-incision time by the four surgery status in Table 2. We can see that the elective cases have relatively short pre-LOS. This is

[2] See page 154 in the training manual: https://www.sts.org/sites/default/files/Training%20Manual%20V2-9%20June%202020.pdf

because most of the elective patients are admitted one day before or on the same day of their operation. We also see a fairly consistent pre-incision time spent in the OR across elective, urgent, and emergent patients. This is likely due to the fact that, for these patients, pre-incision time is very protocol-driven where the patient goes through standard preparation before the surgeon actually cuts the patient. On the other hand, the pre-incision time for salvage patients is shorter. Although the sample size is very small, this may be indicative of the highly time-sensitive nature of these procedures.

Table 2 Summary Statistics of Pre-incision metrics by Surgery Status

Status	Num Obs.	Pre-surgery LOS (days)	Pre-incision time (hours)
Elective	2479	1.16 (2.92)	1.48 (0.28)
Urgent	2488	4.90 (9.83)	1.51 (0.31)
Emergent	374	15.32 (30.11)	1.48 (0.45)
Salvage	11	8.09 (7.54)	1.18 (0.51)
All	5352	3.90 (11.19)	1.49 (0.31)

2.3. Surgery Metrics and Patient Outcomes

We now report the summary statistics for the surgical metrics and outcomes in Table 3. First, we see that the average incision and OR time are longer for the urgent and emergent cases than that for the elective cases. This is not surprising as the non-elective cases tend to be more complicated and thus take longer time to perform. On average, the incision stage consists of 67% of the total OR time. Second, the elective cases have the shortest post-LOS, while the emergent cases have the longest. This reflects the fact that the patients of the elective cases are generally less severe than those of the urgent and emergent cases and follow typical post-surgery protocols. Third, the total time in ICU accounts for both the initial ICU visit and the potential revisits and is longer for the more complex cases. For patient outcomes, we focus on three binary ones: reoperation, readmission to the hospital, and mortality. Reoperation includes all causes for a patient to return to OR: bleeding, valve dysfunction, MI, aortic disease, and other cardiac and non-cardiac reasons. However, we exclude reoperations within 24 hours after operation due to acute bleeding because these acute reoperations increase the surgeon's workload on the same day, but they are not documented in the STS data. (This includes 108 cases, which account for 2% of our sample.) Mortality is defined as death in 30 days after the operation regardless of the location (e.g., in hospital or at home). Not surprisingly, we see that the urgent cases on average are associated with worse outcomes than the elective ones, and the emergent cases have the worst average outcomes among the three categories.

3. Econometric Framework

We now develop the econometric framework for identifying the effect of daily workload on surgery duration and outcomes. For each case i, denote its surgeon and surgery date by s and t respectively. $Workload_{ist}$

Table 3 Summary Statistics of Surgery Metric and Patient Outcomes

	Elective	Urgent	Emergent	Salvage	All
OR time (hours)	6.79 (1.79)	7.24 (1.99)	8.31 (2.59)	8.28 (2.60)	7.11 (1.99)
Incision time (hours)	4.52 (1.56)	4.88 (1.76)	5.80 (2.31)	5.96 (2.27)	4.78 (1.75)
Post-LOS (days)	8.73 (7.85)	13.22 (18.53)	25.60 (22.27)	20.45 (12.15)	12.02 (15.55)
Total ICU (days)	3.61 (5.55)	5.88 (11.36)	13.29 (17.35)	15.59 (13.03)	5.37 (10.09)
Reoperation	0.0347	0.076	0.251	0.364	0.070
Readmission	0.0891	0.104	0.118	0.200	0.098
Mortality	0.0157	0.031	0.061	0.455	0.027
Number	2479	2488	374	11	5352

is the daily workload of case i's surgeon s on day t. Since each case i maps to a unique pair of surgeon s and day t, we use $Workload_i$ and remove the subscripts s and t for all variables for brevity. We consider two measures for the surgeon daily workload. The first measure $NumCases_i$ is the total number of cases excluding i performed by surgeon s on day t. The second measure $SumInc_i$ represents the total incision time of cases excluding i by surgeon s on day t. Note that if the surgeon performs only one case in a day, we would have $NumCases_i$ and $SumInc_i$ equal to zero by their definitions.

The summary statistics of the two daily workload measures are reported in Table 4 below. We see that it is very common for a surgeon to perform multiple cases on the same day: the median of $NumCases_i$ is one for both the full sample and the block sample. That is, for at least half of the cases in our sample, their surgeons perform at least two cases in total on their surgery date. This is also reflected by the average of $SumInc_i$, which is 3.27 (resp 3.14) hours for the full (resp. block) sample. The two daily workload measures are highly correlated. The correlation is 0.91 and 0.92 for the full and block samples, respectively. We choose $NumCases_i$ as the workload measure in the main specification because it is easier to interpret the results. We also conduct the full set of analyses described below using $SumInc_i$ as the workload measure. The results are qualitatively similar and are provided in Online Supplement S.3.

Table 4 Summary Statistics of Daily Workload for Full and Block Sample (Full Sample: N = 5,352, Block Sample: N = 2,492)

	Full Sample			Block Sample		
Workload	Mean	Median	Std	Mean	Median	Std
$NumCases_i$	0.69	1.00	0.69	0.66	1.00	0.65
$SumInc_i$	3.27	3.17	3.58	3.14	3.10	3.38

We control for a variety of demographic, medical, and operative factors as explanatory variables in our estimation as described in Section 2.2. We also include six operational variables: dummy variables for the day of the week, month, and year of the operation, the pre-LOS of patient, the block schedule status, and the cardiac patient census in the hospital. The block schedule status of a case has three possible categories: "in-schedule" if the surgeon has a block assignment on that day (37.6%), "out-of-schedule" if the surgeon does not have a block assignment (9.0%), and "unknown" if the block information is missing for that day (53.4%). We include the block status in our estimation as it may impact the surgical duration and outcome. For example, if a surgeon performs an operation out of her block assignment, the surgeon may have to work with less familiar nurses or medical staff, which can negatively affect the surgical outcomes.

We compute the *cardiac patient census* for each day t as the number of patients in our sample that have been admitted before day t, but have not been discharged from the hospital by day t. We use the cardiac patient census as an approximation for the system congestion level, which has been shown to affect healthcare outcomes (e.g., Kim et al. 2015). We note that the cardiac department in our study operates in a relatively independent manner. Thus, the cardiac patient census serves as a good measure for the congestion level faced by the cardiac department. Specifically, the Cardiac Surgery ICU is not available to non-cardiac surgery patients. As most of the patients (99.5%) go to the Cardiac Surgery ICU after their operations, we expect the cardiac patient census to be highly correlated with the ICU congestion. On the other hand, we can not accurately measure the ICU occupancy level because our data only contains the total time spent in the ICU during the hospitalization, but does not differentiate between initial or subsequent visits. The cardiac patient census has a mean of 62.3 with a standard deviation of 9.3. As an additional check, we also construct a cardiac surgery ICU census using the patients in our sample by assuming they spend their total ICU time in their first ICU visit after OR exit. We find adding this cardiac surgery ICU census does not affect our estimation results.

In total, we have 34 independent variable in our estimation. We provide a detailed description of these independent variables used in Online Supplement S.1. We represent these variables and a constant by X_i for case i. To estimate the effect of daily workload, we consider the following regression models. Let the dependent variable y_i be one of the surgical outcomes or patient outcomes described in Section 2.3. For continuous y_i, we employ the linear model:

$$y_i = X_i\beta + \gamma Workload_i + \varepsilon_i, \tag{1}$$

where ε_i is the error term. For binary y_i, we use the following Probit model:

$$\begin{aligned} y_i^* &= X_i\beta + \gamma Workload_i + \varepsilon_i, \\ y_i &= \mathbf{1}\{y_i^* > 0\}, \end{aligned} \tag{2}$$

where y_i^* is a latent variable, $\mathbf{1}\{\}$ is the indicator function, and the error term ε_i follows a normal distribution. We use equations (1) and (2) to estimate γ, which is the effect of daily workload on y_i averaged across all cases by the surgeon in a day.

As a naive approach, we can estimate the coefficients in (1) and (2) by ordinary least squares (OLS) or maximum likelihood estimation (MLE) and interpret the estimated γ as the effect of daily workload on the dependent variable y_i. However, this approach ignores the endogeneity in the daily workload of surgeons. That is, the surgeon daily workload can be affected by patients' severity factors that are unobserved in the data but are considered by the surgeon (e.g., a patient's cognitive state). For example, the surgeon may schedule more cases in a day if the unobserved severity factors are lower and imply shorter incision times. Consequently, both the dependent variable (e.g., incision time) and the daily workload (e.g., number of other cases) are affected by regressor X_i as well as the unobserved severity factors. If we ignore this endogeneity problem, the estimated coefficients will be biased. In the example described above, ignoring the unobserved severity factor introduces a negative bias to the estimate of γ, as the unobserved severity level is negatively correlated with the daily workload and positively correlated with incision time. Thus, using OLS to estimate (1) may yield a negative γ even if the true effect is positive. To address the endogeneity bias, we employ the instrumental variable (IV) method to obtain consistent estimates of the coefficients. We construct two IVs using the operational data from the cardiac department. We introduce the two IVs and explain their validity in the next section.

We acknowledge the effect of workload in (1) and (2) is likely to vary across cases, e.g., the first and last case by the surgeon in a day. However, as we do not have the information on how cases are scheduled within a day, we cannot effectively control for the potential endogeneity in the sequencing of cases. Thus, we choose to estimate the effect averaged across all cases performed by the surgeon in a day. As a potential future direction, it would be practically important to estimate the effect of surgeon daily workload for individual cases. This requires more data on the scheduling and sequencing of operations within a day.

3.1. Instrumental Variables

We propose two novel IVs using operational data and demonstrate their validity in this section. The two IVs are both computed on the surgeon-day level, i.e., they are the same across the cases performed by a surgeon on the same day. This is inline with our estimation of the average effect in (1) and (2) for all cases by the surgeon in a day.

3.1.1. Total Cases by *Other* Surgeons The first IV we consider is $TotOther_i$, the total number of cases performed by other surgeons on the same day as case i. The relevance condition is satisfied for this IV because the number cases by other surgeons can affect the daily workload of the focal surgeon through *resource sharing* across surgeons. Although the surgeons' schedules are fixed well in advance in terms of the day of the week they are assigned an OR, the exact number of cases they will perform on a given day are usually not finalized until shortly before the day starts. In addition, while the hospital uses a block booking system, the surgeons still share the same pool of resources such as anesthesiologists, medications, and equipment such as ventilators, etc. Thus, more cases performed by other surgeons in a given day mean there are fewer resources available for the focal surgeon and this tends to limit the workload of the focal surgeon on the same day. This resource sharing phenomenon may be why such a large focus of the literature on workload has been at the system/unit level rather than at the individual level as we study. We aim to pick up such variation using the IV. For above reasons, we expect $TotOther_i$ to be negatively correlated with the focal surgeon daily workload. To test the relevance condition, we conduct a simple linear regression of $NumCases_i$ ($SumInc_i$) on $TotOther_i$, while controlling for patient demographic, medical risk, operative, and operational covariates, as well as a second IV introduced in next section. In particular, the cardiac patient census is included to control for the system's congestion level. We find that the coefficient for $TotOther_i$ is -0.078 (-0.385) for $NumCases_i$ ($SumInc_i$) with p-values smaller than 0.0001 (for both). Hence, we find that the total number of cases by other surgeons does, indeed, explain variation in the daily workload of the focal surgeon.

We next consider whether $TotOther_i$ likely satisfies the exclusion restriction. The surgeons in the cardiac department at our study hospital have substantial ownership of their patients and schedules. They rarely coordinate with other surgeons beyond whether there is OR time when scheduling their own cases, and it is entirely the discretion of the focal surgeon which operations to prioritize amongst his/her own patients. That is, an individual surgeon has little control over other surgeons' patients and scheduling. This suggests that the workload of other surgeons should not be directly correlated with the unobserved severity factors of the focal surgeon's patients. We also check whether the IV is correlated with the *observed* measures of severity. Table 17 in Online Supplement S.5 summarizes the correlation between the IV and 21 observed severity factors. The magnitude of all correlations is smaller than 0.1, and the average (absolute) correlation is 0.022. Thus, we conclude that the IV and observable severity factors are unlikely to be correlated.

3.1.2. Gap to Next Block Schedule We construct a second IV using the block schedule data. We consider the following behavioral observation: if a surgeon has to wait for a long period of time for the next scheduled block, the surgeon may be motivated to squeeze more cases into the current day. Thus, we expect the gap to next block (in days) of the focal surgeon, $GapNext_i$, to be positively correlated with the surgeon daily workload. In other words, $GapNext_i$ satisfies the relevance condition. Moreover, the block schedule of each surgeon is fixed in advance and adjusted very infrequently (twice a year). Thus, it is unlikely for $GapNext_i$ to be correlated with the unobserved severity factors of the surgeon's current patients. This suggests that the exogeneity requirement for $GapNext_i$ is likely to be satisfied.

For the periods with the block schedule data available, we can directly compute $GapNext_i$ for each case. However, the block schedule data is missing for a significant proportion of the sample horizon as described in Section 2.1. For the periods without the block schedule data, we construct $GapNext_i$ as follows. First, we impute the block schedule on the surgeon-day level using a logistic regression. Then, we calculate the *expected* $GapNext_i$ based on our imputation. This enables us to maintain the entire sample for estimation. Simply dropping the missing periods would reduce our sample size by 53%, though we consider a robustness check with this subsample in Section 4.3. The results are qualitatively similar.

We describe the details for the block imputation model and the calculation of $GapNext_i$ for the cases without block information in Online Supplement S.2. The imputation model fits the observed block schedule data well. The McFadden's R-squared of the estimated model is 0.31 and the Area Under Curve (AUC) from the model classification is 0.86. In addition, we perform an out-of-sample test using the first 80% (resp. last 20%) of the block sample as the training (resp. test) set. The imputation model leads to an out-of-sample AUC of 0.75 for the test set. We then compute the imputed $GapNext_i$ for the test set and compare it with the actual gap to next block. We find that the correlation between the imputed and actual $GapNext_i$ is 0.87 for the test set. We also conduct a linear regression for the actual gap with the imputed $GapNext_i$ (plus a constant) being the independent variable. The R-squared for this linear regression is 0.75 and the coefficient for the imputed $GapNext_i$ is 0.96 with a p-value smaller than 0.001. These out-of-sample tests further support the effectiveness of our imputation model for calculating $GapNext_i$.

In summary, we calculate $GapNext_i$ by the block schedule data for the cases with block information, or by the imputation model for the cases without. To test the relevance condition, we conduct a simple linear regression of $NumCases_i$ ($SumInc_i$) on $GapNext_i$, while controlling for the patient demographic, medical risk, operative, and operational covariates, as well as the first IV $TotOther_i$. We find that the coefficient for $GapNext_i$ is 0.005 (0.044) for $NumCases_i$ ($SumInc_i$) with a p-value 0.09 (0.005). (The higher statistical significance level of $GapNext_i$ for $SumInc_i$ can be explained as $SumInc_i$ has more variation than $NumCases_i$, which takes integer values only.) Hence, we find that the total number of cases by other surgeons does, indeed, explain the variation in the total workload of the focal surgeon. In addition, as shown by Table 17 in Online Supplement S.5, the correlation between $GapNext_i$ and the *observed* measures of severity is very weak. The largest (average) absolute correlation is 0.084 (0.036).

The summary statistics of the two IVs are shown in Table 5 for both the full sample and the block sample. We see that, in aggregate, the other surgeons on average perform 4.1 cases on the operation day of the focal surgeon and the average (resp. median) gap to next block schedule is 3.46 days (resp. 2.21 days) for the full sample. Moreover, the standard deviation of $GapNext_i$ is 3.04 days, reflecting significant variation. This is because the blocks of each surgeon are distributed unevenly across days. We also notice that the statistics of $GapNext_i$ is very similar for the full sample and the block sample. This further supports the effectiveness of our schedule imputation model as the distribution of $GapNext_i$ imputed by the model is close to that calculated directly from the block data.

3.2. Estimation Methods

We estimate the effect of daily workload in models (1) and (2) using the two IVs introduced above. We describe the estimation methods below. For continuous dependent variable y_i, we estimate the linear model

Table 5 Summary Statistics of the IVs for Full and Block Sample
(Full Sample: N = 5,352, Block Sample: N = 2,492)

IV	Full Sample Mean	Full Sample Median	Full Sample Std	Block Sample Mean	Block Sample Median	Block Sample Std
$TotOther_i$	4.11	4.00	1.78	4.18	4.00	1.77
$GapNext_i$	3.46	2.21	3.04	3.40	2.00	3.22

(1) using the two-stage least squares (TSLS) regression. The TSLS estimation is conducted as follows. In the first stage, we regress the daily workload on the exogenous variables X_i and the two IVs using OLS:

$$Workload_i = X_i\beta + \eta_1 TotOther_i + \eta_2 GapNext_i + \xi_i. \tag{3}$$

The first stage regression measures the impact of the two IVs on a surgeon's daily workload. For the two IVs to affect the daily workload (i.e., the relevance condition), at least one of η_1 and η_2 should be statistically different from zero. In the second stage, we replace $Workload_i$ in (1) with its fitted values from (3) and estimate γ by OLS. Note that the standard errors in the second stage need to be adjusted as we are plugging in estimates of $Workload_i$.

For binary dependent variable y_i, we use the full information maximum likelihood estimation method to estimate the effect γ in the probit model (2) (Woodridge 2010, Cameron and Trivedi 2013). Specifically, the models for the daily workload in (3) and the outcome in (2) are estimated jointly under the assumption that the error terms (ε_i, ξ_i) follow a bivariate normal distribution. To capture the endogeneity in daily workload, we allow ε_i and ξ_i to be correlated. Thus, there can be unobserved severity factors that affect both the surgical outcomes and the daily workload.

We find that the distributions of incision time, post-LOS, and total ICU time have long tails on the right end, thus we winsorize them by their 97.5th percentiles to mitigate the impact of extreme values. This corresponds to 9.2 hours for incision time, 50 days for post-LOS, and 29.7 days for total ICU time. These winsorization choices are quite conservative, and our estimation results are robust to other choices of winsorization levels. In addition, for both the linear and probit models in (1) and (2), we cluster the standard errors by the surgeon's identifier to account for the potential correlation across cases of the same surgeon.

4. Estimation Results

This section provides the main estimation results regarding the effects of surgeon daily workload on surgery duration and outcomes. We first provide the results estimated from the full sample in Section 4.1. Then we analyze the heterogeneity in the effects for elective and non-elective patients in Section 4.2. Here we show the results with surgeon's daily workload measured by $NumCases_i$. In Online Supplement S.3, we briefly discuss the results when we measure surgeon's workload by $SumInc_i$ (total incision time of other cases). The results from the two workload measures are largely consistent in both directions and magnitudes.

4.1. Effect of Daily Workload on Surgery Duration and Patient Outcomes

In Table 6, we report the estimated effects of surgeon daily workload, with $Workload_i$ measured by $NumCases_i$ in (1) and (2). The results using $SumInc_i$ as the daily workload measure are similar and are reported in Table 15 of Online Supplement S.3. Since the IV varies at the surgeon-day level, the estimated γ here captures the effect averaged across all the cases performed by the surgeon in a day. In Section 3.1, we have reported that the coefficients η_1 and η_2 in (3) of the two IVs are statistically significant with expected signs. In addition, a weak instrument test for the hypothesis $\eta_1 = \eta_2 = 0$ in (3) is strongly rejected at the significance level of 0.0001 with an F-statistic of 74.4. These evidence support the validity of the IVs for our estimation.

For the three continuous dependent variables, incision time, post-LOS, and total ICU time, we show the estimated γ in (1) and its standard errors. For the three binary dependent variables, reoperation, readmission, and mortality, we report the estimated average marginal effects (AME) of daily workload as they are easier to interpret. The estimated coefficient γ in (2) for the three binary outcomes are provided in Table 20 of Online Supplement S.5. The Panel A in Table 6 reports the estimation results from the TSLS and full MLE with the two IVs, as described in Section 3.2. For comparison, we also show in Panel B the results when we ignore the endogeneity problem using simple OLS to estimate (1) and MLE to estimate (2).

Table 6 Estimated Effects of Daily Workload (Number of Other Cases) on Surgery Duration and Patient Outcomes: Full Sample

	Continuous y_i: Coefficients			Binary y_i: AME		
	Incision time	Post-LOS	Total ICU time	Reoperation	Readmission	Mortality
Panel A: Full	$0.400^{\dagger}$ (0.209)	1.402^{*} (0.547)	1.074^{*} (0.437)	0.033^{**} (0.010)	$0.066^{\dagger}$ (0.036)	0.010 (0.010)
Num Obs.	5345	5344	5319	5345	5116	5081
Panel B: Full (no IV)	-0.101^{**} (0.039)	-0.052 (0.144)	-0.007 (0.083)	-0.002 (0.004)	-0.001 (0.009)	0.010^{*} (0.004)
Num Obs.	5345	5344	5319	5345	5116	5081

The estimated effects of surgeon daily workload (number of other cases) on surgery duration and patient outcomes for the full sample. We report the estimated coefficients in (1) for the three continuous dependent variables, and the AME from (2) for the three binary dependent variables. Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

First, we see from the "Incision time" column in Panel A of Table 6 that higher daily workload tends to increase the incision time of the cases performed by the focal surgeon. In particular, adding one more case increases the incision time of each case performed by the surgeon by 0.4 hours (24 minutes) on average. This translates to an 8.3% relative increase of the average incision time. The effect is statistically significant at the 10% level with a p-value of 0.056. On the other hand, if we ignore the endogeneity in daily workload and estimate the model by OLS, the effect becomes the opposite as shown in the "Incision time" column in Panel B: the coefficient is negative and statistically significant at the 1% level. This shows that it is essential to address the endogeneity in the daily workload as surgeons may schedule more cases if the unobserved severity factor implies shorter incision times, resulting in a negatively biased estimate of the effect.

A priori, it is not clear how a surgeon's daily workload may impact the incision time. First, surgeons may "speed up" the operations when they have more cases to perform in a day, leading to a shorter incision time. This type of speedup effect is found in, e.g., Kc and Terwiesch (2009). On the other hand, surgeons may take more time to complete their tasks due to fatigue and operational constraints associated with high daily workloads. After addressing the endogeneity bias with proper IVs, our empirical results here support the second mechanism, i.e., higher daily workload of surgeons leads to longer incision times.

A possible explanation for the mechanism is as follows. First, unlike patient transporting and patient LOS studied in Kc and Terwiesch (2009), cardiac operations are complex and delicate procedures, it is very difficult for surgeons to speed up in their operations. Second, given they are such demanding tasks, performing multiple cases a day could cause fatigue of surgeons, both physically and mentally, thus leading to longer incision time. In addition, when a surgeon needs to do multiple cases in a day, some of them may have to be performed outside of normal working hours. Thus, they are more likely to be associated with fewer medical staff and less experienced nurses, as well as constrained resources and delay in bed flow. For example, one "on-call" anesthesiologist may have to cover multiple operations performed in the night. For these reasons, the impact of surgeon fatigue and other operational constraints outweighs the potential channels for speedup, and causes longer incision time on average when more cases are performed. As we do not have the complete operational data in the cardiac department (e.g., shift schedule of nurses and other medical staff), we are not able to fully identify the factors that lead to the longer incision times beyond anecdotal evidence suggested by our clinical collaborators. That said, our results provide clear evidence for the negative impact of high surgeon daily workload.

On the other hand, we note that model (1) suggests that the hospital may be able to achieve substantial improvements in its performance by "smoothing" surgeon workload across days. For example, if we reschedule a surgeon with two cases in one day to two separate days, the estimated effect suggests the total expected incision time for these two cases would decrease by 0.8 hours. We leverage this insight to propose a surgical scheduling model that captures such effects in Section 5.

We now consider the effect of daily workload on patient outcomes, including two continuous outcomes, post-LOS and total ICU time, and three binary outcomes, reoperation, readmission, and mortality. In the "post-LOS" and "total ICU time" columns in Panel A of Table 6, we find that higher daily workload increases the post-LOS and total ICU time. Specifically, adding one more case increases the total ICU time and post-LOS by 1.07 and 1.40 days, respectively, for the cases performed by the surgeon on the same day. This is equivalent to a 12% increase for post-LOS and a 19% increase for total ICU time. Without using the

IVs, the effect is insignificant for both outcomes. The results here suggest increased surgeon daily workload is associated with longer post-surgery recovery time for patients. As we discussed for incision time, the negative effect can be attributed to multiple potential factors, including surgeon fatigue due to long working hours, as well as other constraints due to operating later in the day. For example, patients who are sent to ICU in the night may have to wait longer before extubation, as it is considered safer to keep them asleep until the intensivist or respiratory therapist is available. This may also increase their ICU time.

The effect on total ICU time and post-LOS is important to consider when managing patient flow. Longer post-surgery LOS will result in increased demand for downstream units and resources and reduce the system throughput. This can lead to overcrowding in the perioperative environment and delay in operations (Zenteno et al. 2016). Additionally, the ICU is often congested and extremely expensive to operate (e.g., Halpern 2011). Given almost all patients (99%) in our sample are sent to the ICU after operation, understanding the factors that impact their ICU recovery time provides an important potential solution for managing ICU congestion. As we discussed for incision time, smoothing the surgeon's workload across days can reduce the total expected ICU time and post-LOS, i.e., the time needed for recovery of the patients. For example, our estimation results imply moving two cases performed by a surgeon in a day to two different days leads to a reduction of 2.1 (resp. 2.8) days in total expected ICU time (resp. post-LOS).

Finally, we examine the impact of workload on three binary patient outcomes, reoperation, readmission, and mortality, which are estimated using full information MLE. (The sample size varies for the three binary outcomes. This is because some levels of categorical variables, e.g., specific procedure types, lead to perfect predictions of the binary outcome, thus the corresponding observations are dropped from the estimation.) First, we find that higher daily workload increases the likelihood of reoperation. Specifically, adding one more case leads to a 3.3 percentage points increase in the reoperation probability for each case performed by the surgeon on the same day. The magnitude of such an increase seems large at first, as the original reoperation probability is about 7% (see Table 3). However, we note that the median surgeon daily workload is two cases; thus adding one more case is equivalent to a 50% increase in daily workload. Such large impact of workload on medical outcomes is also observed in the literature. For example, Kc and Terwiesch (2009) find that a 10% increase in overwork increases the morality rate by 2.2 percentage points, which is a 32% relative change in their setting.

Similar negative effect of higher daily workload is also observed for the likelihood of hospital readmission. Specifically, adding one more case increases the readmission probability by 6.6 percentage points when estimated from the full sample. For 30-day mortality, however, the effect of daily workload is not significant. The insignificant effect for mortality seems surprising at first, given the negative impact of daily workload on other outcomes. One possible reason for such difference is that surgeons tend to pay greater attention to the patients with high risk of death, thus the mortality rate is less impacted by surgeon daily workload. In addition, the lack of significance may be because there is not enough variation in this outcome for the IV analysis to be statistically significant. Interestingly, we see that without the IV, using OLS, the coefficient is positive and statistically significant at the 5% level. This may be because if a patient is at high risk of death, he or she is more likely to be operated on even when the OR schedule is full, resulting in an increased risk of death for patients on high workload days on average. Thus, the endogeneity bias may go in the opposite direction for mortality compared to the other outcomes: Surgeons may nominally be able to operate on more low-risk patients in a given day, resulting in low-risk patients being more likely to be operated on during high workload days. However, when a patient with very high risk of death comes along, this patient can also be squeezed into the high workload days.

4.2. Heterogeneous Effects of Surgeon Daily Workload: Elective and Non-elective Patients

In the previous section, we see that increased daily workload leads to longer incision time and worse patient outcomes on average. In this section, we further investigate the impacts of daily workload for elective and non-elective cases separately. The non-elective cases include urgent, emergent, and salvage cases, and account for more than half (53%) of the full sample. The separate analyses allow us to investigate the potential heterogeneity in the impact of daily workload for different types of patients. In particular, we estimate the econometric models (1) and (2) using the subsamples of elective and non-elective cases separately.

The relevance condition of the two IVs is still satisfied for the elective and non-elective subsamples. The estimated η_1 and η_2 in (3) using the elective and non-elective cases are reported in Table 19 of the Online Supplement S.5. We find the two IVs still impact the surgeon daily workload with expected signs in the two subsamples. Moreover, the magnitudes are similar in the two subsamples as in the full sample. These results support the validity of the two IVs for the subsamples of elective and non-elective cases. Furthermore, a weak instrument test for the hypothesis $\eta_1 = \eta_2 = 0$ in (3) is strongly rejected at the significant level of 0.0001 with F-statistics of 26.3 and 180.9 for elective and non-elective cases, respectively.

Table 7 shows the estimated effects of surgeon daily workload (number of other cases) for the elective (Panel A) and non-elective samples (Panel B), respectively. As in Table 6, we report the estimated coefficients for the three continuous dependent variables (incision time, post-LOS, and total ICU time), and the AMEs for the three binary dependent variables (reoperation, readmission, and mortality). The coefficient γ in (2) for the three binary dependent variables, when estimated by the elective and non-elective samples, are given in Table 21 of Online Supplement S.5. The results when we measure surgeon' daily workload by $SumInc_i$ (total incision time of other cases) are qualitatively similar, and are reported in Online Supplement S.3 for elective and non-elective cases.

We find substantial heterogeneity in the effects of surgeon daily workload for elective and non-elective cases. The effect of daily workload on incision time is statistically significant (at 1% level) for elective cases (Panel A), but insignificant for non-elective cases (Panel B). The magnitude of the impact for elective cases is similar to that for the full sample. In particular, performing one more case increases the incision time

Table 7 Estimated Effects of Daily Workload (Number of Other Cases) on Surgery Duration and Patient Outcomes: Elective and Non-elective Sample

	Continuous y_i: Coefficients			Binary y_i: AME		
	Incision time	Post-LOS	Total ICU time	Reoperation	Readmission	Mortality
Panel A: Elec	0.334* (0.134)	−0.224 (1.114)	0.194 (0.794)	0.020 (0.041)	0.033 (0.064)	0.054 (0.049)
Num Obs.	2474	2474	2454	2394	2398	1897
Panel B: Non-elec	0.474 (0.329)	2.796* (1.417)	1.850* (0.908)	0.050* (0.021)	0.077* (0.038)	−0.001 (0.011)
Num Obs.	2871	2870	2865	2871	2697	2769

The estimated effects of surgeon daily workload (number of other cases) on surgery duration and patient outcomes for the elective and non-elective sample. We report the estimated coefficients in (1) for the three continuous dependent variables, and the AME from (2) for the three binary dependent variables. Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

of each elective case by 20 minutes, which is equivalent to an 7% increase on average. The heterogeneity in the impact on incision time may be due to the fact that non-elective cases (urgent and emergent cases) are generally more urgent and time sensitive than elective cases, so their incision time is less impacted by surgeon daily workload.

However, such heterogeneity appears in the opposite direction when we examine the impact of surgeon daily workload on patient outcomes. Specifically, we find that increased surgeon daily workload significantly impacts the surgery outcomes of non-elective patients, but not for elective ones. For example, for the post-LOS, we see that the coefficient of $NumCases_i$ is statistically significant for the non-elective cases, but insignificant for the elective ones ("Post-LOS" column). Moreover, the magnitude of the effect is larger for the non-elective cases than that for the full sample: Adding one more case leads to 2.79 more days in the post-LOS of non-elective patients, which is twice as that for the full sample (1.40 days). Similar heterogeneity is also seen in the total ICU time ("Total ICU time" column). The estimated coefficient of $NumCases_i$ is 1.85 days for the total ICU time of non-elective cases, which is much larger than that for the full sample (1.07 days). On the other hand, the daily workload does not significantly impact the total ICU time of elective cases. Finally, we find that the surgeon daily workload significantly increases the likelihood of reoperation and readmission for non-elective patients, but is insignificant for elective patients. Adding one more case increases the probabilities of reoperation and readmission by 5.0 and 7.7 percentage points for the non-elective cases. Both effects are larger than that for the full sample (3.3 and 6.6 percentage points for reoperation and readmission respectively). These consistent results show that the surgery outcomes of non-elective patients are more negatively affected by surgeon daily workload.

One potential explanation for the heterogeneous effect for patient outcomes is that non-elective cases are generally more urgent and complicated than elective ones, with more severe patients. Thus, the outcomes of non-elective cases may be more sensitive to surgeon fatigue or operational constraints due to high daily workload. On the other hand, the elective patients are on average less severe, and they recover more quickly after the operation. This can be seen by the summary statistics of the patient outcomes in Table 3: the non-elective patients have longer post-LOS and total ICU time, and higher likelihood of reoperation and readmission on average. The standard deviations of total ICU time and post-LOS are also much larger for the non-elective patients, implying there is more variation in their surgical outcomes than the elective patients.

Our results demonstrate the consistent negative impact of surgeon daily workload on multiple patient outcomes. Such effects are particularly significant for non-elective patients, who are generally more severe. Our results provide new empirical evidence for the link between high workload level and worse patient outcomes (see, e.g., Kc and Terwiesch 2009 and Kuntz et al. 2015) – specifically at the individual level. From the managerial perspective, it suggests that when hospitals design their surgery schedules, they should take into account the effects of surgeon daily workload in order to improve patient flow and patient outcomes. We explore this direction in Section 5.

4.3. Robustness Checks

We conduct similar regression analyses under alternative specifications to examine the robustness of our main findings. The results are reported in Tables 22 and 23 of Online Supplement S.5. We briefly discuss the results below. First, as shown in Section 3.1, the $GapNext_i$ (expected gap to next block) is a less powerful IV than $TotOther_i$ (number of cases by other surgeons) in explaining the variation in surgeon daily workload. In addition, imputation is needed to construct $GapNext_i$ for the period without block schedule information. Thus, we run the regressions (1) and (2) using only $TotOther_i$ as the IV. This allows us to keep the entire sample and without using any imputation. The results with one IV are very similar to that in our main specifications. We still find high daily workload is associated with longer surgery duration and worse patient outcomes. The heterogeneity in the estimated effects between elective and non-elective cases also holds.

In our main specification, we winsorize the three continuous outcomes (incision time, post-LOS, and total ICU time) at their 97.5th percentile, respectively. For robustness checks, we also experiment the winsorization level of 95th percentile. We find the results are largely similar, although the magnitudes of the effects become smaller for post-LOS and total ICU time. The change in the magnitude can be explained as follows. The distribution of total ICU time (as well as post-LOS) has a long tail on the right. Thus, different winsorization levels can lead to substantial change of the upper bound of the final sample. For example, the 95th and 97.5th percentile of total ICU time is 19.1 and 29.7 days respectively. As another check, we run the regressions after taking the log transformation of the three continuous dependent variables. The results again are qualitatively similar.

In addition, we perform the regressions on the subsample with complete block schedule. As we mentioned, this will substantially decrease the sample size as the block schedule is missing for more than a half of the horizon. The results are still consistent, although some of them become insignificant. We still observe the negative impact of daily workload on incision time and likelihood of reoperation, as well as on post-LOS and total ICU time for non-elective patients. Finally, we estimate using the sample consisting of eight standard procedures, which consists of 64% of all cases. For these standard procedures, we add their predicted mortality scores as a severity indicator to the exogenous variables X_i. The predicted mortality score measures the patient's likelihood of operative mortality. It is trained and validated on a national database (O'Brien et al. 2018), thus does not hinge on the sample we use. This risk score is only available for the standard procedures. The results for the standard procedure sample are qualitatively consistent. We find that higher surgeon daily workload is associated with longer incision time and total ICU time. The estimated effects of $NumCases_i$ under these alternative specifications are reported in Tables 22 and 23 of Online Supplement S.5. The results of $SumInc_i$ as the daily workload measure are qualitatively similar, and are available upon request.

Finally, we perform an additional robustness check for the post-LOS. Unlike the total ICU time which is measured in hours, the post-LOS can only be computed on a daily basis as we do not have the exact discharge time in the STS data. This may ignore the difference in the total recovery time of patients if they exit the OR at different times on the same day (e.g., 12PM versus 6PM). To account for patients' actual recovery time, we use several more conservative measures of post-LOS by adjusting the original measure using the OR entry or exit time of each case. We find that our estimation results by TSLS are very similar when using these more conservative measures. The definitions of these alternative measures are summarized in Online Supplement S.5. Their estimation results are provided in Table 24 therein.

5. A Surgery Scheduling Model with Impact from Daily Workload

Our empirical analyses show that increased surgeon daily workload can lead to longer surgery duration and worse patient outcomes. In this section, we propose a surgical scheduling model that incorporates such effects. While there is a rich literature on surgery scheduling, most of it assumes exogenous distributions for the surgery duration and patient outcomes (e.g., post-LOS). That is, the impact of surgeon workload is largely ignored. However, as shown in our results, high daily workload of surgeons is associated with multiple negative outcomes. Thus, our model aims to quantify the potential benefits we can obtain by incorporating these effects into surgical scheduling.

Our model considers a relatively small change to the current schedule used by our partner hospital. In particular, we consider the reassignment of the operations in our sample to different days. By changing the surgery dates, we aim to capture the potential benefit of smoothing surgeon daily workload. As we discussed in Section 4, this can decrease surgery duration and improve patient outcomes. On the other hand, we keep the surgeon assigned to each patient unchanged. This allows us to isolate the impacts of optimizing the surgery dates, and thus highlight the benefit of incorporating the effects of surgeon daily workload in surgical scheduling. In addition, it provides an easily implementable potential solution to improve the scheduling in the hospital. We expect the benefit to be larger if we introduce other types of decisions in our model, e.g., the number of ORs assigned to the cardiac department.

To save space, we briefly describe the surgical scheduling model below and include its details in Online Supplement S.4. We solve the model for each week (Sunday to Saturday) in our four year sample. Our optimization model considers the decisions to assign cases to each day. We exclude the cases on the weekends in the scheduling model. The objective in our model is to minimize the total expected incision time, post-LOS, or ICU time of all cases in our sample, which are estimated using our econometric model (1). To focus on the impact of daily workload, we assume the term $X_i\beta$, which primarily depends on the patient's risk and operative factors, remains unchanged in the new schedule. We show that the objective function can be written as

$$\min \sum_{s\in S}\sum_{t\in T} \tilde{n}_{s,t} \cdot \big(\gamma^{(el)}\tilde{n}_{s,t}^{(el)} + \gamma^{(ne)}\tilde{n}_{s,t}^{(ne)}\big). \tag{4}$$

Here we index the day in the week by $t \in T$ and the surgeon by $s \in S$, with T and S being the sets of surgery dates and surgeons for cases in the given week; $\tilde{n}_{s,t}^{(el)}$ and $\tilde{n}_{s,t}^{(ne)}$ denote the number of elective and non-elective cases performed by surgeon s on day t respectively. The coefficients $\gamma^{(el)}$ and $\gamma^{(ne)}$ are the estimated effect γ for the daily workload effect on incision time, post-LOS, or total ICU time – depending on which one we are to minimize – for the elective and non-elective cases respectively, which are reported in Table 7 of Section 4.1. We set the coefficient to be zero if it is not statistically significant at the 10% level. When we ignore the heterogeneity in the impacts of daily workload, we use the average treatment effects in Table 6 with $\gamma^{(el)} = \gamma^{(ne)} = \gamma^{(avg)}$.

We consider multiple feasibility constraints for the new schedule. We allow elective cases to be assigned to any day of the week. On the other hand, we impose that urgent cases can only be assigned to the original date or the adjacent days, while the emergent and salvage cases can only be scheduled on the original dates. These constraints reflect the different levels of time sensitivity for different types of patients. In addition, we impose an upper bound on surgeon daily workload. We assume the surgeon can perform at most $\bar{n}^{(c)}$ cases in a day in the new schedule, unless the surgeon already performs more cases in the original schedule. Finally, we set an upper bound on the number of days worked by each surgeon in a week. This accounts for the fact that surgeons have other responsibilities such as seeing patients in the office, teaching, and attending conferences. We assume the maximum number of working days for a surgeon in a week is $\bar{n}^{(d)}$, unless the surgeon works for more days in the original schedule. We show that the optimization problem can be formulated as a MIQP with binary decision variables, quadratic objective, and linear constraints.

5.1. Results

In this section, we summarize the results from our scheduling model, which demonstrate the benefit of incorporating the effects of surgeon daily workload in surgical scheduling. For our main numerical results, we set $\bar{n}^{(c)} = 3$ and $\bar{n}^{(d)} = 4$, respectively. This means that the surgeon's maximum daily workload is three cases and the maximum number of working days is four days, unless the corresponding quantity is larger in the original schedule. In the numerical analysis, we discuss the feasibility of the new schedules under the two parameter choices. As a further robustness check, we show in Online Supplement S.4 that our scheduling model still leads to substantial improvement even under more restrictive conditions.

We show our main results in Table 8. The first two columns show the variable we are optimizing ("Objective") and the estimated effects ("Effect") we use in our objective function (4), with which we solve the surgical scheduling model. The third and fourth columns ("Obj orig" and "Obj new") report the objective values (4) under the original and new schedules, respectively. The fifth column ("ΔObj") reports the absolute reduction in the objective function, which demonstrates the benefit of applying our surgical scheduling model. The next column ("Number of reduced week") reports the number of weeks (out of the 209 weeks in our sample) that we can achieve reduction in the objective function under the new schedule. Finally, the last column ("Rel. ΔTotal") shows the relative reduction in the sum of the corresponding objective we are optimizing over, i.e., Rel. $\Delta\text{Total} = \Delta\text{Obj}/\text{Total}$, where Total is the sum of observed values (incision time, post-LOS, or total ICU time) from all patients in our sample. As in our estimations, we winsorize the corresponding variables to their 97.5th percentile. Thus, Rel. ΔTotal represents the relative reduction in total incision time, post-LOS, or ICU time. It provides an alternative measure for the benefit from the new surgical schedule.

Table 8 Results of the Surgical Scheduling Model

Objective	Effect	Obj orig	Obj new	ΔObj	Number of improved weeks	Rel. ΔTotal (in % points)
Inc Time	Avg	3626.00	3022.00	604.00	205	2.46%
(in hours)	Het	1447.22	1074.14	373.08	203	1.52%
Post-LOS	Avg	12283.08	10166.06	2117.02	205	3.70%
(in days)	Het	12979.88	9918.26	3061.62	204	5.34%
Total ICU time	All	9735.81	8114.07	1621.74	205	6.68%
(in days)	Avg	8754.2	6728.45	2025.75	204	8.34%

Note that when we use the average effect of daily workload for elective and non-elective cases, i.e., $\gamma^{(el)} = \gamma^{(ne)} = \gamma^{(avg)}$, the optimal schedules from the model are the same for the three variables. To see why this is the case, by (4), the objective with average effect can be expressed as

$$\min \gamma^{(avg)} \cdot \sum_{s \in S} \sum_{t \in T} \left(\tilde{n}_{s,t}\right)^2 . \tag{5}$$

Thus, the coefficient $\gamma^{(avg)}$ does not impact the solution (although it affects the objective value). Similarly, we obtain the same optimal schedule for post-LOS and ICU time when we use the heterogeneous effect of daily workload. In this case, the objective function can be written as

$$\min \gamma^{(ne)} \cdot \sum_{s\in S}\sum_{t\in T} \tilde{n}_{s,t}\tilde{n}_{s,t}^{(ne)}. \tag{6}$$

This is because we have $\gamma^{(el)} = 0$ for post-LOS and ICU time, i.e., surgeon daily workload does not impact the two outcomes of elective cases. Since $\tilde{n}_{s,t} = \tilde{n}_{s,t}^{(el)} + \tilde{n}_{s,t}^{(ne)}$, this can be further written as

$$\min \gamma^{(ne)} \sum_{s\in S}\sum_{t\in T} \left[\left(\tilde{n}_{s,t}^{(ne)}\right)^2 + \tilde{n}_{s,t}^{(el)}\tilde{n}_{s,t}^{(ne)}\right]. \tag{7}$$

We now analyze the results of our surgical scheduling model. In Table 8, we see that our new schedule leads to substantial improvement for all three outcomes using both average and heterogeneous effects of daily workload. For total incision time, the new schedule with average (resp. heterogeneous) effect leads to a 604 (resp. 373) hours decrease in the four-year horizon, which is equivalent to a 2.46% (resp. 1.52%) relative reduction. The OR is an extremely expensive resource with cost up to \$37 per minute (Childers and Maggard-Gibbons 2018). Thus, the reduction in the incision time can save the hospital up to \$335,220 each year. On the other hand, given the average incision time is 4.8 hours in our sample, the hospital may be able to add 31 new cases each year due to the reduction in OR time from the new schedule. With some cardiac operation netting margins of over \$21,000 per case on average (Robinson 2011), this has the potential to translate to an additional \$651,000 in profit for this service. In addition, we find that the new schedule leads to improvement for most (205) of the 209 weeks in our sample. This shows that the benefit of our scheduling model is not limited to a small number of weeks, and the original schedule can be substantially improved.

Our scheduling model also substantially reduces the total expected post-LOS and ICU time when optimized to do so. Recall that when considering the average treatment effect, the resulting schedules are identical, so all three metrics are improved. When considering the heterogeneous treatment effect, the reported savings are for when the schedule is optimized to minimize the particular outcome; the other outcomes will also be improved, but perhaps to a lesser amount than when we specifically optimized for it. Using the average effect, the new schedule decreases the total post-LOS by 2,117 days and the total ICU time by 1,621 days, which translate to a 3.70% and 6.68% relative drop respectively. The benefit is even larger with heterogeneous effect, as the relative reduction for post-LOS (resp. ICU time) increases to 5.34% (resp. 8.34%). The bigger benefit under heterogeneous effect can be explained as follows. First, as shown in Table 7, the surgeon daily workload does not negatively impact the post-LOS and ICU time of elective cases. This gives us more flexibility in adjusting the schedule. Second, the impact of surgeon daily workload on the non-elective cases is much larger than that estimated from the full sample (2.79 vs 1.40 for post-LOS and 1.85 vs 1.07 for ICU time). This leads to the greater benefit from the new schedule when using the heterogeneous effect.

The benefits for post-LOS and ICU time from our new schedule are economically important. To see this, we convert the reduction in the downstream resource to the number of more patients the hospital can accommodate each year, assuming the downstream resource is the only bottleneck. Over the four year horizon in our data set, this can be computed as:

$$\Delta\text{Pat} = \frac{\Delta\text{Obj}}{4 \text{ Years} \times \text{AvgTime}},$$

where AvgTime denotes the average post-LOS or total ICU time. Taking the results with heterogeneous effects as an example, the reduction in the total post-LOS and ICU time translates to 63 and 94 more patients admitted each year, respectively. In addition, we find similar reduction in the average occupancy level in the downstream unit and ICU from the new schedule. With the heterogeneous effect, the average census in the downstream unit and ICU decreases by 5.47% and 8.17% respectively. The above results highlight the potential benefits of our surgical scheduling model in reducing downstream congestion, which can be a bottleneck in the perioperative environment.

5.2. Analysis of the New Schedules

We have shown that our surgical scheduling model can improve the incision time, post-LOS, and ICU time. We now take a closer look at the new schedule to investigate the mechanisms that lead to the improvement. This provides important managerial insights on how hospitals should account for the impact of surgeon daily workload in surgical scheduling.

We consider two schedules from our model. The first is the one using the average effect with the objective given in (5). As mentioned before, the resulting schedule is the same for the three outcomes (incision time, post-LOS, and ICU time). The second is the one using the heterogeneous effect for post-LOS and ICU time, where the surgeon daily workload impacts the non-elective cases but not the elective ones. The corresponding objective is given in (6). In Table 9, we provide some summary statistics of the two schedules (first two rows) as well as the original one (last row). The columns $\tilde{n}_{s,t} = i$ for $i = 1, 2, 3, 4$ report the number of surgeon-day pairs for which the surgeon performs i cases in a day. The next column "Avg n_{day}" shows the average number of days worked by a surgeon in a week, given the surgeon appears at least once in the schedule. The columns $\tilde{n}^{(ne)}_{s,t} = i$ for $i = 1, 2, 3$ report the number of surgeon-day pairs that a surgeon performs i non-elective cases a day. The last column shows the average number of elective cases performed by a surgeon in a day, given the surgeon performs at least one non-elective case in that day. It thus measures the co-occurrence of elective and non-elective cases in the schedule.

We first consider the schedule under the average effect. To minimize the objective function (5), the new schedule should smooth surgeon workload across days, i.e., reducing the number of days with multiple cases. In Table 9, we see that this is indeed achieved by our model as the new schedule significantly reduces the number of days with high workload. Specifically, the number of surgeon-day pairs with $\tilde{n}_{s,t} = 3$, i.e.,

Table 9 Summary Statistics of Schedules

Effect	$\tilde{n}_{s,t}=1$	$\tilde{n}_{s,t}=2$	$\tilde{n}_{s,t}=3$	$\tilde{n}_{s,t}=4$	Avg n_{day}	$\tilde{n}_{s,t}^{(ne)}=1$	$\tilde{n}_{s,t}^{(ne)}=2$	$\tilde{n}_{s,t}^{(ne)}=3$	$\mathsf{E}(\tilde{n}_{s,t}^{(el)} \mid \tilde{n}_{s,t}^{(ne)}>0)$
Avg	3268	980	39	1	3.278	2207	322	7	0.151
Het	2924	727	321	2	3.038	2292	281	6	0.063
Original	2268	1249	177	13	2.834	1872	424	48	0.266

the surgeon performs three cases in the day, decreases from 177 in the original schedule to 39 in the new schedule. Similar reduction is also observed for the surgeon-day pairs with $\tilde{n}_{s,t}=4$ (from 13 to 1). The reduction in these high workload days is mostly made up by the surgeon-day pairs with a single case (i.e., $\tilde{n}_{s,t}=1$), which increases from 2,268 to 3,268 in the new schedule. This result shows that the new schedule effectively smooths surgeon workload across days. In particular, the proportion of cases with surgeon's daily workload no more than two cases increases from 89% in the original schedule to 98% in the new schedule.

Next, we examine the new schedule obtained with heterogeneous effect for post-LOS and ICU time. By (7), the minimization of the objective includes two parts: smoothing surgeon's non-elective workload across days, as well as reducing the *co-occurrence* of elective and non-elective cases. Both of the two aspects are reflected in the new schedule according to Table 9. First, the new schedule significantly reduces the number of surgeon-day pairs with multiple non-elective cases: the number of pairs with $\tilde{n}_{s,t}^{(ne)}=2$ (resp. $\tilde{n}_{s,t}^{(ne)}=3$) decreases from 424 to 281 (resp. 48 to 6) in the new schedule. In addition, the new schedule decreases the co-occurrence of elective and non-elective cases: given the surgeon performs at least one non-elective case, the average number of elective cases in that day drops from 0.27 in the original schedule to 0.06 in the new schedule. This leads to the managerial insight that hospitals should reduce surgeons' workload in the days when they have non-elective cases to perform, as non-elective patients are generally more severe with complicated procedures.

While the solution to our scheduling model leads to substantial improvements, it is important to check the feasibility of the new schedule. The first is the number of working days of surgeons in each week. Although asking a surgeon to work for more days naturally smooths the workload, it might be difficult to implement in practice if the change is too large. Second, as the OR capacity of the cardiac department is limited and hard to expand in short periods, we need to make sure the total number of cases by all surgeons in a day does not increase substantially in the new schedules. To address the first concern, we compute the average number of working days by a surgeon in each week. As shown in the sixth column ("Avg n_{day}") in Table 9, the average number of working days in each week only mildly increases in the new schedules — from 2.83 days in the original schedule to 3.27 under the average effect and 3.02 under the heterogeneous effect. Next, we show in Figure 1 the frequency distributions of total number of cases in a day from the original (black-triangle line) and new schedules (blue-circled line for average effect and red-squared line for heterogeneous effect). We see that the number of days with extremely high total number of cases (e.g.,

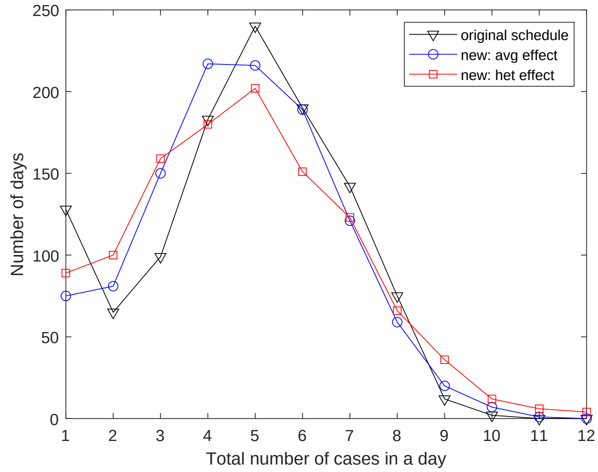

Figure 1 Frequency distributions of total number of cases by all surgeons in a day

more than ten cases in a day) remains small under the new schedules. In addition, the distributions of total number of cases in a day are similar under the original and new schedules. Thus, our new schedules do not lead to a significant increase in the peak OR usage of the cardiac department. This further supports the feasibility of our scheduling model.

6. Conclusion

In many human-run service systems, service time and quality can be endogenously affected by the level of workload. In this work, we focus on the relationship between workload and performance in the context of cardiac surgery. Specifically, we study how surgery duration and patient outcomes are impacted by individual surgeon daily workload. Using a detailed data set of cardiac surgery, we find that higher surgeon daily workload leads to longer surgery duration and worse patient outcomes. We develop two novel IVs using the operational factors in the cardiac department. Our method effectively addresses the endogeneity problem due to unobserved risk factors. Our study provides new evidence for the negative impact of surgeon fatigue or operational constraints due to high daily workload. It suggests that hospitals may be able to improve surgery performance if each surgeon's workload can be smoothed across days.

While our results provide strong evidence of the impact of workload on cardiac surgery outcomes, our study has a number of limitations. First, we only have block schedules from part of our data set and we had to impute the rest of the schedule. While our imputation seems to be reasonable, it would be ideal to have the full schedule available. Second, our data comes from a single hospital. Other hospitals may have different scheduling procedures which may make the IVs more or less appropriate. Finally, as we have conducted an IV analysis, our results only provide insight into cases that *comply* with the IVs. There are some operations that must happen, regardless of shared resources of block schedule, so our analysis does not provide insights into the effect of surgeon workload on these cases.

Based on our findings, we develop a surgery scheduling model that incorporates the effect of surgeon workload. We find that by simply rescheduling operations within a week, with substantial restrictions on how much non-elective operations can be moved, substantial improvements could be achieved for both surgery duration and patient outcomes. Our model demonstrates the potential improvements in patient flow in the OR (via incision time) and post-surgery (via post-LOS and total ICU time) by accounting for the impact of surgeon workload when scheduling surgery. With more flexibility in how/when to schedule surgery, it is possible even larger gains can be achieved. As such, our results suggest that it is important for hospital managers and surgeons to consider the impact of surgeon workload when managing their ORs.

References

Bashir M, Harky A, Fok M, Shaw M, Hickey GL, Grant SW, Uppal R, Oo A (2017) Acute type a aortic dissection in the united kingdom: surgeon volume-outcome relation. *The Journal of Thoracic and Cardiovascular Surgery* 154(2):398–406.

Batt RJ, Terwiesch C (2016) Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Science* 63(11):3531–3551.

Berry Jaeker JA, Tucker AL (2017) Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Science* 63(4):1042–1062.

Cameron AC, Trivedi PK (2013) *Regression analysis of count data*, volume 53 (Cambridge university press).

Chalfin DB, Trzeciak S, Likourezos A, Baumann BM, Dellinger RP, study group DE, et al. (2007) Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Critical Care Medicine* 35(6):1477–1483.

Chan CW, Farias VF, Bambos N, Escobar GJ (2012) Optimizing intensive care unit discharge decisions with patient readmissions. *Operations Research* 60(6):1323–1341.

Childers CP, Maggard-Gibbons M (2018) Understanding costs of care in the operating room. *JAMA Surgery* 153(4):e176233–e176233.

Clark JR, Huckman RS (2012) Broadening focus: Spillovers, complementarities, and specialization in the hospital industry. *Management Science* 58(4):708–722.

Delasay M, Ingolfsson A, Kolfal B (2016) Modeling load and overwork effects in queueing systems with adaptive service rates. *Operations Research* 64(4):867–885.

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

Erdogan SA, Denton BT (2010) Surgery planning and scheduling. *Wiley Encyclopedia of Operations Research and Management Science* .

Freeman M, Savva N, Scholtes S (2017) Gatekeepers at work: An empirical analysis of a maternity unit. *Management Science* 63(10):3147–3167.

Freeman NK, Melouk SH, Mittenthal J (2016) A scenario-based approach for operating theater scheduling under uncertainty. *Manufacturing & Service Operations Management* 18(2):245–261.

George JM, Harrison JM (2001) Dynamic control of a queue with adjustable service rate. *Operations Research* 49(5):720–731.

Govindarajan A, Urbach DR, Kumar M, Li Q, Murray BJ, Juurlink D, Kennedy E, Gagliardi A, Sutradhar R, Baxter NN (2015) Outcomes of daytime procedures performed by attending surgeons after night work. *New England Journal of Medicine* 373(9):845–853.

Green LV, Nguyen V (2001) Strategies for cutting hospital beds: the impact on patient service. *Health Services Research* 36(2):421.

Gul S, Denton BT, Fowler JW, Huschka T (2011) Bi-criteria scheduling of surgical services for an outpatient procedure center. *Production and Operations Management* 20(3):406–417.

Gupta D (2007) Surgical suites' operations management. *Production and Operations Management* 16(6):689–700.

Halpern SD (2011) ICU capacity strain and the quality and allocation of critical care. *Current Opinion in Critical Care* 17(6):648–657.

Hopp WJ, Iravani SM, Yuen GY (2007) Operations systems with discretionary task completion. *Management Science* 53(1):61–77.

Janhofer DE, Lakhiani C, Song DH (2019) Addressing surgeon fatigue: current understanding and strategies for mitigation. *Plastic and Reconstructive Surgery* 144(4):693e–699e.

Kc DS (2014) Does multitasking improve performance? evidence from the emergency department. *Manufacturing & Service Operations Management* 16(2):168–183.

Kc DS, Staats BR (2012) Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manufacturing & Service Operations Management* 14(4):618–633.

Kc DS, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* 55(9):1486–1498.

Kc DS, Terwiesch C (2011) The effects of focus on performance: Evidence from california hospitals. *Management Science* 57(11):1897–1912.

Kc DS, Terwiesch C (2012) An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* 14(1):50–65.

Keskinocak P, Savva N (2020) A review of the healthcare-management (modeling) literature published in manufacturing & service operations management. *Manufacturing & Service Operations Management* 22(1):59–72.

Kim SH, Chan CW, Olivares M, Escobar G (2015) ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* 61(1):19–38.

Kuntz L, Mennicken R, Scholtes S (2015) Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science* 61(4):754–771.

May JH, Spangler WE, Strum DP, Vargas LG (2011) The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management* 20(3):392–405.

McDermott KW, Freeman WJ, Elixhauser A (2017) Overview of operating room procedures during inpatient stays in us hospitals, 2014: statistical brief# 233. *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville: Agency for Healthcare Research and Quality* .

Needleman J, Buerhaus P, Pankratz VS, Leibson CL, Stevens SR, Harris M (2011) Nurse staffing and inpatient hospital mortality. *New England Journal of Medicine* 364(11):1037–1045.

Neuraz A, Guérin C, Payet C, Polazzi S, Aubrun F, Dailler F, Lehot JJ, Piriou V, Neidecker J, Rimmelé T, et al. (2015) Patient mortality is associated with staff resources and workload in the ICU: a multicenter observational study. *Critical Care Medicine* 43(8):1587–1594.

Oliva R, Sterman JD (2001) Cutting corners and working overtime: Quality erosion in the service industry. *Management Science* 47(7):894–914.

Olivares M, Terwiesch C, Cassorla L (2008) Structural estimation of the newsvendor model: an application to reserving operating room time. *Management Science* 54(1):41–55.

O'Brien SM, Feng L, He X, Xian Y, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland Jr JC, Lobdell KW, et al. (2018) The society of thoracic surgeons 2018 adult cardiac surgery risk models: Statistical methods and results. *Annals of Thoracic Surgery* 105(5):1419–1428.

Rath S, Rajaram K (2021) Staff planning for hospitals with implicit cost estimation and stochastic optimization. *Production and Operations Management* .

Robinson JC (2011) Variation in hospital costs, payments, and profitabilty for cardiac valve replacement surgery. *Health Services Research* 46(6pt1):1928–1945.

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

Shanafelt TD, Balch CM, Bechamps G, Russell T, Dyrbye L, Satele D, Collicott P, Novotny PJ, Sloan J, Freischlag J (2010) Burnout and medical errors among american surgeons. *Annals of Surgery* 251(6):995–1000.

Staats BR, Gino F (2012) Specialization and variety in repetitive tasks: Evidence from a japanese bank. *Management science* 58(6):1141–1159.

Tan TF, Netessine S (2014) When does the devil make work? an empirical study of the impact of workload on worker productivity. *Management Science* 60(6):1574–1593.

Thomas M, Allen MS, Wigle DA, Shen KR, Cassivi SD, Nichols III FC, Deschamps C (2012) Does surgeon workload per day affect outcomes after pulmonary lobectomies? *Annals of Thoracic Surgery* 94(3):966–972.

Wang G, Pourghannad B (2020) Matching patients with surgeons: Heterogeneous effects of surgical volume on surgery duration. *Available at SSRN* .

Wolff RW (1989) *Stochastic modeling and the theory of queues* (Pearson College Division).

Woodridge JM (2010) *Econometric Analysis of Cross Section and Panel Data, 2nd Edtion* (The MIT Press).

Zenteno AC, Carnes T, Levi R, Daily BJ, Dunn PF (2016) Systematic or block allocation at a large academic medical center. *Annals of Surgery* 264(6):973–981.

The Impact of Surgeon Daily Workload and its Implications for Operating Room Scheduling: Online Supplement

Yiwen Shen
Information Systems, Business Statistics and Operations Management,
Hong Kong University of Science and Technology, yiwenshen@ust.hk

Carri W. Chan
Decision, Risk, and Operations, Columbia Business School, cwchan@columbia.edu

Fanyin Zheng
Decision, Risk, and Operations, Columbia Business School, fanyin.zheng@columbia.edu

Michael Argenziano
Department of Surgery, New York Presbyterian/Columbia University Irving Medical Center, ma66@cumc.columbia.edu

Paul Kurlansky
Department of Surgery, New York Presbyterian/Columbia University Irving Medical Center, pk2245@cumc.columbia.edu

S.1. Description and Summary Statistics of Independent Variables in (1) and (2)

To control for the effects of patient's characteristics and severity levels, we include a comprehensive list of demographic, risk, operative, and operational factors as independent variables in our estimation models. Some of these factors are already discussed in Section 3. We now provide the description and summary statistics for other independent variables included in X_i for our models (1) and (2).

In Table 10, we document the descriptions, types, and summary statistics of the independent variables. We also provide their locations in the STS data collection form. We handle missing values in the binary and categorical variables as follows: if the number of missing observations is smaller than 100 (1.8% of the sample), we impute their values using the majority from the cases in the same New York Heart Association (NYHA) class. Otherwise, we add a new category "Unknown" to represent the missing values. Summary statistics of the categorical variables are reported in Table 11. Note that the NYHA classification is not available (N/A) if the patient has not experienced heart failure. The Pulmonary Artery (PA) pressure is coded as "High" if it is higher than 55mg and "Low" otherwise. We also include the patient's admission type, which refers to the channels for the patient to be admitted to the hospital. The four admission types and their numbers of observations are: elective (3,777), emergency department (448), transfer-in (1,117), and other (10).

We classify the cases to different surgery types to control for the procedures performed by the surgeons. First, we have eight standard surgery types from the STS data: coronary artery bypass graft (CABG), aortic valve replacement (AVR), mitral valve replacement (MVR), mitral valve repair (MVr), and their combinations CABG+AVR, CABG+MVR, CABG+MVr, and AVR+MVR. For the cases that do not fall into the standard types, we classify their surgery types by the following heuristic rule. We collect from the STS data which of the following four procedures are performed in the operation: coronary artery bypass, valve, other cardiac procedure, and other non-cardiac procedure. If only one of the four procedures is performed, we classify the case as a non-standard isolated type, e.g., "non-standard isolated valve" if only the valve procedure is conducted. If more than one of the procedures are performed, we classify the case as the "non-standard multiple" type. Finally, if none of the four procedures is performed, we classify it as "not identified." In total, we have six types for the non-standard procedures, i.e., four non-standard isolated ones, non-standard multiple, and not identified. The numbers of cases of each type (both standard and non-standard ones) are summarized in Table 12.

Table 10 Description and Summary Statistics of Other Independent Variables in Models (1) and (2)

Variable	Description	Section in STS	Type	Mean
Race	Patient's race	Demographics	Categorical	-
Endocard	Endocarditis	Risk factor	Binary	0.053
PeriAD	Peripheral arterial disease	Risk factor	Binary	0.088
Lung	Lung disease with severity $\geq$ mild	Risk factor	Binary	0.192
Hypertension	Hypertension	Risk factor	Binary	0.777
CaroStenosis	Carotid Stenosis	Risk factor	Binary	0.054
Syncope	Syncope	Risk factor	Binary	0.031
Dialysis	Dialysis for renal failure	Risk factor	Binary	0.030
Diabetes	Insulin control for diabetes	Risk factor	Binary	0.111
Liver	Liver disease	Risk factor	Binary	0.022
Cancer	Cancer within five years	Risk factor	Binary	0.062
Thoracic	Thoracic aorta disease	Risk factor	Binary	0.094
DrugUse	Recent or remote drug use	Risk factor	Binary	0.088
Smoke	Smoke status of patient	Risk factor	Categorical	-
PrevCI	Previous cardiac intervention	Previous Intervention	Binary	0.431
CardShock	Cardiogenic shock	Preoperative	Binary	0.076
MI	Prior MI	Preoperative	Binary	0.120
NYHA	NYHA classification	Preoperative	Categorical	-
Aorta	Aorta procedure performed	Operative	Binary	0.123
Incidence	Non-initial cardiovascular surgery	Operative	Binary	0.188
PA_Pressure	PA systolic pressure	Hemodynamics	Categorical	-
TotCABG	Number of arteries bypassed	Coronary Bypass	Continuous	1.36

In summary, the independent variable X_i in (1) and (2) includes the factors in Table 10, patient's gender and age, surgery status, patient's admission type, procedure type in Table 12, surgeon's identifier, patient's pre-LOS, block schedule status, cardiac patient census, and dummies for weekday, month, and year of the operation.

Table 11 Summary Statistics of Categorical Variables in Table 10

Variable	Category	Num Obs.	Ratio
NYHA	N/A	1933	0.361
	I	516	0.096
	II	998	0.186
	III	991	0.185
	IV	663	0.124
	Unknown	251	0.047
Race	White	4273	0.798
	Asian	590	0.110
	Black	274	0.051
	Other	215	0.040
Smoke	FALSE	2694	0.503
	TRUE	2429	0.454
	Unknown	229	0.043
PA Pressure	High	376	0.070
	Low	2247	0.420
	Unknown	2729	0.510

Table 12 Numbers of Cases by Surgery Types

Surgery Type	Number of Cases	Ratio
Standard (N = 3420)		
CABG	1718	0.321
AVR	683	0.128
MVR	225	0.042
MVr	254	0.047
CABG + AVR	318	0.059
CABG + MVR	57	0.011
CABG + MVr	58	0.011
AVR + MVR	107	0.020
Non-standard (N = 1932)		
Isolated Valve	574	0.107
Isolated CAB	28	0.005
Isolated cardiac	369	0.069
Isolated non-cardiac	15	0.003
Multiple	690	0.129
Not identified	256	0.048

S.2. Block Schedule Imputation Model

We impute the block schedule status for the surgeon-day pairs that appear in our sample, but with missing block schedule information. Let $Y_{s,t}$ be a binary variable which takes value one if surgeon s has an assigned block on day t (i.e., in block schedule), and zero otherwise (i.e., out of block schedule). We estimate $Y_{s,t}$

using a logistic model, i.e.,

$$\ln\left[\frac{\Pr(Y_{s,t}=1|X'_{s,t})}{\Pr(Y_{s,t}=0|X'_{s,t})}\right]=X'_{s,t}\beta+\varepsilon_{s,t}, \tag{8}$$

where $X'_{s,t}$ is a set of independent variables and $\varepsilon_{s,t}$ denotes the error term. The regressor $X'_{s,t}$ contains 23 independent variables (plus a constant term) for surgeon s on day t. For example, it includes the numbers of elective, urgent, and emergent cases by the focal and other surgeons on day t. As we are imputing instead of forecasting $Y_{s,t}$, the regressor $X'_{s,t}$ also contains variables that depend on the "future" information after day t, e.g., the number of days worked by surgeon s in the current calendar week.

To impute the block schedule status, we include multiple operational factors related to the workload of the focal and other surgeons. Table 13 summarizes the variables included in $X'_{s,t}$ (plus a constant term). In particular, $\text{ORTime}_{s,t}$ denotes the sum of OR time of the cases by surgeon s on day t, ignoring overlapping due to surgery parallel. $\text{StartHour}_{s,t}$ and $\text{EndHour}_{s,t}$ are calculated using the OR entry and exit time of the cases by surgeon s on day t; $\text{StartHour}_{s,t}$ (resp. $\text{EndHour}_{s,t}$) corresponds to the OR entry (resp. OR exit) time of the earliest (resp. latest) case, rounded to the nearest hour. $\text{PatRemain}_{s,t}$ is the number of patients remaining in the hospital for surgeon s. This refers to the patients that (1) already admitted to the hospital by day $t-1$, (2) operations have not been performed by day $t-1$, and (3) operations are eventually performed by surgeon s. $\text{WDElecRatio}_{s,t}$ is the proportion of elective cases by surgeon s in $[t-180, t+180]$ that fall on the same weekday as t. This variable is included as surgeons' blocks tend to fall on specific weekdays to reduce the variation in surgeons' schedule, and most elective cases are performed in their surgeons' block. We have in total 23 independent variables (plus a constant term) in the logistic model (8) for imputing the block schedule.

Table 13 Definition of Independent Variables in the Schedule Imputation Model (8)

Variable	Definition
Surg_s	Dummy variable for surgeon identifier
WeekDay_t	Dummy variable for weekday of t
$\text{ElecCur}_{s,t}$, $\text{UrgCur}_{s,t}$, $\text{EmergCur}_{s,t}$	Number of elective/urgent/emergent cases by surgeon s on day t
$\text{ElecOth}_{s,t}$, $\text{UrgOth}_{s,t}$, $\text{EmergOth}_{s,t}$	Number of elective/urgent/emergent cases by other surgeons on day t
$\text{ORTime}_{s,t}$	Total OR time of cases by surgeon s on day t
$\text{StartHour}_{s,t}$, $\text{EndHour}_{s,t}$	Start and end of the cases by surgeon s on day t
$\text{StartLate}_{s,t}$, $\text{EndEarly}_{s,t}$	Indicators for $\text{StartHour}_{s,t} \geq 8$AM and $\text{EndHour}_{s,t} \leq 3$PM
$\text{AdmCur}_{s,t}$	Numbers of patients admitted by surgeon s on day t
$\text{PatRemain}_{s,t}$	Numbers of patients remaining in the hospital for surgeon s
$\text{NumPreDay}_{s,t}$, $\text{NumPostDay}_{s,t}$	Numbers of cases by surgeon s in the previous and next weekday
$\text{WorkPreDay}_{s,t}$, WorkNextDay	Indicators for whether surgeon s works in the previous and next weekday
$\text{NumCurWeek}_{s,t}$	Numbers of days worked by surgeon s in current calendar week
$\text{DistLast}_{s,t}$, $\text{DistNext}_{s,t}$	Number of days from the previous and next working day of surgeon s
$\text{WDElecRatio}_{s,t}$	Proportion of elective cases by surgeon s in $[t-180, t+180]$ that are performed on the same weekday as t

We estimate the logit model (8) using the surgeon-day pairs in the block sample, for which the values of $Y_{s,t}$ are known. We then use the estimated model to impute the block schedule for the periods where the block schedule is missing. With the fitted probability $\Pr(Y_{s,t}=1|X'_{s,t})$, we calculate the expected gap to next block as:

$$GapNext_i = \sum_{l=1}^{T-1}\Big[\prod_{j=1}^{l-1}(1-p_{t+j}^{(blk)})p_{t+l}^{(blk)}\Big]\times l + \prod_{j=1}^{T-1}\big(1-p_{t+j}^{(blk)}\big)\times T, \tag{9}$$

where

$$p_{t+l}^{(blk)} := \Pr(Y_{s,t+l}=1|X'_{s,t+l}).$$

Here t denotes the surgery date of case i; $p_{t+l}^{(blk)}$ is the probability that case i's surgeon has a block on day $t+l$; T is the truncation level for the maximum expected gap. We set it to be 14 days in our computation, as we find from the block schedule data that it is rare for a surgeon to stay idle without any block assignment in two consecutive weeks. The calculation in (9) is based on an implicit assumption that whether a surgeon has a block assignment is independent across days. Thus, the term $\prod_{j=1}^{l-1}(1-p_{t+j}^{(blk)})p_{t+l}^{(blk)}$ represents the probability that the first block assignment after day t occurs on day $t+l$. Note that for the periods with block schedule data, the corresponding $GapNext_i$ can also be computed by (9) with $p_{t+l}^{(blk)}$ set to zero or one according to the actual block schedule.

Imputation Results: We briefly discuss the results for the block schedule imputation model (8). We estimate model (8) using the periods with block schedule information. The weekends are dropped as all blocks are assigned on weekdays. This leaves us with 1680 surgeon-day pairs in the block sample for estimation. We use the McFadden's R-squared to measure the model's performance, which is defined as

$$R^2 = 1 - \frac{\ln(l^{mod})}{\ln(l^{null})},$$

where l^{mod} is the likelihood from the estimated model, l^{null} is the likelihood from the null model with only an intercept. In addition, we also compute the Area Under Curve (AUC) from the fitted model. We find the imputation model fits the data well with a McFadden's R-squared of 0.31 and an AUC of 0.86. As discussed in Section 3.1.2, we find the block imputation model also achieves good out-of-sample performance.

In Table 14, we report the estimated coefficients and average marginal effects (AME) for select variables in model (8). Here we only report the variables with p-value smaller than 0.05 if they are not surgeon or weekday dummies. We make the following observations. First, more elective and urgent cases by other surgeons on the same day (ElecOth$_{s,t}$ and UrgOth$_{s,t}$) decreases the probability that the focal surgeon is assigned a block schedule, i.e., $\Pr(Y_{s,t}=1)$. This reflects the resource sharing among surgeons in the department on the same day, which is discussed for the first IV TotOther$_i$. On the other hand, more elective cases (ElecCur$_{s,t}$) and patients admitted (AdmCur$_{s,t}$) by the focal surgeon increases the probability of being in the block schedule. Next, a late start after 8AM (StartLate$_{s,t}$) by the focal surgeon s decreases the

Table 14 Select Coefficients in the Logistic Model (8)

N = 1,680, R-squared = 0.31, AUC = 0.86

Variable	Coefficient	AME
$\text{ElecOth}_{s,t}$	-0.206^{***}	-0.020^{***}
	(0.059)	(0.006)
$\text{UrgOth}_{s,t}$	-0.140^{*}	-0.014^{*}
	(0.062)	(0.006)
$\text{ElecCur}_{s,t}$	1.297^{***}	0.126^{***}
	(0.300)	(0.028)
$\text{AdmCur}_{s,t}$	0.332^{***}	0.032^{**}
	(0.099)	(0.010)
$\text{NumCurWeek}_{s,t}$	-0.293^{*}	-0.029^{*}
	(0.141)	(0.014)
$\text{DistNext}_{s,t}$	-0.125^{**}	-0.012^{**}
	(0.045)	(0.004)
$\text{StartLate}_{s,t}$	-1.457^{***}	-0.205^{**}
	(0.351)	(0.065)
$\text{WDElecRatio}_{s,t}$	6.598^{***}	0.643^{***}
	(0.790)	(0.081)

Standard errors are reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. Select coefficients for schedule imputation model (8).

probability of being in the block schedule. This suggests that surgeons tend to start operating early when in their block schedules.

The surgeon's recent workload also has explanatory power for $\Pr(Y_{s,t} = 1)$. For example, the number of days worked in the current calendar week ($\text{NumCurWeek}_{s,t}$) and the distance to the next work day ($\text{DistNext}_{s,t}$) are negatively associated with $\Pr(Y_{s,t} = 1)$. Finally, $\text{WDElecRatio}_{s,t}$ denotes the proportion of elective cases by the focal surgeon in $[t - 180, t + 180]$ that fall on the same weekday as t. We see that its coefficient and AME are positive and statistically significant. This can be explained by the fact that the blocks are usually assigned on specific weekdays for each surgeon, and more than 90% of elective cases are performed in their surgeons' block schedule. Thus, a higher ratio $\text{WDElecRatio}_{s,t}$ increases the probability of block schedule for the focal surgeon.

S.3. Estimation Results with Total Incision Time as Workload

In this section, we briefly discuss the estimation results when we use the $SumInc_i$, i.e., total incision time of other cases, as surgeon's daily workload measure. They are largely consistent with our main results in Section 4 with $NumCases_i$ as the daily workload measure. In Section 3.1, we have shown the two IVs satisfy the relevance condition for $SumInc_i$ as daily workload measure.

Table 15 reports the estimation results for the full sample with and without the two IVs. We see that increased surgeon's daily workload is associated with longer surgery duration and worse patient outcomes including post-LOS, total ICU time, as well as likelihood of reoperation and readmission. Specifically, adding one more hour in total incision time of other cases increases the incision time of each case performed by the surgeon by five minutes. Given the average incision time of a case is 4.8 hours, the magnitude of effect here ($5 \times 4.8 = 24$ minutes) is consistent with that estimated by $NumCases_i$ in Table 6. Similarly, we find that one more hour of $SumInc_i$ increases the post-LOS and total ICU time by 0.27 and 0.21 days, as well as increases the likelihood of reoperation and readmission by 0.6 and 1.3 percentage points. The directions and magnitudes of these effects are consistent with those in Table 6 when we measure surgeon's daily workload using the number of cases.

Table 15 Estimated Effects of Daily Workload (Total Incision Time of Other Cases) on Surgery Duration and Patient Outcomes: Full Sample

	Continuous y_i: Coefficients			Binary y_i: AME		
	Incision time	Post-LOS	Total ICU time	Reoperation	Readmission	Mortality
Panel A: Full	0.079^{*} (0.038)	0.271^{**} (0.101)	0.206^{*} (0.081)	0.006^{*} (0.002)	$0.013^{\dagger}$ (0.007)	0.002 (0.002)
Num Obs.	5345	5344	5319	5345	5116	5081
Panel B: Full (no IV)	-0.016 (0.010)	0.008 (0.033)	0.008 (0.016)	-0.001 (0.001)	-0.000 (0.001)	$0.002^{\dagger}$ (0.001)
Num Obs.	5345	5344	5319	5345	5116	5081

The estimated effects of surgeon daily workload (total incision time of other cases) on surgery duration and patient outcomes for the full sample. We report the estimated coefficients in (1) for the three continuous dependent variables, and the AME from (2) for the three binary dependent variables. Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

In Table 16, we show the estimation results with $SumInc_i$ for elective and non-elective cases separately. Here we observe the same heterogeneity in the effects of surgeon's daily workload as in Section 4.2. Specifically, the effect of surgeon's workload $SumInc_i$ on incision time is statistically significant for elective cases, but not for non-elective cases. On the other hand, the effects on patient outcomes are statistically significant only for non-elective patients. Moreover, for patient outcomes, the magnitudes of effects are larger for the non-elective patients than those for the full sample. These observations again are consistent with the results discussed in Section 4.2 when we measure surgeon's daily workload by $NumCases_i$.

S.4. Surgical Scheduling MIQP Formulation

In the following, we formulate the surgical scheduling model used in Section 5. We introduce the notation and decision variables, feasibility constraints, and objective functions. We show the optimization model can be formulated as a Mixed-Integer Quadratic Programming (MIQP) problem.

Table 16 Estimated Effects of Daily Workload (Total Incision Time of Other Cases) on Surgery Duration and Patient Outcomes: Elective and Non-elective Sample

	Continuous y_i: Coefficients			Binary y_i: AME		
	Incision time	Post-LOS	Total ICU time	Reoperation	Readmission	Mortality
Panel A: Elec	0.065^{**} (0.025)	-0.051 (0.203)	0.019 (0.149)	0.003 (0.007)	0.005 (0.012)	0.011 (0.009)
Num Obs.	2474	2474	2454	2394	2398	1897
Panel B: Non-elec	0.096 (0.062)	0.569^{*} (0.253)	0.376^{*} (0.165)	0.010^{*} (0.004)	0.017^{*} (0.008)	-0.000 (0.002)
Num Obs.	2871	2870	2865	2871	2697	2769

The estimated effects of surgeon daily workload (total incision time of other cases) on surgery duration and patient outcomes for the elective and non-elective sample. We report the estimated coefficients in (1) for the three continuous dependent variables, and the AME from (2) for the three binary dependent variables. Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

S.4.1. Decision Variables and Feasibility Constraints

We solve the scheduling model for each calendar week (Sunday to Saturday) in the four year horizon of our sample. For a given week, we index each case by $i \in C$, where C is the set of all cases performed on the weekdays in the given week. We exclude the operations on the weekends (2.8% of sample) in the model, as their times are generally hard to change. According to the surgery status, the set C can be divided into three exclusive subsets C_{el}, C_{ug}, and C_{es}, which represent the elective, urgent, as well as emergent and salvage cases respectively. For each case i, we denote its surgeon by $\tilde{s}(i)$ and original surgery date by $\tilde{t}(i)$.

We index the day in the week by $t \in T$ and the surgeon by $s \in S$, with T and S being the sets of surgery dates and surgeons for cases $i \in C$. We use A_s to denote the set of cases performed by surgeon s. Our optimization model considers which cases to assign to each day. Thus, our decision variables are $x_{i,t}$ for $i \in C$ and $t \in T$. Each $x_{i,t}$ is a binary variable; it takes value one if case i is assigned to day t, and zero otherwise. We formulate the set of constraints to ensure the feasibility of the resulting schedule. First, every case should be assigned one and only one date in the final schedule. This translates to

$$\sum_{t \in T} x_{i,t} = 1, \ \forall i \in C. \tag{10}$$

For each case, we specify its feasible set of surgery dates according to its status. For elective cases, we allow them to be assigned to any day of the week of the original date. This is because elective patients are generally stable, thus each surgeon has high flexibility in scheduling their operations. On the other hand, we impose that urgent cases can only be scheduled on the original date or the adjacent days, while the emergent and salvage cases can only be scheduled on the original date. These constraints reflect the reality that urgent cases are more time sensitive than elective ones as their patients are more severe. In addition, the hospital has little control over the arrival time of emergent and salvage patients. We formulate these constraints as

$$x_{i,t} = 0, \text{ if } i \in C_{ug} \text{ and } |t - \tilde{t}(i)| > 1, \tag{11}$$

and

$$x_{i,t} = 0, \text{ if } i \in C_{es} \text{ and } t \neq \tilde{t}(i). \tag{12}$$

We impose an upper bound on surgeon daily workload, i.e., the number of cases performed by each surgeon in a day, to reflect a physical limit on how much time a surgeon can spend operating. On average, a surgery takes 7.11 hours to complete, with a minimum of 2.15 hours and a median of 6.8 hours. Thus, while surgeons can work overtime and parallelize part of some operations, a reasonable upper bound on the number of cases per day is 2 or 3. As we keep the surgeon assigned to each case unchanged, the number of cases by surgeon s on day t in the new schedule can be expressed as,

$$\tilde{n}_{s,t} = \sum_{i \in A_s} x_{i,t}. \tag{13}$$

The summation on the right-hand side includes all the cases by surgeon s. Then, the constraint on surgeon daily workload can be formulated as

$$\tilde{n}_{s,t} \leq \max\{\bar{n}^{(c)}, n_{s,t}^{(c)}\}, \ \forall t \in T \text{ and } \forall s \in S, \tag{14}$$

where $n_{s,t}^{(c)}$ is the number of cases performed by surgeon s on day t in the original schedule; $\bar{n}^{(c)}$ is a model parameter to be specified. It denotes the maximum daily workload of a surgeon in the new schedule, unless the surgeon already performs more cases in the original schedule.

Finally, we set an upper bound on the number of days worked by each surgeon in a week. Although asking the surgeons to work for more days naturally smooths their daily workload, it would be difficult to implement in reality given their other responsibilities. Note that the surgeon s works on day t in the new schedule if at least one case is performed, i.e., $\tilde{n}_{s,t} > 0$. Thus, we can formulate the constraint as

$$\sum_{t \in T} \mathbf{1}\{\tilde{n}_{s,t} > 0\} \leq \max\{\bar{n}^{(d)}, n_s^{(d)}\}, \ \forall s \in S. \tag{15}$$

where $n_s^{(d)}$ is the number of working days by surgeon s in the original schedule; $\bar{n}^d$ is the model parameter denoting the maximum number of days worked by a surgeon, unless the surgeon works for more days in the original schedule.

The constraint (15) is inconvenient to implement as the indicator function is non-linear. We circumvent this difficulty by proposing the following linear formulation. Let the binary variable $z_{s,t}$ denote whether surgeon s works on day t in the new schedule. We bound it by

$$z_{s,t} \leq M \cdot \tilde{n}_{s,t} \text{ and } z_{s,t} \geq m \cdot \tilde{n}_{s,t}, \tag{16}$$

where M (resp. m) is a sufficiently large (resp. small) constant. In our study, we can set them as $M = 100$ and $m = 0.01$. It is easy to verify by (16) that $z_{s,t}$ takes value one if $\tilde{n}_{s,t} > 0$ and zero if $\tilde{n}_{s,t} = 0$. Thus, it always equals to the indicator function $\mathbf{1}\{\tilde{n}_{s,t} > 0\}$. Then, we can rewrite the constraint (15) in the following linear form as

$$\sum_{t \in T} z_{s,t} \leq \max\{\bar{n}^d, n_s^{(d)}\}, \ \forall s \in S. \tag{17}$$

In the model formulation, we do not impose an upper bound on the total number of cases by all surgeons in a day. After we solve for the optimal schedule, we perform feasibility checks to show that the new schedules do not lead to significant increase in the total number of cases in a day. In summary, our model includes the constraints (10), (11) – (12), (13) – (14), and (16) – (17), all of which are formulated in linear form.

S.4.2. Objective Functions and MIQP Formulation

We now introduce the objective function for our model and show how to formulate the surgical scheduling model as an MIQP problem. We consider three alternative objective functions: minimizing the total expected incision time, post-LOS, or total ICU time. Following our econometric model (1), the expected value of the three variables can be expressed as

$$\hat{y}_i = X_i \beta + \gamma NumCases_i. \tag{18}$$

Here we use the number of other cases as the surgeon daily workload measure. The variable $\hat{y}_i$ is specified as incision time, post-LOS, and total ICU time, respectively. The coefficient γ is reported in Section 4 for each dependent variable. We also allow for the heterogeneity in the impacts of daily workload for elective and non-elective cases.

By (18), the expected value $\hat{y}_i$ can be decomposed to two parts

$$l_i = X_i \beta \text{ and } d_i = \gamma NumCases_i.$$

To focus on the impact of daily workload, we assume the first part l_i, which primarily depends on the patient's risk and operative factors, remains unchanged in the new schedule. However, the second part d_i will be affected by the surgeon's workload in the new schedule. Our objective is to minimize the total expected value $\hat{y}_i$, i.e., $\min \sum_{i \in C} \hat{y}_i$. As we assume l_i does not change, this is equivalent to minimizing the sum of the d_i, which is

$$\min \sum_{i \in C} d_i.$$

We now explicitly express the objective $\min \sum_{i \in C} d_i$ under the new schedule. The total daily workload term for surgeon s on day t is

$$d'_{s,t} = \sum_{i \in A_s} d_i x_{i,t}.$$

It is straightforward to see the summation over all cases in C is equal to that over all surgeons and days:

$$\sum_{i \in C} d_i = \sum_{s \in S} \sum_{t \in T} d'_{s,t}. \tag{19}$$

As such, it is sufficient to write out the objective function using $d'_{s,t}$ instead of d_i.

The number of cases performed by surgeon s on day t is given by $\tilde{n}_{s,t}$ in (13). To account for the heterogeneous effects as discussed in Section 4.2, we need to further obtain the number of elective and non-elective cases. Similar to (13), they are given by

$$\tilde{n}^{(el)}_{s,t} = \sum_{i \in A_s} x_{i,t} \mathbf{1}\{i \in C_{el}\} \text{ and } \tilde{n}^{(ne)}_{s,t} = \sum_{i \in A_s} x_{i,t} \mathbf{1}\{i \notin C_{el}\}.$$

Then, the total impact from daily workload for surgeon s on day t can be expressed as

$$d'_{s,t} = (\tilde{n}_{s,t} - 1) \cdot \big(\gamma^{(el)} \tilde{n}^{(el)}_{s,t} + \gamma^{(ne)} \tilde{n}^{(ne)}_{s,t}\big).$$

Here $\gamma^{(el)}$ and $\gamma^{(ne)}$ are the estimated coefficient γ for the daily workload effect on incision time, post-LOS, or total ICU time – depending on which one we are to minimize – for the elective and non-elective cases respectively, which are reported in Table 7 of Section 4.1. We set the coefficient to be zero if it is not statistically significant at the 10% level. When we ignore the heterogeneity in the impacts of daily workload, we use the average treatment effects in Table 6 with $\gamma^{(el)} = \gamma^{(ne)} = \gamma^{(avg)}$. Plugging $d'_{s,t}$ into (19), the objective function is given by

$$\min \sum_{s \in S} \sum_{t \in T} (\tilde{n}_{s,t} - 1) \cdot \big(\gamma^{(el)} \tilde{n}^{(el)}_{s,t} + \gamma^{(ne)} \tilde{n}^{(ne)}_{s,t}\big),$$

Note that the summation terms $\sum_{t \in T} \sum_{s \in S} \tilde{n}^{(el)}_{s,t}$ and $\sum_{t \in T} \sum_{s \in S} \tilde{n}^{(ne)}_{s,t}$ represent the total numbers of elective and non-elective cases from all surgeons in the week, thus they remain unchanged in the new schedule. Then the above objective can be simplified as

$$\min \sum_{s \in S} \sum_{t \in T} \tilde{n}_{s,t} \cdot \big(\gamma^{(el)} \tilde{n}^{(el)}_{s,t} + \gamma^{(ne)} \tilde{n}^{(ne)}_{s,t}\big). \tag{20}$$

It is easy to verify that the objective function is quadratic in the decision variables $x_{i,t}$. Thus, our model is formulated as an MIQP with quadratic objective (20) and linear constraints (10), (11) – (12), (13) – (14), and (16) – (17). The decision variables are $x_{i,t}$ for $i \in C$ and $t \in T$, as well as $z_{s,t}$ for $s \in S$ and $t \in T$ as introduced in (16). All the decision variables are binary. The final MIQP formulation to minimize total incision time, post-LOS, or ICU time is given below:

$$\min \quad \sum_{s \in S} \sum_{t \in T} \tilde{n}_{s,t} \cdot \big(\gamma^{(el)} \tilde{n}^{(el)}_{s,t} + \gamma^{(ne)} \tilde{n}^{(ne)}_{s,t}\big)$$

such that

$$
\begin{aligned}
&\tilde{n}_{s,t} = \sum_{i \in A_s} x_{i,t}, \\
&\tilde{n}^{(el)}_{s,t} = \sum_{i \in A_s} x_{i,t} \mathbf{1}\{i \in C_{el}\}, \\
&\tilde{n}^{(ne)}_{s,t} = \sum_{i \in A_s} x_{i,t} \mathbf{1}\{i \notin C_{el}\}, \\
&\sum_{t \in T} x_{i,t} = 1, \ \forall i \in C, \\
&x_{i,t} = 0, \ \text{if } i \in C_{ug} \text{ and } |t - \tilde{t}(i)| > 1, \\
&x_{i,t} = 0, \ \text{if } i \in C_{es} \text{ and } t \neq \tilde{t}(i), \\
&\tilde{n}_{s,t} \le \max\{\bar{n}^{(c)}, n^{(c)}_{s,t}\}, \ \forall t \in T \text{ and } \forall s \in S, \\
&z_{s,t} \le M \cdot \tilde{n}_{s,t}, \\
&z_{s,t} \ge m \cdot \tilde{n}_{s,t}, \\
&\sum_{t \in T} z_{s,t} \le \max\{\bar{n}^{d}, n^{(d)}_{s}\}, \\
&x_{i,j}, \, z_{s,t} \in \{0, 1\}.
\end{aligned}
$$

The constants are set as $M = 100$ and $m = 0.01$. The model parameters $\bar{n}^{(c)}$ and $\bar{n}^{(d)}$ denote the upper bound on surgeon's number of cases performed in a day and number of days worked in a week, respectively.

S.4.3. Robustness Check with Restrictive Conditions

As a further robustness check for the results in Section 5, we solve the surgical scheduling model under more restrictive conditions with $\bar{n}^{(c)} = 2$ and $\bar{n}^{(d)} = 3$. That is, we decrease the maximum daily workload of surgeons to two cases a day and the maximum number of working days to three days a week, unless the corresponding levels are larger in the original schedule. These conditions are more restrictive than those considered in Section 5 with $\bar{n}^{(c)} = 3$ and $\bar{n}^{(d)} = 4$.

Under these restrictive conditions, we still find substantial improvements from the resulting schedules, although the magnitude of benefits becomes smaller. Specifically, the corresponding new schedules reduce the total incision time, post-LOS, and ICU time by 1.25%, 1.87%, and 3.38%, respectively when we use the average effects. The reductions change to 0.99%, 3.45%, and 5.39% when we use heterogeneous effects. The benefits are still economically important. For example, the potential profit gain from the reduction of incision time can be as large as $174,500 a year. In addition, we still achieve improvement for most of the 208 weeks in our sample (184 with average effect and 197 with heterogeneous effect). This robustness check further demonstrates the benefit of incorporating the potential effect of surgeon daily workload in surgical scheduling.

S.5. Supplementary Tables

This section includes the supplementary tables. Table 17 summarizes the correlation between the two IVs and 21 observable severity factors of patients. Table 18 shows the coefficients of the two IVs in (3) when we estimate (2) and (3) by full MLE for binary outcomes. Table 19 reports the estimated coefficients of IVs on daily workload from elective and non-elective samples. Tables 20 and 21 show the estimated coefficient γ for the two workload measures in (2) for the three binary outcomes. The robustness check results under different specifications are summarized in Tables 22 and 23. Due to the smaller sample sizes of the block and standard procedure samples, we do not estimate the heterogeneous effects on binary outcomes (i.e., by elective and non-elective cases separately) using those samples. Recall the rates of adverse outcomes are generally low in our sample, especially for standard procedures; thus, we do not have a large enough sample size for these regressions to converge.

In addition, as a robustness check for the estimation results of post-LOS, we use several more conservative measures of post-LOS by adjusting the original measure using the OR entry or exit time of each case. The estimated coefficients of $NumCases_i$ by TSLS are given in Table 24. The first column shows the original results in Tables 6 and 7, in which we compute the post-LOS as the number of days between the OR exit and discharge dates. In the second column ("Entry $\geq$ 3PM"), we subtract a day from the post-LOS if the OR entry time of the case is later than 3PM as there is some evidence that late surgery start times are associated with an increase of LOS by one day. In the third column ("OR Exit Hour"), we subtract the hours elapsed before OR exit on the day of OR exit. In the last two columns, we further subtract a day from the post-LOS if the patient leaves the OR after 12PM and 4PM, respectively. We see from Table 24 that our estimated effects for post-LOS remain similar and robust in all these conservative measures.

Table 17 Correlation between IVs and Observable Severity Factors

Factor	$TotOther_i$	$GapNext_i$
Gender: Male	−0.019	0.003
Status: non-electives	−0.096	0.084
Age	0.054	−0.042
NYHA: III or IV	0.002	0.059
Endocarditis	−0.017	0.035
Previous intervention	−0.031	0.045
Peripheral arterial disease	0.002	0.040
Incidence	−0.030	0.055
Lung disease	0.023	0.038
Hypertension	0.034	−0.025
Drug use	−0.017	0.022
Cancer	0.008	0.009
Carotid stenosis	0.027	0.016
Cardiogenic shock	−0.052	0.097
Syncope	0.001	−0.010
MI	−0.007	0.014
Dialysis	−0.013	0.041
Diabetes	0.005	0.005
Systolic pressure: high	0.009	0.040
Liver disease	0.020	0.035
Thoracic aorta disease	−0.020	−0.049
Smokes	−0.006	0.029

Table 18 Estimated Coefficients of IVs by Full MLE of (3) and (2) (Full Sample)

	Reoperation		Readmission		Mortality	
IV	$NumCases_i$	$SumInc_i$	$NumCases_i$	$SumInc_i$	$NumCases_i$	$SumInc_i$
$TotOther_i$	−0.078***	−0.385***	−0.072***	−0.359***	−0.076***	−0.375***
	(0.008)	(0.051)	(0.008)	(0.052)	(0.008)	(0.052)
$GapNext_i$	0.005	0.042†	0.007†	0.053*	0.007†	0.052*
	(0.004)	(0.022)	(0.004)	(0.026)	(0.004)	(0.026)
Num Obs.	5345	5345	5116	5116	5081	5081

Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. Coefficients and standard errors of two IVs in (3) when we estimate (2) and (3) jointly by full MLE for reoperation, readmission, and mortality.

Table 19 Impact of IVs on Daily Workload (Elective and Non-elective Sample)

	Elective Sample		Non-elective Sample	
IV	$NumCases_i$	$SumInc_i$	$NumCases_i$	$SumInc_i$
$TotOther_i$	-0.067^{***}	-0.342^{***}	-0.087^{***}	-0.418^{***}
	(0.008)	(0.039)	(0.008)	(0.040)
$GapNext_i$	0.006	0.053^{*}	0.005	0.042^{*}
	(0.005)	(0.026)	(0.004)	(0.021)
Num Obs.	2474	2474	2871	2871
Adj R^2	0.123	0.143	0.168	0.179

Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$. Coefficients and standard errors of two IVs in (3) for elective and non-elective samples.

Table 20 Estimated Coefficients of Daily Workload in (2) for Binary Outcomes: Full Sample

	Reoperation		Readmission		Mortality	
	$NumCases$	$SumInc_i$	$NumCases$	$SumInc_i$	$NumCases$	$SumInc_i$
Panel A: Full	0.288^{***}	0.054^{**}	0.379^{*}	0.077^{*}	0.203	0.040
	(0.083)	(0.020)	(0.186)	(0.038)	(0.191)	(0.035)
Num Obs.	5345	5345	5116	5116	5081	5081
Panel B: Full (no IV)	-0.021	-0.005	-0.003	-0.002	0.189^{*}	$0.034^{\dagger}$
	(0.040)	(0.006)	(0.052)	(0.008)	(0.085)	(0.018)
Num Obs.	5345	5345	5116	5116	5081	5081

Estimated coefficient γ in (2) for the full sample. Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Table 21 Estimated Coefficients of Daily Workload in (2) for Binary Outcomes: Elective and Non-elective Sample

	Reoperation		Readmission		Mortality	
	$NumCases$	$SumInc_i$	$NumCases$	$SumInc_i$	$NumCases$	$SumInc_i$
Panel A: Elec	0.290	0.045	0.211	0.030	0.994^{*}	0.196^{*}
	(0.522)	(0.099)	(0.394)	(0.074)	(0.460)	(0.088)
Num Obs.	2,468	2,468	2,406	2,406	1,901	1,901
Panel B: Non-elec	0.332^{*}	0.066^{*}	0.428^{*}	0.094^{*}	-0.010	-0.000
	(0.133)	(0.028)	(0.183)	(0.040)	(0.175)	(0.033)
Num Obs.	2871	2871	2697	2697	2769	2769

Estimated coefficient γ in (2) for the elective and non-elective samples. Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Table 22 Estimated Effects of Daily Workload (Number of Other Cases) on Surgery Duration and Patient Outcomes: Different Specifications

Specification	All	Elective	Non-elec	All	Elective	Non-elec	All	Elective	Non-elec
One IV	0.411†	0.323*	0.485	1.475*	-0.154	2.862†	1.161**	0.360	1.910*
	(0.211)	(0.152)	(0.332)	(0.573)	(1.192)	(1.521)	(0.442)	(0.839)	(0.944)
Winsorize: 95th pct	0.377*	0.341***	0.433	1.014**	0.324	1.798*	0.603*	0.385	0.925*
	(0.191)	(0.095)	(0.306)	(0.369)	(0.924)	(0.820)	(0.286)	(0.601)	(0.442)
Log-transform	0.102*	0.099***	0.113	0.097*	0.057	0.158†	0.096	0.120	0.131
	(0.042)	(0.026)	(0.074)	(0.044)	(0.075)	(0.085)	(0.071)	(0.146)	(0.115)
Num Obs.	5345	2474	2871	5344	2474	2870	5319	2454	2865
Block Sample	0.445†	0.166	0.732	0.896	0.451	3.415**	1.095	0.768	2.368**
	(0.250)	(0.199)	(0.459)	(1.202)	(1.219)	(1.161)	(0.740)	(1.044)	(0.914)
Num Obs.	2489	1182	1307	2489	1182	1307	2482	1177	1305
Standard Types	0.277†	0.458***	0.158	0.785	0.787	0.995	0.737**	1.146†	0.638
	(0.165)	(0.064)	(0.281)	(0.733)	(0.989)	(1.281)	(0.233)	(0.671)	(0.572)
Num Obs.	3416	1717	1699	3416	1717	1699	3407	1709	1698

Estimated coefficient γ in (1) for three continuous dependent variables under different specifications. Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Table 23 Estimated Effects of Daily Workload (Number of Other Cases) on Binary Outcomes: Different Specifications

	Reoperation	Readmission	Mortality
One IV	0.037***	0.064†	0.010
	(0.010)	(0.034)	(0.010)
Num Obs.	5345	5116	5081
Block Sample	0.062†	0.018	-0.001
	(0.034)	(0.054)	(0.024)
Num Obs.	2487	2354	2267
Standard Types	0.020	0.006	0.007
	(0.015)	(0.017)	(0.012)
Num Obs.	3413	3333	3112

Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Table 24 Estimated Effects of Daily Workload (Number of Other Cases) on Post-LOS with Different Adjustments

Original	Entry $\geq$ 3PM	OR Exit Hour	Exit $\geq$ 12PM	Exit $\geq$ 4PM
Panel A: Full Sample				
1.402^{*}	1.679^{**}	1.450^{**}	1.348^{*}	1.548^{**}
(0.547)	(0.555)	(0.537)	(0.542)	(0.498)
Panel B: Non-elective Sample				
2.796^{*}	3.130^{*}	2.848^{*}	$2.733^{\dagger}$	2.941^{*}
(1.417)	(1.428)	(1.416)	(1.421)	(1.374)

Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.