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

In many service systems, an 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 effect of workload has been primarily considered at the system and instantaneous level rather than the individual and cumulative 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 the individual surgeon's daily workload. Using a detailed data set of more than 5,600 cardiac operations in a large hospital, we quantify how individual surgeon daily workload (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 operational factors of the cardiac surgery department, including the regular surgery schedule of surgeons. We find that high daily workload for the focal surgeon is associated with longer OR times and worse patient outcomes. Specifically, a surgeon's higher daily workload leads to longer post-surgery length-of-stay in the ICU and hospital. These results highlight the potential negative impact of high individual surgeon workload. We develop a surgical scheduling model that incorporates the estimated impact of surgeon daily workload. We solve the model by mixed-integer quadratic programming and show that our proposed schedule can substantially reduce total OR time and post-surgery length-of-stay. Our results suggest that hospitals should take into account the effects of individual surgeon 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

Operating rooms are significant cost sinks as well as revenue generators of US hospitals (Rothstein and Raval 2018). They are extremely expensive medical resources, with estimated costs of up to 37 dollars per minute (Childers and Maggard-Gibbons 2018). At the same time, operating room generate nearly 50% of hospitals' total revenue (McDermott et al. 2017). Thus, improving the efficiency of how operating rooms are utilized is essential for hospitals to achieve their clinical and financial goals. Not surprisingly, various efforts have been devoted to operating room management from hospital administrators, policy makers, and healthcare professionals. Common recommendations include increasing capacity, better training of surgeons and staff, as well as managing schedules in order to facilitate more timely care for patients (see, e.g., May et al. 2011). However, most of the literature on surgical scheduling focus on the modeling and algorithmic side without identifying causal determinants of the duration and the outcomes of the operation. In this study, we focus on a novel factor that affects the operating room performance. Specifically, we investigate how individual surgeon's daily workload impacts the surgery duration and outcomes in the context of cardiac operations, and, in turn, how such impacts can be utilized to improve surgery scheduling.

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. However, a growing body of empirical research shows that the service time of human-serviced systems can be endogenously impacted by the overall system 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 is important for 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), emergency departments (Kc 2014, Batt and Terwiesch 2016), as well as paramedic teams (Bavafa and Jónasson 2023). Beyond service time, the effect of system workload on the quality of care, such as mortality and readmission, has also been investigated in both the operations management and medical communities (e.g., Kc and Terwiesch 2009, Kc and Terwiesch 2012, and Needleman et al. 2011). In this paper, we investigate the impact of workload in the setting of operating rooms. Based on our empirical findings, we then show how hospitals can improve surgical scheduling by incorporating such effects.

We consider a novel type of workload using a detailed data set of cardiac operations. In particular, we measure individual surgeon's daily workload by the total number of operations performed on a given day. In most of the existing literature in healthcare settings, the workload is measured at the system level in an instantaneous way. A common example of this type is the hospital unit's bed occupancy at the time of a patient's admission or discharge (e.g., Kc and Terwiesch 2012, Kuntz et al. 2015, and Berry Jaeker and Tucker 2017). In contrast, we focus on the *cumulative* workload at the *individual* surgeon level, which, to our best knowledge, is the first in the field of operations management. For cardiac operations, surgeons often have high ownership of their patients, thus the workload at the individual level would be a more relevant measure than that at the system (department) level. In addition, the effect of workload at the individual level may be subject to more behavioral and operational variations. On the other hand, cardiac operations usually take a long time to complete and are highly demanding for surgeons. Consequently, high cumulative workload in a day may have negative impacts on surgeon's performance. For these reasons, the effect of individual surgeon's cumulative workload may differ from that of the system-level, instantaneous workload. Understanding such effects can provide new operational levers for hospitals to improve their performance. For example, while the department-level workload is hard to change, the hospitals may have more flexibility in changing daily workload of individual surgeons in their scheduling.

In our study hospital, it is common for a cardiac surgeon to perform multiple operations a day: 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 other ancillary resources, such as nurses and post-surgery recovery beds. In this study, our goal is to understand the effect of surgeon workload on operating time and surgery outcomes. Due to data limitations (e.g., lack of shift schedule of ancillary medical staff and nurses), we are not able to fully differentiate the driving factors leading to such effect, e.g., surgeon fatigue versus operational constraints. However, we provide suggestive evidence to show that surgeon fatigue is the more likely driving factor. Based on our empirical findings, we develop and solve a surgical scheduling model that incorporates the estimated effect of surgeon daily workload. We show that our proposed schedule can substantially reduce surgery duration and patient's post-surgery length-of-stay.

We examine the impact of surgeon 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 and cumulative level – on multiple outcomes. First, we examine how surgeon daily workload affects the surgery duration of each case, as measured by its Operating Room (OR) time or incision time. This sheds light on the relationship between individual server's 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. We also examine whether the treatment effect of surgeon workload is heterogeneous for different types of patients, e.g., elective patients and non-elective patients.

Our detailed data set allows us to control for a comprehensive set 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 can be endogenous. This is because there are likely risk factors that are considered by the surgeons when they schedule their cases, but not all of these factors are observed in the data (e.g., patient's cognitive status). These unobserved factors will affect both surgeon daily workload and 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 unobserved in the data, it will generate a negative bias in the estimated effect of surgeon daily workload. Alternatively, a high risk case may be squeezed in so as to avoid delaying care for that patient, resulting in a positive bias in the estimated effect of surgeon daily workload.

We handle the endogeneity bias by utilizing an instrumental variable (IV) approach. A valid IV in our study should influence the surgical outcomes only via the surgeon daily workload, conditioned on the exogenous variables. 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 tend to limit the daily workload of the focal surgeon. The second IV is the proportion of the focal surgeon's elective cases performed on the same weekday as the focal case over a long horizon. Surgeons in our collaborating hospital tend to schedule their operations on certain weekdays in a week, especially for elective patients. The second IV thus captures the impact of such long-term working pattern on surgeon daily workload. We validate the two IVs empirically with our data and show that they indeed significantly impact surgeon's daily workload, after controlling for a comprehensive set of demographic, risk, and operational factors. We additionally provide evidence that the two IVs are unlikely to be correlated with patient's unobservable factors. 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 surgery duration and worse outcomes for their patients. Specifically, we estimate that adding one more case to a surgeon's daily workload increases the OR time by 27.3 minutes for each case performed by the surgeon in the day. This is a 6.5% relative increase. Similar impact is also observed for the incision time. These results highlight how workload affects service time when the workload level is already high, which is consistent with the second tipping point empirically observed in Berry Jaeker and Tucker (2017), albeit in a very different clinical setting. We find that surgeon daily workload also leads to longer post-surgery LOS of patients in both the ICU and the hospital: when the surgeon does one more case in a day, the affected patients are expected to stay in the ICU (resp. hospital) for 1.1 (resp. 1.2) more days after their operations on average. These results highlight the negative impacts of high surgeon daily workload on surgery duration and patient outcomes. We further show that there is substantial heterogeneity in the effect of surgeon daily workload for elective and non-elective (urgent and emergent) patients. We find that the effects of surgeon daily workload on post-surgery LOS are significant only for non-elective patients and have larger magnitudes than those for the full sample. One possible explanation for such heterogeneity is that the non-elective patients are generally more severe, thus their surgical outcomes 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. For the importance of operating rooms, the literature on surgical scheduling is large (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, which is determined by the surgical schedule itself. We thus propose a scheduling model that accounts for such effect. We consider the intervention of changing the surgery dates 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 a cost neutral intervention. We formulate and solve the model as a mixed-integer quadratic programming problem. Using our estimated effects, we find the new schedules from our model can substantially reduce the total OR time and post-LOS, which are economically important for the hospital. This highlights the benefits of accounting for the impact of surgeon daily workload in surgery scheduling. Our work reveals the benefit of load balancing at the *individual* level in the context of cardiac surgery scheduling.

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 primary streams of literature: (1) effect of system workload on service rate and quality, (2) volume-outcome relationship, (3) 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, which focuses on the relationship between system workload and service rate. The dynamic queueing literature has studied the optimal policies and system performance when service rate is adaptive to system state (e.g., George and Harrison 2001, Delasay et al. 2016). 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) 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 can be partially reconciled by an inverted-U shape pattern. 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, such as server speedup (Kc and Terwiesch 2009), task reduction (Oliva and Sterman 2001), multitasking (Freeman et al. 2017), and server fatigue (Kuntz et al. 2015).

There is also a rich literature studying the effect of workload on servers' behavior and quality. 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 settings, 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.

In most of the existing healthcare literature, workload is measured at the system level, e.g. bed occupancy in different hospital units (e.g., Kc and Terwiesch 2012, Kuntz et al. 2015, and Berry Jaeker and Tucker 2017). In contrast, the impact of individual server's workload is relatively understudied. Different from these works, we consider the workload at the individual surgeon level, as it is more relevant in the cardiac surgery setting. On the other hand, the existing literature largely focuses on the *current* workload level using the instantaneous, or at least recent, system state, e.g., the unit's bed occupancy at the time of patient's admission or discharge (e.g., Kc and Terwiesch 2012, Kim et al. 2015, and Berry Jaeker and Tucker 2017). Instead, we study the effect of surgeon's *cumulative* workload using the total number of operations performed in a day. In a survey paper, Delasay et al. (2019) develop a general framework to describe the impact of workload on service times. Our workload measure resembles the *extended load* in their framework, which tracks the history of workload, but is measured at the individual level. We empirically show the negative effect of extended load at the individual server level on service time and outcome.

Two related papers in healthcare operations management also consider the workload at the individual level. Kc (2014) uses operational data at the individual level like in our study. 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 daily workload, where patients are served in sequence. The recent work of Bavafa and Jónasson (2023) measures the fatigue level of paramedic crews using the cumulative number of prior jobs completed during a shift. They show that workers' fatigue increases the mean and uncertainty of the service time. However, their clinical setting and empirical approach are very different from ours (ambulance service versus cardiac operation). In addition to average service time, we also reveal the impact of surgeon daily workload on the surgical outcomes and develop a scheduling model that incorporates such effects.

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 (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). Research in different empirical settings has been conducted to investigate the drivers and mechanisms behind the relationship, e.g., learning and specialization. Relevant works in this area include Kc and Terwiesch (2011), Kc and Staats (2012), Clark and Huckman (2012), and Staats and Gino (2012) among others. Recent work by Wang and Pourghannad (2020) shows that 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., 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 survey in Janhofer et al. 2019). However, the empirical results on the relation between surgeon fatigue and worse patient outcomes are mixed in the medical literature (see, e.g., Thomas et al. 2012, Govindarajan et al. 2015). Our work sheds light on this important problem using a detailed empirical data set of cardiac surgery and rigorous econometric analysis. As a limitation of our study, we do not have direct measure on surgeon fatigue e.g., obtaining the surgeon sleeping time and mental state via survey, thus we acknowledge that there may be other operational factors that contribute and/or explain the negative effects of surgeon daily workload. That said, medical and psychological literature has shown that high cumulative workload can lead to server fatigue (Thomas et al. 2012, Hockey 2013). In addition, the cumulative number of task completions has been commonly used as a measure for worker fatigue (Bavafa and Jónasson 2023). Given the long duration of cardiac operations, we thus regard surgeon fatigue as one plausible explanation for our findings and provide suggestive evidence for this channel. Different from existing medical literature, which focuses primarily on correlational analysis, we use IVs to control for the endogeneity and generate causal estimates.

Finally, we 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 (see May et al. 2011 for a review). Different objectives are considered in operating room scheduling, including minimizing costs, maximizing profit and utilization, as well as smoothing downstream census (e.g., Freeman et al. 2016 and Zenteno et al. 2016). 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 reserve operating room capacity. However, most of the existing literature assumes the surgery duration to be exogenous and independent of surgeon workload. To our best 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 Wang and Pourghannad (2020), in which the surgery duration is affected by the surgeon's past volume. Our work differs from theirs in two important aspects. First, we focus on the impact of surgeon daily workload and apply an IV method to address the endogeneity bias. Next, we consider the assignment of operations to available days, instead of matching patients and surgeons as in their work.

2. Data and Clinical Setting

2.1. Data Selection

In this study, we use cardiac surgery data from a large academic hospital over the period of July 2015 to July 2019. The data is 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 analyzing the impact of daily workload, as measured by the total number of surgical operations per day. 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). We describe the main variables in the STS data in next section.

We start from 5,604 cases from the STS data in the four year horizon. 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 year 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.

2.2. Patient Risk Factors and Surgery Metrics

The STS data set provides a comprehensive set of characteristics of each patient, which allows us to control for the patient's severity and condition. It 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, as well as preoperative conditions such as whether the patient experienced heart failure, cardiogenic shock, or myocardial infarction (MI) before the operation. Table 1 reports the summary statistics of patients' gender, age, and critical status for both the full, elective, and non-elective samples. Specifically, a patient

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

is classified as critical if he or she experiences a cardiogenic shock or syncope before the operation. In Section S.1 of Online Supplement, we provide a detailed description of other risk factors in our econometric framework and their summary statistics.

Table 1 Summary Statistics of Patients for Full, Elective, and Non-elective Samples (Full: N = 5,352, Elective: N = 2,480, Non-elective: N = 2,872)

	Full Sample			Elective Sample			Non-elective Sample		
	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std
Gender: Male	0.675	-	-	0.655	-	-	0.692	-	-
Age	64.73	66.00	12.56	66.08	68.00	12.12	63.59	65.00	12.79
Critical	0.103	-	-	0.031	-	-	0.166	-	-

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 sample. 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. We then classify the operations to different types to control for the differences in their procedures. The classification of surgery type is described in Section S.1 of Online Supplement.

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 calculate the OR time of each case 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

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

relatively high level of participation of the surgeon. Thus, the incision time is a more relevant measure for a surgeon's working time than the OR time.

We present the summary statistics for the OR time and its three stages by the four surgery status in the top of Table 2. The standard deviation is reported in the parenthesis. We 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 stage 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. We also find 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 (4.78 hours) consists of 67% of the OR time (7.11 hours).

Table 2 Summary Statistics of Surgery Metrics and Patient Outcomes

	Elective	Urgent	Emergent	Salvage	All
Pre-incision time (hours)	1.48 (0.28)	1.51 (0.31)	1.48 (0.45)	1.18 (0.51)	1.49 (0.31)
Incision time (hours)	4.52 (1.56)	4.88 (1.76)	5.80 (2.31)	5.96 (2.27)	4.78 (1.75)
Post-incision time (hours)	0.73 (0.38)	0.78 (0.40)	0.87 (0.47)	0.95 (0.49)	0.76 (0.40)
OR time (hours)	6.79 (1.79)	7.24 (1.99)	8.31 (2.59)	8.28 (2.60)	7.11 (1.99)
Pre-surgery LOS (days)	1.16 (2.92)	4.90 (9.83)	15.32 (30.11)	8.09 (7.54)	3.90 (11.19)
Post-surgery LOS (days)	8.73 (7.85)	13.22 (18.53)	25.60 (22.27)	20.45 (12.15)	12.02 (15.55)
Total ICU time (days)	3.61 (5.55)	5.88 (11.36)	13.29 (17.35)	15.59 (13.03)	5.37 (10.09)
Number	2479	2488	374	11	5352

Finally, we 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. We calculate 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. We also have the total time a patient spends in ICU after the operation, including both the initial ICU visit and the potential revisits. Their summary statistics are reported in the bottom of Table 2. First, we can see that the elective cases have relatively short pre-LOS. This is because most of the elective patients are admitted one day before or on the same day of their operation, reflecting the flexibility in their schedule. Second, the elective cases have the shortest post-LOS and total ICU time, 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 usually follow typical post-surgery protocols.

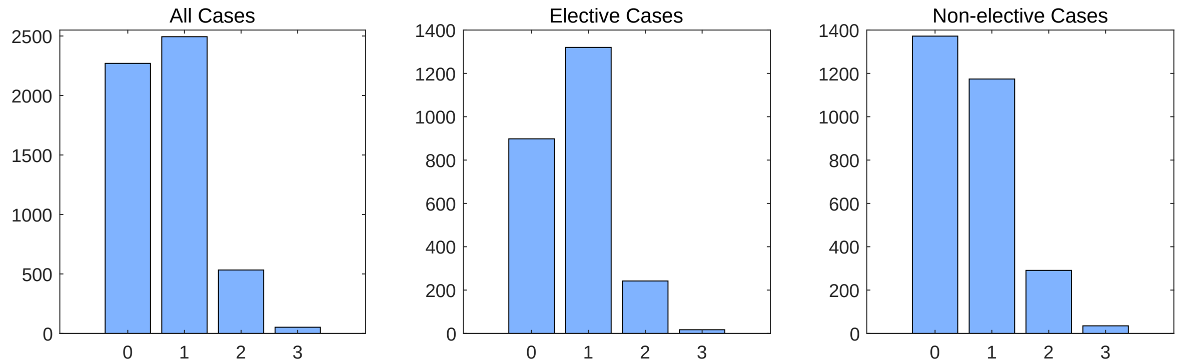

Figure 1 Frequency distributions of $Workload_i$ (number of additional cases excluding focal case i) for full (left), elective (middle), and non-elective (right) samples

3. Econometric Framework

We now develop the econometric framework for identifying the effect of surgeon daily workload on surgery duration and outcomes. For each case i, denote its surgeon and surgery date by s and t, respectively. $Workload_{ist}$ is the daily workload of case i's surgeon s on day t. We measure $Workload_{ist}$ by the number of cases in addition to i performed by surgeon s on day t.Note that if the surgeon performs only one case in a day, we would have $Workload_{ist}$ equal to zero by its definition. 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 note that in addition to the operations recorded in the STS data, the surgeon may have other duties which may also contribute to their overall daily workload. These duties are not included in the data, and thus are not accounted in our study. That said, we expect our measure, i.e., number of operations performed, captures the primary workload of surgeons, as cardiac surgeries are extremely complicated procedures and take a long time to complete. As a robustness check, we also measure $Workload_{ist}$ by the total incision time of other cases performed by surgeon s on day t, which takes continuous values. The two workload measures are highly correlated. The estimation results are very similar and available upon requests.

Figure 1 shows the distribution of $Workload_i$ (number of additional cases) for the full, elective, and non-elective samples. We see that it is very common for a surgeon to perform multiple cases – i.e., $Workload_i \geq 1$, in a day. For at least half of the cases, their surgeons perform at least two cases in total on their surgery date. This holds for the full sample as well as the elective and non-elective samples. Moreover, there is a substantial proportion of cases for which the surgeon's total daily workload is three or four cases. Such high daily workload of surgeons may affect surgery duration and patient outcomes.

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 five operational variables: dummy variables for the day of the week, month, and year of the operation, the pre-LOS of patient, and the cardiac patient census in the hospital. The *cardiac patient census* is calculated 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 to measure 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 also expect the cardiac patient census to be highly correlated with the downstream cardiac ICU congestion. The cardiac patient census has a mean of 62.3 patients 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, under a reasonable assumption that they spend their total ICU time in their first ICU visit after exiting the OR. We find that including the cardiac surgery ICU census in the analysis does not affect our estimation results.

In total, we have 82 covariates in our model, including demographic, medical, and operative factors, as well as system congestion level, and dummies for time and surgeon fixed effects. We provide a detailed description of these independent variables used in Section S.1 of Online Supplement. We represent these variables and a constant by X_i for case i. To estimate the effect of daily workload, we employ the linear model for a dependent variable y_i:

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

where the error term ε_i follows a normal distribution. We use equation (1) to estimate γ, which is the effect of daily workload on y_i averaged across all cases by the surgeon in a day. We consider four dependent variables related to surgical metrics and patient outcomes: OR time and incision time of the surgery, as well as the post-LOS and total ICU time of the patient. We do not estimate the effect of surgeon daily workload on common binary outcomes of an operation such as mortality and reoperation. This is due to the size of our data set and the extremely low occurrence of these adverse outcomes. For example, the mortality and reoperation probability in our sample are 2.7% and 5.9%, respectively. In addition, since our model includes many categorical variables for patient's risk and operative factors, some of them can lead to perfect prediction for the binary outcomes, thus prohibiting estimation of the effect. We expect accurately estimating the effect on the binary outcomes would require a much larger sample which provides more variation in the outcomes.

As a naive approach, we can estimate the coefficients in (1) by ordinary least squares (OLS) 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 levels are lower and imply shorter OR times. Consequently, both the dependent variable (e.g., OR time) and the daily workload (e.g., number of additional 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 OR time. Thus, using OLS to estimate (1) may yield a negative γ even if the true effect is positive. The opposite bias may also be observed if the unobserved severity factors are positively with the surgeon's daily workload, e.g., when surgeons need to pack in more cases due to the deterioration of patients. To address the endogeneity bias, we employ an instrumental variable (IV) method to obtain consistent estimates of the coefficients. We construct two IVs using the operational data from the cardiac department and discuss their validity in the next section.

We acknowledge the effect of workload in (1) is likely to vary across cases within the same day, e.g., the first and last case by the surgeon in a day. However, as we do not have the information on the decision-making process for how cases are scheduled within a day, we cannot effectively control for the potential endogeneity in the sequencing of cases. More importantly, given the limited observations where a surgeon performs three or four cases in a day, it is not possible to estimate the heterogeneous effects for surgeon's nth case in a day. Thus, in our main specification, we choose to estimate the effect averaged across all cases performed by the surgeon in a day. As a robustness check, we find consistent results when estimating the effect for being the non-first case of the surgeon as well as the impact of surgeon's prior incision time in a day. We discuss these analysis in Section 4.3.

3.1. Instrumental Variables

We now propose two novel IVs and discuss their validity. The two IVs are both defined on the surgeon-day level, i.e., they are the same for all of the cases by a surgeon on the same day. This is inline with our estimation of the average effect in (1) 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. It satisfies the relevance condition because the number cases by other surgeons can affect the daily workload of the focal surgeon through *resource sharing* across surgeons. In our study hospital, although the surgeons' schedules are typically 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 is usually not finalized until shortly before the day starts. In addition, while the hospital uses a block booking system, it only serves as a loose guidance for how surgeons schedule their cases,[3] and many resources are still shared by the surgeons within a day (e.g., staff, medications, equipment). For example, we find that a perfusionist can work for multiple surgeons in a day when the department workload

[3] Using the periods with block information, we find that roughly 20% of cases are performed out of the focal surgeon's blocks.

is high. Additionally, if the other surgeons are performing many operations whose patients must stay in the ICU post-surgery, this could impact the ICU bed-availability for the focal surgeon's patients. Thus, more cases performed by other surgeons in a given day could translate into fewer resources available for the focal surgeon. This tends to limit the workload of the focal surgeon on the same day. The 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. By above discussion, we expect $TotOther_i$ to be negatively correlated with the focal surgeon's daily workload.

We next consider whether $TotOther_i$ 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 available OR time and relevant resources when scheduling their own cases, and it is entirely up to the discretion of the focal surgeon which operations to prioritize amongst his/her own patients. Thus, an individual surgeon has little control over other surgeons' patients and schedules. This suggests that $TotOther_i$ should not be directly correlated with the unobserved severity factors of the focal surgeon's patients. A remaining concern is that the other surgeons' workload may be correlated with the downstream congestion, which can affect the surgical outcomes. We emphasize that the IV we use is the same-day workload. In addition, we control for cardiac patient census in our model (1) in order to account for the downstream congestion. Similar approaches of controlling for downstream hospital/unit congestion have been used in a number of studies when congestion has been used as an IV (e.g. Kim et al. 2015, Freeman et al. 2021, Soltani et al. 2022).

In Section 3.3, we provide empirical evidence for the relevance and exogeneity conditions of the IV, including a weak instrument test and Sargan's overidentification test. The statistics of $TotOther_i$ is reported in Table S.6 of the Online Supplement. We find that there are on average about four cases performed by other surgeons in a day. There is substantial variation in $TotOther_i$, as shown by its standard deviation of 1.79 cases. This holds for both the full sample as well as the elective and non-elective samples.

3.1.2. Ratio of Elective Cases on Current Weekday We construct a second IV that captures the frequency a surgeon performs elective cases on the focal weekday. This is based on the following operational feature within the cardiac surgery service. Cardiac surgeons usually have multiple responsibilities in addition to performing operations, including seeing patients in the office, teaching, and attending conferences. Thus, the hospital tends to allocate OR time for each surgeon on specific weekdays to reduce the uncertainty in his/her schedule. These block schedules are often set weeks and even months in advance. Such approach is especially significant for elective cases, which are typically scheduled at least two weeks in advance. Thus, if a given weekday (e.g., Tuesday) is associated with a large proportion of surgeon s's recent elective cases, it is likely to be one of the typical days surgeon s is allocated in the block schedule. This leads to higher expected workload for surgeon s on the days that fall on the same weekday.

To capture the long-term working pattern of surgeons, we construct a second IV as the ratio of focal surgeon's elective cases that fall on the current weekday, $ElecRatioWD_i$. For case i, it is calculated as:

$$ElecRatioWD_i = \frac{\text{Num. of elective cases by } s \text{ in } [t-L, t+L] \text{ performed on } WD_i}{\text{Total number of elective cases by } s \text{ in } [t-L, t+L]}, \tag{2}$$

where WD_i denotes the weekday on which case i is performed; t and s are the surgery date and surgeon of case i; L is a parameter determining the horizon around current date, i.e., L days before and after day t. We exclude the elective cases done in the current week in both the numerator and denominator of (2). In rare cases, we set $ElecRatioWD_i$ to be zero if the total number of elective cases in the denominator is smaller than ten, as it suggests that the surgeon was not working regularly for the period considered. (Our results are robust if we drop these cases entirely in the identification.) IVs based on long-term time average of workload have been widely used in OM literature (see, e.g., Tan and Netessine 2014, Freeman et al. 2017, Soltani et al. 2022). Our second IV follows a similar idea.

The operational factor captured by the $ElecRatioWD_i$ is reflected by the block booking system used in the hospital, in which a surgeon is assigned with fixed time slots (blocks) and resources to perform their operations. We obtained the surgeons' block schedule for 22 out of the 48 months in our sample; the remaining months were unavailable due an administrative-leave of the scheduling staff. However, we find that surgeons' blocks indeed tend to fall on specific weekdays. For example, 71% of Surgeon 3's blocks fall on Wednesday and Friday. On the other hand, 92% of elective cases are performed within a surgeon's blocks. These factors suggest that the elective ratio in (2) is an effective indicator for the surgeon's working pattern among weekdays. We note that the elective ratio in (2) can be calculated for the entire sample, while we only have the block information for less than a half of our sample. Thus, using $ElecRatioWD_i$ as an IV, instead of the actual block schedule, avoids significant data loss.

Since the surgeons' schedules are adjusted very infrequently (e.g., twice a year), we use a large horizon length L in (2) for calculating the ratio of elective cases. We consider three candidate values: $L = 60$, 90, or 180 days. Using the period with surgeon's block status, we find that setting $L = 180$ days leads to the highest correlation between $ElecRatioWD_i$ and the indicator for a case's block status. So we use $L = 180$ days to calculate $ElecRatioWD_i$ in our main analysis. As robustness checks, the results with $L = 60$ and 90 days in (2) are largely similar (see Section 4.3). The summary statistics of $ElecRatioWD_i$ are reported in Table S.6. The mean of $ElecRatioWD_i$ is 22% for the full sample. There is also substantial variation in $ElecRatioWD_i$ that can be leveraged for identification: its standard deviation is 12%, which is relatively large compared with the mean.

In summary, we expect the $ElecRatioWD_i$ to be positively correlated with the surgeon daily workload, satisfying the relevance condition as a valid IV. We also expect $ElecRatioWD_i$ to satisfy the exclusion restriction. First, in the calculation of $ElecRatioWD_i$, we use a relatively long horizon and have dropped the focal surgeon's cases (including the focal case) in the current week. Second, a surgeon's block schedule is usually fixed well in advance and adjusted very infrequently. Thus, $ElecRatioWD_i$ tends to reflect the surgeon's long-term working pattern over weekdays and is unlikely to be correlated with the severity factors of patients treated by the surgeon in the current day. Section 3.3 provides further empirical evidence for the relevance and exogeneity conditions of this IV.

We also note that the IV is not mechanically correlated with surgeon's workload on current day or the status of focal case: the correlation between $ElecRatioWD_i$ and the daily workload of focal surgeon (resp. an indicator for focal case being a non-elective one) is only 0.194 (resp. -0.203). These suggest that surgeons tend to do more operations and schedule elective cases on their usual operation days, which aligns with our interpretation. However, the correlations are relatively low, suggesting the IV reflects surgeon's long-term behaviors.

Although we have argued that the IV $ElecRatioWD_i$ is unlikely to be correlated with the unobserved risk factors as it is based on the surgeon's long-term working pattern, one might be concerned that a case performed on a non-regular working day may be more complex and time sensitive and these unobserved factors may jeopardize the validity of our IV. While one can never fully rule out all potential threats to identification, the potential bias introduced by such unobservables would only attenuate our estimated negative effects of daily workload. This is because the surgeon's daily workload tends to be high on the regular working days with large $ElecRatioWD_i$. If the cases on a surgeon's non-regular working day (low $ElecRatioWD_i$) were on average more severe, the days with low surgeon workload would be associated with more unobserved risk. This relationship implies that the true effects of surgeon daily workload are actually more negative than we could estimate. Thus, such a violation of the exclusion criteria would make it *harder* to estimate a negative effect of surgeon daily workload and our results should be interpreted as conservative estimates.

3.2. Estimation Methods

We estimate the effect of daily workload in model (1) using the two IVs introduced above. We estimate the linear model (1) using the two-stage least squares (TSLS) regression (Woodridge 2010). 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 *together* using OLS:

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

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. The standard errors in the second stage need to be adjusted as we are plugging in estimates of $Workload_i$.

We find that the distributions of OR time, 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 11.9 hours for OR time, 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, as discussed in Section 4.3. In addition, in model (1), we cluster the standard errors by the surgeon's identifier to account for the potential correlation across cases of the same surgeon.

3.3. Validity Tests of the IVs

In this section, we discuss the testing results for the relevance and exogeneity conditions for the two IVs. To test the relevance condition, we run the first stage regression (3) to check how the two IVs affect the surgeon daily workload. The estimated coefficients are reported in Table 3. We run regression (3) for the full, elective, and non-elective samples, respectively.

Table 3 First Stage Regression: Estimated Effects of IVs on Surgeon Daily Workload

	Full	Elective	Non-elective
$TotOther_i$	-0.075^{***}	-0.065^{***}	-0.084^{***}
	(0.006)	(0.008)	(0.008)
$ElecRatioWD_i$	0.993^{***}	0.871^{***}	1.105^{***}
	(0.087)	(0.139)	(0.115)
Num. Obs	5345	2474	2871

The estimated effects of the two IVs on surgeon daily workload. We report the estimated coefficients η_1 and η_2 in (3). Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

We see the two IVs are statistically significant with expected signs for both workload measures and the three samples. For the full sample, the coefficient of $TotOther_i$ and $ElecRatioWD_i$ is -0.075 and 0.993 respectively, both of which are significant at the 0.1% level. This shows the two IVs indeed explain variation in the focal surgeon's daily workload, thus satisfying the relevance condition. Specifically, increasing the $TotOther_i$ (decreasing $ElecRatioWD_i$) by one standard deviation is associated with 0.13 (resp. 0.12) fewer cases for the focal surgeon in a day. This translates to a nearly 20% standard deviation change in the average surgeon daily workload, which is a substantial effect. Similar observations hold for the elective and non-elective samples. A weak instrument test for the hypothesis $\eta_1 = \eta_2 = 0$ is strongly rejected at the significant level of 0.1% with F-statistics of 56.8, 31.2, and 86.6, for the full, elective, and non-elective sample, respectively.

Next, we consider the exogeneity conditions of the two IVs. Although it is impossible to verify the exogeneity condition directly, we provide some empirical evidence for why this is likely to hold. First, we check whether the two IVs are correlated with *observed* measures of patient severity. Table S.7 of the

Online Supplement summarizes the correlation between the two IVs and 21 observed severity factors. The correlations are generally small. For $TotOther_i$ (resp. $ElecRatioWD_i$) , the correlation is smaller than 0.1 for all the 21 (resp. 20 of the 21) patient's risk factors. The average absolute correlation is only 0.019 and 0.044 for $TotOther_i$ and $ElecRatioWD_i$, respectively. Thus, we conclude that the two IVs are unlikely to be correlated with the patient severity factors.

We then conduct the Sargan's overidentification test for the exogeneity of the two IVs (see Woodridge 2010). This is possible because we have two IVs but only one endogenous variable ($Workload_i$). The Sargan's test mainly consists of two steps. The first step is to estimate the model (1) by TSLS with the two IVs. Then, we regress the residuals from the TSLS model on the two IVs (and the exogenous variables X_i). The intuition is that under the null hypothesis that the IVs are valid, they should have no significant explanatory power for the residuals. We find this is indeed the case for all the four dependent variables (OR time, incision time, post-LOS, total ICU time) and three samples (full, electives, non-electives) in our study. The p-values from the Sargan's test are reported in Table S.8 of the Online Supplement. The p-values are all greater than 0.1. Thus, we cannot reject the null hypothesis that the two IVs are valid.

4. Estimation Results

This section provides the main estimation results regarding the effects of surgeon daily workload (number of additional cases) on surgery duration and outcomes. We first show 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.

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

We report the estimated effects of surgeon daily workload in (1), which is measured by the number of additional cases performed by the focal surgeon. For the four dependent variables (OR time, incision time, post-LOS, and total ICU time), we provide the estimated γ and its standard errors for the full sample in Table 4. Panel A shows the estimated effects from the TSLS with the two IVs, as described in Section 3.2. For comparison, Panel B reports the results when we ignore the endogeneity problem and estimate (1) by a simple OLS. The estimated coefficients of other select covariates are reported in Table S.11 of the Online Supplement. By model (1), the estimated γ captures the effect averaged across all the cases performed by the surgeon in a day.

By the "OR time" column in Panel A, we see that higher daily workload tends to increase the OR time of the cases performed by the focal surgeon. In particular, adding one more case increases the OR time of each case performed by the surgeon by 0.455 hours (27 minutes) on average. This translates to a 6.5% relative increase of the average OR time. The effect is statistically significant at the 0.1% level. Similar impact is observed for the incision time, with the estimated effect being 0.458 hours (28 minutes). 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 "OR time" column in Panel B: the coefficient is negative (-0.149) and statistically significant at the 0.1% level. Thus, it is essential to address the endogeneity in the daily workload as surgeons may schedule more cases if the unobserved severity factors imply shorter OR 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 surgery duration. First, surgeons may "speed up" the operations when they have more cases to perform in a day, leading to a shorter OR or 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 under high daily workload. Possible reasons include surgeon fatigue (both physically and mentally) due to long working hours, as well as operational constraints in the related resources when more patients are serviced (e.g., shared staff, less experienced nurses, or delay in bed flows). After addressing the endogeneity bias with proper IVs, our empirical results support the second effect, i.e., surgeon fatigue and operational constraints outweigh the potential channels for speedup and cause longer surgery duration on average when more cases are performed.

Table 4 Estimated Effects of Surgeon Daily Workload (Number of Additional Cases) on Surgery Duration and Patient Outcomes: Full Sample

	Panel A: With IV				Panel B: Without IV			
	OR time (in hours)	Incision time (in hours)	Post-LOS (in days)	ICU time (in days)	OR time (in hours)	Incision Time (in hours)	Post-LOS (in days)	ICU time (in days)
Estimated γ	0.455^{***}	0.458^{***}	1.109^{***}	1.166^{***}	-0.149^{***}	-0.100^{**}	-0.077	-0.018
	(0.137)	(0.101)	(0.335)	(0.240)	(0.036)	(0.038)	(0.129)	(0.086)
Patient demographics	Y	Y	Y	Y	Y	Y	Y	Y
Risk/operative factors	Y	Y	Y	Y	Y	Y	Y	Y
System occupancy	Y	Y	Y	Y	Y	Y	Y	Y
Surgeon fixed effect	Y	Y	Y	Y	Y	Y	Y	Y
Time fixed effect	Y	Y	Y	Y	Y	Y	Y	Y
Num. Obs	5345	5345	5344	5319	5345	5345	5344	5319

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

We now consider the effect of daily workload on patient outcomes, including post-LOS and total ICU time. By the "Post-LOS" and "Total ICU" columns in Panel A, we find that higher daily workload increases the post-LOS and total ICU time when estimating the model using TSLS. Specifically, adding one more case increases the total ICU time and post-LOS by 1.17 and 1.11 days, respectively, for the cases performed by the surgeon on the same day. This is equivalent to a 9.3% increase for post-LOS and a 21.7% increase for total ICU time. Without using the IVs, the effect is insignificant for both outcomes, as shown in the last two columns in Panel B. The results here suggest that increased surgeon daily workload is associated with longer post-surgery recovery time for patients. As we discussed above, 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 the 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 (Kim et al. 2015). 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 a potential solution for managing ICU congestion.

We note that model (1) suggests that the hospital may be able to improve 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 that the total expected OR time (resp. post-LOS) of these two cases would decrease by $0.455 \times 2 = 0.91$ hours (resp. 2.21 days). We leverage this insight to propose a surgical scheduling model that captures such effects in Section 5.

Our results provide clear evidence for the negative impact of high surgeon daily workload. One potential mechanism is surgeon fatigue, which we are not able to directly measure due to data limitations. That said, we perform several additional tests to shed light on the potential channels. First, we find that the surgeon daily workload has no significant impact on the total time of pre and post-incision stages of the focal case, which mainly include routine tasks that can be completed by nurses or other staffs.[4] Second, we run our regression after controlling for the supporting staff's workload using the total number of cases by other surgeons on the same day. That is, we include $TotOther_i$ as an exogenous variable in (1) and only use $ElecRatioWD_i$ as the IV. We obtain similar effect of surgeon daily workload in this set-up. Third, instead of estimating the average effect for all cases in a given day, we estimate the effect of being a non-first case of a surgeon in a day (sorted by the OR entry time). We find that being a non-first case and having longer prior incision time are associated with longer surgery duration and recovery time. Finally, we show that the effect of surgeon's daily workload on surgery duration is still significant even after we control for the OR entry hour of the case. These results are further discussed in Section 4.3.

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 OR/incision time and patient's post-surgery recovery time on average. We now further investigate the impacts of workload for elective and non-elective cases separately. The non-elective cases include urgent, emergent, and salvage cases, and

[4] This is consistent with the findings from Bavafa and Jónasson (2023) in the ambulance paramedics setting: the effect of workload is more significant for the sub-processes that require complex knowledge work.

account for more than half (53%) of the full sample. Specifically, we estimate the econometric model (1) using the subsamples of elective and non-elective cases respectively.[5] This allows us to reveal the potential heterogeneity in the impact of daily workload for different types of patients. In each regression, the workload is still measured as the number of all cases performed by the surgeon in a day. By Table 3, the two IVs still impact the surgeons daily workload with expected signs in the two subsamples.

Table 5 Estimated Effects of Surgeon Daily Workload (Number of Additional Cases) on Surgery Duration and Patient Outcomes: Elective and Non-elective Sample

	Panel A: Elective				Panel B: Non-elective			
	OR time (in hours)	Incision time (in hours)	Post-LOS (in days)	ICU time (in days)	OR time (in hours)	Incision Time (in hours)	Post-LOS (in days)	ICU time (in days)
Estimated γ	0.567^{*} (0.256)	0.531^{**} (0.187)	-0.155 (0.867)	0.101 (0.620)	0.377^{*} (0.158)	0.395^{*} (0.170)	1.605^{*} (0.807)	1.647^{***} (0.367)
Patient demographics	Y	Y	Y	Y	Y	Y	Y	Y
Risk/operative factors	Y	Y	Y	Y	Y	Y	Y	Y
System occupancy	Y	Y	Y	Y	Y	Y	Y	Y
Surgeon fixed effect	Y	Y	Y	Y	Y	Y	Y	Y
Time fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Num. Obs	2474	2474	2474	2454	2871	2871	2870	2865

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

Table 5 reports the estimated effects of surgeon daily workload for the elective (Panel A) and non-elective samples (Panel B), respectively. By the "OR time" columns, we see the surgeon daily workload significantly increases the OR time for both elective and non-elective cases. Both effects are significant at the 5% level. The magnitude of impact is larger for the elective cases than that for the non-elective cases. In particular, performing one more case in a day increases the OR time of each elective and non-elective case by 34 and 22 minutes, respectively. Similar observations hold for the incision time, with the impact being 30 minutes for the elective cases and 24 minutes for the non-elective ones. The results suggest that the surgeon daily workload has a greater impact on surgery length for elective cases, though the differences are not statistically significant. Such heterogeneity may be explained by the fact that non-elective cases (urgent and emergent cases) are generally more time sensitive than elective ones, so their OR times are less impacted by surgeon daily workload.

We find substantial heterogeneity in the effects of surgeon daily workload on patient's outcomes for elective and non-elective cases. Specifically, increased surgeon daily workload significantly impacts the surgery

[5] An alternative way to study the heterogeneous effects is to run a regression with an interaction term between workload and surgery status. However, this would restrict the coefficients of other controls to be the same for elective and non-elective patients. We believe allowing the controls to have different coefficients in the stratified regressions is more appropriate. Indeed, we find some controls have very different coefficients for elective and non-elective patients, such as NYHA class and endocarditis indicator.

outcomes of non-elective patients, but not for elective ones. For the post-LOS, the coefficient of $Workload_i$ is statistically significant for the non-elective cases, but insignificant for the elective ones ("Post-LOS" columns). 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 1.61 more days in the post-LOS of non-elective patients, which is 45% larger than that for the full sample (1.11 days). Similar heterogeneity is also seen in the total ICU time ("Total ICU" columns). The estimated coefficient of $Workload_i$ is 1.65 days for the total ICU time of non-elective cases, larger than that for the full sample (1.17 days). On the other hand, the daily workload does not significantly impact the total ICU time of elective cases. A z-test shows that the difference in coefficients is significant at the 5% level for total ICU time but insignificant for the post-LOS. This may be attributed to the large standard errors of the post-LOS estimates due to sample size.

One potential explanation for the heterogeneous effects 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 2: the non-elective patients have longer post-LOS and total ICU time 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.

Our results demonstrate the consistent negative impact of surgeon daily workload on surgical outcomes. Such effects are particularly significant for non-elective patients, who are generally more severe. This provides 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 cumulative and 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 outcomes. We explore this direction in Section 5.

4.3. Robustness Checks and Mechanisms of Effects

We conduct similar regression analyses under various alternative specifications to examine the robustness of our main findings and shed light on the mechanisms of the key estimated effects. The results are discussed in Section S.4 of the Online Supplement. We briefly discuss them below. First, we control for the OR entry time of each case by adding an exogenous dummy variable in (1) representing whether the OR entry time is in the morning (7AM to 12PM), in the afternoon (12PM to 5PM), or in other times.[6] The estimated effects are reported in Panel A of Table S.9. The effects are very similar to those under our main specification,

[6] The results are similar when we use a continuous variable for OR entry hour.

though the effect on OR time becomes smaller (27 to 20 minutes for the full sample) after adding the OR entry time dummies.

We then run the regression (1) using only $TotOther_i$ or $ElecRatioWD_i$ as the IV. This allows us to check the effectiveness of the two IVs separately. As shown in Panel B of Table S.9, the results with one IV are qualitatively similar to those in our main specifications. We still find high daily workload is associated with longer surgery duration and worse patient outcomes. We then check the impact when we use alternative specifications to construct the second IV $ElecRatioWD_i$. First, we calculate $ElecRatioWD_i$ using only the period prior to the current day, i.e., using interval $[t-L,t)$ in (2). Second, we change the window length L in (2) from 180 days to 60 or 90 days. The results are reported in Panel C of Table S.9. We find the direction and magnitude of the estimated effects are again similar to our main findings.

Next, we control for supporting staff's workload in our model using the total number of cases by other surgeons on the same day ($TotOther_i$) as a proxy. That is, we include $TotOther_i$ as an exogenous variable in (1) and only use the ratio of elective cases on current weekday ($ElecRatioWD_i$) as the IV. The estimation results are reported in Panel D of Table S.9. The estimated effects for the full sample are qualitatively similar to those under our main specification in Table 4, except that the effect of surgeon daily workload on post-LOS becomes insignificant.

In our main specification, we winsorize the four dependent variables at their 97.5th percentile, respectively. For robustness checks, we also experiment the winsorization level of 95th or 99th percentiles. The results are given in Panel E of Table S.9. We find the results are largely similar, although the magnitudes of the effects become smaller (resp. larger) for post-LOS and total ICU time when we use a lower (resp. higher) winsorization level.[7] As another check, we run the regressions after taking the log transformation of the dependent variables, and again obtain qualitatively similar results. The magnitude of impact from the regression with log transformation is consistent with those in our main specification.

We then 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 differences in the total recovery time of patients – measured by hours – 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 conservative measures of post-LOS by adjusting the original measure using the OR entry or exit time of each case. We find that the estimated effect are very similar when using these alternative measures. The definitions of these measures are summarized in Section S.4 of the Online Supplement. The estimated results are provided in Table S.10 therein.

Instead of focusing on the average effect of surgeon daily workload, we also estimate the effect of not being the first case of a surgeon in a day. First, we introduce a binary variable $NonFirst_i$ for each case,

[7] The change in the magnitude can be explained as follows. The distribution of total ICU time (and 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.

denoting whether it is a non-first case of its surgeon on the surgery day. Second, we calculate the surgeon's total incision time prior to the start of focal case, $PriorInc_i$. We then estimate the effect of $NonFirst_i$ or $PriorInc_i$ on the four dependent variables using the two IVs. We find that being a non-first case in a day or having longer prior incision time is associated with longer surgery duration and worse patient outcomes. The magnitudes of the effects are consistent with our main findings on the average effect in model (1). The model set-up and estimation results are discussed in Section S.3. This provides suggestive evidence for the fatigue channel of surgeon daily workload.

In unreported results, we also find that our estimated effects are robust when we drop the emergent or salvage cases from our sample. These patients may have special medical conditions or requests, making the IVs less applicable. On the other hand, we cannot construct IVs based on the emergent or salvage cases as they are very infrequent in our sample ($< 10\%$).

5. A Surgery Scheduling Model with Impact from Daily Workload

In this section, we propose a surgical scheduling model that accounts for the effect of surgeon daily workload. 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). In some recent work, the surgeon's long-term volume (e.g., number of cases performed in past year) is shown to affect the surgery duration and outcomes, and such volume-outcome relationship is recently accounted in the surgical scheduling (Wang and Pourghannad 2020). However, to our best knowledge, the impact of surgeon daily workload is largely ignored in current literature. As shown in our empirical analyses, increased surgeon daily workload can lead to longer surgery duration and worse patient 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 study 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 across days. As we discussed in Section 4, this can decrease surgery duration and improve patient outcomes. On the other hand, we keep the patient cohort and the surgeon assigned to each patient unchanged. With these constraints, the total arrival to the cardiac department and each surgeon's cumulative volume are largely unchanged. This allows us to isolate the impacts from surgeon daily workload in surgical scheduling by optimizing the surgery dates. In addition, it provides an easily implementable potential solution to improve the current scheduling in the hospital. We expect to gain larger benefit if we introduce other types of interventions in our model, e.g., increasing 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 Section S.2. We solve the model for each week (Sunday to Saturday) in our four year sample. Our model considers the decisions to assign cases to each day. Note that the cases are still scheduled within the week they were actually performed in the data. Thus, the changes from our model are relatively small, supporting the feasibility of the resulting schedule. The weekends are excluded in the scheduling model, as surgeons rarely work on weekends except for emergent and salvage cases.

The objective in our model is to minimize the total expected OR time, post-LOS, or ICU time of all cases in our sample,[8] as 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. In Section S.2, we show that the objective function can be written as

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

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; $\tilde{n}^{(el)}_{s,t}$ and $\tilde{n}^{(ne)}_{s,t}$ denote the number of elective and non-elective cases performed by surgeon s on day t in the new schedule respectively. The coefficients $\gamma^{(el)}$ and $\gamma^{(ne)}$ are the estimated effect γ for the daily workload effect on OR time, post-LOS, or total ICU time – depending on which one we are to minimize – for the elective and non-elective cases respectively. They are reported in Table 5 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 4 with $\gamma^{(el)} = \gamma^{(ne)} = \gamma^{(avg)}$. As shown in Section S.2 of the Online Supplement, the optimal schedules from the model are the same for the three variables under average effect of daily workload. Similarly, we obtain the same optimal schedule for post-LOS and ICU time when using the heterogeneous effect.

In our empirical set-up, the average effects show the impact of workload on all cases using a larger sample size, while the heterogeneous effects reveal the difference in impact for elective and non-elective cases using the smaller, more granular data sets. As we discuss below, the managerial insights can be different under the two types of effects. Choosing which effect to account for in scheduling depends on the concrete challenges and management goals of the hospital, which can vary in different situations. For example, heterogeneous effects may be more proper if a hospital finds its elective cases are relatively easy and not affected by surgeon's daily workload.

We consider multiple feasibility constraints for the new schedule. We allow elective cases to be assigned to any weekday 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 accordance with the practice of our partner hospital. In addition, we impose an upper bound

[8] Here we use OR time instead of incision time as it is more related to hospital's OR cost. We obtain similar results when we use total incision time as objective.

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 in their work days, 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 in Online Supplement Section S.2 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 in below, 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 in a week is four days, unless the corresponding quantities are larger in the original schedule. Our scheduling model still leads to economically substantial improvement even under more restrictive conditions. The results are available upon request.

We show our main results in Table 6. 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 improved 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 (total OR time, post-LOS, or ICU time) we are optimizing over, i.e., Rel. $\Delta\text{Total} = \Delta\text{Obj}/\text{Total}$, where Total is the sum of observed values from all patients in our sample.[9] Thus, the relative reduction Rel. ΔTotal provides an alternative measure for the benefit from the new surgical schedule.

In Table 6, 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 OR time, the new schedule with average (resp. heterogeneous) effect leads to a 687 (resp. 768) hours decrease in the four-year horizon, which is equivalent to a 1.82% (resp. 2.03%) relative reduction. The OR is an extremely expensive resource with cost estimated to be up to \$37 per minute (Childers and Maggard-Gibbons 2018). Thus, the reduction in the OR time could save the hospital up to \$426,240 each year. In addition, we find that the new schedule leads to

[9] As in our estimations, we winsorize the corresponding variables to their 97.5th percentile.

Table 6 Estimated Effects of the Surgical Scheduling Model

Objective	Effect	Obj orig	Obj new	ΔObj	Number of improved weeks	Rel. ΔTotal (in % points)
OR Time (in hours)	Avg	4124.58	3437.53	687.05	205	1.82%
	Het	4240.78	3472.96	767.82	208	2.03%
Post-LOS (in days)	Avg	10053.09	8378.50	1674.59	205	2.78%
	Het	7594.86	5837.39	1757.48	204	2.92%
Total ICU time (in days)	Avg	10569.79	8809.13	1760.66	205	6.87%
	Het	7793.60	5990.14	1803.47	204	7.04%

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. Using the average effect, the new schedule decreases the total post-LOS by 1,675 days and the total ICU time by 1,761 days, which translate to a 2.78% and 6.87% relative drop respectively. The benefit is slightly larger with heterogeneous effect, with the relative reduction for post-LOS (resp. ICU time) increasing to 2.92% (resp. 7.04%). The benefits from our new schedule are economically important. To see this, we can 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 $\Delta\text{Pat} = \Delta\text{Obj}/(4 \text{ Years} \times \text{AvgLOS})$, where AvgLOS denotes the average post-LOS or total ICU time after surgery. Taking the results with heterogeneous effects as an example, the reduction in the total post-LOS and ICU time translates to 37 and 83 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 decreases by 5.82%. 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 (see, e.g., Zenteno et al. 2016).

We note that the resulting schedules can lead to improvement for multiple objectives simultaneously, instead of improving one at the cost of another. For example, when considering the average effects, the resulting schedules are identical, so all the three objectives are improved by the amounts reported in Table 6. On the other hand, when considering the heterogeneous effects, the resulting schedule is the same when optimizing for total post-LOS and ICU time. Thus, both post-surgery metrics are reduced simultaneously.

5.2. Managerial Insights from the New Schedules

We have shown that our surgical scheduling model can improve the OR 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 $\gamma^{(el)} = \gamma^{(ne)} = \gamma^{(avg)}$ in (4). As mentioned before, the resulting schedule is the same for the three outcomes (OR time, post-LOS, and ICU time).[10] 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. In Table 7, 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 $i \in \{1, 2, 3, 4\}$ 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 $\tilde{n}_{s,t}^{(ne)} = i$ for $i \in \{1, 2, 3\}$ 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.

Table 7 Summary Statistics of Schedules

Effect	$\tilde{n}_{s,t}=1$	$\tilde{n}_{s,t}=2$	$\tilde{n}_{s,t}=3$	$\tilde{n}_{s,t}=4$	$\bar{n}_{day}$	$\tilde{n}_{s,t}^{(ne)}=1$	$\tilde{n}_{s,t}^{(ne)}=2$	$\tilde{n}_{s,t}^{(ne)}=3$	$\mathsf{E}(\tilde{n}_{s,t}^{(el)} \mid \tilde{n}_{s,t}^{(ne)} > 0)$
Avg	3268	980	39	1	3.28	2214	317	8	0.15
Het (LOS)	2955	737	304	2	3.06	2298	278	6	0.06
Original	2268	1249	177	13	2.83	1872	424	48	0.27

We first consider the schedule under the average effect. To minimize the objective function, the new schedule should smooth surgeon workload across days, i.e., reducing the number of days with multiple cases. In first row of Table 7, 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 $\tilde{n}_{s,t} = 2$ (resp. $\tilde{n}_{s,t} = 3$), i.e., the surgeon performs two (resp. three) cases in a day, decreases from 1249 in the original schedule to 980 (resp. 177 to 39) in the new schedule. In addition, the new schedule almost fully eliminates the situation where a surgeon performs four cases in a day (from 13 to 1). The reduction in these high workload days is 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.

Next, we examine the new schedule obtained with heterogeneous effect for post-LOS or ICU time. In this case, the minimization of the objective suggests that the hospital should reduce the workload on the days with non-elective cases. We see this is indeed achieved in the new schedule according to Table 7. 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 278 (resp. 48 to 6) in the new

[10] The schedule optimizing for OR time with heterogeneous effect is qualitatively similar to that using the average effect.

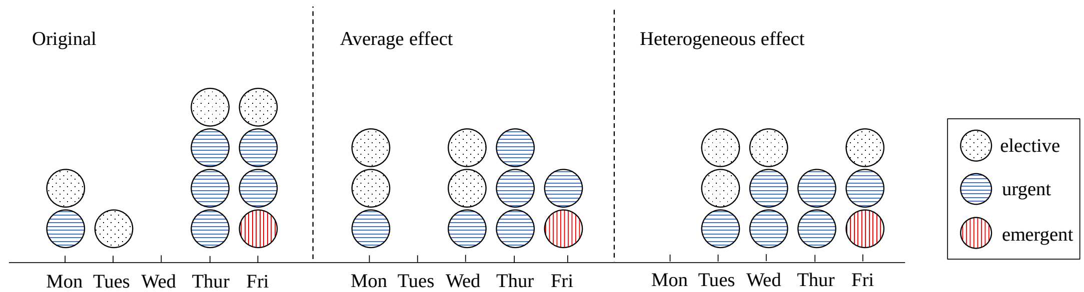

Figure 2 Example surgical schedules from our optimization model: Surgeon 2, Week 39

schedule. In addition, the new schedule decreases the workload from elective cases when a surgeon needs to do 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 the non-elective cases to perform, as these patients are generally more severe.

We further use a concrete example to show how our scheduling model works. We look at the schedule for Surgeon 2 in Week 39 of our sample. This is an extremely busy week for the surgeon, with a total of eleven cases performed across four weekdays (four elective, six urgent, and one emergent cases). The original schedule is shown in the left panel of Figure 2. We see surgeon daily workload is highly unbalanced across days: there are two and one cases on Monday and Tuesday, but four cases on both Thursday and Friday. The middle and right panels show the new schedules from our model with average and heterogeneous effects, respectively. We find the surgeon's workload is smoothed under both schedules, and the maximum daily workload is now three cases. Moreover, we see the schedule under the heterogeneous effect further smooths the number of non-elective cases across days: the maximum number of non-elective cases in a day decreases from three to two. This is because only the non-elective cases are negatively affected by high daily workload when we use heterogeneous effect for post-LOS or ICU time.

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 peak department daily workload, i.e, 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 "$\bar{n}_{day}$" column in Table 7, 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 3.28 (resp. 3.06) under the average (resp. heterogeneous) effect. Next, Figure S.1 of the Online Supplement shows the frequency distributions of total number of cases in a day from the original and new schedules. We see that the number of days with extremely high department workload (e.g., more than ten cases in a day) remains small under the new schedules. In addition, the distributions of department daily workload 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.

Our study focuses on a new aspect in surgical scheduling — the cumulative daily workload of individual surgeons. On the other hand, the surgeon's past volume and the system-level workload, which are more commonly considered in existing literature, are largely unchanged in the new schedules. First, we keep the matching between patients and surgeons fixed and only adjust the schedule within each week, thus the surgeon's volume over a longer period (e.g., past year) is not affected. Second, by Figure S.1, the distribution of daily department workload is similar under the original and new schedules. Thus, our model isolates the impact of individual surgeon daily workload. We expect the benefits from our model to be larger if we had more flexibility in the scheduling, e.g., changing the assignment between surgeon and patient.

While our scheduling model does not capture the full scope of surgical scheduling and also takes a retrospective view, it can still provide useful managerial insights. First, when feasible, hospitals should smooth surgeon's workload to reduce the frequency of very busy days, e.g., with three or four cases. Second, hospitals should try to control the workload of surgeons on the days when they have to perform more complicated non-elective cases. We believe such insights can be useful for setting surgical schedules to improve patient flow, OR time, and patient's post-surgery outcomes.

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 such relationship 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, 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. The IVs effectively addresses the endogeneity problem due to unobserved risk factors. This provides new evidence for the negative impact of surgeon fatigue or operational constraints due to high daily workload.

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 practical 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 OR time) and post-surgery (via post-LOS and total ICU time) by accounting for the impact of surgeon workload when scheduling surgery. 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.

While our results provide strong evidence of the impact of workload on cardiac surgery outcomes, our study has a number of limitations. First, our data comes from a single hospital. Other hospitals may have different scheduling procedures which may make the IVs more or less appropriate. Second, 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. We defer these to future research.

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. *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.

Bavafa H, Jónasson JO (2023) The distributional impact of fatigue on performance. *Management Science* .

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.

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.

Delasay M, Ingolfsson A, Kolfal B, Schultz K (2019) Load effect on service times. *European Journal of Operational Research* 279(3):673–686.

Freeman M, Robinson S, Scholtes S (2021) Gatekeeping, fast and slow: An empirical study of referral errors in the emergency department. *Management Science* 67(7):4209–4232.

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.

Hockey R (2013) *The psychology of fatigue: Work, effort and control* (Cambridge University Press).

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.

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

Rothstein DH, Raval MV (2018) Operating room efficiency. *Seminars in Pediatric Surgery*, volume 27, 79–85.

Soltani M, Batt RJ, Bavafa H, Patterson BW (2022) Does what happens in the ed stay in the ed? the effects of emergency department physician workload on post-ed care use. *Manufacturing & Service Operations Management* 24(6):3079–3098.

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* .

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)

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).

In Table S.1, 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 S.2. 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 "others". In total, we have six types for the non-standard procedures, i.e., four non-standard isolated ones, non-standard multiple, and others. The numbers of cases of each type (both standard and non-standard ones) are summarized in Table S.3.

Table S.1 Description and Summary Statistics of Other Independent Variables in Model (1)

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
PAPressure	Systolic pressure	Hemodynamics	Categorical	-
TotCABG	Number of arteries bypassed	Coronary Bypass	Continuous	1.36

In summary, the independent variable X_i in (1) includes the factors in Table S.1, patient's gender and age, surgery status, patient's admission type, procedure type in Table S.3, surgeon's identifier, patient's pre-LOS, cardiac patient census, and dummies for weekday, month, and year of the operation.

Table S.2 Summary Statistics of Categorical Variables in Table S.1

Variable	Category	Num Obs.	Ratio
NYHA	N/A	1933	36.1%
	I	516	9.6%
	II	998	18.6%
	III	991	18.5%
	IV	663	12.4%
	Unknown	251	4.7%
Race	White	4273	79.8%
	Asian	590	11.0%
	Black	274	5.1%
	Other	215	4.0%
Smoke	FALSE	2694	50.3%
	TRUE	2429	45.4%
	Unknown	229	4.3%
PA Pressure	High	376	7.0%
	Low	2247	42.0%
	Unknown	2729	51.0%

Table S.3 Numbers of Cases by Surgery Types

Surgery Type	Num Obs.	Ratio
Standard (N = 3420)		
CABG	1718	32.1%
AVR	683	12.8%
MVR	225	4.2%
MVr	254	4.7%
CABG + AVR	318	5.9%
CABG + MVR	57	1.1%
CABG + MVr	58	1.1%
AVR + MVR	107	2.0%
Non-standard (N = 1932)		
Isolated Valve	574	10.7%
Isolated CAB	28	0.5%
Isolated cardiac	369	6.9%
Isolated non-cardiac	15	0.3%
Multiple	690	12.9%
Others	256	4.8%

S.2. 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.

S.2.1. 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

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

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

$$x_{i,t} = 0, \text{ if } i \in C_{ug} \text{ and } |t - \tilde{t}(i)| > 1, \tag{S.2}$$

and

$$x_{i,t} = 0, \text{ if } i \in C_{es} \text{ and } t \neq \tilde{t}(i). \tag{S.3}$$

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,

$$\tilde{n}_{s,t} = \sum_{i \in A_s} x_{i,t}. \tag{S.4}$$

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

$$\tilde{n}_{s,t} \leq \max\{\bar{n}^{(c)}, n_{s,t}^{(c)}\}, \ \forall t \in T \text{ and } \forall s \in S, \tag{S.5}$$

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

$$\sum_{t \in T} \mathbf{1}\{\tilde{n}_{s,t} > 0\} \leq \max\{\bar{n}^{(d)}, n_s^{(d)}\}, \ \forall s \in S. \tag{S.6}$$

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 (S.6) 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

$$z_{s,t} \leq M \cdot \tilde{n}_{s,t} \text{ and } z_{s,t} \geq m \cdot \tilde{n}_{s,t}, \tag{S.7}$$

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 (S.7) 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 (S.6) in the following linear form as

$$\sum_{t \in T} z_{s,t} \leq \max\{\bar{n}^d, n_s^{(d)}\}, \ \forall s \in S. \tag{S.8}$$

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 (S.1), (S.2) – (S.3), (S.4) – (S.5), and (S.7) – (S.8), all of which are formulated in linear form.

S.2.2. 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 OR time, post-LOS, or total ICU time. Following our econometric model (1), the expected value of the three variables can be expressed as

$$\hat{y}_i = X_i\beta + \gamma Workload_i. \tag{S.9}$$

Here we measure workload as the number of other cases performed by the focal surgeon on the day. The variable $\hat{y}_i$ is specified as OR 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 (S.9), the expected value $\hat{y}_i$ can be decomposed to two parts

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

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:

$$\sum_{i \in C} d_i = \sum_{s \in S} \sum_{t \in T} d'_{s,t}. \tag{S.10}$$

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 (S.4). 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 (S.4), they are given by

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

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

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

Here $\gamma^{(el)}$ and $\gamma^{(ne)}$ are the estimated coefficient γ for the daily workload effect on OR 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 5 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 4 with $\gamma^{(el)} = \gamma^{(ne)} = \gamma^{(avg)}$. Plugging $d'_{s,t}$ into (S.10), the objective function is given by

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

Note that the summation terms $\sum_{t \in T} \sum_{s \in S} \tilde{n}_{s,t}^{(el)}$ and $\sum_{t \in T} \sum_{s \in S} \tilde{n}_{s,t}^{(ne)}$ 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

$$\min \sum_{s \in S} \sum_{t \in T} \tilde{n}_{s,t} \cdot \left(\gamma^{(el)} \tilde{n}_{s,t}^{(el)} + \gamma^{(ne)} \tilde{n}_{s,t}^{(ne)}\right). \tag{S.11}$$

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 (S.11) and linear constraints (S.1), (S.2) – (S.3), (S.4) – (S.5), and (S.7) – (S.8). 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 (S.7). All the decision variables are binary. The final MIQP formulation to minimize total OR time, post-LOS, or ICU time is given below:

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

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.

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 this, by (4), the objective with average effect can be expressed as

$$\min \gamma^{(avg)}\cdot\sum_{s\in S}\sum_{t\in T}\left(\tilde{n}_{s,t}\right)^2. \tag{S.12}$$

Thus, the coefficient $\gamma^{(avg)}$ does not impact the solution and can be dropped from (S.12) (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

$$\min \gamma^{(ne)}\cdot\sum_{s\in S}\sum_{t\in T}\tilde{n}^{(ne)}_{s,t}\left(\tilde{n}^{(ne)}_{s,t} + \tilde{n}^{(el)}_{s,t}\right). \tag{S.13}$$

This follows by plugging $\tilde{n}_{s,t} = \tilde{n}^{(el)}_{s,t} + \tilde{n}^{(ne)}_{s,t}$ in (4) and using the fact $\gamma^{(el)} = 0$ for post-LOS and ICU time, i.e., surgeon daily workload does not impact the two outcomes of elective cases.

S.3. Effect of Being a Non-first Case and Prior Incision Time

In this section, we estimate the effect of being a non-first case of a surgeon in a day (i.e. the second case or later). This sheds lights on the potential mechanism for the impact of surgeon daily workload. We estimate the model:

$$y_i = X_i\beta + \gamma NonFirst_i + \varepsilon_i, \tag{S.14}$$

where $NonFirst_i$ is a binary variable, indicating if case i is not the first case of its surgeon on its surgery day. Here we sort the cases by their OR entry time. In our sample, we have 1642 (30.6%) cases with $NonFirst_i = 1$, i.e., being non-first ones. Out of these non-first cases, 88% (resp. 11.6%) are the second (resp. third) case of the surgeon in a day.

We estimate model (S.14) using TSLS with the two IVs described in Section 3.1. Since our two IVs are constructed at the daily level, the variation in $NonFirst_i$ picked up by the IVs in the first-stage is also at the daily level. That is, we cannot differentiate the treatment effect between the second case versus the third or fourth case. The two IVs still significantly impacts the endogenous variable $NonFirst_i$ with expected signs: The coefficients for $ToTOther_i$ and $ElecRatioWD_i$ are -0.027 and 0.505, respectively. Both are statistically significant at the 0.1% level. This shows that the relevance condition of the two IVs is still satisfied.

Table S.4 Estimated Effects of Being a Non-first Case in a Day: Full Sample

	OR Time (in hours)	Incision Time (in hours)	Post-LOS (in days)	Total ICU (in days)
Effect of non-first case	1.075^{***} (0.296)	1.093^{***} (0.209)	2.453^{**} (0.860)	2.773^{***} (0.544)
Patient demographics	Y	Y	Y	Y
Risk/operative factors	Y	Y	Y	Y
System occupancy	Y	Y	Y	Y
Surgeon fixed effect	Y	Y	Y	Y
Time fixed effects	Y	Y	Y	Y
Num. Obs	5345	5345	5344	5319

The estimated effects of being a non-first case on surgery duration and patient outcomes for the full sample. Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

We report the estimated effects of being a non-first case in a day in Table S.4. We see that being a non-first case in a day is associated with longer surgery duration and worse patient outcomes (longer ICU time and post-LOS). This is inline with our findings on the average effect of surgeon daily workload. The magnitudes of effects are also consistent. Recall that the estimates in Table 4 represent the average effect for all cases performed by a surgeon in a day, while the coefficients in Table S.4 are the effects only for the non-first cases. On the other hand, most non-first cases (88%) are the second case of the surgeon in our sample. Thus, the fact that the effects on non-first cases (in Table S.4) are about twice of the average effects for all cases (in Table 4) suggests that the the magnitudes of effects are largely consistent.

As an addition check, we also estimate the effect of a surgeon's total prior incision time in a day before the start of the focal case. The total prior incision time, $PriorInc_i$, includes the incision time of all cases by the focal surgeon that started before the OR entry time of case i. If case i is the first case of the focal surgeon in a day, its prior incision time will be zero. We replace $NonFirst_i$ by $PriorInc_i$ in (S.14) and still estimate the model by TSLS with the two IVs. The relevance condition of the two IVs still hold for the endogenous variable $PriorInc_i$: in the first stage regression, the coefficients of $ToTOther_i$ and $ElecRatioWD_i$ are -0.059 and 2.413 respectively, both significant at the 0.1% level. Table S.5 reports the estimated effects of the total prior incision time for the four dependent variables. Again, the results are largely consistent with those given in Table S.4 as well as our main specification. We find longer prior incision time leads to longer surgery duration and post-surgery LOS. Moreover, given the average prior incision time for non-first cases is roughly four hours, the magnitudes of the estimates in Tables S.4 and S.5 are also consistent.

Table S.5 Estimated Effects of Total Prior Incision Time: Full Sample

	OR Time (in hours)	Incision Time (in hours)	Post-LOS (in days)	Total ICU (in days)
Effect of total prior incision time	0.258^{**} (0.086)	0.268^{***} (0.045)	0.509^{**} (0.197)	0.670^{***} (0.115)
Patient demographics	Y	Y	Y	Y
Risk/operative factors	Y	Y	Y	Y
System occupancy	Y	Y	Y	Y
Surgeon fixed effect	Y	Y	Y	Y
Time fixed effects	Y	Y	Y	Y
Num. Obs	5345	5345	5344	5319

The estimated effects of total prior incision time on surgery duration and patient outcomes for the full sample. Robust standard error is reported in parenthesis; $^{\dagger}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

S.4. Supplementary Tables

This section includes the supplementary tables. Table S.6 reports the summary statistics of the two IVs. Table S.7 summarizes the correlation between the two IVs and 21 observable severity factors of patients. Table S.8 reports the p-values from the Sargan's test for the validity of the IVs. Table S.9 reports the estimation results for various robustness checks. To save space, here we report the estimated effects on OR time, post-LOS, and total ICU time. The results for incision time are similar to those for OR time, and are available upon request. For ease of comparison, we include in the first two rows ("Base Model") the estimation results from our main specification in Tables 4 and 5. Panel A shows the results when we add a dummy controlling the OR entry hour of each case. In Panel B, we provide the estimated effects when we only use one of the two IVs. In Panel C, we use different specifications for calculating the second IV $ElecRatioWD_i$. Panel D reports the results when we include $TotOther_i$ as an exogenous variable in our model (1) and only use $ElecRatioWD_i$ as the IV. Finally, Panel E examines the impact when we use different winsorization levels as well as a log transformation for the dependent variables. The estimated effects under these set-ups are qualitatively similar to our main analysis and are discussed in Section 4.3.

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 $Workload_i$ by TSLS are given in Table S.10. The first column shows the original results in Tables 4 and 5, 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 S.10 that our estimated effects for post-LOS remain similar and robust in all these conservative measures.

Next, Figure S.1 shows 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).

Finally, we report in Table S.11 the estimated coefficients of selected patient demographic, risk, operative, and operational factors in our regressions. The coefficients are estimated by TSLS using the two IVs on the full sample.

Table S.6 Summary Statistics of IV

	Full Sample			Elective Sample			Non-elective Sample		
IV	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std
$TotOther_i$	4.11	4.00	1.79	4.30	4.00	1.67	3.95	4.00	1.87
$ElecRatioWD_i$	0.22	0.23	0.12	0.25	0.25	0.11	0.20	0.21	0.13

Table S.7 Correlation between IVs and Observable Severity Factors

	$TotOther_i$	$ElecRatioWD_i$
Gender: Male	−0.019	−0.026
Age	0.054	0.072
Smoke	−0.006	−0.003
Drug use	−0.017	−0.015
NYHA: III or IV	0.002	−0.102
Endocarditis	−0.017	−0.060
Peripheral arterial disease	0.002	−0.042
Incidence	−0.030	−0.084
Lung disease	0.023	0.015
Liver disease	0.020	−0.011
Thoracic aorta disease	−0.020	0.005
Dialysis	−0.013	−0.017
Diabetes	0.005	−0.040
Cancer	0.008	0.011
Hypertension	0.034	0.043
Previous intervention	−0.031	−0.077
Carotid stenosis	0.027	0.013
Cardiogenic shock	−0.052	−0.185
Syncope	0.001	−0.007
MI	−0.007	−0.068
Systolic pressure: high	0.009	−0.027

Table S.8 p-values from Sargan's Test for the two IV's Validity

OR time			Incision time			Post-LOS			Total ICU time		
Full	Elec	Non-elec	Full	Elec	Non-elec	Full	Elec	Non-elec	Full	Elec	Non-elec
0.751	0.186	0.405	0.535	0.134	0.544	0.632	0.951	0.207	0.827	0.566	0.887

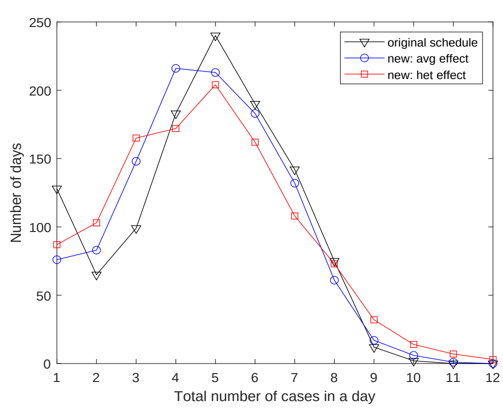

Figure S.1 Frequency distributions of total number of cases by all surgeons in a day

Table S.9 Robustness Checks: Estimated Effects of Daily Workload (Number of Other Cases)

Panel A: Controlling for OR entry time

	OR time (in hours)			Post-LOS (in days)			Total ICU (in days)		
	Full	Elec	Non-elec	Full	Elec	Non-elec	Full	Elec	Non-elec
Base model	0.455*** (0.137)	0.567* (0.256)	0.377* (0.158)	1.109*** (0.335)	−0.155 (0.867)	1.605* (0.807)	1.166*** (0.240)	0.101 (0.620)	1.647*** (0.367)
Dummy for OR entry time	0.326* (0.131)	0.394* (0.187)	0.286† (0.165)	1.118*** (0.328)	0.003 (0.820	1.575* (0.741)	1.081*** (0.228)	0.142 (0.573)	1.527*** (0.333)

Panel B: Estimation with one IV

	OR time (in hours)			Post-LOS (in days)			Total ICU (in days)		
	Full	Elec	Non−elec	Full	Elec	Non−elec	Full	Elec	Non−elec
$TotOther_i$	0.424* (0.181)	0.356*** (0.090)	0.477 (0.291)	1.370** (0.528)	−0.110 (1.152)	2.559† (1.347)	1.090** (0.413)	0.371 (0.800)	1.716* (0.846)
$ElecRatioWD_i$	0.498* (0.254)	0.902† (0.529)	0.246 (0.184)	0.738 (0.547)	−0.225 (1.239)	0.350 (1.170)	1.272*** (0.193)	−0.314 (0.930)	1.559*** (0.468)

Panel C: Other specifications for $ElecRatioWD_i$

	OR time (in hours)			Post-LOS (in days)			Total ICU (in days)		
	Full	Elec	Non-elec	Full	Elec	Non-elec	Full	Elec	Non-elec
Using prior cases in $[t-L,t)$	0.423*** (0.127)	0.560** (0.192)	0.367* (0.175)	1.160* (0.514)	0.470 (0.876)	1.516† (0.789)	1.193*** (0.300)	0.330 (0.673)	1.630*** (0.327)
$L = 60$ days	0.428** (0.159)	0.431† (0.236)	0.420† (0.226)	1.583*** (0.322)	0.627 (0.872)	2.076** (0.747)	1.231*** (0.207)	0.692 (0.639)	1.502*** (0.414)
$L = 90$ days	0.448** (0.152)	0.479† (0.280)	0.419* (0.176)	1.281*** (0.376)	0.249 (0.864)	1.730* (0.759)	1.091*** (0.232)	0.317 (0.660)	1.437*** (0.324)

Panel D: Controlling for $TotOther_i$ as an exogenous variable

	OR time (in hours)			Post-LOS (in days)			Total ICU (in days)		
	Full	Elec	Non−elec	Full	Elec	Non−elec	Full	Elec	Non−elec
Control for $TotOther_i$	0.507† (0.289)	0.953 (0.592)	0.214 (0.218)	0.666 (0.612)	−0.236 (1.355)	0.044 (1.333)	1.292*** (0.228)	0.373 (1.011)	1.537* (0.615)

Panel E: Other winsorization levels

	OR time (in hours)			Post-LOS (in days)			Total ICU (in days)		
	Full	Elec	Non-elec	Full	Elec	Non-elec	Full	Elec	Non-elec
Winsorize: 95th pct	0.491*** (0.109)	0.556** (0.204)	0.434** (0.181)	0.863* (0.336)	0.238 (0.758)	1.065† (0.611)	0.768*** (0.165)	0.218 (0.448)	0.984*** (0.227)
Winsorize: 99th pct	0.525*** (0.124)	0.576** (0.210)	0.472* (0.199)	1.564* (0.639)	−0.186 (0.910)	2.248† (1.300)	1.453*** (0.399)	−0.061 (0.857)	2.161** (0.676)
Log transform	0.081** (0.026)	0.106* (0.045)	0.064* (0.027)	0.080* (0.033)	0.062 (0.060)	0.082 (0.062)	0.119* (0.060)	0.042 (0.102)	0.153† (0.091)

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

Table S.10 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
Full Sample	1.109*** (0.335)	1.275*** (0.316)	1.140*** (0.325)	1.077** (0.333)	1.205*** (0.296)
Non-elective Sample	1.605* (0.807)	1.789* (0.797)	1.635* (0.801)	1.589* (0.803)	1.678* (0.769)

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

Table S.11 Coefficients of Control Variables in Model (1): Full Sample

	OR time	Post-LOS	Tot ICU time
Gender: Male	0.136*	−0.051	0.098
	(0.067)	(0.423)	(0.225)
Age	−0.005	0.050***	0.040***
	(0.003)	(0.015)	(0.007)
Race: Asian	0.097	1.247*	0.179
	(0.122)	(0.520)	(0.259)
Race: Black	0.148	0.401	0.15
	(0.103)	(0.529)	(0.310)
Race: Other	0.158	−0.314	−0.097
	(0.135)	(0.383)	(0.350)
Endocarditis	0.363***	1.431*	0.147
	(0.076)	(0.622)	(0.390)
Lung disease	0.238***	1.487***	0.779***
	(0.062)	(0.247)	(0.190)
Peripheral arterial disease	0.177	1.077***	0.605*
	(0.119)	(0.269)	(0.261)
Liver disease	0.207	2.167†	1.190*
	(0.131)	(1.108)	(0.483)
Thoracic aorta disease	0.346**	0.406	0.491
	(0.123)	(0.678)	(0.540)
Cancer	0.06	0.515*	0.485
	(0.099)	(0.237)	(0.340)
Hypertension	0.095*	0.396	0.374†
	(0.037)	(0.393)	(0.216)
Recent drug use	−0.016	0.990†	0.384
	(0.089)	(0.576)	(0.300)
Dialysis/renal failure	0.398**	6.119***	4.875***
	(0.148)	(0.888)	(0.402)
Diabetes	0.175**	2.167***	1.007***
	(0.065)	(0.246)	(0.255)
Smoke	−0.016	−0.041	0.116
	(0.026)	(0.172)	(0.080)
Carotid stenosis	0.068	0.001	−0.242
	(0.058)	(0.489)	(0.424)
Syncope	−0.176†	−0.265	−0.629
	(0.105)	(0.538)	(0.413)
Previous intervention	0.104*	1.128*	0.844***
	(0.046)	(0.458)	(0.211)
Cardiogenic shock	0.350	4.132***	2.011***
	(0.231)	(0.623)	(0.424)
Prior MI	0.023	0.350	0.477†
	(0.057)	(0.418)	(0.268)
Aorta procedure	1.407***	−0.229	−0.143
	(0.337)	(1.276)	(0.783)

	OR time	Post-LOS	Tot ICU time
Status: Urgent	0.096*	0.976***	0.511**
	(0.046)	(0.284)	(0.188)
Status: Emergent	0.512*	2.515	1.820**
	(0.241)	(1.748)	(0.700)
Status: Salvage	0.253	1.52	5.105***
	(0.675)	(2.933)	(1.142)
Adm Type: ED	0.031	1.160*	0.473†
	(0.142)	(0.531)	(0.246)
Adm Type: Transfer in	−0.053	1.751***	1.118**
	(0.086)	(0.391)	(0.346)
Adm Type: Other	−0.875***	1.394	0.441
	(0.263)	(1.781)	(1.210)
Pre-LOS	0.002	0.040**	0.004
	(0.002)	(0.015)	(0.007)
Department census	0.007***	0.001	0.007
	(0.001)	(0.008)	(0.008)
Num. of arteries bypassed	0.418***	−0.122	0.167
	(0.044)	(0.297)	(0.173)
NYHA: N/A	0.077	-0.001	-0.104
	(0.050)	(0.153)	(0.129)
NYHA: II	0.038	0.037	−0.236
	(0.073)	(0.186)	(0.185)
NYHA: III	0.214***	0.988*	0.683*
	(0.056)	(0.386)	(0.281)
NYHA: IV	0.354***	5.491***	2.643***
	(0.105)	(0.956)	(0.544)
NYHA: Unknown	0.214	1.735**	0.52
	(0.130)	(0.656)	(0.456)
Non-initial surgery	1.486***	0.582	0.312
	(0.132)	(0.468)	(0.252)
Systolic pressure: Low	−0.072	−1.303***	−0.853**
	(0.127)	(0.323)	(0.280)
Systolic pressure: Unknown	−0.108	−1.634***	−1.003**
	(0.123)	(0.319)	(0.307)
Num Cases	0.455***	1.109***	1.166***
	(0.137)	(0.335)	(0.240)
Constant	4.276***	3.583†	−2.075*
	(0.479)	(1.841)	(1.021)
Procedure type	Y	Y	Y
Surgeon fixed effect	Y	Y	Y
Weekday fixed effect	Y	Y	Y
Monthly fixed effect	Y	Y	Y
Year fixed effect	Y	Y	Y
Num. obs	5,345	5,344	5,319