Adaptive Server Behavior to Schedule Deviations and Its Consequences: Evidence from Operating Rooms

Yiwen Jin
Sauder School of Business, University of British Columbia, Vancouver, BC Canada V6T 1Z2, yiwen.jin@sauder.ubc.ca

Yichuan Ding
Desautels Faculty of Management, McGill University, Montreal, QC Canada H3A 1G5, daniel.ding@mcgill.ca

Steven M. Shechter
Sauder School of Business, University of British Columbia, Vancouver, BC Canada V6T 1Z2, steven.shechter@sauder.ubc.ca

Jugpal S. Arneja
Department of Surgery, University of British Columbia, Vancouver, BC Canada V5Z 1M9 jarneja@cw.bc.ca

Problem Definition: We study how clinical teams adaptively respond to real-time deviations from the planned operating room (OR) schedules and the associated consequences of these responses. Specifically, we investigate whether clinical personnel adjust their service speed when they are ahead of or behind the original schedule and whether this affects patient readmissions and reoperations.

Methodology: We empirically explore these questions using a unique surgery data set that includes actual and scheduled surgery time stamps. We construct a dynamic panel model and utilize the Arellano-Bond approach to identify adaptive behavior. The adaptive service behavior renders an instrumental variable (IV) for us to address endogeneity issues in estimating the impact of surgical speed on surgical quality.

Results: We identify a new type of adaptive server behavior within schedules, which complements the existing scheduling and behavioral queueing literature. We also contribute to the literature on surgical speed and quality. We find that surgical and cleaning teams speed up when they fall behind schedule and slow down when they get ahead of schedule, with the slowdown exhibiting a stronger effect. Quantitatively, surgical teams expedite the next surgery by an average of 5.6% when facing one standard deviation (SD) delay in the planned start for that surgery, whereas they take on average 10.5% longer when they are one SD ahead of schedule. In the turnover times, cleaning teams accelerate by 10.3% (slow down by 22.1%) on average when they are one SD ahead of (behind) schedule. We then leverage the deviation from the scheduled start as an IV. We present a causal study that a faster-than-scheduled procedure duration erodes surgical quality by increasing 30-day readmission and reoperation probabilities.

Managerial implications: Our research unveils the responsive patterns of surgical and cleaning teams when confronted with deviations from schedules in ORs. We further find that faster-than-scheduled speedup in surgeries is detrimental to surgical quality. Understanding this behavioral phenomenon can help hospital managers predict end-of-shift (EOS) times better. Through a counterfactual analysis, we quantify and obtain a convex relationship between readmissions/reoperations and EOS time differences. This can assist managers in scheduling surgeries to achieve desired efficiency-quality trade-offs.

Key words: operating room scheduling; behavioral queueing; speed-quality tradeoff

1. Introduction

Although service providers strive to complete scheduled activities on time, deviations are inevitable, and the actual progress may run behind or ahead of schedule. For example, 40% - 70% of construction projects have undergone delays in different countries (Soliman and Alrasheed 2021), while up to 28% of projects have been completed earlier than the due dates (Wolfe 2020). In the airline industry, approximately 20% of flights are delayed (Eltoukhy et al. 2017, Dand et al. 2019). Delays in medical appointments have been widely reported (Ahnood et al. 2018, Qi 2017), and the situation was recently exacerbated by the COVID-19 pandemic (McSheffrey 2022).

Service providers' responses to schedule deviations may vary, and we categorize them into three types using terminologies borrowed from biology (Cinquin et al. 2002). The first type is negative (balancing) feedback, which describes service providers who attempt to reverse schedule deviations by speeding up when facing delays and slowing down when ahead of schedule. For example, in project and airline management, where schedule disruptions arise frequently, responses include expediting activities to catch up when behind schedule (Li et al. 2000, Montet 2014, Bubalo and Gaggero 2015). The second type of feedback mechanism is positive feedback, which describes service providers who react to deviations by further amplifying them. One example is Brooks' Law in software project management, where delayed progress of software development leads to further delays (Chernoguz 2011). The third possibility is that there does not appear an apparent response; i.e., service providers adhere to their original tempo in each task, regardless if they are running behind or ahead of schedule.

The three response types have different implications for total service duration and system throughput. We refer to this immediate effect as the first-order impact. Moreover, service duration changes may affect the quality of service delivered. We refer to this as the second-order impact. For instance, pilots accelerate more to save time when flights are delayed at the cost of increased fuel burn and carbon emissions (Marla et al. 2017). Construction projects can also be rushed to catch up with the project deadlines, but often at the cost of poorer project quality, such as rework (Gao and Low 2014) or reduced safety of the workers (Cheah 2007, Nepal et al. 2006).

In this paper, we explore service providers' adaptive behavior and their first and second order impact in a critical healthcare setting–the operating room (OR). Delays are pervasive in ORs, with approximately 70% of surgical procedures experiencing them (Qi 2017). OR delays lead to longer patient wait times and possible cancellations, which greatly inconvenience patients and their caregivers. On the other hand, surgeries that finish earlier than scheduled may result in idle surgery teams, wasting costly healthcare resources (Childers and Maggard-Gibbons 2018).

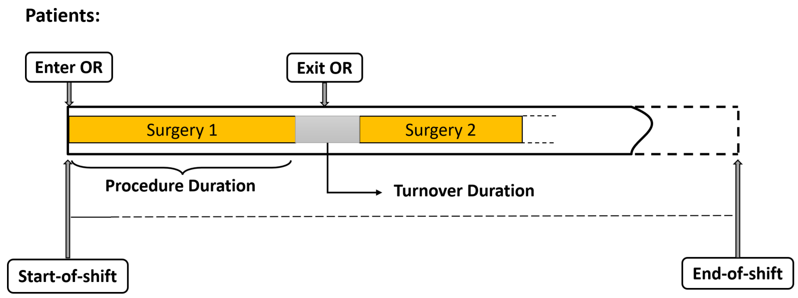

Figure 1 **OR Shift Structure**

The most significant difference between ORs and other healthcare service sites is that there is almost no chance for walk-in patients because most operations are pre-arranged appointments made several months earlier. Emergent cases can be added to shifts at the last minute, but that only happens for about 6.8% of all OR shifts from our data. As a result, the service providers in ORs, which include the surgical teams and the cleaning teams (with some overlap in their staff), are readily aware of schedule deviations by comparing the actual and planned surgery start times.

Knowing how the surgical and cleaning teams respond to schedule deviations can help OR managers better predict procedure duration, improve patient experience, and estimate the potential risk associated with the responses (Lapierre et al. 1999). Researchers have attempted to understand surgical teams' response mechanisms by surveying or interviewing surgeons. Nevertheless, no consensus has been reached – there are results in support of positive feedback (Moulton et al. 2010), negative feedback (Orri et al. 2015), and no feedback (Dexter et al. 2006).

Instead of relying on subjective survey results, we take an econometric approach to investigate adaptive behavior and its effects. Using a unique surgical data set including both the actual and planned enter and exit OR times, we were able to calculate the actual and planned procedure duration (from enter to exit) and the turnover duration (from exit to next case's enter) as shown in Figure 1. We develop empirical methods to investigate how surgical and cleaning teams respond to schedule deviations. The responses by surgical and cleaning teams can be quantified by changes in the procedure or turnover duration (i.e., cleaning and preparation), respectively, as a result of the difference between actual and planned start and end times. To the best of our knowledge, this is the first study to use surgical data and econometric methods to investigate the response behavior of surgical and cleaning teams to real-time schedule updates in the OR. We find that both surgical and cleaning teams speed up when falling behind and slow down when getting ahead of schedule, with the slowdown exhibiting a stronger effect.

While keeping schedules on time is often desirable, attempts to do so might come with a cost, such as compromised service quality. A rich literature finds a positive correlation between the duration

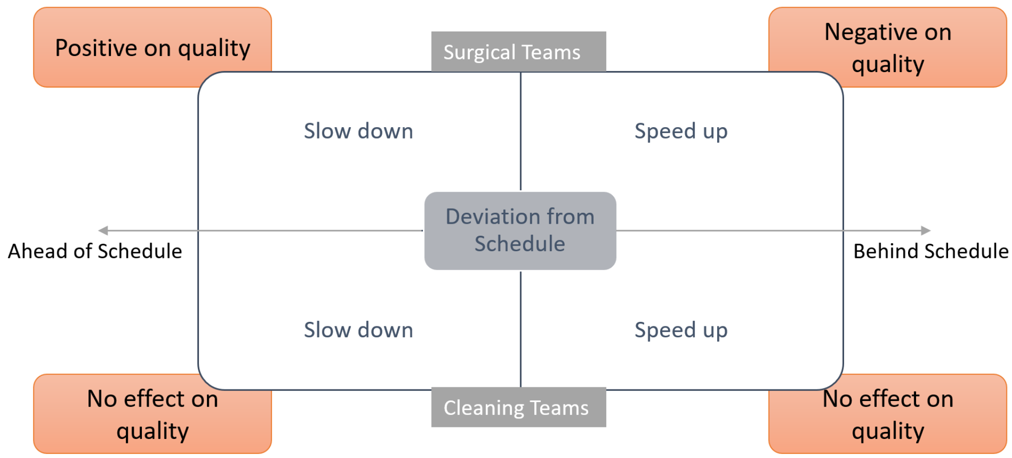

Figure 2 **Clinical Personnel Responses to Real-time Schedule Deviations in OR Shifts and Subsequent Impact on Quality**

of surgical procedures and post-surgical complication rates (Cheng et al. 2018). However, most studies only demonstrate a *correlative* instead of a *causal* relationship between these measures. Furthermore, researchers acknowledge that the findings of such correlation are limited due to the inability to adjust for potential confounders (Jackson et al. 2011). This leaves the identification of a causal effect an open question. For example, complications during surgeries tend to both increase the surgical duration as well as increase post-surgical complication risks. In addition, certain surgical procedures, such as distraction osteogenesis (Kempton et al. 2014), are confounders as they correlate with longer procedure duration and higher complication rates. Using data on 30-day post-surgical readmissions and reoperations, we address the endogeneity issue by designing an instrumental variable (IV) that helps identify the causal impact of procedure duration on surgical quality. The validity of our IV builds on the surgical teams' adaptive behavior that we identify. After correcting for the bias with the IV, we find that faster surgical procedures lead to higher readmission and reoperation probabilities. Figure 2 summarizes the different responsive behavior of surgical and cleaning teams in ORs and the subsequent influence on surgical quality.

To investigate these two questions, we leverage a comprehensive administrative data set that includes 7,868 surgery cases from April, 2018, to October, 2019, in a primary children's hospital in Canada. The data records each surgery case's actual and scheduled start and end times. Surgeons and the booking office of the hospital take charge of the scheduling process of each case at most times (Johnston et al. 2019), and the surgical teams are informed of the planned start and end times of all cases at least one day earlier. By comparing the actual and planned time schedules for each surgical case, we obtain (1) relative procedure duration (RPD), which is the difference between the actual procedure duration and the planned one, (2) deviation from the scheduled start (DSS), which is the difference between the actual start time of the surgery and the planned one, (3) relative turnover duration (RTD) and (4) deviation from the scheduled end (DSE) of the procedure. Overall 62% of the operations experience delayed start compared with schedules, which corresponds to positive DSS values. Based on these quantities, we estimate the impact of DSS on subsequent surgical duration (i.e., RPD) and the impact of DSE on subsequent turnover duration (i.e., RTD).

We adopt several empirical strategies in this study. First, to identify how surgical and cleaning teams respond to real-time deviations from the original schedule, the standard ordinary least squares (OLS) method does not work. Because the DSS for the focal surgery is correlated with the RPD of the previous surgeries in the same shift, the OLS estimator is biased (Nickell 1981). A similar rationale applies to DSE and RTD. To address this issue, we formulate an auto-regression in which a linear transformation addresses the above correlation. The general moment method (GMM) estimator of dynamic panel models is applied in which the high-order lagged regressors serve as instruments to derive an unbiased estimator (Arellano and Bond 1991).

The GMM estimator shows that surgical and cleaning teams speed up when they fall behind and slow down when they are ahead of schedule. Specifically, when a surgery case has one standard deviation (SD) delayed start than planned (23.03 minutes), the surgical team would expedite the focal surgery by 5.6% on average relative to the median procedure duration. For one SD delayed start (15.88 minutes), the focal surgery will be prolonged by 10.5%. At the turnover stage, the cleaning teams speed up by 10.3% when facing a SD (26.87 minutes) delayed exit and slow down by 22.1% when facing a SD (18.89 minutes) early exit. The responses of cleaning teams in the turnover stage are more intense than surgical teams and the slowdown effect is stronger. Second, to address the potential endogeneity in estimating the impact of surgical duration on quality, we utilize DSS as an IV. According to the IV-probit model, an increment of procedure duration results in lower 30-day readmission (reoperation) probabilities. We did not use IV when estimating the impact of the preceding RTD on the quality of the subsequent surgery because there is less endogeneity concern in the estimation. The estimation results show that the preceding RTD has no significant impact on quality.

To gain a comprehensive understanding of the influence of surgery speedup and slowdown on OR shift end times and surgical quality, we conduct a counterfactual analysis. We unveil a convex relation between these two quantities, which depicts the efficiency-quality trade-off frontier in the current OR. We conclude that reducing the gap between actual and planned shift end times can be achieved at the cost of a higher average 30-day readmission and reoperation probabilities or vice-versa.

To develop a comprehensive understanding of the impact of surgery speedup and slowdown on OR end-of-shift (EOS) times and surgical quality, we conduct a counterfactual analysis. We reveal a convex relationship between these two factors, illustrating the efficiency-quality trade-off frontier within the current OR. We conclude that narrowing the gap between the actual and planned EOS can be accomplished, but at the expense of increased 30-day readmission and reoperation probabilities, or vice versa.

1.1. Contribution

Our study contributes to healthcare management and service operations literature in the following aspects.

First, we identify healthcare providers' responsive behavior to real-time schedule deviations in the OR setting, which complements the vast literature on adaptive server behavior in operations management. We find that the balancing feedback behavior exists for both surgical and cleaning teams. Based on discussions with our clinical co-authors, reasons for this may include wanting to avoid potential add-on cases or working overtime. This finding can help surgeons and OR managers, especially those working in a socialized healthcare system like Canada, better predict surgery completion times and improve shift planning. Our findings encourage policymakers to reevaluate and develop incentive schemes for surgical teams within a socialized healthcare system, where they currently receive a nearly fixed salary.

Second, our study provides new insights regarding the causal relationship between surgical duration and quality, which is known to be difficult due to various confounders that bring endogeneity to the estimation process. We find that after addressing endogeneity, faster-than-scheduled surgeries lead to increased post-surgical complication rates, which is contrary to the correlative findings in the medical literature that longer surgical duration is associated with poorer quality outcomes. We propose an IV to address this challenge, which has a high level of generality and the potential to be applied to other service settings to study the quality-speed relation.

Finally, we propose an auto-regressive model to address the serial correlation of procedure duration within the same shift. We utilize the GMM estimator from dynamic panel models with the high-order lagged regressors as instruments to derive an unbiased estimator. This allows us to apply these statistical methods to explore the behavioral characteristics of service providers. These methods may find other applications in queueing models, where serial correlations are common when customers are served in sequence.

The rest of the paper is organized as follows. Section 2 reviews the literature related to our work. Section 3 discusses the formation of hypotheses. Section 4 describes the healthcare context and structure of the data set, and Section 5 formulates the model and the GMM estimation method. Section 6 presents the estimation results and heterogeneity analysis. Section 7 discusses the impact on surgical quality and the IV-probit estimation results. Section 8 summarizes the counterfactual analysis. Finally, Section 9 concludes the findings and provides managerial insights.

2. Literature Background

2.1. Healthcare Provider Productivity

Our research contributes to the expanding body of empirical literature that explores the relationship between operational factors and healthcare providers' productivity. KC and Terwiesch (2009) find that hospital workers accelerate the service rate as the load increases initially, while they decrease the service rate as the load continues to increase or remains high after a certain period. They also provide evidence that such a high load of overwork curtails the quality of care. Chan et al. (2017) propose a queueing model to study how delay experienced by a patient impacts service times in intensive care units. Berry Jaeker and Tucker (2017) explore the effect of hospital occupancy level on patient length-of-stay (LOS). They identify a "saturation effect" for hospital employees, as they cannot speed up to overcome the workload when the occupancy level is high. Johnston et al. (2019) conduct an exploratory study of the surgeons' decision process of scheduling surgery cases and identify different priority structures of decision making. Deo and Jain (2019) study how patient length-of-stay (LOS) in a clinic varies with hour-of-day, and they find the LOS tends to be larger early in the day while reduces later. Their findings diverge from ours because our study focuses on how service duration is affected by deviations from the original schedule, as opposed to examining the time of day. Shen et al. (2021) also find a negative impact of surgeon workload on both the efficiency and quality of cardiac surgeries and propose a mixed integer program to incorporate these factors in surgery scheduling. Other papers discuss how behavioral factors influence the productivity and performance of service providers Kim et al. (2015), Kc (2019), Wang and Pourghannad (2020), Ibanez and Toffel (2020). Researchers also investigate how scheduled workload impacts turnover times between surgeries (Wang et al. 2013) whereas they find no significant relationship.

Besides the influence of workload and other healthcare sites' environmental factors, doctors' experience and team structure have also been investigated. Kc and Staats (2012) explore how surgeons' experience in focal and related tasks influence their performance. They find that a surgeon's focal experience has a greater effect on surgeon performance. Avgerinos and Gokpinar (2017) study the influence of surgery team's familiarity on team productivity. They find that a high dispersion of pairwise team familiarity (measured by the standard deviation of the number of past collaborations) lowers productivity in cardiac surgeries. Ramdas et al. (2018) study surgeons' learning and forgetting with new surgical device versions and the resulting impact on surgery duration and quality. They identify significant productivity and quality costs associated with the first use of new devices. Ibanez et al. (2018) explore how radiological doctors deviate from the prescribed task sequence at discretion when they read images, which erodes productivity.

Our research adds to this stream of research by examining how real-time deviations from schedules influence both the efficiency and quality of surgeries. We analyze the effects of real-time discrepancies between actual occurrences and planned schedules, which sets our study apart from other research in the field. By dissecting the dynamics of shifts, we offer a novel perspective for understanding operating room behavior.

2.2. Surgical Speed and Quality

Our research also relates to the general topic of healthcare efficiency-quality trade-offs. In operations management, Song and Veeraraghavan (2018) review papers on healthcare delivery quality and summarize the most commonly used service quality metrics in healthcare settings. They show that a critical combination of quality measures is structure, process, and outcome. Roth et al. (2019) study the relationship among efficiency, quality, and patient experience in hospitals.

In the medical literature, a vast body of research is associated with this topic. A correlative study by Daley et al. (2015) utilizing National Surgical Quality Improvement Project (NSQIP) data shows a correlation between surgical duration and complications. Longer surgery time than the statewide established standard is associated with higher risk. Wan et al. (2020) explore post-surgical complications of plastic surgery and find that increased operative time is a common risk factor for plastic surgeries. Christakis et al. (2019) find that operation duration is correlated with complications for posterior retroperitoneoscopic adrenalectomy procedures. A survey paper based on a literature review and meta-analysis by Cheng et al. (2018) propose that prolonged surgery duration leads to an increased risk of complications.

While the aforementioned medical examples indicate that longer operation durations are correlated with poorer patient outcomes, the literature also presents other findings on this topic. In a retrospective analysis of abdominal and colorectal surgeries, Liverani et al. (1994) find that longer surgery duration does not lead to a significant increase in complications. Schliemann et al. (2009) find no correlation between the surgery duration and the incidence of complications in an orthopedic procedure. Kempton et al. (2014) review 30 papers on hand surgery and explore the duration of treatment and complication rates. They find that the duration does not have a significant odds ratio for post-surgical complications. McDermott et al. (2015) systematically review the preoperative and intraoperative factors associated with post-surgical complication risks and also influence procedure duration.

Most of the studies in this thread focus on a *correlative* relationship instead of a *causal* connection. We contribute to them by proposing an IV to tackle unobservable confounders when carrying out a causal analysis between surgical duration and quality. Our approach enables us to address this question with a causal perspective.

3. Hypothesis Development

Schedules are designed to enhance productivity and punctuality. However, discrepancies between actual and planned start/end times are bound to arise once schedules are announced. In ORs, the clinical teams review the schedules a day in advance and remain cognizant of their progress in relation to the planned timeline throughout the day. In our analysis, we draw upon firsthand insights from surgeons and corroborating evidence from the literature to formulate plausible hypotheses regarding their responses to deviations between actual and planned start times.

The firsthand experience comes from the interviews with our surgeon collaborator and the supporting staff. We verify the following practices in the hospital. 1) There is a culture that strives to avoid canceling surgeries when running late, as patients have typically waited for a year or longer. Consequently, the surgical and cleaning teams have to work overtime, instead of canceling the case at the end of shift when they cannot finish on time. This practice supports the conjecture that both teams speed up when faced with delays to circumvent working overtime. 2) On the flip side, besides the scheduled elective cases, emergent add-on cases are likely to be assigned to shifts that are finishing or have finished earlier than scheduled by the manager. However, such extra operative cases do not result in increased financial compensation under the socialized health system. From an incentive perspective, this supports both team's slowdown behavior, meaning they decelerate when ahead of schedule to avoid extra add-on cases.

We also find corroborating evidence from the literature that supports the surgical teams' responsive behavior as described above. According to interviews conducted with 23 surgeons, Orri et al. (2015) show that surgeons feel obliged to demonstrate good time management of ORs when faced with time constraints such as the scheduled shift. They also find a psychological reason that surgeons want to build a good reputation for "good sense of time and schedules." This finding reveals another motive that explains the delay-speedup behavior. Conversely, if they are ahead of schedule, they feel less pressured and naturally slow down to stay close to the planned schedule. This phenomenon is referred to as "situationally responsive slowing-down moment" (Moulton et al. 2010), which complements the incentive reason behind the slowing down behavior of surgical teams. Combining these findings, we propose the following hypothesis on surgical teams' responsive behavior.

Hypothesis 1A. *Surgical teams speed up when they fall behind schedule, and they slow down when they are ahead of schedule in the surgery stage.*

For cleaning teams, we also learn that they typically receive indications from surgical teams when falling behind or ahead of schedule at the end of a procedure. The cleaning teams then act accordingly during the turnover stage to minimize deviations from schedules, demonstrating both speedup (when delayed) and slowdown (when getting ahead). Additionally, cleaning teams share concerns about not canceling surgeries and avoiding extra add-on cases, similar to surgical teams. Overall, we propose the following hypothesis on cleaning teams' responsive behavior.

Hypothesis 1B. *Cleaning teams speed up when falling behind and slow down when being ahead in the turnover stage.*

In line with Hypotheses 1A and 1B, both surgical and cleaning teams exhibit a balancing feedback response in their service speed when encountering deviations from the pre-established schedule.

Addressing our second research question, we consider the consequences of altering surgery durations. Substantial evidence in medical research suggests that extended operative durations correlate with increased post-surgical complication risks. Cheng et al. (2018) review 66 related studies and conclude that the likelihood of complications rises significantly with prolonged operative durations. Specifically, they find a 14% increase in the likelihood of complications for every additional 30 minutes of operating time. However, the underlying mechanism for this finding remains uncertain since various factors in prolonged surgeries can lead to post-surgical complications. Common examples include increasing microbial exposure and associated infections, tissue retraction, and bleeding. In light of this literature, we propose the following hypothesis.

Hypothesis 2A. *All else equal, a longer procedure time is associated with an increased post-surgical risk.*

However, the analyses in the existing literature on this topic have a crucial flaw: they are primarily *correlative* instead of *causal*. As mentioned in Cheng et al. (2018), correlative analyses suffer from the inability to adjust for potential confounders, which may lead to biased results. For example, a specific procedure called distraction osteogenesis is associated with both longer surgical duration and higher complication rates, making it an endogenous confounder between procedure duration and complications (Kempton et al. 2014). The anesthesia step of mechanical ventilation is another potential confounder that influences certain post-surgical complications while adding to procedure duration (Güldner et al. 2015). We deduce that there are multiple sources of confounders in the relationship between surgical speed and quality, and they cannot be ignored. We also conclude that these confounders tend to increase both surgical duration and complication risk levels.

Apart from the influence of confounders, we observe that performing procedures faster may increase the likelihood of making mistakes in surgery procedures (Arora et al. 2009, 2010), which may further lead to rising post-surgical complication risks. Young et al. (2014) study the impact of operating time on patient outcomes in knee surgeries and find that shortening procedures is associated with an increased risk of post-surgical revisions. The reason is that surgeons aim to reduce operative time by improving efficiency, potentially leading to compromises in surgical technique. Consequently, we propose the following opposite hypothesis.

Hypothesis 2B. *All else equal, a shorter procedure time leads to an increased post-surgical risk when endogeneity is addressed.*

4. Clinical Setting and Data

Our study is carried out at a prominent children's hospital in Canada. The hospital serves approximately 9,700 patients annually and employs 119 surgeons across ten different clinical departments. The hospital has 13 available operating rooms, with nine being utilized simultaneously. The primary working hours of the hospital are on weekdays, although some emergency cases may be performed during weekends.

4.1. Surgery Stages and Shift Scheduling

Before delving into the details of the data set, we first introduce the composition of the surgical and cleaning teams, the surgery process, and the scheduling of an OR shift. The surgical team consists of medical staff in different areas, with the mix depending on the procedure type. The team often includes a surgeon, an anesthesiologist, a physician assistant, and a few nurses. The surgeon leads the team while the other supporting professionals assist the surgeon before, during, and after the surgery. The cleaning team, comprising cleaning staff, nurses, and technicians, is responsible for the turnover stage between two consecutive cases. They clean the OR thoroughly after the previous patient exits and prepare the room for the next surgery.

Figure 1 depicts the breakdown of a typical surgical procedure and the patient pathway. Upon arrival on the surgery day (the epoch of entering the department in the diagram), nurses attend to the patient, assessing their condition for surgery by measuring vital signs and performing other preparations. The patient is then taken to the designated OR. The procedure duration refers to the total time a patient spends in the OR. Once the procedure is completed and the patient exits, the cleaning team starts working. This period (from the previous exit to the next patient's entrance) constitutes the turnover connecting two consecutive surgeries.

The shift's start time, marked by the commencement of the first surgery, is typically scheduled for either 7:45 am or 8:20 am. Occasionally, shifts may begin later due to additional training or administrative tasks. Prior to the shift's start, the surgical team receives a comprehensive briefing on the day's surgical schedule, even though the schedule is usually sent to the team several days in advance. The shift is usually scheduled to end between 3:15 pm - 3:45 pm. Patients with operations scheduled for that day will be notified a few days earlier and typically arrive about two to three hours before their scheduled surgery start times.

Scheduling surgeries in a hospital is a complex and lengthy process managed by a centralized booking office. There are nine ORs for regular shifts and two additional ORs for emergent add-on cases. Usually, each OR is assigned to a sole surgical team from one department throughout the day (shift). Each shift typically includes four surgeries. The booking office then contacts the patient and the surgeon to determine which surgeries will be performed on that date , ensuring the total shift length close to eight hours. The booking office then drafts an *OR slate* for each surgery

OR SLATE

From 12-Jan-2022 to 13-Jan-2022

Total # Cases: 3

Case #	Surgeons	Start Time	End Time	PIR (min)	Patient Name	MRN	Procedure	Bed
January 12, 2022								
1234567	Surgeon A	7:45	10:00	135	DOE, JOHN	15645611	P1	20-ADP-PICU
9876541	Surgeon A	10:30	12:30	120	SMITH, JACK	25654894	P1	10-SDC
9852642	Surgeon A	13:00	15:00	120	TURNER, MARY	34948480	P3	30-INPT

Figure 3 An Example of an OR Slate

day in an online system. A surgery slate is a spreadsheet containing essential information for each surgery case listed in each row. This information includes planned start and end times, planned procedure duration, and a brief procedure description. It also contains other identifications such as patient name, a medical record number (MRN in Figure 3), bed number, and the surgeon in charge. Figure 3 presents an example of an OR slate with three surgeries scheduled for January 12, 2022.

We learn from our surgeon collaborators that surgeries are scheduled jointly by the booking office and the surgeons. The booking office first forecasts the duration of each surgical case's procedure based on the average of all procedures of the same type in the previous calendar year. Surgeons then make adjustments to these forecasts, taking into account the individual patient's clinical conditions. The booking office also sets the estimated turnover duration to either 20, 25, or 30 minutes, depending on the procedure type. A schedule is then created by sequencing the surgeries and placing them back-to-back, ensuring the total duration, including procedure time and turnover duration, fits roughly into an 8-hour shift.

The completed day-shift schedule is distributed to the surgical and cleaning teams at least one day before the surgery day so that they are well-informed about each surgery's planned start and end time. The finalized schedules are then recorded in the system and appear in the dataset. Occasionally, some emergency cases are added to the shift on the surgery day without notifying the surgical team in advance. These emergent add-on cases are not recorded on the slate since they have no advanced schedule. However, they are still recorded in the surgical database. About 7.6% of the shifts contain add-on cases. We discuss how to handle shifts with add-on cases in Section 4.2.

4.2. Data

The surgery dataset we use is extracted from the hospital's system, documenting each case's progress in the OR and hospital from April 2018 to October 2019. The dataset includes both the planned and actual start and end times of each surgery. It also contains patients' demographic information, such as case type, patient type, age, and surgical procedures.

By comparing the actual and planned time schedules for each surgical case, we obtain the difference between the actual and planned procedure duration, which we refer to as the *relative procedure duration* (RPD). A positive (negative) RPD indicates that the surgery is performed slower (faster) than planned. A similar measure has been used in the literature to gauge a surgical team's speed (Pandit et al. 2009). We define the *deviation from scheduled start* (DSS) of surgery as the difference between the actual start time and the planned one. Thus, a positive (negative) DSS suggests a late (early) start of the focal case compared to the scheduled. We then obtain the positive and negative parts of DSS, namely DSS^+ and DSS^-, to capture the extent of delayed and early start, respectively. For the turnover stage, we define relative turnover duration (RTD) as the difference between the actual and scheduled turnover duration, following our previous definitions. A positive RTD indicates that the cleaning team is slower than the planned turnover duration and vice versa. We then define the *deviation from scheduled end* (DSE) to measure the difference between the actual and planned end time of the focal surgery. Note that as the turnover starts immediately after a surgery's end (see Figure 1), the end time of the focal surgery is the start time of the turnover stage. Similar to DSS, we also compute the positive and negative parts of DSE (DSE^+ and DSE^-) to measure the scale of delayed (early) end of a surgery.

We can also identify sequence shuffling and cancellations by comparing the scheduled sequence with the actual one. Sequence shuffling indicates that the actual sequence differs from the scheduled, which is usually caused by a patient's failure to arrive on time. Some surgeries are canceled mid-shift and thus will not have actual start and end times, but they can be identified from the slate.

We apply the following exclusion criteria before proceeding with the empirical analysis. First, we exclude all cases from the diagnostic department, as they are primarily medical examinations instead of surgeries and do not fit the context of this study. Second, following our surgeon collaborator's suggestions, we focus on shifts with at least three surgeries, as they contain enough variation in the shift composition. The remaining departments are listed in Table 1. Third, apart from elective surgeries, emergent add-on cases may occur during the shift. Sometimes, add-on cases will be performed in different ORs, disrupting the original schedules. On urgent occasions, the surgeon may leave the focal surgery (with the surgical team taking over) and attend to the inserted emergent add-on case. To eliminate the interference of add-on cases, we exclude all emergent add-on cases and the subsequent impacted cases, accounting for 3.6% of the total cases. Similarly, to eliminate the interference of cancellations, we exclude surgeries following cancellations in shifts (4.7% of all). To test the robustness of our findings, we conducted additional analyses in the Online Appendix Section EC.2.2, excluding entire shifts containing at least one add-on case and entire shifts with at least one cancellation. The results obtained from these robustness tests are consistent with our estimates. Finally, to avoid potential extra training or interruptions of shifts, we exclude shifts that do not start during the normal period (i.e., before 9 am) and surgeries occurring outside typical shift times (before 7 am or after 7 pm), which account for 4.8%.

For the turnover stage, it is worth noting that when the next patient is not present or not ready yet, the subsequent surgery must be postponed, leading to an artificially longer recorded turnover duration (as turnover is the period from the prior end to the next start). To reduce this possibility, we only include turnover samples in which the second surgery's patient arrives at least 30 minutes earlier than the recorded start time of the surgery (ensuring that the patient should be ready for surgery as soon as the turnover finishes). Additionally, the turnover duration following the final surgery in each shift is not included in our study.

The final sample consists of 7,868 surgery cases across various departments from April 2018 to October 2019. For the turnover stage, we obtain 6,210 final observations. The sample includes 1,928 shifts, with the number of surgeries ranging from three to eight for each shift. There are 111 different procedure codes across all departments. The sample features 69 surgeons with an average of 18 years of experience, and 75% of them have more than nine years of experience. Overall, 65.3% of the surgeries in the sample experience delayed starts compared to schedules. Table 1 summarizes the variables. Most surgeries take approximately 45 to 99 minutes, as shown by the 25% and 75% quartiles and the average procedure duration is 78.53 minutes. As measures of the relative speed of surgical teams, the average RPD and RTD in the samples are -1.048 minutes and 1.432 minutes, respectively, which are much smaller than the average realized procedure and

Table 1 Summary Statistics

Panel A: Continuous Variables						
Variables	Mean	Median	Std	IQR	25% Q.	75% Q.
Surgery part:						
Real procedure duration*	78.53	67	50.47	54	45	99
Relative procedure duration (RPD)*	-1.048	-3	25.62	24	-14	10
Deviation from scheduled start (DSS)*	7.746	7	32.02	29	-9	20
Delayed start (DSS$^+$)*	15.55	7	23.03	20	0	20
Early start (DSS$^-$)*	7.803	0	15.88	9	0	9
Age	7.561	6.571	5.197	8.798	3.014	11.81
Elective surgery indicator	0.999		0.037			
Sequence shuffling indicator	0.037		0.188			
N	7,868					
Turnover part:						
Real turnover duration*	27.33	25	11.49	10	21	31
Relative turnover duration (RTD)*	1.432	-1	11.49	10	-5	5
Deviation from scheduled exit (DSE)*	7.779	6	38.26	42	-15	27
Delayed exit (DSE$^+$)*	18.29	6	26.87	27	0	27
Early exit (DSE$^-$)*	10.51	0	18.89	15	0	15
Different procedure in next case indicator	0.534		0.499			
N	6,210					
Post-surgical risk part:						
30-day readmission	0.034		0.182			
30-day reoperation	0.029		0.168			

Panel B: Categorical Variables										
Variables	Categories									
Patient type	ADP	SDC	IN	OUT						
	13.75%	84.06%	1.932%	0.254%						
Patient severity	I	II a	II b	III	IV	V	VI			
	0.267%	2.758%	7.855%	19.38%	44.65%	20.45%	4.639%			
Department	CVS	DDS	ENT	GAST	GENL	NEUR	OPTH	ORTH	PLAS	UROL
	0.203%	20.69%	10.17%	13.17%	11.27%	0.038%	12.44%	10.55%	6.215%	15.25%

Notes: Variables with $\star$ are measured in minutes. ADP: admitted patients, SDC: surgical daycare, IN: inpatient, OUT: outpatient. For service departments, we have CVS for cardiovascular, NEU for neurology, DDS for dental, ENT for otolaryngology, GAST for gastrointestinal, GENL for general, OPTH for ophthalmology, ORTH for orthopedics, PLAS for plastic and UROL for urology.

turnover durations. This demonstrates that the scheduled procedure and turnover durations are almost unbiased estimators of the actual durations. Sequence shuffling occurs in 3.7% of total cases. We control for it by creating a dummy variable to label those cases. Since the turnover stage connects two surgeries, we also include a dummy indicating whether the two surgeries are of the same procedure type. Different procedure combinations account for 53.4% of the turnover sample.

For patient demographics, the average age of patients is 7.6 years in the sample. Elective cases constitute most of the sample, with only 0.1% of the sample representing emergent cases. As for post-surgical risk measures, the average probability for 30-day readmission is 0.034, and 0.029 for reoperation.

In Panel B of Table 1, we report the distribution of departments and patient types. Most cases are surgical daycare (SDC) patients who rarely need to stay in the hospital overnight. Admitted cases (ADP) account for 13.27% and require post-surgical observation. From the patient severity distribution, we observe that most patients belong to the moderate categories: level III, IV, and V, while severe cases (I, II a, and II b) account for about 10% of all cases. Regarding department distribution, dental, urology, gastrointestinal, ophthalmology, general, and otolaryngology are the major departments, each constituting more than 10% of all cases.

5. Empirical Methods for Clinical Teams' Responsive Behaviour

The surgery data set consists of multiple shifts, each consisting of surgeries performed by the same surgical team in the shift. We use $i = 1, \ldots, N$ to index the shifts, with the total number of shifts being $N = 1,928$. In each shift, the sequence of surgical cases is indexed by $t = 1, \ldots, T_i$ by their actual start time, with T_i denoting the number of surgeries in shift i. Different shifts may have different numbers of surgeries. As a result of our exclusion criteria, $T_i \geq 3$, while the maximum value of T_i is 8 in our sample. Since T_i varies across shifts, we obtain an unbalanced panel. By slightly abusing notation, we use RPD_{i0} to denote the difference between the actual and planned start time of the first case in shift i. Let S_{it} denote the deviation from the scheduled finish time of case t in shift i including turnover stage, or equivalently, the deviation from the scheduled start time (DSS) of case $t+1$, which has the following expression

$$S_{it} = \mathrm{RPD}_{i0} + \sum_{k=1}^{t} (\mathrm{RPD}_{ik} + \mathrm{RTD}_{ik}) \;\; t = 1, \ldots, T_i. \tag{1}$$

As shown in the above expression, S_{it} is the cumulative deviation from the schedule, beginning with the deviation (if any) of the start time of the first surgery y_{i0} and including all relative procedure and turnover durations through the first t cases.

To estimate the impact of a positive DSS (delayed start) or a negative DSS (earlier start) on the RPD, we formulate the following specification:

$$\mathrm{RPD}_{it} = \gamma_n \cdot (S_{i,t-1})^- + \gamma_p \cdot (S_{i,t-1})^+ + \mathbf{x}_{it}^\top \beta + \delta_t + \alpha_i + u_{it}, \;\; t = 1, \ldots, T_i \tag{2}$$

where the negative and positive parts of $S_{i,t-1}$ are included. The coefficients γ_n and γ_p represent how surgical teams respond when getting ahead and falling behind, respectively. The unobserved fixed (case-independent) effect for shift i is represented by α_i, and δ_t denotes the unobserved case-specific effect that varies with t. β stands for the coefficient vector of the regressors $\mathbf{x}_{it}$, which includes patient demographics and case type information. u_{it} denotes the model's error term.

Since S_{it} is a linear function of $y_{i0}, y_{i1}, \ldots, y_{i,t-1}$ as given in Equation (1), multiple lagged response variables $y_{ik}(k=1,\ldots,t-1)$ appear on the right-hand-side (RHS) of the regression equation for y_{it}. If we use the OLS method to estimate the coefficients, the errors in different regression equations will be correlated, which violates the basic assumption of OLS (Nickell 1981). Thus, we reformulate the problem as an auto-regressive specification using a linear transformation. To that end, Equation (1) renders the following relation between y_{it} and S_{it},

$$\mathrm{RPD}_{it} = S_{it} - S_{i,t-1} - \mathrm{RTD}_{it},\ t = 1, 2, \ldots, T_i \tag{3}$$

Plugging the above expression of y_{it} into the left-hand-side of Equation (2) with a simple transformation, we have

$$S_{it} - \mathrm{RTD}_{it} = (\gamma_n - 1) \cdot (S_{i,t-1})^- + (\gamma_p + 1) \cdot (S_{i,t-1})^+ + \mathbf{x}_{it}^\top \beta + \delta_t + \alpha_i + u_{it} \tag{4}$$

We thus obtain an auto-regressive model where the coefficients of interest, γ_n, and γ_p, are absorbed into the coefficient of the positive and negative part of the auto-regressive term $S_{i,t-1}$. Such an auto-regressive formulation is often employed in dynamic panels, in which the regressors include one or more lagged dependent variables, allowing for the modeling of dynamic behavior across time series (Pesaran 2015).

For the relative turnover duration (RTD), with a slight abuse of notation, we have the following similar specification:

$$\mathrm{RTD}_{it} = \nu_n \cdot (\mathrm{DSE}_{i,t-1})^- + \nu_p \cdot (\mathrm{DSE}_{i,t-1})^+ + \mathbf{x}_{it}^\top \beta_1 + \mathbf{n}_{it}^\top \beta_2 + \tilde{\delta}_t + \tilde{\alpha}_i + \tilde{u}_{it},\ t = 1, \ldots, T_i \tag{5}$$

where DSE stands for deviation from the scheduled exit time, which is exactly the scheduled start time of the turnover stage. Apart from the same regressors $\mathbf{x}_{it}$ as in Equation (2), we also include $\mathbf{n}_{it}$ for RTD that comprises 1) indicator of whether case t and case $t+1$ belong to the same type and 2) the patient type of case $i+1$. This is because the turnover stage consists of both cleaning and preparation steps for two consecutive operations. Like DSS, a positive DSE means that the focal surgery finishes late (thus the subsequent turnover stage starts late) and vice versa. The coefficients of interest are ν_n and ν_p for the negative and positive parts, respectively. The relationship between DSE and RTD is $\mathrm{DSE}_{i,t} = \mathrm{DSE}_{i,t-1} + \mathrm{RTD}_{i,t} + \mathrm{RPD}_{i,t+1}$. After similar transformation processes, we obtain the following auto-regressive model for RTD:

$$\mathrm{DSE}_{i,t} - \mathrm{RPD}_{i,t+1} = (\nu_n - 1) \cdot (\mathrm{DSE}_{i,t-1})^- + (\nu_p + 1) \cdot (\mathrm{DSE}_{i,t-1})^+ + \mathbf{x}_{it}^\top \beta_1 + \mathbf{n}_{it}^\top \beta_2 + \tilde{\delta}_t + \tilde{\alpha}_i + \tilde{u}_{it}, t = 1, \ldots, T_i \tag{6}$$

To obtain unbiased estimates of γ_n, γ_p in Equation (4) and ν_n, ν_p in Equation (6), we employ the Arellano-Bond (A-B) generalized method of moments (GMM) for dynamic panel data (Arellano and Bond 1991). Essentially, the A-B method utilizes higher-order lagged regressors as instrumental variables (IVs) that induce a series of moment conditions for GMM estimation. It is important to note that the directly obtained estimates are $\hat{\gamma}_n - 1$ and $\hat{\gamma}_p + 1$ for RPD ($\hat{\nu}_n - 1$ and $\hat{\nu}_p + 1$ for RTD), while we report $\hat{\gamma}_n, \hat{\gamma}_p, \hat{\nu}_n, \hat{\nu}_p$ after adding (subtracting) one. The details of the estimation process can be found in the Online Appendix Section EC.1.

6. Estimation Results and Interpretations

6.1. Main Results

Table 2 displays the GMM estimation results. Note again that the directly obtained estimates are $\hat{\gamma}_n - 1$ and $\hat{\gamma}_p + 1$ for RPD ($\hat{\nu}_n - 1$ and $\hat{\nu}_p + 1$ for RTD) and we report $\hat{\gamma}_n, \hat{\gamma}_p, \hat{\nu}_n, \hat{\nu}_p$ after adding (subtracting) one. We control for other regressors that vary with case sequence t in the same shift, such as patient type, patient severity, procedure type, and the sequence fixed effect t itself, which will not be eliminated after first-difference (FD) or forward-orthogonal deviation (FOD) transformations in the GMM estimation process (see Section EC.1).

From Table 2 columns (1) and (2), we see that in the surgical stage, both FD and FOD GMM estimates are significant for both early and delayed start where $\widehat{\gamma}_n$ by FOD-GMM is 0.444 (0.617 by FD-GMM) and $\hat{\gamma}_p$ equals -0.164 (-0.219 by FD-GMM). To gain a general understanding of these estimates, we consider a scenario where a surgical team faces an early start by one standard deviation (SD) of the focal case (15.88 minutes). From FOD-GMM estimates, we know that the surgical team would *slow down* the focal surgery by $0.444 \times 15.88 = 7.05$ minutes (10.5% of the median procedure duration) compared to the scheduled duration, ceteris paribus. Conversely, for one SD delayed start of a case (23.03 minutes), the surgical team would *expedite* the focal surgery by $0.164 \times 23.03 = 3.78$ minutes (5.6% of the median procedure duration).

We observe the behavior of cleaning teams during the turnover stage from columns (3) and (4). We find that both FD and FOD yield significant estimates for both early and delayed start, where $\hat{\nu}_n$ by FOD is 0.292 (0.190 by FD) and $\hat{\nu}_p$ by FOD is -0.096 (-0.121 by FD). If there is a one SD early exit of the preceding surgery (18.89 minutes), then the following turnover will be extended by 5.52 minutes (22.1% of the median turnover duration) than planned. On the other hand, if the exit is delayed by one SD (26.87 minutes), the turnover will be shortened by 2.58 minutes (10.3%). These results illustrate how cleaning teams adaptively manage the turnover stage to minimize the deviation from schedules in both directions. We also find that the scale of response in turnover is greater compared to that of surgical teams in surgeries.

Table 2 GMM Estimation Results

	Surgery		Turnover	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Early Start	0.617***	0.444***	0.190***	0.292***
	(0.159)	(0.126)	(0.051)	(0.047)
Delayed Start	-0.219*	-0.164*	-0.121**	-0.096*
	(0.106)	(0.081)	(0.045)	(0.045)
Elective case	13.78	-25.54	4.505	-20.74
	(60.37)	(58.13)	(2.843)	(19.96)
Sequence shuffling	-1.557	5.274	-7.235***	-9.074***
	(5.549)	(5.274)	(1.864)	(2.284)
Age	0.280	0.277*	0.361*	0.025
	(0.148)	(0.138)	(0.163)	(0.172)
Different procedure			0.976	1.748**
			(0.588)	(0.543)
Next case type			Included	Included
Patient type	Included	Included	Included	Included
Patient severity	Included	Included	Included	Included
Procedure type	Included	Included	Included	Included
Time fixed effect	Included	Included	Included	Included
Number of instruments	253	253	258	253
Number of observations	7,868	7,868	6,210	6,210

Notes: Both GMM models are estimated using the Arellano-Bond method. In column (1) and (2), we report $\hat{\gamma}_n$ and $\hat{\gamma}_p$ for RPD. In column (3) and (4), we report $\hat{\nu}_n$ and $\hat{\nu}_p$ for RTD. Time fixed effect t stands for the sequence of cases in shift i while the shift fixed effect is zeroed out in the FD (FOD) processes. Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

In summary, we find that surgical and cleaning teams exhibit both speedup and slowdown behavior, supporting Hypothesis 1A and 1B, which reflects the presence of balancing feedback behavior for both teams. We illustrate the estimates visually in Figure 4, where the slopes of the four lines represent the coefficients in Table 2. Although the coefficients for the surgery stage are larger than those for the turnover stage, the responses in terms of percentage scale are more pronounced in the turnover stage. For both surgical and cleaning teams, the magnitude of slowing down behavior is greater than speedup. Based on our discussions with surgeon collaborators, we understand that it is more feasible for surgical teams to slow down rather than speed up, as the latter poses potential safety concerns. Additionally, surgical teams have incentives to slow down because when they are ahead of schedule, the OR administrator is more likely to assign extra add-on cases to their shift with little to no extra financial compensation due to the fixed salary scheme. During the turnover stage, cleaning teams face fewer concerns regarding clinical quality and thus have more flexibility to accommodate responses in both directions.

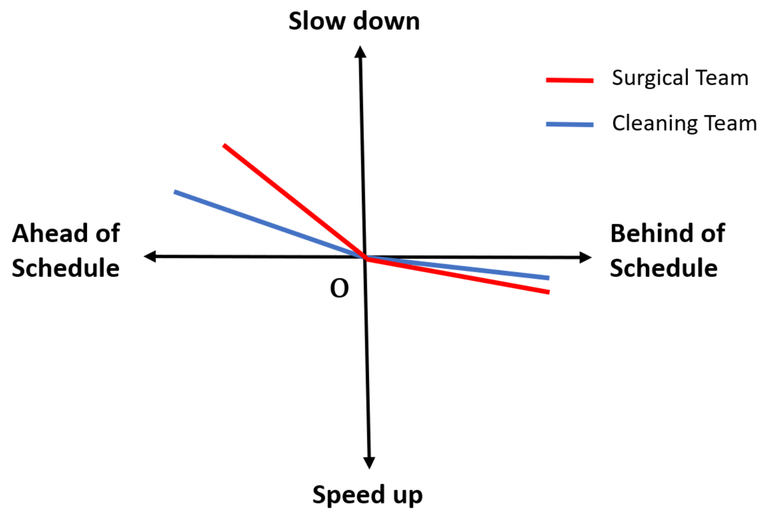

Figure 4 **Surgical and Cleaning Teams' Responsive Behavior**

In Table EC.1, we also report the FE-OLS estimation results of Equation (2) and Equation (5). We find that the estimates differ in scale from the GMM estimates due to Nickel bias (Nickell 1981), as mentioned in Section 5, although the signs of the estimates are consistent.

6.2. Heterogeneity Across Surgeon Seniority

In this section, we explore the responsive behavior across different surgeon seniority groups and we focus on the RPD. As the leader of the surgical team during an operation, a surgeon's experience is crucial for guiding other team members. There is extensive literature on the relationship between experience and performance, particularly in healthcare settings (Kc and Staats 2012, Staats et al. 2018, Ibanez et al. 2018). Senior surgeons, with more experience, are better acquainted with procedures and are therefore aware of both the challenging and more manageable aspects of a procedure.

To investigate how surgical teams' behavior varies by surgeon seniority, we perform a subsample study for junior and senior surgeon groups. Based on discussions with our surgeon collaborators, we choose 20 years after obtaining the Doctor of Medicine (M.D.) degree as the threshold for surgeon seniority. In our setting, the junior (< 20) and senior (≥ 20) groups comprise 3,793 and 4,075 surgical cases, respectively.

Table 3 reveals that the senior surgeons' group yields significant estimates for both early and delayed starts. To understand the scale of such adaptive behavior, we again consider one SD delayed (early) start as an example. On average, senior surgeons speed up by $0.319 \times 23.03 = 7.35$ minutes (10.97% of median procedure duration) or slow down by $0.505 \times 15.88 = 8.02$ minutes (11.97% of median procedure duration) in the focal case. Surgical teams led by senior surgeons tend to expedite when facing delays and slow down when ahead of schedule at a similar pace, that is, the balancing

Table 3 Heterogeneity of Responses in Surgery Stage between Surgeon Seniority Groups

	Senior		Junior	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Early Start	0.679***	0.505***	0.368	0.404
	(0.150)	(0.141)	(0.265)	(0.210)
Delayed Start	-0.359*	-0.319**	-0.313*	-0.234
	(0.155)	(0.113)	(0.154)	(0.136)
Elective case	59.10*	42.09	15.79	-13.77
	(23.89)	(37.86)	(65.03)	(48.12)
Sequence shuffling	-5.889	-4.485	0.166	-3.363
	(7.337)	(5.698)	(10.28)	(8.709)
Age	-0.255	-0.129	0.324	0.410
	(0.273)	(0.189)	(0.247)	(0.224)
Patient type	Included	Included	Included	Included
Patient severity	Included	Included	Included	Included
Procedure type	Included	Included	Included	Included
Time fixed effect	Included	Included	Included	Included
Number of instruments	244	244	227	227
Number of observations	3,793	3,793	4,075	4,075

Notes: Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

feedback. In contrast, junior surgeons do not consistently exhibit speedup or slowdown behavior. As our surgeon collaborators explain, one reason for such discrepancies between their responsive patterns could be that, compared to junior surgeons, senior surgeons may have trainees in the OR. Senior surgeons would be faster and more likely to allow trainees to participate in procedures for better practice. When running late, the senior surgeon may take over from the trainee to steer the schedule back on track. Junior surgeons, on the other hand, might perform the work themselves without substantially involving the trainees.

7. Exploration on Surgical Quality

7.1. Empirical Methods on Quality

In this section, we investigate the impact of procedure duration speedup/slowdown on patients' post-surgical risks (refer to Hypotheses 2A and 2B in Section 3). This analysis allows us to examine the second-order effects of surgical teams' responses to delayed and early starts. We consider two indicators: post-surgical 30-day readmissions and 30-day reoperations. These indicators are widely used to measure patient risk and clinical quality (Morris et al. 2007, Maali et al. 2018).

To ensure the accuracy of our analysis, we exclude all dental (DDS) and gastrointestinal (GAST) surgeries based on our surgeon collaborator's recommendations. These two service types typically do not involve readmissions or reoperations due to the nature of the procedures (0.61% and 1.17%). After removing these two types of surgeries, we are left with 5,204 observations.

We first consider a probit model to analyze the relationship between the readmission/reoperation binary variable and the RPD of a case:

$$r_{it}^* = \rho \cdot \mathrm{RPD}_{it} + \zeta \cdot \mathrm{RTD}_{i,t-1} + \mathbf{s}_{it}^\top \beta_1 + \mathbf{F}_i^\top \beta_2 + \epsilon_{it} \tag{7}$$

$$P\left(r_{it} = 1\right) = P\left(r_{it}^* > 0\right) = \Phi\left(\rho \cdot \mathrm{RPD}_{it} + \zeta \cdot \mathrm{RTD}_{i,t-1} + \mathbf{s}_{it}^\top \beta_1 + \mathbf{F}_i^\top \beta_2\right) \tag{8}$$

$$P\left(r_{it} = 0\right) = P\left(r_{it}^* \leq 0\right) = 1 - \Phi\left(\rho \cdot \mathrm{RPD}_{it} + \zeta \cdot \mathrm{RTD}_{i,t-1} + \mathbf{s}_{it}^\top \beta_1 + \mathbf{F}_i^\top \beta_2\right) \tag{9}$$

where r_{it} denotes a binary indicator for the post-surgical risk event for case t in shift i, which could be either 30-day readmission or reoperation; r_{it}^* stands for the associated latent variable; and ρ is the coefficient of interest, (i.e., how the RPD of the surgery influences the probability of a patient's post-surgical readmission). We also control for the RTD of the preceding surgery (i.e., $\mathrm{RTD}_{i,t-1}$) in the model to explore whether the change in the turnover duration before the focal case starts influences post-surgical risks and ζ stands for the associated coefficient. We include other explanatory variables $\mathbf{s}_{it}$ and fixed effects $\mathbf{F}_i$ such as day-of-week and surgeon. β_1 and β_2 denote the coefficients for $\mathbf{s}_{it}$ and $\mathbf{F}_i$ respectively. ϵ_{it} is the error term that follows a standard normal distribution, and $\Phi(\cdot)$ denotes the cumulative standard normal distribution function.

The above probit model, however, may yield biased estimates for ρ. Various factors influence a patient's post-surgical readmission/reoperation likelihood. Although we have controlled for patient demographics, surgeon fixed effects, and time fixed effects, other unobservables could still be included in the error term ϵ_{it} and correlate with the critical explanatory variable, RPD. For example, the surgical team may encounter unexpected complications during surgery, which may take longer to complete and result in a larger RPD. Meanwhile, unexpected complications during the surgery are associated with a higher risk of developing post-surgical complications. Thus, even when RPD and post-surgical complication rates are positively correlated, it does not imply causality since unobservables, such as complications during the procedure, can trigger both. In other words, a direct estimate of ρ may have overlooked the existence of confounders and lead to an upward bias.

We search for an appropriate IV to address the endogeneity in Equation (7) to obtain unbiased estimates. Based on the previous section's analyses of surgical teams' adaptive behavior, we utilize DSS as our IV here. First, DSS satisfies the relevance condition, as it leads to a change in the RPD of the focal case. Second, regarding the exogeneity condition, DSS reflects the cumulative deviations from the schedule based on earlier cases in the same shift (see Equation (1) in Section 5) and should be uncorrelated with the post-surgical complication rate of the current surgery. We have performed robustness tests to rule out other potential aspects by which the focal surgery might be influenced by DSS, such as fatigue (see Appendix EC.2). Therefore, the exogeneity condition is also satisfied, lending support that our IVs are valid for the estimation of ρ. Here, we adopt the *IV-probit* model for identification. The structural equations are as follows:

$$\begin{aligned} \mathrm{RPD}_{it} &= \gamma \cdot \mathrm{DSS}_{it} + \mathbf{x}_{it}^{\top}\omega_1 + \mathbf{F}_i^{\top}\omega_2 + \xi_{it} \\ r_{it} &= 1\left\{\rho \cdot \mathrm{RPD}_{it} + \zeta \cdot \mathrm{RTD}_{i,t-1} + \mathbf{s}_{it}^{\top}\beta_1 + \mathbf{F}_i^{\top}\beta_2 + \epsilon_{it} > 0\right\}, \end{aligned} \tag{10}$$

where the first equation stands for the relation between endogenous RPD and instruments. $\mathbf{x}_{it}$ represents exogenous variables. ω_1 and ω_2 represent the associated coefficients of the control variables. ξ_{it} denotes the error term where ξ_{it} and ϵ_{it} follow multivariate normal distribution. We utilize the conditional maximum likelihood (CMLE) method to obtain the estimates from the joint equations mentioned above. The CMLE estimator is unbiased and offers two advantages compared to the traditional two-step control function (CF) approach: (1) it is more efficient, and (2) it simplifies the computation of partial effects (Wooldridge 2010). The estimated correlation between the two error terms in Equation (10) is used to examine the existence of endogeneity.

We employ 30-day readmission/reoperation as proxy measures for surgical quality. We do not have detailed data on the readmissions, so we cannot distinguish those related to the surgery, which would otherwise provide greater accuracy. Despite the limitation of using readmissions and reoperations as proxies for service quality, our analysis offers an IV-based approach to identify the potential impact of service speed on service quality, which is challenging due to various potential confounding factors (Johnston 1995).

7.2. Estimation Results

In Table 4, we present the estimation results from IV-probit and plain probit models based on 30-day readmissions and reoperations, respectively. From columns (3) and (4), we find that for 30-day readmission and reoperation, the RPD of focal surgery is positively significant as $\hat{\rho}_{rdms30}^{nonIV} = 0.0043$ and $\hat{\rho}_{reop30}^{nonIV} = 0.0052$. The RTD of the preceding surgery (connects the previous and focal surgery) is not significant under the plain probit model.

The non-IV probit model seems to imply that speedup in surgeries (negative RPD) decreases the patient's 30-day readmission and reoperation probability. This suggests that longer procedure duration is correlated with increased 30-day readmission and reoperation rates, which is consistent with the correlative findings in the literature (Cheng et al. 2018). As we mentioned earlier, however, the observed positive correlation does not necessarily imply causality. Instead, it is likely driven by confounders that are positively correlated with both the RPD and the readmission/reoperation

Table 4 Impact of Relative Procedure and Turnover Duration on Postsurgical Risks

	Panel A: IV-Probit		Panel B: NonIV-Probit	
Variables	(1) 30-Read	(2) 30-Reop	(3) 30-Read	(4) 30-Reop
RPD of the focal surgery	-0.0316***	-0.0332***	0.0043**	0.0052**
	(0.0066)	(0.0051)	(0.0014)	(0.0018)
RTD of the preceding surgery	0.0007	0.0016	0.0020	0.0050
	(0.0025)	(0.0022)	(0.0053)	(0.0050)
Sequence shuffling	Included	Included	Included	Included
Age	Included	Included	Included	Included
Patient type	Included	Included	Included	Included
Patient severity	Included	Included	Included	Included
Procedure type	Included	Included	Included	Included
Time fixed effect	Included	Included	Included	Included
Day-of-week fixed effect	Included	Included	Included	Included
Surgeon fixed effect	Included	Included	Included	Included
Correlation of errors	0.8842***	0.9292***		
N	5,204	5,204	5,204	5,204

Notes: Estimated by the maximum likelihood method. Dependent variables: 30-day readmission, 30-day reoperation indicators. We report the correlation of the errors of the two equations in the IV-probit models in the final part. Robust standard errors are shown in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

probability. This speculation is further supported by the second-to-last row in the table, where all IV-probit models yield positively significant-from-zero correlations of the two errors (ξ_{it} and ϵ_{it} as shown in Equation (10)).

The IV-probit model yields estimates $\hat{\rho}^{IV}_{rdms30} = -0.0316$ for readmissions and $\hat{\rho}^{IV}_{reop30} = -0.0332$ for reoperations. All of them are negatively significant. Comparing estimates from the IV-Probit and plain probit models, it is clear that the estimates become negative when the upward bias is eliminated. Thus, speedup (slowdown) in the surgery procedure leads to an increase (decrease) in both 30-day readmission and reoperation probabilities, supporting Hypothesis 2B and rejecting Hypothesis 2A. However, under the IV-probit model, the RTD of the preceding surgery remains insignificant for either risk measure.

The coefficients in Table 4 do not directly quantify the impact on readmissions and reoperations due to the nonlinear probit function. To better understand the scale of the RPD's impact on post-surgical risks, we compute how marginal changes in the RPD lead to different readmission and reoperation probabilities. We perturb the RPD in each observation of the sample by plus or minus one minute and compute the associated readmission and reoperation probabilities based on the fitted structural model eq. (10), holding other variables constant. We then take the average of all samples' fitted values. We find that expediting the relative speed of surgeries by one minute increases the average 30-day readmission probability by 3.973 $\times 10^{-3}$ (and increases the average

30-day reoperation probability by 3.994 $\times 10^{-3}$), while prolonging the RPD of surgeries by one minute decreases the average readmission probability by 3.629 $\times 10^{-3}$ (and decreases the average reoperation probability by 3.587 $\times 10^{-3}$).

8. Counterfactual Study

The results in the previous section reveal the trade-off between surgical speed and quality in operating rooms (ORs). We find that faster-than-scheduled operations lead to increased post-surgical readmission/reoperation risks and vice versa. This section presents a counterfactual study to quantitatively measure the trade-offs between efficiency and quality, providing managerial insights to surgeons and hospital managers. Specifically, the counterfactual study aims to demonstrate the relationship between the disparity between the actual and planned end-of-shift times (EOS) and post-surgical readmissions and reoperations. Surgical teams are likely concerned about minimizing the difference between the actual and scheduled EOS, which refers to the finishing time of the last surgery in a shift. If the shift ends later, it means that the surgical team has to work overtime, whereas if the shift ends significantly earlier than scheduled, another add-on surgery may be assigned to the team.

Using the actual timestamps of surgeries, we construct six counterfactual scenarios where each operation is prolonged or shortened by 5, 10, and 15 minutes, respectively. We then obtain the average fitted probabilities of 30-day readmissions and 30-day reoperations based on the updated RPD values through the structural model Equation (10). In addition to the readmission and reoperation rates as quality metrics, we compute the counterfactual EOS time based on the counterfactual RPD of cases in the same shift. We then obtain the average difference between each shift's counterfactual EOS and planned EOS.

The analysis is based on the quality exploration sample we adopt, where dental and gastrointestinal surgeries are not included since these departments have scarce readmissions and reoperations. To analyze the effects on the EOS, we select 1,584 unabridged shifts to conduct counterfactual analysis (we exclude those cases after an urgent add-on case or a cancellation, as explained in Section 4). The final case sample size is 4,429.

Figure 5 and Table 5 summarize the relationship between surgical quality and the differences between the actual and planned EOS under six scenarios. The x-axis represents the change in EOS compared to the current scenario, which serves as a baseline, depicted using a dashed vertical line. The y-axis represents quality measures, the average fitted 30-day readmission/reoperation probabilities. Overall, their relationship presents a convex shape. In the sample, the average difference to the planned EOS is approximately 10 minutes (i.e., on average, a shift ends later than planned),

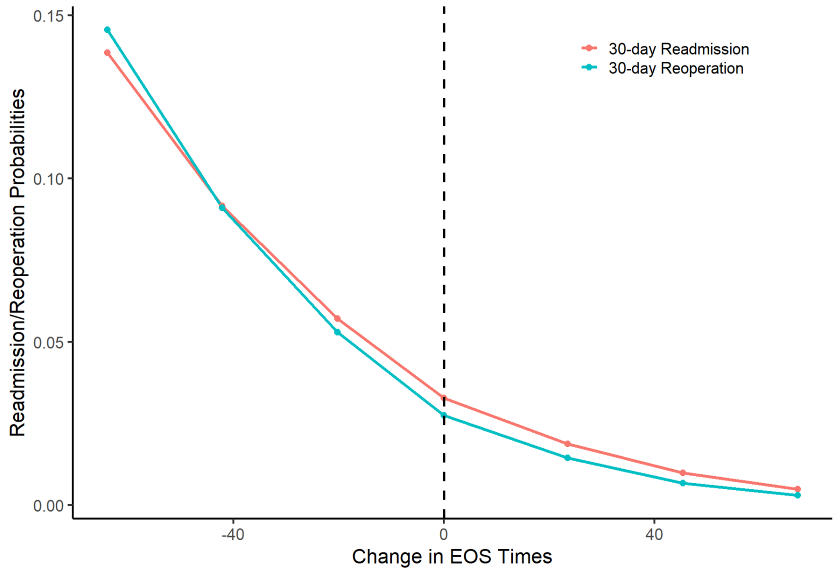

Figure 5 **Surgical Quality and End-of-Shift Time**

with a 0.0327 average 30-day readmission probability and 0.0275 average reoperation probability. If we extend each operation by 5 minutes, we obtain reduced average readmission and reoperation probabilities equal to 0.0188 and 0.0144, and the difference to the planned EOS increases to 33.51 minutes. On the other hand, when each surgery is expedited by 5 minutes, we obtain a reduced difference to the planned EOS (-10.2 minutes) and increased 30-day readmission/reoperation probabilities (0.0571 and 0.0530). Overall, this relationship provides guidance for hospital managers to achieve desired quality goals when creating OR schedules.

Table 5 **Results of the Counterfactual Analysis**

Scenario	Readmission	Reoperation	Diff EOS
-15 min	0.1385	0.1455	-53.92
-10 min	0.0915	0.0911	-32.06
-5 min	0.0571	0.0530	-10.20
Original	0.0327	0.0275	10.03
+5 min	0.0188	0.0144	33.51
+10 min	0.0099	0.0068	55.37
+15 min	0.0049	0.0030	77.23
Observations	4,429	4,429	1,584

Notes: Readmission and reoperation columns present the fitted average rates based on changing RPD (plus or minus 5, 10, 15 minutes) for each case using the estimated Equation (10). The Diff EOS column present the averaged difference to planned EOS time when RPD is changed.

9. Conclusions and Future Research

Operating rooms exemplify service sites where deviations from prescribed schedules frequently occur, often leading to physical and financial repercussions. Although abundant literature exists on this topic, two questions remain open. First, while many researchers have studied surgical team responses to schedule deviations, they primarily rely on surveys to explore response mechanisms instead of using real-time and scheduled surgery data. Second, concerning the relationship between surgical speed and quality, many researchers conduct correlative analyses, which may be biased due to various confounders in surgical procedures. In this paper, we address both questions by investigating surgical teams' immediate responses to early and late starts of surgeries, cleaning teams' responses in the turnover stage, and exploring how such responses further influence post-surgical complication risks. We apply econometric identification techniques to a unique and comprehensive surgical dataset to achieve this goal.

To answer the first question, we employ a series of empirical strategies. The primary challenge is that the ordinary least squares (OLS) estimator is biased since the general deviation from the scheduled start (DSS) is correlated with the relative procedure duration (RPD) of prior cases. We propose an auto-regressive model to address the serial correlation of procedure durations within the same shift. We then utilize the dynamic panel model and the associated GMM estimators to identify the parameters. We find that surgical and cleaning teams speed up when they fall behind schedule and slow down when they get ahead of schedule, with the slowdown exhibiting a stronger effect. That is, they exhibit the balancing feedback behavior. Quantitatively, surgical teams expedite the next surgery by an average of 5.6% when facing a one standard deviation (SD) delay in the planned start for that surgery, whereas they take on average 10.5% longer when they are one SD ahead of schedule. In the turnover times, cleaning teams accelerate by 10.3% (slow down by 22.1%) on average when they are one SD ahead of (behind) schedule. This insight is valuable for OR managers, as understanding the responsive behavior of clinical teams can further improve OR scheduling practices and potentially enhance patient satisfaction. Our findings are particularly relevant to ORs in a socialized healthcare system. The identified phenomenon likely results from the lack of incentives in the socialized healthcare system, which may motivate the redesign of incentives for surgical teams within such a system. Additionally, we explore the heterogeneity of behavior among different surgeon seniority groups for the surgery stage, finding that the speedup/slowdown phenomenon occurs only in the senior group.

On the second question, we explore how the speedup/slowdown of an operation influences a patient's post-surgical risks, measured by 30-day post-surgical readmissions and reoperations. Estimating the impact of service speed on service quality is challenging due to endogeneity issues. To address this, we propose an instrumental variable (IV) based on our observation that the early/late start (DSS) of a case may impact surgical quality only through the procedure duration. After correcting the upward bias, we find that speeding up procedures increases 30-day readmission and reoperation probabilities, contrary to the correlative findings of existing literature. Through a counterfactual analysis, we reveal a convex relationship between the readmission/reoperation rate and the difference between real and planned end-of-shift (EOS) times. This curve relates to the fundamental topic of efficiency-quality trade-offs in service operations. Surgical teams can adjust their paces to reduce the difference in scheduled EOS time; however, this may lead to increases in 30-day post-surgical readmission and reoperation risks, as captured in Figure 5. This finding has important implications for healthcare providers, particularly surgeons and hospital managers. Methodologically, we propose an IV to address endogeneity issues in estimation, which is shown to work effectively. This method can be extended to analyses in other service sectors, such as appointment scheduling and project management.

We conclude our paper by suggesting some directions for future research. Our paper focuses on how clinical teams respond to schedule deviations and the subsequent impacts on post-surgical risks. Future studies can consider prescriptive data-driven models that incorporate such behavior to optimize daily surgical schedules. We also note that Canada's socialized healthcare system differs significantly from the diversified health insurance systems in the U.S. Therefore, another interesting angle relates to the incentive scheme design of healthcare systems and how surgical teams behave under different payment modes. For example, the OR manager may provide appropriate incentives for clinical teams to work overtime. Researchers can also investigate similar mechanisms in project management, transportation, and any service areas where time schedules for tasks are made in advance.

References

Ahnood D, Souriti A, Williams GS (2018) Assessing patient acceptance of virtual clinics for diabetic retinopathy: a large scale postal survey. *Canadian Journal of Ophthalmology* 53(3):207–209.

Anderson TW, Hsiao C (1981) Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76(375):598–606.

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies* 58(2):277–297.

Arellano M, Bover O (1995) Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68(1):29–51.

Arora S, Sevdalis N, Nestel D, Tierney T, Woloshynowych M, Kneebone R (2009) Managing intraoperative stress: what do surgeons want from a crisis training program? *The American Journal of Surgery* 197(4):537–543.

Arora S, Sevdalis N, Nestel D, Woloshynowych M, Darzi A, Kneebone R (2010) The impact of stress on surgical performance: a systematic review of the literature. *Surgery* 147(3):318–330.

Avgerinos E, Gokpinar B (2017) Team familiarity and productivity in cardiac surgery operations: The effect of dispersion, bottlenecks, and task complexity. *Manufacturing & Service Operations Management* 19(1):19–35.

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.

Bubalo B, Gaggero AA (2015) Low-cost carrier competition and airline service quality in europe. *Transport Policy* 43:23–31.

Chan CW, Farias VF, Escobar GJ (2017) The impact of delays on service times in the intensive care unit. *Management Science* 63(7):2049–2072.

Cheah CY (2007) Construction safety and health factors at the industry level: The case of singapore. *Journal of Construction in Developing Countries* 12(2).

Cheng H, Clymer JW, Chen BPH, Sadeghirad B, Ferko NC, Cameron CG, Hinoul P (2018) Prolonged operative duration is associated with complications: a systematic review and meta-analysis. *Journal of Surgical Research* 229:134–144.

Chernoguz DG (2011) The system dynamics of brooks' law in team production. *Simulation* 87(11):947–975.

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

Christakis I, Ng CS, Chen C, Yiin YH, Grubbs EG, Perrier ND, Lee JE, Graham PH (2019) Operation duration and adrenal gland size, but not bmi, are correlated with complication rate for posterior retroperitoneoscopic adrenalectomy for benign diseases. *Surgery* 165(3):637–643.

Cinquin O, et al. (2002) Positive and negative feedback: striking a balance between necessary antagonists. *Journal of Theoretical Biology* 216(2):229–241.

Daley BJ, Cecil W, Clarke PC, Cofer JB, Guillamondegui OD (2015) How slow is too slow? correlation of operative time to complications: an analysis from the tennessee surgical quality collaborative. *Journal of the American College of Surgeons* 220(4):550–558.

Dand A, Yildirim B, Saeed K (2019) Analysis of airline and passenger delay costs using faa data. *IIE Annual Conference.Proceedings* 735–741.

Deo S, Jain A (2019) Slow first, fast later: Temporal speed-up in service episodes of finite duration. *Production and Operations Management* 28(5):1061–1081.

Dexter F, Weih LS, Gustafson RK, Stegura LF, Oldenkamp MJ, Wachtel RE (2006) Observational study of operating room times for knee and hip replacement surgery at nine us community hospitals. *Health Care Management Science* 9(4):325–339.

Eltoukhy AE, Chan FT, Chung SH (2017) Airline schedule planning: A review and future directions. *Industrial Management & Data Systems* .

Gao S, Low SP (2014) The last planner system in china's construction industry—a swot analysis on implementation. *International Journal of Project Management* 32(7):1260–1272.

Güldner A, Kiss T, Serpa Neto A, Hemmes SN, Canet J, Spieth PM, Rocco PR, Schultz MJ, Pelosi P, Gama de Abreu M (2015) Intraoperative protective mechanical ventilation for prevention of postoperative pulmonary complications: a comprehensive review of the role of tidal volume, positive end-expiratory pressure, and lung recruitment maneuvers. *Anesthesiology* 123(3):692–713.

Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society* 1029–1054.

Ibanez MR, Clark JR, Huckman RS, Staats BR (2018) Discretionary Task Ordering: Queue Management in Radiological Services. *Management Science* 64(9):4389–4407.

Ibanez MR, Toffel MW (2020) How scheduling can bias quality assessment: Evidence from food-safety inspections. *Management Science* 66(6):2396–2416.

Jackson TD, Wannares JJ, Lancaster RT, Rattner DW, Hutter MM (2011) Does speed matter? the impact of operative time on outcome in laparoscopic surgery. *Surgical Endoscopy* 25(7):2288–2295.

Johnston D, Diamant A, Quereshy F (2019) Why do surgeons schedule their own surgeries? *Journal of Operations Management* 65(3):262–281.

Johnston R (1995) The determinants of service quality: satisfiers and dissatisfiers. *International Journal of Service Industry Management* .

Kc DS (2019) Worker productivity in operations management. *Available at SSRN 3466947* .

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.

Kempton SJ, McCarthy JE, Afifi AM (2014) A systematic review of distraction osteogenesis in hand surgery: what are the benefits, complication rates, and duration of treatment? *Plastic and Reconstructive Surgery* 133(5):1120–1130.

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.

Kripfganz S, et al. (2019) Generalized method of moments estimation of linear dynamic panel data models. *London Stata Conference*, number 17.

Lapierre SD, Batson C, McCaskey S (1999) Improving on-time performance in health care organizations: a case study. *Health Care Management Science* 2(1):27–34.

Li H, Love P, Drew D (2000) Effects of overtime work and additional resources on project cost and quality. *Engineering Construction and Architectural Management* 7(3):211–220.

Liverani A, Chiarot M, Bezzi M, Angelini L (1994) Is surgery duration really a complication factor? *Minerva Chirurgica* 49(9):747–750.

Maali Y, Perez-Concha O, Coiera E, Roffe D, Day RO, Gallego B (2018) Predicting 7-day, 30-day and 60-day all-cause unplanned readmission: a case study of a sydney hospital. *BMC Medical Informatics and Decision Making* 18(1):1–11.

Marla L, Vaaben B, Barnhart C (2017) Integrated disruption management and flight planning to trade off delays and fuel burn. *Transportation Science* 51(1):88–111.

McDermott F, Heeney A, Kelly M, Steele R, Carlson G, Winter D (2015) Systematic review of preoperative, intraoperative and postoperative risk factors for colorectal anastomotic leaks. *Journal of British Surgery* 102(5):462–479.

McSheffrey E (2022) More than 500 surgeries postponed in b.c. last week due to covid-19, minister says. URL `globalnews.ca/news/8504204/bc-covid-surgery-delay-update-january-2022/`.

Montet B (2014) Flight delayed? your pilot really can make up the time in the air. fivethirtyeight.com/features/flight-delayed-your-pilot-really-can-make-up-the-time-in-the-air/, accessed: 2021-11-13.

Morris AM, Baldwin LM, Matthews B, Dominitz JA, Barlow WE, Dobie SA, Billingsley KG (2007) Reoperation as a quality indicator in colorectal surgery: a population-based analysis. *Annals of Surgery* 245(1):73.

Moulton Ca, Regehr G, Lingard L, Merritt C, MacRae H (2010) 'slowing down when you should': initiators and influences of the transition from the routine to the effortful. *Journal of Gastrointestinal Surgery* 14(6):1019–1026.

Nepal MP, Park M, Son B (2006) Effects of schedule pressure on construction performance. *Journal of Construction Engineering and Management* 132(2):182–188.

Nickell S (1981) Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society* 1417–1426.

Orri M, Revah-Lévy A, Farges O (2015) Surgeons' emotional experience of their everyday practice-a qualitative study. *PLoS One* 10(11):e0143763.

Pandit J, Stubbs D, Pandit M (2009) Measuring the quantitative performance of surgical operating lists: theoretical modelling of 'productive potential'and 'efficiency'. *Anaesthesia* 64(5):473–486.

Pesaran MH (2015) *Time Series and Panel Data Econometrics* (Oxford University Press).

Qi J (2017) Mitigating delays and unfairness in appointment systems. *Management Science* 63(2):566–583.

Ramdas K, Saleh K, Stern S, Liu H (2018) Variety and experience: Learning and forgetting in the use of surgical devices. *Management Science* 64(6):2590–2608.

Roth A, Tucker AL, Venkataraman S, Chilingerian J (2019) Being on the productivity frontier: Identifying "triple aim performance" hospitals. *Production and Operations Management* 28(9):2165–2183.

Schliemann B, Seybold D, Gessmann J, Fehmer T, Schildhauer T, Muhr G (2009) Bipolar hemiarthroplasty in femoral neck fractures–impact of duration of surgery, time of day and the surgeon's experience on the complication rate. *Zeitschrift Fur Orthopadie Und Unfallchirurgie* 147(6):689–693.

Shen Y, Chan CW, Zheng F, Argenziano M, Kurlansky P (2021) The impact of surgeon daily workload and its implications for operating room scheduling. *Working Paper* .

Soliman E, Alrasheed K (2021) Model for construction project delay occurrence (pdo). *Journal of Engineering Research* 9(3 B):144–167.

Song H, Veeraraghavan S (2018) Quality of Care. *Handbook of Healthcare Analytics: Theoretical Minimum for Conducting 21st Century Research on Healthcare Operations* 79–108.

Staats BR, Kc DS, Gino F (2018) Maintaining beliefs in the face of negative news: The moderating role of experience. *Management Science* 64(2):804–824.

Wan M, Zhang JX, Ding Y, Jin Y, Bedford J, Nagarajan M, Bucevska M, Courtemanche DJ, Arneja JS (2020) High-risk plastic surgery: An analysis of 108,303 cases from the american college of surgeons national surgical quality improvement program (acs nsqip). *Plastic Surgery* 28(1):57–66.

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

Wang J, Dexter F, Yang K (2013) A behavioral study of daily mean turnover times and first case of the day start tardiness. *Anesthesia & Analgesia* 116(6):1333–1341.

Windmeijer F (2005) A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics* 126(1):25–51.

Wolfe JS (2020) 2020 report: Construction suffers from wasted time slow payment. www.levelset.com/blog/2020-report-construction-wasted-time-slow-payment/, accessed: 2020-04-30.

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel data* (MIT press).

Young SW, Mutu-Grigg J, Frampton CM, Cullen J (2014) Does speed matter? revision rates and functional outcomes in tka in relation to duration of surgery. *The Journal of Arthroplasty* 29(7):1473–1477.

Online Appendix

EC.1. GMM Estimation

For dynamic panel data, the classic fixed-effect ordinary-least-square (FE-OLS) estimator yields biased estimators, especially when the panel data has large cross-sectional units N and short time periods T (in our case, $N = 1,928$ and $T \in [3, 8]$). This bias is known as the *Nickell bias* (Nickell 1981). The bias arises because the lagged response variable is one of the explanatory variables. Therefore after the demeaning process (a step of the FE estimator), there will be a correlation between the autoregressive regressor and the error term. Thus, to achieve unbiased estimates of γ_n and γ_p from Equation (4), we adopt the Arellano-Bond (A-B) method (Arellano and Bond 1991) of dynamic panel models. The A-B approach utilizes higher-order lagged regressors as instrumental variables (IV) that induce a series of moment conditions for GMM estimation. The IV selection criterion depends on the properties of the explanatory variables. We now take the model of RPD as an example to elaborate the GMM estimation process. For RTD, the estimation process is largely the same. First, we transform the original model Equation (4) via first-difference (FD) or forward-orthogonal deviation (FOD). We try both approaches to reach a robust estimate. FD is a simple and direct method by taking the first order difference between two consecutive cases, which is also applied to regular static panel models (Anderson and Hsiao 1981, Arellano and Bond 1991).

Arellano and Bover (1995) proposed FOD method for dynamic panel data and the estimator is called the FOD-GMM estimator. The FOD transformation operator is as follows

$$\tilde{\Delta}_t u_{it} = \sqrt{\frac{T-t+1}{T-t}} \left(u_{it} - \frac{1}{T-t+1} \sum_{s=0}^{T-t} u_{i,t+s} \right)$$

Under the assumption that u_{it} is serially uncorrelated, we obtain that $\mathrm{Corr}\left(\tilde{\Delta}_t u_{it}, \tilde{\Delta}_t u_{i,t-1}\right) = 0$. The advantage that the FOD-GMM estimator has compared with diff-GMM is the higher data utilization rate for unbalanced panel data sets. Note that we set $S_{i,0}$, the DSS of the first case in a shift as the difference between the real start time and scheduled start time so that more data can be utilized.

After transformation, we obtain the following two equations for them, respectively,

$$\begin{aligned} \mathrm{FD}: \ \Delta(S_{it} - \mathrm{RTD}_{it}) &= (\gamma_n - 1)\Delta(S_{i,t-1})^- + (\gamma_p + 1)\Delta(S_{i,t-1})^+ + \Delta\mathbf{x}_{it}^\top\beta + \Delta\delta_t + \Delta u_{it} \\ \mathrm{FOD}: \ \tilde{\Delta}_t(S_{it} - \mathrm{RTD}_{it}) &= (\gamma_n - 1)\tilde{\Delta}_t(S_{i,t-1})^- + (\gamma_p + 1)\tilde{\Delta}_t(S_{i,t-1})^+ + \tilde{\Delta}_t\mathbf{x}_{it}^\top\beta + \tilde{\Delta}_t\delta_t\tilde{\Delta}_t + u_{it}, \end{aligned} \tag{EC.1}$$

where Δ and $\tilde{\Delta}_t$ stand for the FD and FOD transformation operators, respectively. Both FD and FOD transformations eliminate all time-invariant fixed effects such as α_i in Equation (2). We control for other variables that vary with case index (time) t in the same shift, such as patient type, case type, and case index, by including them in $\mathbf{x}_{it}$, which remain in the regression equations after the FD or FOD transformation.

Second, we need to determine the appropriate lag order of all explanatory variables on the RHS of Equation (EC.1), which leads to the moment conditions that are indispensable for GMM estimation in the A-B method. Generally speaking, the right-hand-side regressors $y_{i,t-1}$ and $\mathbf{x}_{it}$ can be categorized into four types further for which the moment conditions are different. (i). *Lagged dependent variable*:

$$E\left[y_{i,t-s}\Delta u_{it}\right]=0, \quad s=2,3,\ldots,t;$$

(ii). *Strictly exogenous regressors*:

$$E\left[\mathbf{x}_{i,t-s}\Delta u_{it}\right]=\mathbf{0}, \quad t-s=1,\ldots,T;$$

(iii). *Predetermined regressors*:

$$E\left[\mathbf{x}_{i,t-s}\Delta u_{it}\right]=\mathbf{0}, \quad s=1,2,\ldots,t;$$

(iv). *Endogenous regressors*:

$$E\left[\mathbf{x}_{i,t-s}\Delta u_{it}\right]=\mathbf{0}, \quad s=2,3,\ldots,t.$$

The criterion that determines the category of a regressor is condition independence where strictly exogenous regressors satisfy $E\left[u_{it}\,|\,\mathbf{x}_{i0},\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}\right]=0$, that is, all of them are conditionally independent of the current error term at t. Predetermined regressors should satisfy $E\left[u_{it}\,|\,\mathbf{x}_{i0},\mathbf{x}_{i1},\ldots,\mathbf{x}_{it}\right]=0$ and endogenous ones satisfy $E\left[u_{it}\,|\,\mathbf{x}_{i0},\mathbf{x}_{i1},\ldots,\mathbf{x}_{i,t-1}\right]=0$.

Here we consider higher lags of the auto-regressive term, namely $(S_{i,t-2})^-,(S_{i,t-3})^-$ and $(S_{i,t-2})^+,(S_{i,t-3})^+$, as instruments for $\Delta(S_{i,t-1})^-$ and $\Delta(S_{i,t-1})^+$ in Equation (EC.1). The associated moment conditions are formulated in Equation (EC.2). Regressors in $\mathbf{x}_{it}$ include exogenous variables of case t such as the patient age, surgery procedure, and case types, which are uncorrelated with u_{it}. We select these explanatory variables and first-order lagged terms as instruments and formulate their moment conditions. The moment conditions for the FD transformed model are presented in Equation (EC.2) below:

$$E\left[(S_{i,t-s})^-\Delta u_{it}\right]=0, E\left[(S_{i,t-s})^+\Delta u_{it}\right]=0, s=2,3; E\left[\mathbf{x}_{i,t-s}\Delta u_{it}\right]=0, s=1,2 \tag{EC.2}$$

In the FOD model, by subtracting forward means from the current levels, we can utilize the first-order lagged terms as instruments, and the associated moment conditions are as follows:

$$E\left[(S_{i,t-s})^-\tilde{\Delta}_t u_{it}\right]=0, E\left[(S_{i,t-s})^+\tilde{\Delta}_t u_{it}\right]=0, s=1,2; E\left[\mathbf{x}_{i,t-s}\tilde{\Delta}_t u_{it}\right]=0, s=0,1 \tag{EC.3}$$

In fact, we can adopt further lagged terms as instruments. Nevertheless, too many instruments may lead to over-identification issues in estimation. Thus, for endogenous variables, we only use their lags of the second and third order for the FD-GMM model, and we use lags of the first and second order for the FOD-GMM model. While for exogenous variables, we always use themselves and first-order lagged terms as instruments. We conduct a series of specification tests to validate the choice of lagged terms (see Section EC.2).

We then conduct GMM estimation based on the above moment conditions. After stacking the moment conditions, we obtain the vector-form moment conditions for GMM estimation:

$$E\left[\mathbf{m}_i(\boldsymbol{\theta})\right] = \mathbf{0}, \quad \boldsymbol{\theta} = (\gamma, \boldsymbol{\beta})$$

where $\boldsymbol{\theta}$ is the vector of all estimands in Equation (EC.1), $\mathbf{m}_i(\cdot), i = 1, 2, \ldots, N, N = 7,868$ denotes the left-hand-side of the moment conditions (EC.2) applied to a case i in the FD-GMM model (or Equation (EC.3) in the FOD-GMM model). The GMM estimators are obtained through the following two-step minimization:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\mathbf{b}} \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i(\mathbf{b})\right)^{\top} \mathbf{W}_0 \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i(\mathbf{b})\right)$$
$$\hat{\boldsymbol{\theta}}_{step2} = \arg\min_{\mathbf{b}} \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i(\mathbf{b})\right)^{\top} \mathbf{W}(\hat{\boldsymbol{\theta}}) \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i(\mathbf{b})\right), \mathbf{W}(\hat{\boldsymbol{\theta}}) = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i(\hat{\boldsymbol{\theta}})\mathbf{m}_i(\hat{\boldsymbol{\theta}})^{\top}\right)^{-1}.$$

In the first step of the GMM, we utilize an initial (suboptimal) weighting matrix $\mathbf{W}_0$, and the purpose is to obtain a GMM estimate of $\boldsymbol{\theta}$. We then obtain the consistent estimate of the inverse of the asymptotic covariance matrix of $\mathbf{m}(\hat{\boldsymbol{\theta}})$, as shown in the second set of equations. We finally update $\mathbf{W}_0$ by $\mathbf{W}(\hat{\boldsymbol{\theta}})$ and obtain the two-step GMM estimator. More details regarding GMM computation can be found inArellano and Bond (1991), Arellano and Bover (1995), Windmeijer (2005), Kripfganz et al. (2019).

EC.2. Robustness Checks

We present the normal FE-OLS estimation results of Equation (2) and Equation (5) in Table EC.1 as a reference.

EC.2.1. Dynamic Panel Model Specification Tests

Serial Correlation Test. Recall that we utilize higher-order lagged regressors as instruments in Equation (EC.2) for GMM estimation of the dynamic panel model. Under such an IV configuration, for these lagged explanatory variables to be valid instruments, the absence of higher-order serial

Table EC.1 FE-OLS Estimation Results

Variables	Surgery	Turnover
Early start	0.430***	0.283***
	(0.035)	(0.018)
Delayed start	-0.533***	-0.059***
	(0.023)	(0.013)
Elective case	13.49	6.419*
	(10.25)	(2.948)
Sequence shuffling	-2.353	-5.208**
	(2.217)	(1.589)
Age	0.348***	0.060
	(0.091)	(0.046)
Different procedure		1.175**
		(0.435)
Next case type		Included
Patient type	Included	Included
Patient severity	Included	Included
Procedure type	Included	Included
Shift fixed effect	Included	Included
Sequence fixed effect	Included	Included
R^2	0.496	0.531
N	7,787	6,120

Notes: The linear model Equation (2) and Equation (5) are estimated by OLS. Estimated by FE-OLS. Robust standard errors clustered by shift are shown in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

correlation between error terms Δu_{it} and $\Delta u_{i,t-s}, (s = 2, 3)$ is indispensable. Here the formal null hypothesis is

$$H_0 : \text{Corr}\left(\Delta u_{it}, \Delta u_{i,t-j}\right) = 0, \;\; j = 2, 3. \tag{EC.4}$$

We thus adopt the Arellano-Bond specification test, also known as the AR(p) test, to examine the above validity requirements, where p stands for the lagged order. To be specific, we perform AR(1) to AR(3) tests for both Models (1) and (2) to identify the starting level of the lag structure of the instruments. From the p values in the last part of Table EC.2 we find that the AR(1) test is rejected, since applying FD transformation to Equation (2) naturally brings first order serial correlation between Δu_{it} and $\Delta u_{i,t-1}$. We cannot reject the null hypotheses of AR(2) and AR(3), which means there is no higher order serial correlation in Equation (EC.1). Therefore, our null hypothesis Equation (EC.4) is supported and the IV configuration we choose is valid for identification. For moment equations of the FOD-GMM model we also conduct similar tests and the results also cannot reject the null hypothesis.

Table EC.2 Specification Tests for Main Model

	Surgery		Turnover	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Early Start	0.617***	0.444***	0.190***	0.292***
	(0.159)	(0.126)	(0.051)	(0.047)
Delayed Start	-0.219*	-0.164*	-0.121**	-0.096*
	(0.106)	(0.081)	(0.045)	(0.045)
Elective case	13.78	-25.54	4.505	-20.74
	(60.37)	(58.13)	(2.843)	(19.96)
Sequence shuffling	-1.557	5.274	-7.235***	-9.074***
	(5.549)	(5.274)	(1.864)	(2.284)
Age	0.280	0.277*	0.361*	0.025
	(0.148)	(0.138)	(0.163)	(0.172)
Different procedure			0.976	1.748**
			(0.588)	(0.543)
Next case type			Included	Included
Patient type	Included	Included	Included	Included
Patient severity	Included	Included	Included	Included
Procedure type	Included	Included	Included	Included
Time fixed effect	Included	Included	Included	Included
Specification tests:				
Arellano–Bond AR(1) p value	0.000	0.000	0.000	0.000
Arellano–Bond AR(2) p value	0.593	0.299	0.205	0.589
Arellano–Bond AR(3) p value	0.950	0.868	0.951	0.436
Hansen J-test χ^2	129.2	124.6	135.2	108.6
Hansen J-test p value	0.287	0.392	0.163	0.649
Number of instruments	253	253	258	253
Number of observations	7,868	7,868	6,210	6,210

Notes: Both of the GMM models are estimated using the Arellano-Bond method. Sequence fixed effect is exactly the time effect t in shift i. We control for other fixed effects that vary with case index in the same shift, such as patient type, case type and case index. Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

We also conduct the Hansen J-test (Hansen 1982) of model over-identification, as the number of instruments is relatively large. The null hypothesis is

$$H_0 : \text{Over-identifying restrictions are valid.}$$

We see that both the FD-GMM and FOD-GMM models yield non-significant χ^2 values. Thus they pass the Hansen test and the IV configurations for identification are supported.

Apart from the main model of Table EC.2, we also conduct AR(p) serial correlation tests and Hansen J-tests for the seniority exploration. As shown in the third part of Table EC.3, both regressions pass the serial correlation tests and Hansen J-tests, which validate the estimation results.

Table EC.3 Estimates for Different Surgeon Seniority Group

	Senior		Junior	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Early Start	0.679***	0.505***	0.368	0.404
	(0.150)	(0.141)	(0.265)	(0.210)
Delayed Start	-0.359*	-0.319**	-0.313*	-0.234
	(0.155)	(0.113)	(0.154)	(0.136)
Elective case	59.10*	42.09	15.79	-13.77
	(23.89)	(37.86)	(65.03)	(48.12)
Sequence shuffling	-5.889	-4.485	0.166	-3.363
	(7.337)	(5.698)	(10.28)	(8.709)
Age	-0.255	-0.129	0.324	0.410
	(0.273)	(0.189)	(0.247)	(0.224)
Patient type	Included	Included	Included	Included
Patient severity	Included	Included	Included	Included
Procedure type	Included	Included	Included	Included
Time fixed effect	Included	Included	Included	Included
Specification tests:				
Arellano–Bond AR(1) p value	0.000	0.000	0.000	0.000
Arellano–Bond AR(2) p value	0.355	0.078	0.279	0.496
Arellano–Bond AR(3) p value	0.978	0.353	0.477	0.990
Hansen J-test χ^2	105.3	112.4	109.6	108.0
Hansen J-test p value	0.731	0.550	0.309	0.350
Number of instruments	244	244	227	227
Number of observations	3,793	3,793	4,075	4,075

Notes: Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses. We omit other regressors in $\mathbf{x}_{it}$. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

EC.2.2. Sample Inclusion Criterion

Remove all shifts with add-on cases. We remove the shifts that include add-on cases from our sample, which yields 7,484 cases (5,827 for RTD) in all. The results are shown in Table EC.4. We find that the estimates for both surgery stage (RPD) and turnover stage (RTD) are consistent with Table 2 and significant. Both surgical and cleaning teams respond to both directions.

Remove all shifts affected by cancellations. We remove all shifts that are affected by cancellations, which yields 7,777 cases in the refined sample (6,122 for RTD). The GMM estimation results are shown in Table EC.5. We find that the estimates are still consistent. These results support our findings of the adaptive behavior by surgical and cleaning teams.

EC.2.3. IV Specification Tests

The validity of using DSS as an IV relies on the underlying assumption that they will impact readmission/reoperation risk only through impacting the RPD. One may argue that DSS is correlated

Table EC.4 Robustness with Add-on Cases Estimation Results

	Surgery		Turnover	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Early Start	0.590***	0.424**	0.250***	0.306***
	(0.166)	(0.134)	(0.057)	(0.057)
Delayed Start	-0.242*	-0.167*	-0.120*	-0.088†
	(0.110)	(0.082)	(0.048)	(0.050)
Next case type			Included	Included
Elective case	Included	Included	Included	Included
Sequence shuffling	Included	Included	Included	Included
Age	Included	Included	Included	Included
Patient type	Included	Included	Included	Included
Patient severity	Included	Included	Included	Included
Procedure type	Included	Included	Included	Included
Time fixed effect	Included	Included	Included	Included
Specification tests:				
Arellano–Bond AR(1) p value	0.000	0.000	0.000	0.000
Arellano–Bond AR(2) p value	0.627	0.367	0.430	0.514
Arellano–Bond AR(3) p value	0.812	0.610	0.858	0.488
Hansen J-test χ^2	129.3	121.0	123.5	119.9
Hansen J-test p value	0.286	0.482	0.394	0.435
Number of instruments	253	253	256	252
Number of observations	7,484	7,484	5,827	5,827

Notes: Both of the two GMM models are estimated using Arellano-Bond method. The time fixed effect is t in shift i. Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses (†$p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

with the fatigue level of the surgical team, which further impacts the readmission/reoperation risk. To address this concern, we control for the hour-of-shift (HOS) of each case, where HOS n stands for the n-th hour after the focal shift starts. The HOS well captures the surgical team's cumulative working hours since the start of shift, and hence provides a proxy for the fatigue level. The regression results with HOS as an extra control are presented in Table EC.6. We find that the IV-Probit estimates of the focal RPD still remain negatively significant for both 30-day readmissions and reoperations. That is, a faster-than-scheduled surgery leads to increased post-surgical risks. We also find that HOS indicators are mostly insignificant across the four models, which manifests that post-surgical readmission and reoperations are not influenced by the fatigue level.

To further verify that DSS is not correlated with the unobserved errors (so they are valid), we run a regression by controlling them and RPD in a plain probit model for 30-day readmission and reoperation. The results, as summarized in Table EC.7, show that the estimates of DSS are not significant. This means that DSS is only correlated with the RPD but not with other unobserved variables. Thus, it is valid to serve as an IV.

Table EC.5 Robustness with Cancellations Estimation Results

	Surgery		Turnover	
Variables	(1) FD	(2) FOD	(3) FD	(4) FOD
Early Start	0.614***	0.499***	0.203***	0.291***
	(0.159)	(0.131)	(0.048)	(0.045)
Delayed Start	-0.212*	-0.170*	-0.098*	-0.081†
	(0.105)	(0.084)	(0.043)	(0.043)
Next case type			Included	Included
Elective case	Included	Included	Included	Included
Sequence shuffling	Included	Included	Included	Included
Age	Included	Included	Included	Included
Patient type	Included	Included	Included	Included
Patient severity	Included	Included	Included	Included
Procedure type	Included	Included	Included	Included
Time fixed effect	Included	Included	Included	Included
Specification tests:				
Arellano–Bond AR(1) p value	0.000	0.000	0.000	0.000
Arellano–Bond AR(2) p value	0.630	0.551	0.194	0.581
Arellano–Bond AR(3) p value	0.977	0.608	0.806	0.519
Hansen J-test χ^2	128.2	120.3	133.8	108.5
Hansen J-test p value	0.311	0.500	0.184	0.653
Number of instruments	253	253	258	253
Number of observations	7,777	7,777	6,122	6,122

Notes: Both of the GMM models are estimated using Arellano-Bond method. The time fixed effect is t in shift i. Robust standard errors clustered by shift i with the Windmeijer (2005) correction are shown in parentheses ($\dagger p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Table EC.6 Effect of RPD on Readmissions and Reoperations with HOS

	Panel A: IV-Probit		Panel B: NonIV-Probit	
Variables	(1) 30-Read	(2) 30-Reop	(3) 30-Read	(4) 30-Reop
RPD of the focal surgery	-0.0316***	-0.0332***	0.0043**	0.0052**
	(0.0066)	(0.0051)	(0.0014)	(0.0018)
RTD of the preceding surgery	0.0007	0.0016	0.0020	0.0050
	(0.0025)	(0.0022)	(0.0053)	(0.0050)
HOS: 2	-0.0917	-0.1480	-0.2150	-0.4366*
	(0.0971)	(0.1100)	(0.2620)	(0.2050)
HOS: 3	-0.0970	-0.1744	-0.1983	-0.4851*
	(0.1070)	(0.1118)	(0.2826)	(0.2226)
HOS: 4	-0.2333	-0.2371	-0.5403	-0.6356*
	(0.1264)	(0.1436)	(0.3121)	(0.3013)
HOS: 5	-0.1996	-0.1669	-0.3980	-0.3576
	(0.1127)	(0.1186)	(0.2600)	(0.2829)
HOS: 6	-0.1616	-0.2093	-0.2629	-0.4316
	(0.1184)	(0.1298)	(0.3183)	(0.3212)
HOS: 7	-0.2037*	-0.2190	-0.3084	-0.3846
	(0.1014)	(0.1198)	(0.2505)	(0.2490)
Sequence shuffling	Included	Included	Included	Included
Age	Included	Included	Included	Included
Patient type	Included	Included	Included	Included
Patient severity	Included	Included	Included	Included
Procedure type	Included	Included	Included	Included
Time fixed effect	Included	Included	Included	Included
Day-of-week fixed effect	Included	Included	Included	Included
Surgeon fixed effect	Included	Included	Included	Included
Correlation of errors	0.8842***	0.9292***		
Number of observations	5,204	5,204	5,204	5,204

Notes: Estimated by the maximum likelihood method. Dependent variables: 30-day readmission, 30-day reoperation indicators. We report the correlation of the errors of the two equations in the IV-probit models in the final part. Robust standard errors are shown in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Table EC.7 Effect of RPD on Readmissions and Reoperations with DSS

	30-day Readmission	30-day Reoperation
RPD of the focal surgery	0.0044***	0.0053***
	(0.0013)	(0.0015)
DSS	0.0018	0.0023
	(0.0011)	(0.0013)
RTD of the preceding surgery	0.0015	0.0042
	(0.0043)	(0.0046)
Sequence shuffling	Included	Included
Age	Included	Included
Patient type	Included	Included
Patient severity	Included	Included
Procedure type	Included	Included
Time fixed effect	Included	Included
Day-of-week fixed effect	Included	Included
Surgeon fixed effect	Included	Included
Number of observations	5,204	5,204

Notes: Estimated by the maximum likelihood method. Dependent variables: 30-day readmission and reoperation indicators. Robust standard errors are shown in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).