8-2023

Leveraging the Granularity of Healthcare Data: Essays on Operating Room Scheduling for Productivity and Nurse Retention

Jaeyoung Kim
jaeyouk@clemson.edu

Leveraging the Granularity of Healthcare Data: Essays on Operating Room Scheduling for Productivity and Nurse Retention

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Management

by
Jaeyoung Kim
August 2023

Accepted by:
Dr. Lawrence Fredendall, Committee Co-Chair
Dr. Ahmet Colak, Committee Co-Chair
Dr. Babur De los Santos
Dr. Benjamin Grant

Abstract

The primary objective of this dissertation is to provide insights for healthcare practitioners to leverage the granularity of their healthcare data. In particular, leveraging the granularity of healthcare data using data analytics helps practitioners to manage operating room scheduling for productivity and nurse retention. This dissertation addresses the practical challenges of operating room (OR) scheduling by combining the existing insights from the prior literature through various tools in data analytics. In doing so, this dissertation consists of three chapters that operationally quantify the operational characteristics of the operating room and surgical team scheduling to improve operating room outcomes, including OR planning and OR nurse retention. This dissertation contributes to healthcare operations research and practice by emphasizing the importance of using granular information from hospitals' electronic health records. While the prior research suggests that different team compositions affect OR productivity and OR time prediction, the empirical insights on how the team composition information can be utilized in practice are limited. We fill this gap by presenting data-driven approaches to use this information for OR time prediction and nurse retention. The first and third chapters deal with OR time prediction with the granular procedure, patient, and detailed team information to improve the OR scheduling. The second chapter deals with the OR nurse retention problem under OR nurses' unique work scheduling environment.

The first chapter, which is a joint work with Ahmet Colak, Lawrence Fredendall, and Robert Allen, examines drivers of OR time and their impact on OR time allocation mismatches (i.e., deviations of scheduled OR time from the realized OR time). Building on contemporary health care and empirical methodologies, the chapter identifies two mechanisms that spur scheduling mismatches: (i) OR time allocations that take place before team selections and (ii) OR time allocations that do not incorporate granular team and case data inputs. Using a two-stage estimation framework, the chapter shows how under- and over-allocation of OR times could be mitigated in a newsvendor

setting using improved OR time predictions for the mean and variance estimates. The chapter's empirical findings indicate that scheduling methods and the resulting scheduling mismatches have a significant impact on team performance, and deploying granular data inputs about teams—such as dyadic team experience, workload, and back-to-back case assignments—and updating OR times at the time of team selection improve OR time predictions significantly. In particular, the chapter estimates a 32% reduction in absolute mismatch times and a more than 20% reduction in OR costs.

The second chapter, which is a joint work with Ahmet Colak and Lawrence Fredendall, addresses the turnover of OR nurses who work with various partners to perform various surgical procedures. Using an instrumental variable approach, the chapter identifies the causal relationship between OR nurses' work scheduling and their turnover. To quantify the work scheduling characteristics—procedure, partner, and workload assignments, the chapter leverages the granularity of the OR nurse work scheduling data. Because unobserved personal reasons of OR nurses may lead to a potential endogeneity of schedule characteristics, the chapter instruments for the schedule characteristics using nurse peers' average characteristics. The results suggest that there are significant connections between nurse departure probability and how procedures, partners, and workload are configured in nurses' schedules. Nurses' propensity to quit increases with high exposure and diversity to new procedures and partners and with high workload volatility while decreasing with the workload in their schedules. Furthermore, these effects are significantly moderated by the seniority of nurses in the hospital. The chapter also offers several explanations of what might drive these results. The chapter provides strategic reasoning for why hospitals must pay attention to designing the procedure, partner, and workload assignments in nurse scheduling to increase the retention rate in the ongoing nursing shortage and high nurse turnover in the U.S.

The third chapter, which is a joint work with Ahmet Colak, Lawrence Fredendall, Babur De los Santos, and Benjamin Grant, systematically reviews the literature to gain more insights into addressing the challenges in OR scheduling to utilize granular team information for OR time prediction. Research in OR scheduling—allocating time to surgical procedures—is entering a new phase of research direction. Recent studies indicate that utilizing team information in OR scheduling can significantly improve the prediction accuracy of OR time, reducing the total cost of idle time and overtime. Despite the importance, utilizing granular team information is challenging due to the multiple decision-makers in surgical team scheduling and the presence of hierarchical structure in surgical teams. Some studies provide some insights on the relative influence of team members, which partly helps address these challenges, but there are still limited insights on which decision-maker has the greatest influence on OR time prediction and how hierarchy is aligned with the relative impact of surgical team members. In its findings, the chapter confirms that there are limited empirical insights in the existing literature. Based on the prior insights and the most recent development in this domain, this chapter proposes several empirical strategies that would help address these challenges and determine the key decision-makers to use granular team information of the most importance.

Dedication

To my parents, Ilya, and everyone who has helped me get to this point.

Acknowledgments

First and foremost, words cannot express my gratitude to Dr. Lawrence Fredendall, the chair of my committee, for his invaluable guidance throughout my doctoral education. His research insights, support, and generous feedback kept me motivated to be not only a better researcher but also a better person. I also could not have undertaken this journey without Dr. Ahmet Colak, the co-chair of my committee. His enthusiasm, research expertise, and novel research insights kept challenging me to reach my full potential on the journey. Also, I am very grateful for the valuable guidance and feedback from my other committee members, Dr. Babur De los Santos and Dr. Benjamin Grant. Their keen research insights and willingness to help inspired me a lot to develop the necessary expertise and skills.

I would like to sincerely thank the faculties in the management department for their support and feedback during my research presentations and other steps of my doctoral education. I especially thank Dr. Aleda Roth who shared her seasoned insights on various supply chain and operations management topics and methodologies in her doctoral seminar courses, which was indispensable and gave me a high standard to live up to as a scholar. Furthermore, I'd like to acknowledge the support and advice from Dr. Maggie Zhang since the first year of my doctoral education.

In addition, I am deeply indebted to Brandon Lee, a graduate of the Management department and a faculty at the University of Dayton, for his help and mentoring. I owe thanks to the faculties at the University of Texas at Dallas—Dr. Shawn Alborz, Dr. Elena Katok, Dr. Ernan Haruvy, Dr. Dorothee Honhon, and Dr. Anyan Qi—who have mentored me at critical times during and after my master's education without hesitation.

Finally, I owe great thanks to my family for their support and unconditional love and to all my friends at Clemson—especially Giyoon, Amin, and Sara—for their support and encouragement at every step of my doctoral education.

Table of Contents

List of Tables

List of Figures

Chapter 1

Reducing OR Time Mismatches Using Data-Driven Operations Scheduling

1.1 Introduction

Managing operating room (OR) time allocations is a fundamental managerial process for surgical health care practitioners. When allocated surgical times differ substantially from actual realized surgical times on the day of the surgery, scheduling mismatches occur. These mismatches occur in two forms: over-allocation of OR times increases idle times; under-allocation of OR times increases overtime and can increase delays in key surgical resources becoming available for others (Olivares et al. 2008). The importance to surgical practitioners of reducing both over-allocation and under-allocation of OR times requires a comprehensive managerial understanding of mismatch causes and mitigation strategies. Before formulating our research questions regarding OR time mismatches, we identify two primary and understudied managerial gaps in the OR management literature and in practice, which motivate our research.

First, our literature review identified limited research insights and empirical evidence on predictive methods and decision models for OR time allocation. The existing OR literature largely focuses on causal inferences about operational and organizational factors that influence OR times and under-explores predictive OR time decision-making models and especially does not consider OR team effects. The ideal OR time allocation decision would use granular data for different procedures, patients, OR teams, and surgical workload and advanced *predictive* data-driven methods to forecast the mean and variance of OR times. Also, while Olivares et al.'s (2008) seminal *newsvendor* application has opened the door for OM researchers to reduce OR time mismatch costs, the use of empirical newsvendor models in the current OR scheduling literature is still rare. The traditional operations management (OM) newsvendor models enable decision-makers to account for both the mean and variance of the underlying processes in determining an optimal safety capacity with given under- and over-allocation cost structures. Ideally, the newsvendor models could incorporate the surgical time variance predicted by the predictive methods to minimize total OR costs.

Second, from a practitioner perspective, a large number of major hospitals use a limited set of the empirical tools and mechanisms to allocate OR times, which constraints their capability to use data in decision-making processes. These substantial gaps and opportunities in health care practice motivate our research to improve both data usage and decision-making capabilities. Specifically, in our research, we partner with a leading South Carolina hospital to study available data and decision-making processes that produce OR time scheduling mismatches. Our partner hospital uses a sequential scheduling policy for OR time allocations and staffing. First, they assign OR times using a moving-average base policy (based on the last ten same-procedure surgeries of the lead surgeon) and auxiliary ad hoc adjustments (based on the lead surgeon's subjective estimates) to predefined blocks of OR time. Next, anesthesiology and nursing schedulers make staffing decisions. As a result of this policy, our partner hospital uses only a small portion of electronic health records (EHR) data available before team selections for OR time decisions and does not consider newsvendor models. This policy limits the ability of the hospital's decision-making processes to include granular data about the procedure, the team, and historical OR time variability.

Based on these two research motivations about the current scheduling policies of our partner hospital and common practice, we raise two research questions in parallel with our research motivations: Do assigning surgery times before team selections, and without granular EHR data inputs, increase OR time mismatches and affect team performance? If so, could using more granular data and models to predict the mean and variance of OR times reduce hospitals' newsvendor mismatch costs under different under- and over-allocation cost structures? The surgery scheduling literature has not fully investigated these questions. These two under-studied managerial questions—representing aspects of OR time complexity and uncertainty—motivate our research to develop a novel predictive OR time estimation approach and to apply a newsvendor decision-making model to reduce OR time mismatches and costs for our partner hospital.

To study our research questions, we obtained restricted access to a large surgical database that contains more than 80,000 unique surgical cases from 27 specialties and 39 ORs. This database contains detailed information about 4.49 million OR team dyads—pairs of OR team members (e.g., surgeon and circulating nurse, surgeon and anesthesiologist, etc.). For each surgery, we collect both scheduled and realized OR timestamps for patient in, procedure start, procedure end, and patient out events, which are critical to evaluating the magnitudes of mismatches across distinct surgical stages in ORs.

Using this database, we make four novel contributions to the OR management literature: (i) in our empirical strategy, we decouple expected OR times before and after team selections and show that OR time expectations at these two time points differ substantially from one another; (ii) we incorporate more granular inputs in the model—incorporating surgical team characteristics into the analysis as dyads and incorporating procedure descriptions into the analysis using principal component analysis (PCA); (iii) we use the mean and variance estimates from our predictive OR time models as inputs to the underlying OM newsvendor model and show that our improved OR time predictions with newsvendor application can significantly reduce total mismatch costs, and lastly; (iv) we empirically document the effects of scheduling mismatches on team performance. In the following paragraphs, we explain each contribution relating to our empirical frameworks and the results.

Our first research contribution relates to incorporating OR team effects into OR time predictions by explicitly decoupling OR time expectations before and after team selections. While many previous causal inference studies in the medical and OM literature have identified the effect of team characteristics and team processes on task completion times due to the differentiated coordination among team members (Faraj and Xiao 2006, Gittell et al. 2006, KC and Terwiesch 2009), practice-driven applications of OR time allocations mostly do not incorporate such causal team effects—such as familiarity, diversity, and workload effects (Reagans et al. 2005, Avgerinos and Gokpinar 2017, Huckman and Staats 2011, Akşin et al. 2021)—on OR time. Similarly, our partner hospital considers only a limited set of OR team data inputs that become available before team selections. To fill this gap, we develop a two-stage estimation framework that estimates OR times both before and after team selections to incorporate OR team data inputs. Specifically, we decompose the expected OR time into two structural components: the ex ante OR time that captures the average OR time expectation before team selection—i.e., OR time that is conditional on pre-team-selection data inputs but unconditional on post-team-selection data inputs; and the ex post OR time that captures the change in the OR time expectation after team selection—i.e., the positive or negative OR time adjustment that is conditional on post-team-selection data inputs. The pre-team-selection data inputs include procedure, patient, surgeon, and overall OR workload variables; the post-team-selection data inputs include team experience and workload variables.

Our two-stage estimations show that ex ante and ex post OR times significantly differ from one another—post-team-selection data inputs play a key role in explaining surgical times. Adding post-team-selection data inputs reduces mean squared error (MSE) of the logged OR time mismatches by 49% compared to the prediction model that uses all available pre-team-selection data inputs. These empirical findings reveal that allocating OR times before team selections can significantly reduce the accuracy of scheduled OR times, resulting in OR time mismatches. Our findings imply that practitioners should update their OR time estimates after team selections to avoid excessive scheduling mismatches.

Our second research contribution relates to our novel approaches to incorporating more granular data inputs and models into predicting OR times. In particular, we increase the granularity for two data input categories: surgical procedures and team member interactions. To capture a large variety of surgical procedures, we use a principal component analysis (PCA) of current procedural terminology (CPT) codes. While distinct procedures affect OR times differently, there is little empirical research that uses a large number of distinct procedures as a control. PCA of the word description in each CPT code reduces the dimensionality of the various procedures while minimizing information loss. Next, to capture the granularity of team member interactions, we take advantage of methodological advancements in empirical dyadic network analysis in supply chain and organizations research—which explore unique firm-to-firm or person-to-person effects in complex collaboration networks. The application of such dyadic network analysis to surgical teams in health care research is still scant, and we use this dyadic approach to capture more granular surgical team member interactions. Specifically, in our dyadic approach, we measure team interaction variables across distinct dyadic role-pairs (for example, a nurse-surgeon pair) rather than aggregating the variables to a team average (Reagans et al. 2005). For example, nurse-surgeon familiarity on a team could be higher or lower than the average familiarity of all pairs or significantly differ from surgeon-anesthesiologist familiarity, which could affect OR time uniquely. Also, we capture both shared experience (a pair's joint dyad experiences) and non-shared experience (a pair's different individual experiences outside the dyad) of each dyadic pair, capturing the similarity and diversity within surgical teams.

We obtain four mismatch estimates that decrease as we add more granular data inputs for OR time predictions: (i) the base OR time mismatch of scheduled (i.e., using the current combined moving average and ad hoc policies) OR times compared to realized OR times, (ii) the ex ante OR time mismatch of our pre-team-selection OR time predictions compared to realized OR times, (iii) the ex post team-level OR time mismatch of our post-team-selection team-level OR time predictions compared to realized OR times, and (iv) the ex post dyad-level OR time mismatch of our post-team-selection dyad-level OR time predictions compared to realized OR times. We find that using more granular procedure data inputs and models through our PCA and dyadic approaches improves the OR time predictions substantially. Specifically, our PCA approach improves ex ante OR time predictions (in stage 1 estimation), and our dyadic ex post model outperforms the aggregate team-level model in our ex post OR time predictions (in stage 2 estimation). The inclusion of granular data inputs both before (PCA analysis) and after team selections (dyadic analysis) would decrease MSE of the OR time predictions by more than 50% compared to the current policy. Our novel empirical approaches and the insights contribute to both health care literature and OR scheduling practice.

Our third research contribution relates to using predictive OR time mean and variance estimates in a traditional OM decision-making model: the newsvendor model. In particular, the inherent variability of OR times—in terms of mean levels and variance uncertainty spreads—creates a newsvendor environment (Lehtonen et al. 2013) where the realized OR times commonly follow a log-normal distribution with frequent under- and over-allocation outcomes in practice. While the analytical OM literature established the newsvendor model as a core topic to make optimal safety capacity time under uncertainty of demand (Olivares et al. 2008), its empirical applications are limited and rarely used for surgical scheduling in practice. Similarly, our partner hospital does not make use of the historical variance of surgery times[1] for OR time decision-making even though it is

[1]OR time variances often are not estimated or used for statistical analysis in practice. In our work, we estimate both the mean and variance of OR times for distinct procedures, patients, OR teams, and workloads. Our mean and variance estimates are particularly useful in the newsvendor context.

a common practice to assign ad hoc safety capacity on OR times in the form of increased idle times (Venkataraman et al. 2018).

Hence, our study compares this ad hoc safety capacity decision in practice (at our partner hospital) with the statistical newsvendor decision in allocating safety capacity OR time. We estimate the mean and variance of OR times and use these estimates to compute the optimal newsvendor OR time reservation under different overage and underage cost scenarios where the overtime (under-allocation) cost equals or exceeds the idle time (over-allocation) cost (Venkataraman et al. 2018)[2]. From the analysis, we find that improved empirical predictions of OR times significantly outperform the existing ad hoc policies with respect to the total mismatch costs, especially when under-allocation is more costly than over-allocation (i.e., when overtime costs significantly exceed idle time costs). Therefore, our findings imply that hospital managers should consider using newsvendor models together with improved OR time predictions (both for mean and variance estimates) to reduce their total combined overtime and idle time costs. Moreover, hospital managers need to have clear cost estimates for overtime and idle time to mitigate the impact of OR time uncertainty through the newsvendor model.

Lastly, our fourth contribution relates to the evidence for scheduling effects on team performance. A recent stream of work scheduling literature has found that work scheduling has a significant impact on worker performance by affecting worker behavior. For example, scheduling volatility can have adverse effects on psychological distress (Schneider and Harknett 2019), productivity (Allen et al. 2013), and turnover (Choper et al. 2019, Bergman et al. 2023, 2022); workers tend to change their work standards in response to the given schedule (Ibanez and Toffel 2020, KC and Terwiesch 2009). Motivated by these studies, in our two-stage estimation framework, we explore the effect of ex ante scheduling mismatches on ex post team performance. More specifically, we consider two scheduling characteristic variables: ex ante scheduling mismatch—the difference between the scheduled and predicted ex ante OR time, representing the difference between planned task length and the realistic task length given the task characteristics—and ex ante task diversity—the difference between the expected and assigned ex ante OR time conditional on team characteristics, representing the deviation of the assigned ex ante OR time from the average OR time expectation for the given team characteristics. Our findings show that these two scheduling characteristics have a significant impact on team performance. By learning these scheduling variables from the stage

[2]Venkataraman et al. (2018) find that hospital managers are more averse to overtime than to idle time.

1 ex ante OR time estimation, our ex post team performance model captures the scheduling effects on team performance and updates the OR time to reduce the eventual scheduling mismatches. Also, our team performance findings with respect to scheduling mismatches provide clear evidence that schedule characteristics have significant behavioral effects on surgical teams, future studies to investigate these effects in depth are called for.

There are six sections in the remainder of this manuscript: §1.2 reviews the relevant background literature for scheduling inputs, decision-making, and outcomes. §1.3 presents our surgery data collection. §1.4 shows variable construction details for pre- and post-team-selection data inputs. §1.5 presents our empirical OR time model and estimation strategy. §1.6 introduces our results for ex ante and ex post OR time models. Lastly, §1.7 concludes with managerial remarks and discussion.

1.2 Related Literature

Empirical research in surgery scheduling is an emerging area in health care operations management (OM) (Roth et al. 2019, KC et al. 2020b, Terwiesch et al. 2020). While there is a large array of empirical and analytical topics on surgical scheduling and OR time allocations, including block scheduling (May et al. 2011), we focus on three research streams for OR environments: (i) scheduling data inputs, (ii) scheduling decision-making, and (iii) scheduling outcomes. Specifically, the first subsection details data inputs that become available before and after team selections for scheduling decisions, the second subsection reviews the decision-making processes for OR time allocations, and the third subsection focuses on OR outcome measures commonly used in previous research. Also, within each subsection, we discuss the foundational literature motivations and contributions that relate to our work.

1.2.1 Scheduling Inputs

This subsection reviews the scheduling data inputs that can help predict OR times. In particular, we consider two categories of inputs that have strong connections to OR times: inputs that do not depend on team formation (patient, procedure, and OR system workload factors) and inputs that depend on team formation (team composition, team experience, and scheduling factors). While previous health care causal inference studies identified these scheduling inputs as predictors of OR time, these inputs have not yet been integrated into a predictive OR time model. In the following paragraphs, we highlight these two major input categories sequentially.

First, previous literature has examined the scheduling data inputs that relate to procedures, patients, and overall OR workloads. More specifically, surgical times can vary by types and number of procedures and patient characteristics (e.g., demographics, severity, and comorbidity) that affect task complexity of the focal surgery (Strum et al. 1999, May et al. 2011); surgeon characteristics (training and experience) that create variability in surgical times (Strum et al. 2000); and overall OR workload—dependent on the month, day, and time of the surgery—which influences resource utilization and productivity (Kelz et al. 2009, Anderson et al. 2014, Berry Jaeker and Tucker 2017, Song and Huckman 2018). In our study, these data inputs relate to ex ante factors before team selections. For pre-team-selection inputs, we integrate these previous findings into our prediction model and extend the literature with a novel approach to account for granular procedure characteristics: principal component analysis (PCA) on procedure description words.

Second, previous literature has examined the team data inputs that relate to surgical team productivity. Particularly, characteristics of individuals, pairs of individuals, and teams as a whole influence task completion time (KC et al. 2020b); greater individual experience reduces task duration (Pisano et al. 2001, Reagans et al. 2005); shared experience among team members increases collective learning and improves their performance (Gittell et al. 2006, Akşin et al. 2021); collective team learning outcomes are more impactful in complex tasks (Faraj and Xiao 2006, Avgerinos and Gokpinar 2017); distinct role-pairs (such as nurse-surgeon pair) have differentiated joint learning curves (Kim et al. 2018); team diversity in experience helps team members manage changes in tasks (Huckman and Staats 2011); and back-to-back team member joint assignment between two consecutive surgeries improves team performance (Stepaniak et al. 2010). In our study, these data inputs relate to the ex post factors known after team selections. Similarly with pre-team-selection inputs, for post-team-selection inputs, we build on these existing findings but extend the literature by incorporating a detailed dyadic analysis to account for dyadic team interaction variables—which are commonly missed in OR time decision-making. Our empirical findings indicate that using dyadic team variables in OR time allocations can significantly improve predictive power. Combined, our findings indicate that hospitals and surgery care practitioners could significantly benefit from using a wider range of granular procedure and team data inputs.

1.2.2 Scheduling Decisions

This subsection reviews OR time allocation decision-making processes. Specifically, we consider three decision-making areas that relate to OR scheduling: the current OR time allocation policies used in practice, emerging data-driven approaches to OR time predictions, and empirical newsvendor models for OR time allocations. In our study, we compare the first two existing areas and suggest the potential use of empirical newsvendor models for data-driven OR time allocations under uncertainty. In the following paragraphs, we review each area in more depth.

First, we discuss the limitations of the current OR time allocation policies in practice, especially for the utilization of data inputs. Historically and until now, the primary OR time allocation policy for large-scale hospitals has involved two main components: moving-average OR time estimates based on the surgeon and procedure type and/or subjective ad hoc estimates by surgeon or nurse (Hosseini et al. 2015). However, we find notable criticisms in the literature of these existing approaches: Eijkemans et al. (2010) discussed that simply averaging historical durations limits prediction capability for OR time and uses a limited portion of EHR data, and Zhou et al. (2016) and Eijkemans et al. (2010) found that the estimates by expert opinions (i.e., surgeon and nurse) are often inferior to estimates from data-driven models. Clearly, the limitations of the current practice motivate our study to improve hospitals' decision-making processes by improving OR time prediction models.

Second, an emerging empirical health care research stream uses contemporary data analytics tools to improve OR time prediction accuracy. For instance, Bartek et al. (2019) and Zhou et al. (2016) used machine learning models to incorporate patient, procedure, and surgeon information; Stepaniak et al. (2009) estimated the statistical distribution of surgical durations by considering anesthesia and surgeon types; and Eijkemans et al. (2010) used a mixed linear model that considers surgeon, anesthesiologist, patient, and procedure characteristics. However, there are much more available data about team characteristics that the earlier data-driven prediction studies have not investigated. For example, a large number of EHR data elements that are found useful to explain OR time in previous causal inference studies, such as team familiarity (Reagans et al. 2005), have not been factored in to empirical predictive analysis as data inputs.

And third, we discuss empirical newsvendor applications in the context of OR time allocation. Even though the OM newsvendor model can be a powerful OR time allocation tool to reduce total costs by balancing overtime and idle time costs under uncertainty regarding surgical times, their practical use in modern medical practice and health care studies is rare and limited in scope. Ideally, OR time allocation decisions should be able to not only predict the mean OR time but also deal with OR time variability from a safety capacity perspective (Olivares et al. 2008, Lehtonen et al. 2013). For this reason, the newsvendor model can be helpful in deriving the optimal OR time reservation given two conditions (Olivares et al. 2008, Cachon and Terwiesch 2012): the mean and variance of the underlying OR time distribution and the ratio of under-allocation and over-allocation costs for OR times. Despite this opportunity, the consideration of OR time variability in practice is implicit rather than explicit; previous health care studies have suggested that often, OR managers assign ad hoc safety time to avoid overtime costs (under-allocation costs) more than idle time costs (overtime costs) because they perceive overtime to be more costly (Strum et al. 1999, Venkataraman et al. 2018). However, most of the predictive research on OR time allocation has not explicitly modeled the newsvendor model to find the optimal safety time. Therefore, in our study, we apply the newsvendor model using the mean and variance estimates of our model and show that improved OR prediction methods can reduce the total OR costs and that the cost ratio between under- and over-allocation of OR times plays a significant role in total OR time mismatch costs.

Our study integrates these existing insights from practice, emerging OR scheduling literature, and OM literature and show that the use of diverse and granular scheduling inputs about the procedure and team significantly contributes to reducing scheduling mismatches and corresponding mismatch costs. In particular, we bridge the health care literature and OM literature by documenting the importance of more powerful OR time predictions in the OR context as newsvendor settings.[3]

1.2.3 Scheduling Outcomes

In this subsection, we review the commonly used scheduling outcome measures in OR practice. In the short term, OR scheduling policies including team selection could directly affect three major outcomes: OR costs, team productivity, and patient outcomes. First, OR studies have found that under- and over-allocations of surgical times have a direct impact on OR costs by affecting delays, idle times and overtimes, resource management, and OR utilization (Olivares et al. 2008,

[3]Indeed, our study shows that with improvements in granular OR time prediction techniques, OR managers can significantly outperform the existing moving-average and ad hoc–based OR time allocation policies when considering both the mean OR time and the total combined newsvendor OR time costs for overtime and idle time.

Eijkemans et al. 2010, Zhou et al. 2016, Bartek et al. 2019). Second, worker productivity studies suggest that OR schedules have a significant impact on worker performance (KC and Terwiesch 2009) and team performance overall (Reagans et al. 2005, Xiao et al. 2015). Third, OR studies discuss that OR schedules significantly affect patient outcomes such as patient flow time and length of stay (LOS) (KC and Terwiesch 2009, Xiao et al. 2015). In the long term, OR scheduling policies could directly affect a broad range of outcomes: long-term hospital profitability and costs, staff motivation and turnover, and patient satisfaction (Cardoen et al. 2010, May et al. 2011). Our findings together with the above literature's insights on OR outcomes imply that improvements in OR time allocation decisions can help hospitals manage both these short- and long-term outcomes.

1.3 Data Access and Overview

This section provides an overview of our data collection, cleaning, and organization methods. We collaborate with a major South Carolina hospital to access granular surgical data sets. Our study received Institutional Review Board–exempt approval for human subjects research, and we access our de-identified surgical database in an encrypted private data folder after removing identifiable patient and staff information. Specifically, our database contains five data sets that provide detailed raw information for 81,967 surgeries between March 2016 and June 2019: (i) a patient data set provides demographic, medical status, anesthesia type, and length of stay information for patients; (ii) a surgeon data set provides the records of medical education, training, and previous professional affiliations for 307 surgeons; (iii) a staff data set provides information on surgical team compositions and roles of the team members, including 1,259 unique individuals; (iv) a case data set provides information on detailed timestamps for patient flow (patient in, procedure start, procedure end, patient out, patient prep, and wrap-up), pre- and post-surgery turnover, and information on surgery dates, times, rooms, and the surgical specialty areas; and (v) a procedure data set provides information on planned and performed surgical procedures for each surgery—the surgical descriptions and procedure codes classified by the CPT of the American Medical Association.

In cleaning our raw data sources for analysis, we drop 3,887 cases with missing and inconsistent timestamps (potentially due to data collection and storage errors). Next, we reconstruct our cleaned raw data sources into three sample data sets with distinct granularity levels for our analysis: (i) a sample case-team data set consists of 78,080 observations, (ii) a sample case-agent data set consists of 615,993 observations, and (iii) a sample case-dyad data set consists of 4,490,696 bilateral observations. Within each sample data set, we categorize our sample variables into three groups: OR time outcome, pre-team-selection input, and post-team-selection input variables. Table 1.1 presents an overview of each raw data source and reconstructed sample data set. OR time outcome variables contain surgical durations, including patient, procedure, and turnover times; pre-team-selection variables contain patient, procedure, surgeon, and overall OR workload information; and post-team-selection variables contain back-to-back (B2B) case assignments, daily workloads, and previous surgical experience of individuals, pairs of individuals, and OR teams as a whole.

Figure 1.1 provides an overview of our timestamp variables and the constructed duration variables. We use six scheduled and realized timestamps to construct six duration variables to track the times of previous-patient out, patient in, procedure start, procedure end, patient out, and next-patient in. Using these timestamps, we generate surgical outcome durations, including pre-turnover (previous-patient out to current-patient in), preparation, procedure, wrap-up, total patient, and post-turnover (current-patient out to next-patient in) times. During each surgical stage, OR teams work together in pairs and sub-teams, and as a whole team, coordinating and communicating with each other in dyadic and team roles. Building on such diverse surgical interactions among OR team members, we formally construct our study variables in the following section.

1.4 Variable Construction

This section details our variable construction steps. Specifically, we construct three major sets of variable categories: (i) pre-team-selection data inputs for patients, procedures, overall OR workloads, and scheduled durations, (ii) post-team-selection data inputs for team characteristics (such as back-to-back case assignments, daily workloads, and previous surgical experiences), and (iii) OR outcome variables with respect to the scheduled and realized times. We differentiate the data elements that become available before and after team assignments since post-team-selection data inputs become available only after team assignments. To present our main variables, we use three sets of descriptive tables: (i) Table 1.2 presents the definitions and summary statics of our pre-team-selection continuous variables, and Table 1.3 shows the levels of our pre-team-selection categorical variables; (ii) Table 1.4, Table 1.5, Table 1.6, and Table 1.7 describe the variable construction and the variations for our post-team-selection variables; and (iii) Table 1.8 presents the definitions and summary statistics for our realized outcome variables, and Table 1.9 describes the variations in the mismatch-related outcome variables across our pre-team-selection categorical variable levels. In the following three subsections, we explain and discuss each data variable category in detail.

1.4.1 Pre-Team-Selection Inputs

There are six categories of pre-team-selection data input variables in our study: scheduled duration, patient, surgeon, OR workload, non-textual procedure information, and textual procedure information variables. Table 1.2 displays continuous and binary pre-team-selection variables and the summary statistics. Table 1.3 displays the major categorical pre-team-selection variables. For the first category, we construct scheduled durations for turnover, patient, and procedure times to capture the effect of schedule characteristics using our detailed timestamps for planned patient flow. The second category, patient characteristics, includes one continuous variable, patient age, and five categorical varaiables—anesthesia type, American Society of Anesthesiology (ASA) severity score, race, gender, and class. The third category, surgeon characteristics, includes two continuous variables: surgeon's tenure in years since residency and the number of previous medical affiliations. The fourth category, OR workload characteristics, includes multiple categorical and continuous variables—room-level, specialty-level, and overall OR suite workload (the scheduled workloads in hours) variables and additional variables that can affect resource utilization in ORs. For room-level workloads, we use daily scheduled room workload, surgery order, idle pre-turnover (binary variable indicating turnover time greater than one hour before the surgery), idle post-turnover (binary variable indicating turnover time greater than one hour after the surgery), and previous add-on case indicator (binary variable indicating cases after an add-on case); for daily specialty-level workloads, we use daily scheduled workload in the given specialty; and for overall OR suite workloads, we use daily scheduled OR suite workload, daily proportion of add-on cases in the OR, month and day fixed effects, and holiday indicators.

The fifth category, non-textual procedure characteristics, includes eight continuous and binary variables: the count of CPT procedures, the CPT procedure text length, the count of co-surgeons[4], robotic surgery and add-on case indicators, OR team size, and overall specialty area experience of the hospital. We capture the last category, textual procedure information, using a

[4]We use the term 'co-surgeon' to refer to a surgeon who is not a primary surgeon but performs part of the procedures in the surgery.

dimensionality reduction approach, principal component analysis (PCA). Since using 1,518 unique CPT codes as fixed effects in our analyses can lead to the curse of dimensionality (Chen 2009) and overfitting (Hawkins 2004) problems, we construct a PCA that makes use of information from the CPT descriptions. Specifically, after casting the unique words from the CPT descriptions into dummy variables, we identify principal components of these dummy variables—see Jolliffe and Cadima (2016), Jiang et al. (2021), Choi et al. (2022) for recent PCA applications. Using Kaiser's criterion, we retain the top 94 PCs whose eigenvalues are greater than one (Kaiser 1960, Ferré 1995, Braeken and Van Assen 2017)—Figure 1.2 visualizes the relationship between the number of principal components (PCs) and the cumulative proportion of sample variability explained by PCs. In the following section, we expand on our pre-team-selection data inputs and construct empirical measures for post-team-selection data inputs.

1.4.2 Post-Team-Selection Inputs

In this section, we describe how we construct our post-team-selection data input variables. The post-team-selection data input variables we measure across three time spans (back-to-back surgery-level, daily, six-month history) are back-to-back case assignment indicator (B2B), daily workload, and 6-month historical surgical experience. We measure each of the post-team-selection variables at three measurement levels: individual, team, and dyad. This construction of variables captures the rich dynamics of team characteristics. The individual-level measurement captures the variations in each data input variable across individuals in OR teams—for example, a scrub nurse might have more experience than another scrub nurse. The team-level measurement captures the variations in each data input variable across OR teams as a whole—for example, one OR team might have more collective team familiarity than another OR team (Reagans et al. 2005, KC and Terwiesch 2009, Stepaniak et al. 2010). To measure team-level variables, we build on Reagans et al.'s (2005) and Avgerinos and Gokpinar's (2017) seminal studies, which use an average of all dyadic pairs (e.g., nurse-surgeon, anesthesiologist-surgeon, etc.) for each team characteristic. Our study extends this prior research examining team characteristics to include daily workload and B2B variables.

Lastly, the dyad-level measurement allows us to capture the variations in each data input variable across distinct dyads in OR teams—for example, a nurse-surgeon dyad might have a different familiarity when compared to an anesthesiologist-surgeon dyad, resulting in different productivity outcomes. This dyadic measurement expands the team-level measurement by capturing more granular team member interactions, contributing to the literature by exploring rich dyadic OR team member relationships. While in this stream of literature, Avgerinos and Gokpinar (2017) explore the within-team dispersion of team familiarity at the team level, our study explicitly specifies dyadic similarities and differences: for each dyadic pair (e.g., a given nurse-surgeon pair), we capture (i) the shared experiences (i.e., dyad-Λ variables) via their joint historical interactions and (ii) the non-shared experience differences (i.e., dyad-Δ variables) via the historical experiences that are separately performed. For example, for a surgeon-nurse pair where the focal agent is the surgeon who performed a total of 120 bariatric surgeries, of which 20 surgeries were performed jointly with the dyad partner (nurse) and 100 without the partner, the shared experience (Λ) is 20. If the nurse performed 60 other bariatric surgeries without that surgeon, the dyad difference experience (Δ) would be the difference in total experiences of these two individuals within the dyad (120 for the surgeon and 80 for the nurse), which equals 40.

This approach allows us to capture more variations across dyads than previous research that aggregated each team variable to the team level (e.g., Reagans et al. 2005). Table 1.4 shows the diversity of the dyad shared (Λ) and difference (Δ) experiences that we capture. For example, on average, a focal agent and their dyad partner in a unique role-pair have different individual specialty experience levels or share different levels of previous collective experiences as compared to other pairs of the team on average, displaying unique and shared experience levels across role-pairs. Furthermore, there is a large diversity in shared experience even within one dyadic role-pair. For example, Figure 1.3 visualizes the dyadic surgery network for the two top-performing surgeons in our data and the CRNAs they work with. It shows that an individual in an OR role has differentiated interactions with various individuals in another role even within the same dyadic role-pair—for instance, while the first surgeon maintains much less frequent interactions with a larger number of CRNAs, the second surgeon maintains more frequent interactions with fewer CRNAs.

We construct three post-team-selection data inputs—B2B case assignments, scheduled workloads, and past experiences—for each of these three variable measurement levels. Table 1.5 gives an overview of our post-team-selection variable categories at each measurement level: dyad (shared, Λ; non-shared, Δ), individual, or team as a whole. In particular, for B2B, we create a binary variable to indicate whether the individual or dyad worked back-to-back within a 30-minute time window from the previous surgery before a given surgery. We consider five scenarios where the individual or dyad could work back-to-back: (i) B2B assignments with the same setup (the same room and same specialty), (ii) B2B assignments with only the same room (with a different specialty), (iii) B2B assignments with only the same specialty (with a different room), (iv) B2B assignments with a different setup (a different room and different specialty), and (v) non-B2B or idle pre-turnover cases where there is no previous surgery.

For workloads for individuals and dyads, we sum the scheduled durations of the surgeries for each individual or dyad for the day. In particular, we observe that an average agent is scheduled for 8.26 hours during a daily shift and that an average team would have dyads who share 3.86 hours of joint-scheduled surgery workload. We measure the last variable, experience, by a rolling total of surgeries over the last six months before the focal surgery for each individual or dyad. For both individuals and dyads, we capture four different types of experience: (i) specialty: experiences in same-specialty surgeries, (ii) non-specialty different-specialty surgeries, (ii) room: same-room surgeries, and (iv) people: surgeries with the rest of the assigned OR team members[5]. For all three post-team-selection variables (B2B, workload, and experience), we measure the team-level post-team-selection variable following Reagans et al. (2005), as stated earlier. In our data, the statistics of these post-team-selection experiences indicate that individual-level and dyad-level post-team-selection input variables exhibit larger variations than team-level variables, which motivates us to explore the effect of granular dyadic inputs on OR time prediction.

Furthermore, these post-team-selection variables show large variations across distinct dyadic role-pairs and specialty areas. Specifically, first, Table 1.6 displays the means for all shared (Λ) experience variables by role-pair; high mean value for each variable implies that the role-pair shares a high level of dyadic interactions on average. Next, Table 1.7 compares the variations of our specialty experience variables at the dyad (shared and difference), agent, and team level across different surgical specialty areas in terms of coefficient of variation (COV). The table illustrates that the relative variation of specialty experience within the respective specialty varies across different specialty areas. Especially, the COV of our dyad shared specialty experience variable shows more variations across specialties. Overall, large variations in our post-team-selection variables illustrate the importance of capturing the variations in OR time predictions. In our empirical strategy, we capture these variations using the granular team inputs we specified in this section. In the following section, we introduce our outcome variables, show the descriptive statistics associated with different

[5]For individuals, the fourth type of experience would be the average number of surgeries an individual performed with the rest of the team members excluding the dyadic partner; for dyads, the average number of surgeries a dyad performed as a pair with the rest of the team members.

categorical variable levels, and discuss how our outcome variables vary across distinct specialty areas.

1.4.3 OR Time Outcome Variables

We create our three types of major OR outcomes variables: OR time duration, mismatch, and patient length of stay (LOS) variables. First, using surgical timestamps, we construct pre-turnover, prep, procedure, wrap-up, patient, and post-turnover durations. Second, using the scheduled and realized time differences, we compute starting and ending delays (i.e., when the realized starting or ending patient-in time is later than the scheduled time), under-allocations (i.e., procedure and patient overtimes when realized duration exceeds the scheduled duration), and over-allocations (i.e., procedure and patient idle times when the scheduled duration exceeds the realized duration). Furthermore, we combine over-allocations and under-allocations to calculate the absolute mismatch measures. Lastly, we capture the patient length of stay (LOS) from our patient data set as an extended OR outcome measure for post-surgery patient conditions. Table 1.8 displays summary statistics and descriptions of the major OR outcome variables (excluding the absolute mismatch measure, which is implied by the sum of overtime and idle time), specialty, room, role-pair, and CPT code used in our analysis. From our mismatch variables, we observe significant differences between scheduled and realized outcomes: on average, a surgery would experience 0.27 hours of starting delay and 0.51 hours of ending delay, 0.55 hours of absolute mismatch in procedure time, and 0.56 absolute hours of absolute mismatch in patient time. Also, on average, we observe a higher amount of overtime than idle time for procedure time and a higher amount of idle time than overtime for patient time. Next, Table 1.9 shows our OR mismatch variable variations across levels within our pre-team-selection categorical variables. We observe significant outcome differences among these levels within each categorical variable, motivating us to incorporate these categorical pre-team-selection inputs in our analysis as contextual OR time drivers.

Lastly, before we proceed to our empirical strategy, we build upon our variable construction and discuss how our pre-team-selection variables, post-team-selection variables, and outcome variables vary across distinct specialty areas grouped by high and low process variability (PV)[6].

[6]We measure the process variability (PV) by the within-procedure variability of procedures averaged within each specialty. Specifically, we compute the standard deviation of procedure times within each procedure (i.e., within-procedure variability) and get the weighted average of within-procedure variability under each specialty—with the weight as the number of procedure observations over total observations within the specialty. Then, we define this weighted average as process variability of a specialty. Using this measure, a specialty with a PV value higher than the mean PV is classified as a high PV group, and vice versa.

Table 1.10 compares the mean differences in selective pre-team-selection, post-team-selection, and realized-duration variables—across distinct surgical specialty areas that are grouped by high and low PV. The table indicates that low PV specialties, on average, have a higher number of procedures in one case; involve older patients; and have longer procedure time, patient time, and LOS. Furthermore, the table shows significant diversity in each variable from three major sets of our variable categories across specialty areas, implying that specialty area is an important data input for our analysis. The following section describes our empirical strategy for taking advantage of the variations observed in our constructed variables.

1.5 Empirical Strategy

In this section, we construct our empirical estimation strategy for OR times and mismatches. Specifically, we construct a two-stage estimation framework for OR times that detaches and isolates the impact of OR teams separately from other OR time drivers. Our structural OR time model has two components: (i) the ex ante OR time that captures the pre-team-selection data inputs (such as procedure, patient, and overall workload effects) unconditional on team selections; and (ii) the ex post OR time adjustment that captures the post-team-selection data inputs (such as team experience, workloads, and back-to-back case assignment variables) conditional on team selections. We build our ex post OR time adjustment specification under two granularity levels: aggregate team level and role-pair dyadic level; the dyadic ex post model captures granular role-pair relationships between surgical team members.

In addition to estimating the mean levels of OR times, we also estimate the OR time variance (i.e., the squared difference between the predicted mean level and realized OR time) for each case in our sample data. Then, using our mean and variance estimates for OR times, we use the classical newsvendor model to calculate the optimal OR time reservations under different scenarios for overtime and idle time costs (i.e., under- and over-allocation costs, respectively). Hence, our empirical strategy enables us to perform a comprehensive cost comparison analysis of different OR time prediction methods: specifically, the current policy of our partner hospital, our ex ante OR time predictions that exclude team effects, and our ex post OR time predictions that include either aggregate (uniform) team effects or more granular (non-uniform) dyadic team effects between distinct OR role-pairs.

We further use our first-stage ex ante OR time estimates as benchmarks to compare against our partner hospital's scheduled OR time allocations observed in our database. Since our partner hospital assigns OR times before team selections, we define the ex ante scheduling mismatch variable between the predicted and scheduled OR times before team selections—i.e., between our ex ante OR time estimates and the hospital's OR time allocations. Then, in our second-stage team performance model, we use these ex ante scheduling mismatch estimates and explore the impact of scheduling mismatches on team performance. Also, we further control for task diversity between the assigned ex ante OR times and expected ex ante OR times based on team characteristics. Accordingly, our task diversity measure helps us update the OR time prediction after team assignment by taking the historical information about scheduling effects on teams—a team's response to the assigned time that deviates from expected task time conditional on team characteristics. In the following subsections, we formally introduce our micro-econometric OR time models.

1.5.1 Ex Ante OR Time

In this subsection, we build our ex ante OR time model before team selections. Our ex ante OR time model relies mainly on pre-team-selection data inputs: patient, surgeon, overall OR workload, and procedure information. The residual between our ex ante OR time estimates and realized OR times motivates our ex post OR time estimations for team performance. More formally, we decompose the realized OR time into two structural components:

$$\underbrace{T_{sdt}^{Realized}}_{\text{Realized OR Time}} = \underbrace{\alpha_0 + \boldsymbol{\alpha}_1 \cdot \boldsymbol{I}_{sd}^{Pre-team}}_{\text{Ex Ante OR Time}} + \underbrace{\Delta T_{sdt}^{ExPost}}_{\text{Ex Post OR Time Residual}}, \tag{1.1}$$

where s denotes a specific surgery, d denotes a specific surgery day, t denotes a specific surgery team, $T_{sdt}^{Realized}$ is the realized surgery time for surgery s, $\boldsymbol{I}_{sd}^{Pre-team}$ denotes the pre-team-selection data inputs before team selections, $\boldsymbol{\alpha}$ is the ex ante OR time parameter vector, and ΔT_{sdt}^{expost} is the ex ante OR time residual that represents OR time adjustment that depends on team effects. Without loss of generality, we denote the predicted ex ante OR time as follows:

$$\underbrace{\widehat{T}_{sd}^{ExAnte}}_{\text{Predicted Ex Ante OR Time}} = \underbrace{\widehat{\alpha}_0 + \widehat{\boldsymbol{\alpha}}_1 \cdot \boldsymbol{I}_{sd}^{Pre-team}}_{\text{Ex Ante OR Time Parameter Estimates}}, \tag{1.2}$$

where the $\widehat{\cdot}$ operator represents the fitted values from our estimations. In our estimations, we use three alternatives for the main OR outcomes of interest—patient, procedure, and post-turnover times—that capture distinct substages of OR operations. In addition, we log-scale our OR time variables to increase the prediction power of our empirical models (May et al. 2011).

Also, since our partner hospital assigns OR times before team selections, we use our predicted ex ante OR times to control for scheduling mismatches that stem from omitting team effects. We hypothesize that these ex ante scheduling mismatches could influence team performance by affecting teams' behavior toward the task completion—more stringent or loose schedules could influence surgical team performance through their motivations and expectations (Jung et al. 2010), and volatile schedules can have adverse effects on productivity due to psychological distress (Allen et al. 2013, Schneider and Harknett 2019). Therefore, we control for pre-team-selection scheduling time mismatches in our second-stage team performance model and compute the scheduling mismatch by:

$$\underbrace{S_{sd}^{Mismatch}}_{\text{Ex Ante Scheduling Time Mismatch}} = \underbrace{T_{sd}^{Scheduled}}_{\text{Originally Scheduled OR Time}} - \underbrace{\widehat{T}_{sd}^{ExAnte}}_{\text{Predicted Ex Ante OR Time}}, \tag{1.3}$$

where $S_{sd}^{Mismatch}$ is the ex ante OR time scheduling mismatch—the deviation between the scheduled OR times and our predicted ex ante OR times, $T_{sd}^{Scheduled}$, before team selections vis-à-vis our data-driven predictions.

1.5.2 Ex Post Team Performance

This subsection constructs our ex post team performance model after team choices, building on our first-stage OR time predictions and mismatch calculations. In particular, we capture how a given OR team performs on a given surgery relative to the ex ante OR time expectations (unconditional on team formations). More formally, our ex post team performance model is given as follows:

$$\underbrace{\Delta T_{sdt}^{ExPost}}_{\text{Ex Post OR Time Performance}} = \beta_0 + \underbrace{\boldsymbol{\beta}_1 \cdot \boldsymbol{S}_{sdt}^{Mismatch}}_{\text{Scheduling Mismatch Effects}} + \underbrace{\boldsymbol{\beta}_2 \cdot \boldsymbol{I}_{sdt}^{Post-team}}_{\text{Ex Post-Team-Selection Inputs}} + \nu_{sdt}, \tag{1.4}$$

where $\boldsymbol{\beta}$ is a vector of team performance variables, and $\boldsymbol{S}_{sdt}^{Mismatch}$ is a vector of scheduling mismatch variables: ex ante over-allocation (idle time) and ex ante under-allocation (overtime). In equation 3.2, post-team-selection inputs $\boldsymbol{I}_{sdt}^{Post-team}$, contain our team characteristics variables that relate to dyadic role-pairs in terms of back-to-back, workload, and experiences variables. Furthermore, we run equation 3.2 under two granularity levels for post-team-selection inputs: (i) incorporating team-level average characteristic effects and (ii) incorporating dyad-level distinct role-pair characteristic effects. Our granular dyadic specifications consider non-uniform relationships between team members and expand the literature on team familiarity (Reagans et al. 2005, Avgerinos and Gokpinar 2017) that studies team aggregated dyadic relationship variables for team variables. Also, we add interaction variables between role-pair and post-team-selection inputs inside the $\boldsymbol{I}_{sdt}^{Post-team}$ vector for the dyadic team performance model and drop potentially multi-colinear and rare variables from our analysis.

In addition, we consider another variable that captures the effect of scheduling on surgical team performance: task diversity. Some hospitals make ad hoc decision to adjust OR time based on team characteristics, depending on the staffing manager's informal knowledge about team characteristics. As a result, there may be some selection of team members for the given task, but the decision is not perfect because the manager makes the decision having incomplete team characteristics knowledge and because team members might be assigned to more diverse surgeries due to training requirements, employee shortages, staff turnovers, or add-on (just-in-time scheduled) cases. Thus, there still exists task diversity for similar teams in surgery assignments. To measure this task diversity, we frame the OR time duration expectations for a given OR team as

$$\underbrace{\widehat{T}_{sd}^{ExAnte}}_{\text{Predicted Ex Ante OR Time}} = \underbrace{\kappa_0 + \boldsymbol{\kappa}_1 \cdot \boldsymbol{I}_{sdt}^{Post-team}}_{\text{Team Characteristic Variables}} + \underbrace{S_{sdt}^{Diversity}}_{\text{Task Diversity Residual}}, \tag{1.5}$$

where $\boldsymbol{\kappa}$ is a vector of ex post parameters that relate to task expectations and $S_{sdt}^{Diversity}$ is the task diversity residual. $S_{sdt}^{Diversity}$ measures the deviation from prior expectations of OR teams about the task completion time (i.e., OR time) given their team member characteristics. Using fitted parameter estimates of equation 1.5, we compute the ex post task diversity values—$\widehat{S}_{sdt}^{Diversity}$. We hypothesize that task diversity will affect team performance as a result of responses of team members to the unfulfilled expectation. Hence, we control for the OR task diversity by adding the variable to our ex post team performance model—equation 3.2. As a result, our ex post team performance model captures the scheduling effect of task diversity and helps us update the OR time model after team selection.

Building on our estimation steps, we generate four OR time benchmark measures and compare their ability to predict realized OR times—i.e., $T_{sdt}^{Realized}$:

1. Scheduled OR time: $T_{sd}^{Scheduled}$—default OR time in our surgery database for our partner hospital that relies mainly on moving-average and ad hoc OR time allocation policies.

2. Ex ante predicted OR time: $\widehat{T}_{sd}^{ExAnte}$—data-driven ex ante OR time before team selections. Our ex ante OR time predictions indicate average task completion times independent of team effect.

3. Ex post predicted OR time at the team level, $\widehat{T}_{sd}^{ExAnte} + \Delta\widehat{T}_{sdt}^{ExPost}$(Team level)—data-driven ex post OR time updates OR times after team selections, assuming uniform dyadic interaction effects. Our team-level specifications recalibrate expected OR times based on selected teams.

4. Ex post predicted OR time at the dyad level, $\widehat{T}_{sd}^{ExAnte} + \Delta\widehat{T}_{sdt}^{ExPost}$(Dyad level)—data-driven ex post OR time updates the OR time after team selections, assuming non-uniform dyadic interaction effects. Our dyad-level specifications recalibrate expected OR times based on selected teams and non-uniform role-pair relationships.

1.5.3 Newsvendor Model

While sections 1.5.1 and 1.5.2 consider only the mean values of OR times, in practice, surgical OR processes exhibit significant variability in time around the predicted mean time. Because of the variability, it is common for hospitals to assign ad hoc OR time to avoid excessive amounts of costly overtime (Lehtonen et al. 2013, Venkataraman et al. 2018), but the use of a statistical model such as the newsvendor model is rare. To account for this variability in OR time using the newsvendor model, we develop a variance estimation framework using the following equation:

$$\underbrace{|T_{sdt}^{Realized} - \widehat{T}_{sd}^{ExAnte}|^2}_{\text{Squared OR Time Residuals}} = \underbrace{\gamma_0 + \boldsymbol{\gamma}_1 \cdot \boldsymbol{I}_{sd}^{Pre-team}}_{\text{Ex Ante OR Time Variance}} + \epsilon_{sdt}, \tag{1.6}$$

where $\boldsymbol{\gamma}$ is the ex ante OR time variance parameter vector. We replicate equation 1.6 for each mean OR time benchmark estimate mentioned in §1.5.2. Next, using both the mean and variance estimates of our OR times, we model the optimal OR time reservation according to the newsvendor model (Olivares et al. 2008, Cachon and Terwiesch 2012):

$$\underbrace{T_{sd}^{Newsvendor}}_{\text{Newsvendor OR Time Reservation}} = F^{-1}\left(\frac{C_u}{C_u + C_o}\right) \tag{1.7}$$

$$= \underbrace{\widehat{T}_{sd}^{ExAnte} + \widehat{T}_{sd}^{ExAnte}(\sigma) * Z^{-1}\left(\frac{C_u}{C_u + C_o}\right)}_{\text{Newsvendor Formula Given the Estimated OR Time Distribution and Cost Parameters}}, \tag{1.8}$$

where $T_{sdt}^{Newsvendor}$ is the optimal newsvendor OR time reservation, $\widehat{T}_{sd}^{ExAnte}(\sigma)$ is the predicted OR time standard deviation according to equation 1.6, $F^{-1}(.)$ is the z-score spread coming from the inverse standard normal distribution, C_u is the under-allocation cost parameter for overtime, and C_o is the over-allocation cost parameter for idle time.

Following the previous evidence that OR managers perceive overtime as more costly than idle time (Strum et al. 1999, Olivares et al. 2008, Venkataraman et al. 2018), we estimate our newsvendor model under three different cost scenarios: when overtime and idle time cost ratios are equal to one, two, and three, respectively. Hence, we explore newsvendor environments where under-allocation (overtime) costs are at least as much as over-allocation (idle time) costs. Furthermore, following Childers and Maggard-Gibbons's (2018) OR time cost estimate, we calibrate the over-allocation cost at \$14.4 per minute—40% of the per-minute total OR cost spent for indirect costs. Hence, our newsvendor approach explores the impact of improved OR time predictions from a risk mitigation perspective against the uncertainty. By computing the optimal safety time using our data-driven OR time predictions and newsvendor model under different cost scenarios, we compare the hospital's ad hoc safety capacity cost performance to the cost performance across our data-driven OR time prediction models. Building on our estimation framework, we show our empirical findings and results in the next section.

1.6 Results

This section presents our main empirical findings and results in four steps: (i) our ex ante OR time analysis using pre-team-selection data inputs, (ii) our ex post team performance analysis using our post-team-selection data inputs, (iii) our main results from our empirical newsvendor models, and (iv) more detailed OR time mismatch comparisons between scheduled, realized, and data-driven OR times. In the following subsections, we sequentially introduce our estimation results.

1.6.1 Ex Ante OR Time

This subsection presents our ex ante OR time estimation results for pre-team-selection data inputs—patient, procedure, surgeon, and overall OR workload variables. We consider three major OR times for our ex ante OR time estimations: patient, procedure, and post-turnover times. Specifically, Table 1.11 reports our ex ante OR time estimates with two specifications: with and without PCA for each OR time model. From the results, we find that pre-team-selection data inputs, together with PCA, explain 73% of patient time variations, 68% of procedure time variations, and 40% of post-surgery turnover time variations. Especially, including principal components (PCs) about granular procedure information has a substantial impact in explaining ex ante patient and procedure times. Hence, we find that surgery time expectations depend significantly on our pre-team-selection data inputs: for example, the number of procedures, team sizes, and patient age significantly increase ex ante OR times. In the following section, we use the residual estimates of our first-stage analysis alongside scheduling mismatch and task diversity measures (that we construct using our first-stage analysis) to estimate the team performance factored into ex post OR times.

1.6.2 Ex Post Team Performance

In this subsection, we present four tables that document the impact of team variables on OR times: Table 1.12 shows our team-level ex post model, Table 1.13 shows our dyad-level ex post model for patient time, Table 1.14 shows our dyad-level ex post model for alternative dependent variables, and Table 1.15 supplements our main analysis by showing our ex post task diversity estimation results. First, Table 1.12 presents our ex post team-level estimates for patient, procedure, and post-turnover times in high and low process variability (PV) specialty groups. Our findings indicate that average team characteristics have a substantial impact on all duration variables: on average, team variables explain around 31% of the remaining patient time residual variations (from the first stage of our estimations), 24% of the remaining procedure time residual variations, and 24% of the remaining post-turnover time residual variations. Furthermore, we find that a large fraction of our back-to-back, workload, and experience variables are significant in predicting OR times. We also find that in all team-level ex post models, the effects of ex ante scheduling mismatches and task diversity on team performance are evident. This means that capturing these behavioral effects using the stage 1 ex ante OR time to update the predictions in the stage 2 ex post OR time indeed contributes to more accurate predictions of OR times. Overall, our post-team-selection data inputs explain the OR outcomes slightly more significantly for the low PV group than for the high PV group.

Second, we run a dyad-level model to explore the impact of dyadic team effects on ex post OR time. In our dyadic approach, we increase the granularity of team interaction variables from the team aggregates to dyadic specifications and consider non-uniform effects among distinct role-pair dyadic relationships. Table 1.13 presents our ex post dyad-level OR time estimation results across different specialty areas. We find that our dyadic analysis substantially improves from the team-level model: in most specialties (25 of 27), the goodness of fit of the dyadic models—in terms of R-squared—is significantly greater than the respective value of the team level models, implying that the model captures a significant amount of outcome variation by unique dyadic relationships. The R-squared also tends to be higher for smaller niche specialties that are less frequent and have less variety in their surgical operation domains. Also, we find most of our main results (on team and mismatch variable categories) consistent between dyadic and team models, while there are some differences across specialty areas.

Third, Table 1.14 presents the dyadic estimation results for our alternative outcome variables—procedure time, post-turnover time, and length of stay (LOS)—with weighted average coefficients across surgical specialty areas. Consistent with the patient time models, we observe improvement in the R-squared for ex post times. The improvement is more significant for post-turnover time prediction, implying that taking granular team information may be more beneficial to explain the coordination of support activities in turnover. And lastly, Table 1.15 displays our task diversity estimates between expected and assigned OR times conditional on team characteristics: we find that while the predicted ex ante OR time significantly correlates with team variables with around a 29% R-squared in patient and procedure times, there are still substantial variations in how diverse tasks are assigned to similar teams—vice versa, tasks with similar characteristics are assigned to diverse teams. Hence, we control for this task diversity by our ex post OR time predictions. In the next subsection, we build on our estimation results to compute the overall newsvendor costs and evaluate the impact of different predictive methods in our study on the newsvendor costs.

1.6.3 Newsvendor Model

This subsection compares the overall cost impact of distinct predictive methods in our study using the seminal newsvendor model (Olivares et al. 2008). Specifically, we compare the current scheduling policy of our hospital to our ex ante and ex post OR time predictions in terms of the total newsvendor cost.[7] Based on the observations of Olivares et al. (2008) and Venkataraman et al. (2018) that hospital managers perceive under-allocation cost (C_u, overtime cost) as more expensive than over-allocation cost (C_o, idle time cost), we consider three cost scenarios that sequentially increase aversion to overtime costs, where under-allocation cost (i) equals over-allocation cost, (ii) is two times over-allocation cost, and (iii) is three times over-allocation cost. We present our main results of the empirical newsvendor model application in Table 1.16 and Table 1.17.

First, from Table 1.16, we observe that our dyadic ex post model significantly outperforms the current policy and our other predictive methods with respect to the total mismatch costs. Specifically, we observe that using our dyadic approach, hospital managers can reduce the total combined mismatch costs by 20% in the most conservative scenario ($C_u = C_o$) and 38% in the least conservative scenario ($C_u = 3 \times C_o$). Second, we observe that the impact of our predictive methods on cost reduction is more significant for more costly under-allocation environments. This result implies that our predictive methods gain more importance when there is more variability in OR times—when OR times deviate from the mean OR times estimates more often—because our newsvendor model can better capture the optimal safety capacity of OR times to avoid costly over-allocation cost. Third, we observe that even our ex ante OR time model (which excludes all post-team-selection inputs) outperforms the current scheduling policy of our partner hospital in each scenario although similar in the most conservative scenario. Lastly, from Table 1.17, we find that some specialties (e.g., plastic and otolaryngology, colorectal) require more granular team inputs (dyadic specifications) to be effective in reducing the newsvendor cost.

Altogether, our findings imply that our partner hospital can improve OR costs in two major ways: first, the hospital can use more team data inputs in its OR time allocation, and second, even without team data inputs, the hospital can use the newsvendor model to account for the OR time variance to balance its under- and over-allocation costs using the newsvendor optimal safety

[7] We use the actual OR time mismatches between scheduled and realized outcomes as the underlying benchmark. We contrast the actual OR time mismatches with the mismatches from our three prediction models: (i) ex ante OR times that integrate only the pre-team-selection data inputs, (ii) team-level ex post OR times that also integrate aggregate post-team-selection data inputs, and (iii) dyad-level ex post OR times that integrate (further) granular post-team-selection data inputs.

capacity decision. Furthermore, the hospital can use our results to identify the specialties in which the newsvendor model will be the most effective.

1.6.4 Mismatch Comparisons

In this subsection, we make more detailed comparisons of OR time mismatches across different prediction models in our study with four exhibits that thoroughly compare these mismatch estimates. First, Table 1.18 presents an overview of the actual and dyadic ex post mismatch levels with respect to patient times—the dyadic ex post model achieves around 32% mismatch reductions in the absolute mean and 38% reduction in the standard deviation compared to the current policy. Especially we observe that the likelihood of costly under-allocation drops significantly from the current policy to our dyadic ex post model. Second, Figure 1.4 plots the distributions of current, ex ante, team- and dyad-level ex post, and realized patient times. The plot illustrates sequential improvements from the current policy to dyad ex post prediction to predict OR times (with regard to realized OR times) as data granularity increases with pre-team- and post-team-selection data inputs. Third, Figure 1.5 displays the mean squared error (MSE) in logged patient times from the current policy and the three OR time prediction models of our estimation framework. We find that the current policy results in an MSE twice as high as the MSE of our dyad-level ex post prediction and that the reduction in MSE is significant for each step from the current policy to dyad ex post time for both high and low process variability specialty groups. Lastly, Figure 1.6 visualizes the comparison between the actual and ex post dyad-level mismatch levels with respect to patient times across specialty areas. All specialties are located below the 45-degree line—where dyad ex post scheduling mismatch equals current policy scheduling mismatch—indicating that most specialty areas could benefit from data-driven OR time allocation policies. Also, as they are further away from the 45-degree line, they might gain more benefits from our predictive models. In the following section, we conclude our paper with managerial remarks reflecting on our main results.

1.7 Conclusion

The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 enables hospitals to collect large-scale electronic health records (EHR) data—in 2017, 95% of U.S. hospitals had adopted EHR data generation (Parasrampuria 2019). The strong trend for health care data collection persists globally: in 2013, preliminary global health care data estimates were around 153 exabytes and were expected to grow more than tenfold by 2020 (Stewart 2020). At the same time, analytical and theoretical developments in health care research highlight a large number of data impact areas for hospitals, including care efficiency and patient monitoring (Mišić and Perakis 2020), and indeed, a 2013 McKinsey report estimates that big data revolutions in health care could save around $450 billion annually—17% of U.S. healthcare costs (Groves et al. 2013).

Despite its potential to improve hospitals' decision making processes, large portions of available health care data are not used. In fact, many hospitals use only a limited set of EHR data elements available for OR time allocation. Specifically, even though there is a substantial literature in numerous health care and OM fields that document the importance of team effects in OR environments, many major hospitals assign OR times before team selection. Consequently, hospitals use only a limited portion of team characteristics data in OR scheduling decisions. On top of data limitations, hospital managers also face their limitation in using appropriate theoretical decision-making policies in practice: a case in point is that newsvendor models capturing both the mean and variability of OR times are rarely used in practice.

To address the policy and data drivers that influence OR time mismatches between scheduled and realized OR times, we partner with a leading South Carolina hospital and construct a structural two-stage framework that decomposes OR times into two components: (i) OR time expectations before team selections (unconditional on surgery teams), and (ii) changes in OR time expectations after team selections (conditional on surgery team characteristics). Then, we use our OR time predictions from the two-stage framework to compute the newsvendor OR time reservation and costs, which can mitigate the total combined overtime and idle time costs of the hospital.

Our findings indicate that OR time mismatches significantly drop when we update OR time expectations after team selections and use granular data elements about procedure and teams in the model. Specifically, these findings indicate that using granular EHR data elements on procedure and team characteristics together can reduce more than 50% MSE of the logged OR times compared to the current policy. While we find that the significant effects of these team variables in explaining OR times differ by dependent variable and specialty, addressing the exact mechanisms of team interaction effects is a promising avenue for future studies. Also, our empirical newsvendor applications show that improved OR time predictions mitigate the impact of OR time variability and reduce the likelihood of costly under-allocation much more effectively by assigning more buffer time than the current policy assigns to cases with high OR time variability. This, in turn, substantially reduces overall mismatch costs under various cost-structure scenarios by balancing the costs of over-allocation and under-allocation (idle time and overtime costs).

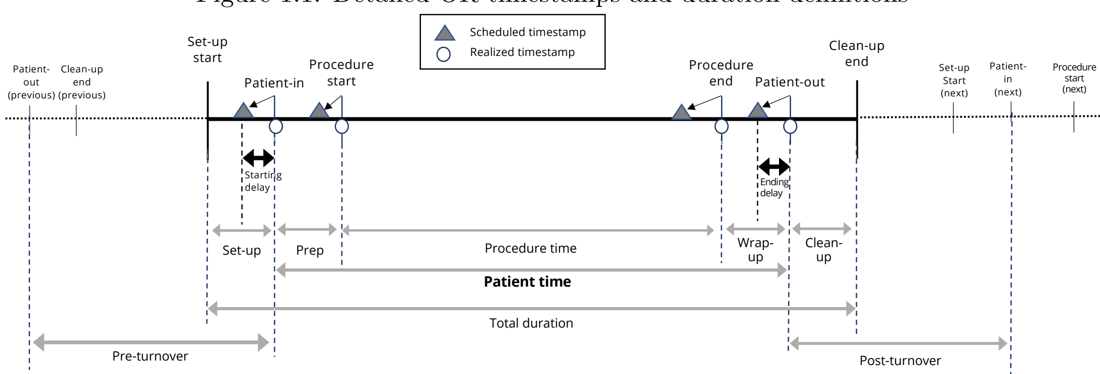
We conclude our paper with three practical considerations for health care leaders. First, hospital managers can check the correlation between their scheduling mismatches and team characteristics to see if they can use team inputs or improve the way they use any team input in their OR time predictions. If the correlation is significant, hospitals have an opportunity to increase their data capabilities behind OR time decision-making by referring to this study's data collection efforts and by considering the data inputs used in this study. Second, OR managers can also use the newsvendor model to consider OR time variability—estimated by our predictive models—to improve their safety capacity decision-making in balancing over-allocation and under-allocation costs; especially, the newsvendor model can build a powerful safety net when overtime cost significantly exceeds idle time cost. And lastly, hospitals can try to assess how OR time mismatches impact their OR teams due to the behavioral impacts of excessive overtime and idle time. In our study, we find that OR time mismatches affect team performance—and knowing these behavioral effects after team selection and updating the eventual OR time reservation can further help reduce OR time mismatches. We believe that the potential uses of our novel framework and insights are broad and can be extended by future research in various related fields, such as team management, forecasting, and capacity planning.

Table 1.1: Raw data sources and reconstructed sample data sets

				Post-team-selection variables			Pre-team-selection variables				
		Observations	Durations	B2B	Workload	Experience	Patient	Procedure	Surgeon	Workload	Description
Raw	Surgeon	307	–	–	–	✓	–	–	✓	–	Surgeon-level education and training history
Data	Patient	79,971	✓	–	–	–	✓	–	–	–	Patient conditions and information for each case
Source	Case	81,967	✓	✓	✓	✓	✓	✓	–	✓	Surgical case information and timestamps
	Procedures	155,418	–	–	–	–	–	✓	–	–	Detailed procedure codes and descriptions
	Staff	790,165	–	✓	✓	✓	–	–	✓	✓	Staff identifiers associated with each case
Reconstructed	Cases	78,080	✓	✓	✓	✓	✓	✓	✓	✓	Combined study variables at the case level
Data Set by	Agents	615,993	✓	✓	✓	✓	✓	✓	✓	✓	Combined study variables at the case-agent level
Granularity	Dyads	4,490,696	✓	✓	✓	✓	✓	✓	✓	✓	Combined study variables at the case-dyad level

This table provides an overview of our raw data sources and reconstructed sample data sets: we cleaned and merged our raw data sources to construct the sample data sets. In each row, we display how each data set relates to the major variable categories in our study (durations, pre- and post-team-selection variables) with a brief description. The pre-team-selection variables relate to patient, procedure, surgeon, and overall OR workloads; the post-team-selection variables relate to back-to-back (B2B) assignments, workload, and shared experience of OR teams. Furthermore, we reconstruct our sample data sets at three distinct granularity levels: case-team (Cases), case-agent (Agents), and case-dyad (Dyads). The cases data set contains team-aggregate dyadic post-team-selection input variables. For instance, we calculate the average team variable on a given post-team-selection input variable as $\sum_{i=1}^{n}\sum_{j=i+1}^{n} \text{Input}_{ij} \,/\, (n(n-1)/2)$, where n is the total number of OR team members and Input_{ij} is a post-team-selection input variable between dyadic focal agent i and the dyad partner j (e.g., the total number of jointly performed surgeries) (Reagans et al. 2005). We drop team members who are inactive or irregular in the OR from our data. After cleaning and reconstruction, our OR samples contain 78,080 cases for 4.5 million dyadic pairs between March 2016 and June 2019.

Figure 1.1: Detailed OR timestamps and duration definitions



This figure illustrates key surgical timestamps and durations in our data: we use the timestamps to construct duration variables for patient, procedure, pre-turnover (previous-patient out to current-patient in), and post-turnover (current-patient out to next-patient in) times for both scheduled and realized events. Triangles on the timeline represent scheduled OR times; circles represent realized OR times. We also construct OR time mismatch variables using the differences between scheduled and realized outcomes: delays (starting and ending), overtime, and idle time. In each phase in the OR, team members engage in unique dyadic and team interactions. For example, in the patient setup phase, the circulator and surgical tech set up the necessary supplies and equipment; in the patient prep phase, the circulator brings the patient into the OR room with anesthesiologist and a CRNA, and the anesthesiologist and CRNA complete the anesthesia induction; during the procedure, while the surgeon performs surgery and leads surgical procedures with the assistance of the circulator, surgical tech, and CRNA, and the anesthesiologist monitors patient status during anesthesia; in the wrap-up phase, the anesthesiologist, CRNA, and circulator transfer the patient to the post-anesthesia care unit (PACU); and in the clean-up phase, the circulator, surgical tech, and cleaning staff clean and prepare the OR for the next surgery.

Table 1.2: Pre-team-selection continuous variable descriptions and summary statistics

		Min	Q1	Q2	Q3	Max	Mean	SD	Description
Scheduled Durations	Pre-turnover	0.00	0.50	0.50	0.50	0.92	0.50	0.13	Scheduled turnover time prior the patient entry
	Patient time	0.00	1.00	1.50	2.33	6.75	1.83	1.23	Scheduled patient-room time
	Procedure time	0.00	0.75	1.17	2.00	6.00	1.55	1.17	Scheduled procedure duration
	Post-turnover	0.00	0.50	0.50	0.50	0.92	0.50	0.13	Scheduled turnover time prior the next patient
Pre-Team Selection Variables	Patient age	0.00	30.00	51.00	66.00	107.00	46.78	23.88	Patient age in years
	Procedures	0.00	1.00	1.00	2.00	14.00	1.77	1.14	Number of case procedures
	Co-surgeons	1.00	1.00	1.00	1.00	5.00	1.04	0.20	Number of co-surgeons in the case
	Robotic	0.00	0.00	0.00	0.00	100.00	2.45	15.46	Robotic case indicator (%)
	Add-on case	0.00	0.00	0.00	100.00	100.00	26.55	44.16	Add-on case indicator (%)
	Team size	3.00	6.00	7.00	9.00	28.00	7.79	2.18	Surgical team size
	Specialty experience	0.69	6.99	7.93	8.53	9.58	7.64	1.30	Cumulative log case count for the specialty
	Surgeon tenure	1.61	2.48	3.00	3.30	3.76	2.93	0.49	Surgoen tenure in years since residency
	Surgeon experience	0.69	1.10	1.10	1.39	2.20	1.12	0.30	Count of the surgeon's prior medical affications
	Room scheduled load	0.00	4.25	5.58	7.00	21.33	5.65	2.12	Scheduled daily room load
	Room order	1.00	1.00	2.00	3.00	12.00	2.27	1.31	Daily order rank of the case in the room
	Idle pre-turnover	0.00	0.00	0.00	100.00	100.00	25.85	43.78	Idle pre-turnover indicator (%)
	Idle post-turnover	0.00	0.00	0.00	100.00	100.00	25.90	43.81	Idle post-turnover indicator (%)
	Add-on previous	0.00	0.00	0.00	0.00	100.00	12.86	33.48	Previous add-on case indicator (%)
	Specialty scheduled	0.00	8.17	13.25	19.58	60.67	15.21	9.82	Scheduled daily specialty load
	OR scheduled load	5.50	141.67	158.92	175.50	210.75	151.03	41.07	Scheduled daily OR load
	OR add-on	8.86	18.28	21.59	25.27	100.00	26.67	19.39	Daily ratio of add-on cases (%)
	Holiday	0.00	0.00	0.00	0.00	100.00	0.44	6.65	Holiday indicator (%)

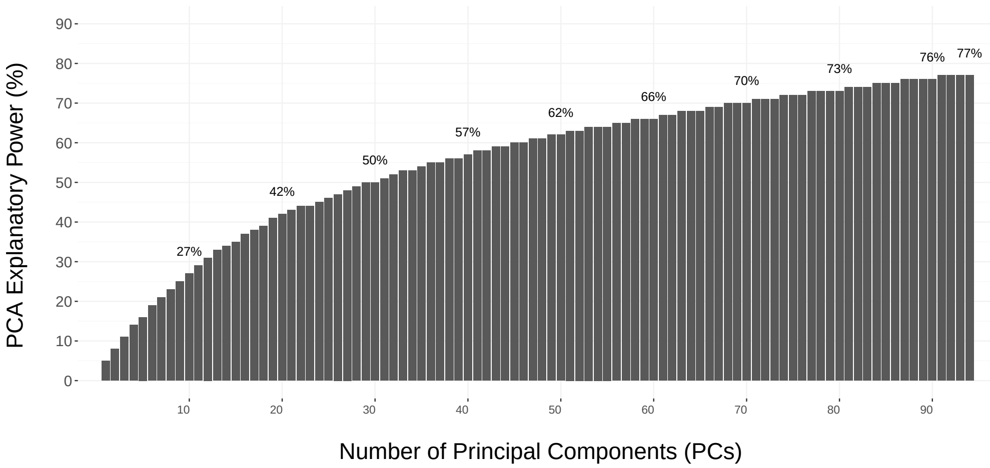
In this table, we present summary statistics and descriptions of the pre-team-selection continuous variables in our study. The first column specifies the variable categories: pre-team-selection variables include case-team level data inputs that are known before team selections, including procedure, surgeon, overall OR workload (in hours) information; scheduled duration variables include our major scheduled duration variables in hours. We winsorize the scheduled duration variables at their 1% tail ends. Note that add-on cases are cases scheduled at the last minute due to their urgency.

Table 1.3: Pre-team-selection categorical variables and the levels

	Levels	Count
Anesthesia Type	General, General Hybrid, None, Other, Regional	5
ASA Score	None, Score 1, Score 2, Score 3, Score 4, Score 5	6
Day of Week	Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday	7
Month	January, February, March, April, May, June, July, and 5 other levels	12
Patient Class	Emergency, Inpatient, Outpatient Bed, Surgery Admit, None, Hospital Outpatient Surgery	6
Gender	Female, Male, None	3
Race	Asian, Black, Hispanic, Mixed, None, Other, White	7
Specialty	General, Urology, Gynecology, Orthopedic, Vascular, Pediatric, Otolaryngology, and 20 other levels	27

In this table, we present our major pre-team-selection categorical variables in the first column, the associated levels in the second column, and the count of levels within each variable in the last column. ASA stands for American Society of Anesthesiology severity score, from 1, indicating a normal health patient, to 5, indicating a moribund patient who is not expected to survive without a surgical procedure (Owens et al. 1978). For brevity, we do not show the current procedural terminology (CPT) codes used in our study. We show only seven or fewer levels for each variable. The italicized 'and more' at the end of levels list indicates that there are more levels not shown in this table.

Figure 1.2: Cumulative proportion of the CPT words variability explained by PCs



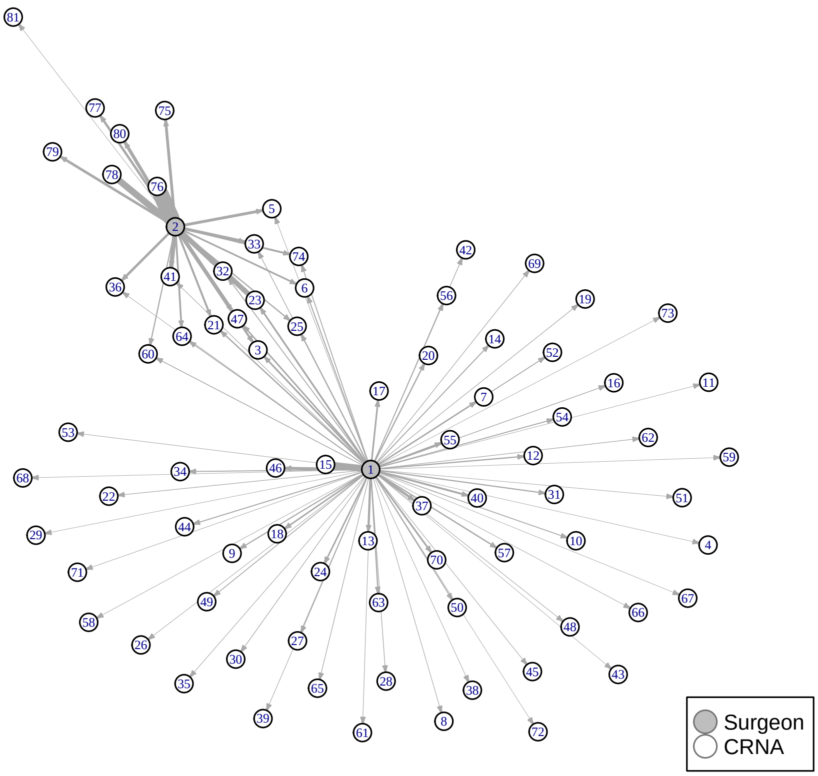
This figure shows cumulative principal component (PC) explanatory power for CPT description words' sample variability by the top PCs from our principal component analysis (PCA) for pre-team-selection input. Specifically, we first generate PCs with a PCA by extracting unique words from 1,518 distinct procedure descriptions classified by CPT—the standard AMA classification used in reporting medical, surgical, and diagnostic procedures (Dotson 2013). Then, using Kaiser's criteria—dropping PCs that have less than a unit in eigenvalues (Kaiser 1960)—we retain the top 94 PCs, which explain over 77% of the variation in words. Such textual mining is an emerging topic in the empirical health care OM field; there is one notable study: Xu et al. (2021) scrape patient reviews to explain physician demand.

Table 1.4: Illustration of diversity in individual and dyadic post-team-selection specialty experience by role-pair

		Mean Specialty Experience			
		Focal Agent	Dyad Partner	Dyad Shared (Λ)	Dyad Difference (Δ)
Anesthesiologist	Anesthesiologist	33.85	33.85	0.39	24.27
	Circulator	48.42	85.59	6.33	61.14
	CRNA	42.67	25.74	1.65	27.83
	Physician	48.13	115.05	5.69	82.80
	Surgical Tech	46.63	89.39	5.68	65.57
Circulator	Anesthesiologist	85.59	48.42	6.33	61.14
	Circulator	76.84	76.81	5.89	73.20
	CRNA	81.59	27.48	3.08	63.88
	Physician	94.52	122.79	13.40	80.37
	Surgical Tech	92.74	99.39	15.24	62.31
CRNA	Anesthesiologist	25.74	42.67	1.65	27.83
	Circulator	27.48	81.57	3.08	63.87
	CRNA	23.00	23.00	0.46	17.04
	Physician	28.24	112.11	4.05	88.34
	Surgical Tech	28.11	87.00	3.51	67.93
Physician	Anesthesiologist	115.05	48.13	5.69	82.80
	Circulator	122.79	94.50	13.39	80.38
	CRNA	112.11	28.24	4.05	88.34
	Physician	81.31	81.36	6.28	70.77
	Surgical Tech	124.28	100.20	15.25	78.16
Surgical Tech	Anesthesiologist	89.39	46.63	5.68	65.57
	Circulator	99.38	92.71	15.23	62.30
	CRNA	87.00	28.11	3.51	67.92
	Physician	100.19	124.28	15.25	78.15
	Surgical Tech	80.21	80.21	8.44	67.21

The table illustrates our post-team-selection variable construction by providing the example of specialty experience by role-pair for the focal agent, the dyad partner, and the dyad (shared, Λ; difference, Δ). Specifically, the first two columns specify the focal agent and the dyad partner within each role-pair, followed by the mean individual specialty experience of the focal agent and dyad partner, dyad shared (Λ) experience, and the dyad difference (Δ) in individual experiences within the role-pair dyad. we measure each experience by the number of surgeries that an agent or a dyad performed in the six months before the focal surgery.

Figure 1.3: Surgeon-CRNA dyadic surgery network for two top surgeons and CRNAs



This figure shows the dyadic surgery network between the top two surgeons and CRNAs in our sample data. In particular, each node represents a unique surgeon or CRNA, each line between two nodes indicates the dyad interaction (i.e., performing surgeries together as a dyad) between two nodes, and the thickness of each line increases with the degree of interaction for the dyad (i.e., the number of surgeries the dyad performed together in the three-year time horizon in our entire sample period). We follow dyadic supply chain research insights (Peng et al. 2022, Yang et al. 2009) to capture diverse role-pair interactions among OR team members.

Table 1.5: Overview statistics of post-team-selection variables

		Dyad-Λ		Dyad-Δ		Individual		Team	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
B2B	Different setup	0.47	6.86	29.72	45.70	17.68	38.15	0.4	1.71
	Same room only	0.27	5.23	0.56	7.43	0.57	7.50	0.3	3.48
	Same specialty only	0.59	7.63	10.01	30.01	5.51	22.81	0.51	2.8
	Same setup	5.02	21.83	5.99	23.74	9.03	28.66	6.54	18.13
Workload	Scheduled	3.86	2.27	6.10	7.51	8.26	6.89	3.64	1.55
Experience	Specialty	6.96	17.63	63.17	70.95	66.94	79.04	7.93	10.63
	Non-specialty	3.65	7.67	126.44	103.67	142.56	119.36	3.65	2.79
	Room	4.64	14.30	128.93	105.13	209.50	127.30	5.46	8.33
	People	10.61	19.72	9.50	12.69	10.97	13.31	11.58	10.63

This table gives an overview of statistics for our post-team-selection variable categories at each measurement level: dyad shared (shared, Λ; difference, Δ), individual, or team as a whole. Specifically, we have three sets of post-team-selection variables: (i) B2B in different setups depending on their room and specialty continuity—such as B2B assignments with different setup (a different room and different specialty) and B2B assignments with same setup (the same room and same specialty), (ii) daily scheduled workloads, and (iii) experience in different contexts—such as experiences in same-specialty surgeries, different-specialty surgeries, same-room surgeries, and surgeries with the rest of the assigned OR team members, excluding the dyadic partner.

Table 1.6: Variations in shared experiences (Λ) by surgical role-pair

			Dyad-Λ B2B				Dyad-Λ Workload	Dyad-Λ Experience			
		Frequency Ratio (%)	Different Setup	Room	Specialty	Same Setup	Scheduled	Specialty	Non-Specialty	Room	People
Anesthesiologist	Circulator	4.59	1.18	0.11	0.75	1.88	4.29	6.33	5.54	5.17	11.87
	Surgical Tech	4.26	0.90	0.10	0.63	1.78	4.28	5.68	5.33	4.45	11.01
	CRNA	3.93	1.06	0.10	0.32	1.77	4.17	1.65	5.08	1.25	6.72
	Physician	2.93	0.24	0.05	0.95	2.20	3.69	5.69	0.65	3.23	6.34
	Anesthesiologist	0.94	4.63	0.00	1.04	0.00	3.86	0.39	2.73	0.19	3.11
Circulator	Surgical Tech	6.29	0.79	0.47	0.72	6.74	4.15	15.24	10.81	11.29	26.04
	CRNA	5.68	0.37	0.45	0.23	6.57	3.84	3.08	3.31	2.23	6.39
	Anesthesiologist	4.59	1.18	0.11	0.75	1.88	4.29	6.33	5.54	5.17	11.87
	Physician	4.21	0.08	0.27	0.58	8.26	3.58	13.40	1.21	7.93	14.61
	Circulator	3.85	0.69	0.26	0.44	4.14	3.59	5.89	5.78	3.90	11.67
CRNA		5.68	0.37	0.45	0.23	6.57	3.84	3.08	3.31	2.23	6.39
	Surgical Tech	5.29	0.26	0.45	0.23	6.27	3.89	3.51	3.29	2.58	6.80
	Anesthesiologist	3.93	1.06	0.10	0.32	1.77	4.17	1.65	5.08	1.25	6.72
	Physician	3.58	0.04	0.24	0.25	7.66	3.43	4.05	0.48	2.29	4.52
	CRNA	2.21	0.11	0.02	0.08	0.40	3.03	0.46	1.52	0.27	1.98
Physician	Circulator	4.21	0.08	0.27	0.58	8.26	3.58	13.39	1.21	7.93	14.60
	Surgical Tech	3.88	0.09	0.27	0.60	7.98	3.63	15.25	1.58	9.38	16.82
	CRNA	3.58	0.04	0.24	0.25	7.66	3.43	4.05	0.48	2.29	4.52
	Anesthesiologist	2.93	0.24	0.05	0.95	2.20	3.69	5.69	0.65	3.23	6.34
	Physician	0.54	0.28	0.09	1.34	2.92	3.58	6.28	0.46	2.34	6.74
Surgical Tech	Circulator	6.29	0.79	0.47	0.72	6.75	4.15	15.23	10.81	11.28	26.03
	CRNA	5.29	0.26	0.45	0.23	6.27	3.89	3.51	3.29	2.58	6.80
	Anesthesiologist	4.26	0.90	0.10	0.63	1.78	4.28	5.68	5.33	4.45	11.01
	Physician	3.88	0.09	0.27	0.60	7.98	3.63	15.25	1.58	9.38	16.82
	Surgical Tech	3.2	0.42	0.25	0.70	2.98	3.86	8.44	6.63	5.68	15.07

In this table, we display the mean values for all shared (Λ) experience variables by role-pair. The first row shows our major post-team-selection input categories (B2B, workload, experience) and the next row shows the different setups for each major variable category for each role-pair, specified in the first two columns. B2B is back-to-back assignment, and we multiply the mean B2B values by 100 to show up to the ten-thousandths decimal place. A high mean value of any variable implies that the role-pair shares a high level of dyadic interactions on average. We rank each role-pair category in descending order of frequency ratio after dropping team members who are inactive or non-primary in the OR—our sample data contain nine distinct OR team member roles and 33 unique major role-pairs. The primary team roles include anesthesiologist, circulator, certified registered nurse anesthetist (CRNA), physician, and surgical tech.

Table 1.7: Variations in specialty experience by specialty

	Coefficient of Variation (COV)			
	Dyad-Λ	Dyad-Δ	Agent	Team
Bariatric	2.69	1.46	1.65	1.73
Cardiothoracic	1.23	0.82	0.77	0.46
Colorectal	1.81	1.16	1.21	0.66
Edocrinolgoy	1.94	1.03	1.17	0.75
Gastrointestinal	3.33	1.44	1.74	1.80
Gen-Trauma	2.06	1.66	2.09	0.86
General	2.08	0.92	0.81	0.82
Gynecology	1.96	1.06	1.02	0.77
Neurology	1.87	0.90	1.04	0.66
Oncology	2.26	1.38	1.47	0.87
Ophthalmology	2.81	1.22	1.45	1.48
Oral	1.76	1.27	1.38	0.94
Ortho-FootAnkle	1.84	1.41	1.51	1.07
Ortho-Hand	2.72	1.24	1.37	1.16
Ortho-Oncology	2.70	1.10	1.43	0.82
Ortho-Ped	2.11	1.14	1.39	0.76
Ortho-Spine	2.17	0.95	1.17	0.68
Ortho-Trauma	1.90	0.95	1.06	0.76
Orthopedic	1.83	0.94	0.96	0.79
Other	3.47	1.51	1.89	1.86
Otolaryngology	1.99	0.91	1.02	0.90
Pediatric	2.13	0.93	1.08	0.88
Plastic	2.97	1.19	1.38	1.14
Radiology	2.54	1.65	2.06	1.52
Thoracic	1.90	0.99	1.16	0.71
Urology	2.61	1.00	1.10	1.11
Vascular	1.64	0.82	0.96	0.62

This table compares the variations of our specialty experience variables at dyad (shared, Λ; difference, Δ), agent, and team level across different surgical specialty areas in terms of coefficient of variation (COV). The first column specifies specialties, followed by dyad- (shared and difference), agent-, and team-level specialty experience columns. The COV, the ratio of the standard deviation to the mean, measures the level of dispersion around the mean, which represents the relative variation of a variable. For example, the COV for the dyad difference (Δ) variable in bariatric surgeries is the standard deviation of dyad differences within bariatric surgeries over the mean. A high level of COV indicates that the respective specialty experience has relatively high variation within the specialty compared to other specialties with a high level of COV.

Table 1.8: Summary statistics and descriptions of outcome variables

		Min	Q1	Q2	Q3	Max	Mean	SD	Description
Realized Duration	Pre-turnover time	0.00	0.50	0.65	1.03	6.53	1.09	1.19	Realized turnover time prior the patient entry
	Procedure time	0.02	0.48	0.95	1.83	6.73	1.40	1.32	Realized procedure duration
	Patient time	0.08	1.03	1.58	2.58	7.92	2.05	1.49	Realized patient-room time
	Post-turnover time	0.00	0.50	0.65	1.03	6.57	1.09	1.20	Realized turnover time prior the next patient
	LOS	0.00	0.00	0.00	4.00	69.00	4.63	10.86	Patient length of stay in days
Realized Mismatch	Starting delay	0.00	0.00	0.00	0.33	2.52	0.27	0.50	Starting delay for the patient entry
	Procedure overtime	0.00	0.00	0.18	0.55	2.38	0.35	0.47	Procedure delay duration
	Procedure idle time	0.00	0.00	0.00	0.13	2.73	0.20	0.47	Procedure idle duration
	Patient overtime	0.00	0.00	0.00	0.20	1.90	0.17	0.34	Pateint-room delay duration
	Patient idle time	0.00	0.00	0.12	0.52	3.45	0.39	0.64	Pateint-room idle duration
	Ending delay	0.00	0.00	0.08	0.75	3.82	0.51	0.79	Ending delay for the patient exit

This table presents summary statistics and descriptions of our major OR outcome variables, excluding the absolute mismatch measure (which is implied by the sum of overtime and idle time). The first column specifies the major variable categories in our study. First, realized durations include pre-turnover, procedure, patient, and post-turnover times. Second, we measure a patient-related outcome as an alternative outcome, length of stay (LOS), but we include LOS in durations in this table for brevity. Third and mismatch variables, include starting and ending delay, procedure and patient overtime, and procedure and patient idle time; the sum of overtime and idle time is the absolute mismatch measure. Durations are in hours—we winsorize the duration variables at their 1% tail ends.

Table 1.9: Mismatch variable variations within categorical variables

		Ratio (%)	Starting Delay	Procedure Overtime	Procedure Idle Time	Patient Overtime	Patient Idle Time	Ending Delay
Anesthesia Type	General	93.69	0.27	0.36	0.20	0.17	0.40	0.52
	Regional	4.69	0.24	0.25	0.10	0.12	0.22	0.35
	General Hybrid	1.48	0.25	0.36	0.16	0.15	0.34	0.45
	Other	0.12	0.23	0.46	0.10	0.25	0.27	0.36
	None	0.02	0.12	0.46	0.64	0.28	0.74	0.34
ASA Score	Score 3	45.70	0.28	0.36	0.21	0.18	0.41	0.53
	Score 2	36.50	0.26	0.33	0.15	0.15	0.32	0.44
	Score 4	10.90	0.27	0.39	0.35	0.20	0.64	0.71
	Score 1	6.34	0.24	0.33	0.12	0.16	0.26	0.38
	Score 5	0.49	0.27	0.65	0.38	0.40	0.59	0.68
	None	0.06	0.47	0.72	0.39	0.59	0.61	0.55
Day of Week	Thursday	19.90	0.27	0.33	0.19	0.15	0.39	0.51
	Friday	19.39	0.26	0.35	0.19	0.17	0.38	0.48
	Tuesday	17.96	0.25	0.36	0.21	0.17	0.42	0.51
	Wednesday	17.85	0.26	0.34	0.20	0.16	0.40	0.51
	Monday	17.56	0.28	0.35	0.20	0.17	0.40	0.52
	Saturday	4.08	0.33	0.45	0.18	0.23	0.33	0.54
	Sunday	3.27	0.32	0.46	0.18	0.25	0.33	0.52
Month	May	10.39	0.28	0.35	0.20	0.18	0.38	0.51
	March	10.18	0.23	0.34	0.20	0.17	0.39	0.46
	June	10.10	0.30	0.35	0.20	0.16	0.39	0.54
	April	9.63	0.26	0.36	0.18	0.19	0.36	0.47
	August	7.94	0.27	0.38	0.19	0.16	0.41	0.52
	October	7.77	0.25	0.35	0.20	0.15	0.42	0.51
	December	7.61	0.26	0.36	0.18	0.16	0.41	0.51
	November	7.46	0.27	0.35	0.19	0.16	0.43	0.55
	July	7.41	0.29	0.39	0.19	0.15	0.42	0.56
	September	7.31	0.28	0.33	0.23	0.16	0.42	0.55
	January	7.11	0.25	0.35	0.20	0.19	0.37	0.48
	February	7.09	0.26	0.34	0.20	0.21	0.32	0.46
Patient Class	Outpatient	38.30	0.25	0.28	0.10	0.12	0.24	0.37
	Inpatient	30.58	0.32	0.42	0.25	0.21	0.46	0.61
	Outpatient Bed	14.29	0.25	0.34	0.21	0.16	0.42	0.51
	Surgery Admit	12.90	0.23	0.43	0.34	0.21	0.66	0.68
	Emergency	3.90	0.27	0.45	0.23	0.25	0.37	0.51
	None	0.03	0.13	0.49	0.50	0.29	0.60	0.28
Gender	Female	50.63	0.27	0.37	0.18	0.18	0.37	0.49
	Male	49.34	0.27	0.34	0.22	0.16	0.42	0.53
	None	0.02	0.12	0.46	0.64	0.28	0.74	0.34
Race	White	74.80	0.26	0.36	0.19	0.17	0.39	0.50
	Black	16.12	0.28	0.34	0.22	0.16	0.43	0.56
	Hispanic	4.27	0.30	0.34	0.18	0.16	0.35	0.52
	None	2.44	0.27	0.38	0.21	0.16	0.42	0.53
	Mixed	0.83	0.27	0.32	0.14	0.15	0.32	0.46
	Other	0.83	0.27	0.29	0.18	0.15	0.36	0.49
	Asian	0.70	0.29	0.32	0.20	0.14	0.40	0.55

This table shows our OR mismatch variable variations across levels within our pre-team-selection categorical variables. The first and second columns specify the categorical variable and the respective categories within each variable in descending order of frequency ratio within the variable. ASA stands for American Society of Anesthesiology severity score, from 1, indicating a normal health patient, to 5, indicating a moribund patient who is not expected to survive without a surgical procedure (Owens et al. 1978). Note that the average total absolute mismatch, while not shown here, is the sum of overtime and idle time. Also note that our analysis considers other categorical variables, including specialty, room, role-pair, and CPT code to estimate OR times, though for brevity they are not shown in this table. All mismatch variables are in hours—we winsorize the duration variables at their 1% tail ends.

Table 1.10: Summary statistics differences by process variability group and specialty

		Pre-Team-Selection Variables		Post-Team-Selection Dyad-Λ		Realized Durations		
		Procedures	Patient age	B2B same setup	Experience people	Procedure time	Patient time	LOS
High PV	General	1.51	52.86	4.32	7.01	1.33	1.96	7.37
	Orthopedic	1.56	46.34	3.14	9.03	1.71	2.44	5.90
	Thoracic	1.52	62.22	5.64	11.67	1.31	1.96	5.11
	Ortho-Trauma	1.47	51.95	3.01	11.83	1.94	2.71	8.56
	Cardiothoracic	3.11	63.14	0.10	17.25	4.70	5.95	13.34
	Plastic	1.78	38.39	3.63	8.82	1.90	2.56	3.30
	Ortho-Spine	3.47	57.16	2.68	12.28	2.55	3.48	4.43
	Oncology	1.96	56.03	2.95	7.91	1.64	2.29	2.87
	Ortho-Oncology	1.34	48.81	3.61	8.74	1.87	2.61	2.14
	Ortho-Hand	1.56	46.45	0.93	5.50	1.36	1.99	4.88
	Bariatric	1.72	53.05	0.00	3.56	2.00	2.67	10.53
	Gen-Trauma	1.25	50.62	1.19	4.25	1.17	1.75	13.23
Low PV	Urology	1.93	54.70	12.48	18.75	0.85	1.41	0.90
	Gynecology	1.73	44.17	5.12	6.86	1.22	1.85	0.54
	Pediatric	1.31	7.83	6.40	12.71	0.87	1.48	6.68
	Neurology	2.17	50.73	4.24	10.51	1.55	2.38	5.40
	Vascular	1.48	60.23	4.11	10.79	1.38	2.04	5.63
	Otolaryngology	2.46	26.83	7.86	14.07	1.16	1.70	2.65
	Colorectal	1.75	55.13	4.12	6.54	1.42	2.03	3.09
	Ortho-Ped	1.31	16.06	4.75	8.42	1.21	1.78	2.88
	Edocrinolgoy	2.11	33.31	2.37	6.15	1.54	2.15	0.13
	Oral	1.12	13.64	7.32	5.97	1.03	1.58	0.77
	Ophthalmology	1.51	47.58	1.40	6.32	1.15	1.67	2.00
	Gastrointestinal	1.43	20.61	1.47	6.09	0.53	0.96	4.64
	Other	1.46	26.87	0.49	7.60	0.87	1.40	6.76
	Ortho-FootAnkle	1.46	50.67	3.01	4.26	1.33	1.91	3.43
	Radiology	1.05	52.60	2.55	6.10	0.79	1.41	4.63

This table compares mean differences in selective pre-team-selection, post-team-selection, and realized-duration variables across distinct surgical specialty areas grouped by high and low process variability (PV). The first two columns specify the PV group and the respective specialties. B2B is back-to-back assignment. Procedure and patient times are in hours—we winsorize the duration variables at their 1% tail ends. Length of stay (LOS) is in days. To split the surgical specialties into low and high PV groups, we consider the within-procedure variability of procedures within each specialty. Specifically, we compute the standard deviation of procedure times within each procedure (i.e., within-procedure variability) and get the weighted average of within-procedure variability under each specialty, calculated as the number of procedure observations over total observations within the specialty. We define this weighted average as process variability (PV) of a specialty. Using this measure, we classify specialties with a PV value higher than the mean PV as a high PV group, and vice versa. We have 12 specialties in the high PV group accounting for 49.90% of the dyadic observations and 15 specialties in the low PV group accounting for 50.10% of the dyadic observations. On average, high PV specialties have a longer procedure time, longer patient time, higher patient age, and longer length of stay (LOS), representing high complexity of procedures within the high PV specialty group.

Table 1.11: Ex ante OR time estimation with pre-team-selection inputs (stage one)

		Patient Time				Procedure Time				Post-Turnover Time			
		Without PCA		With PCA		Without PCA		With PCA		Without PCA		With PCA	
		Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Baseline	Intercept	4.069**	0.117	4.205**	0.107	3.141**	0.179	3.352**	0.165	1.433**	0.478	1.167**	0.478
Pre-Team	Patient age	0.079**	0.003	0.058**	0.002	0.130**	0.004	0.102**	0.004	−0.014	0.011	0.009	0.011
Selection	Male	0.010**	0.003	0.026**	0.003	0.024**	0.004	0.043**	0.004	−0.022*	0.012	−0.030**	0.012
Variables	Procedures	0.261**	0.005	0.354**	0.007	0.412**	0.008	0.542**	0.011	−0.293**	0.020	−0.279**	0.032
	Robotic	0.147**	0.010	0.068**	0.011	0.190**	0.016	0.095**	0.016	−0.445**	0.042	−0.336**	0.048
	Panels	−0.015*	0.008	−0.012*	0.007	−0.024*	0.012	−0.026*	0.011	−0.120**	0.032	−0.107**	0.033
	Add-on case	−0.100**	0.005	−0.104**	0.005	−0.156**	0.008	−0.162**	0.008	−0.122**	0.022	−0.131**	0.023
	Team size	0.124**	0.001	0.100**	0.001	0.167**	0.001	0.135**	0.001	−0.123**	0.003	−0.098**	0.003
	Specialty experience	−0.027**	0.002	−0.025**	0.001	−0.035**	0.002	−0.033**	0.002	−0.034**	0.007	−0.033**	0.007
	Surgeon tenure	0.000	0.001	0.000	0.001	0.002	0.001	0.002	0.001	−0.004	0.003	−0.004	0.003
	Surgeon experience	0.024**	0.008	0.017*	0.007	0.041**	0.012	0.032**	0.011	−0.001	0.033	−0.008	0.033
	Room scheduled load	0.014**	0.001	0.003**	0.001	0.021**	0.001	0.007**	0.001	0.291**	0.003	0.302**	0.003
	Room order	−0.026**	0.002	−0.012**	0.002	−0.031**	0.003	−0.016**	0.003	−0.273**	0.008	−0.284**	0.008
	Idle pre-turnover	0.059**	0.004	0.046**	0.004	0.080**	0.006	0.063**	0.006	−0.383**	0.017	−0.364**	0.017
	Add-on previous	0.002	0.005	0.003	0.004	−0.001	0.007	0.003	0.007	−0.055**	0.020	−0.057**	0.020
	Specialty scheduled	0.001**	0.000	0.001**	0.000	0.002**	0.000	0.002**	0.000	−0.004**	0.001	−0.004**	0.001
	OR scheduled load	0.000*	0.000	0.000**	0.000	0.000**	0.000	0.001**	0.000	−0.003**	0.000	−0.003**	0.000
	Holiday	0.024	0.022	0.040*	0.020	0.090**	0.034	0.113**	0.031	−0.615**	0.091	−0.617**	0.091
Controls	Fixed effects	Yes		Yes		Yes		Yes		Yes		Yes	
	PCs	No		Yes		No		Yes		No		Yes	
Statistics	R-squared	0.675		0.7319		0.6172		0.6788		0.3938		0.4015	
	Observations (K)	77.877		77.877		77.877		77.877		77.877		77.877	

The table shows our ex ante OR time estimates with pre-team-selection inputs. In particular, we use patient, procedure, and post-turnover times as dependent variables. Within each specification, we run our estimation with and without principal components (PCs) from our principal component analysis (PCA) on CPT description words. Our analysis includes fixed-effect variables, including patient class, patient gender, patient race, anesthesia type, ASA score, room, specialty, lead surgeon, day of the week, time of the day, though for brevity they are not shown in the table. Also, each analysis includes additional controls for previous durations before the start of each dependent variable duration. We use log-minutes for surgical duration variables. Two stars (**) indicates significance at 1%, and one star (*) indicates significance at 5%.

Table 1.12: Ex post OR time estimation with post-team-selection inputs (team-level stage two)

		Patient Time				Procedure Time				Post-Turnover Time			
		High PV		Low PV		High PV		Low PV		High PV		Low PV	
		Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Baseline	Intercept	0.397**	0.096	0.164	0.313	1.715**	0.467	1.870	1.539	1.684**	0.469	1.734	1.537
Behavioral	SM over-allocation	0.627**	0.010	0.749**	0.009	0.132**	0.030	0.139**	0.024	0.014	0.017	0.094**	0.015
	SM under-allocation	−0.547**	0.007	−0.469**	0.006	−0.074*	0.036	−0.031	0.030	−0.082**	0.012	−0.026**	0.011
	TD over-allocation	0.245**	0.010	0.093**	0.032	0.068*	0.036	−0.032	0.121	−0.018	0.084	−0.086	0.068
	TD under-allocation	−0.177**	0.013	−0.059*	0.031	−0.369**	0.049	0.001	0.119	0.169*	0.083	0.181**	0.069
Dyad-Λ	B2B different setup	0.171	0.108	0.016	0.094	−0.947*	0.526	−0.462	0.459	−0.703	0.683	0.173	0.569
	B2B same setup	−0.107**	0.014	−0.098**	0.023	−0.613**	0.068	−0.411**	0.112	−0.473**	0.064	−0.459**	0.053
	Workload scheduled	−0.110**	0.011	−0.140**	0.022	0.195**	0.052	−0.181	0.117	0.025	0.092	−0.184**	0.077
	Experience non-specialty	0.010*	0.006	0.010*	0.004	−0.019	0.027	0.018	0.023	−0.016	0.029	0.039*	0.023
Controls	Post-team controls	Yes		Yes		Yes		Yes		Yes		Yes	
	Pre-team controls	Yes		Yes		Yes		Yes		Yes		Yes	
	Precedent durations	Yes		Yes		Yes		Yes		Yes		Yes	
Statistics	R-squared	0.307		0.3241		0.239		0.2428		0.2396		0.2436	
	Observations (K)	33.340		44.537		33.340		44.537		33.340		44.537	

This table shows our main results for team performance model estimation with team aggregate post-team-selection inputs—as opposed to dyadic inputs. Specifically, we estimate patient, procedure, and post-turnover times for each process variability (PV) specialty group. We drop variables that are highly correlated with other variables in our sample to avoid a multi-colinearity problem and, for brevity, list only selective dyad shared (Λ) variables for post-team-selection inputs. Also, while not shown, each analysis includes additional controls for pre-team-selection inputs and previous durations before the start of each dependent variable duration. We use log minutes for surgical duration variables. Two stars (**) indicates significance at 1%, and one star (*) indicates significance at 5%.

Table 1.13: Ex post OR time estimation with post-team-selection inputs (dyad-level and specialty-level stage two)

		Baseline	Behavioral				Dyad-Λ				Statistics	
		Intercept	SM Over-Allocation	SM Under-Allocation	TD Over-Allocation	TD Under-Allocation	B2B Different Setup	B2B Same Setup	Workload Scheduled	Experience Non-specialty	R^2	Obs
High PV	Bariatric	0.834**	−0.739**	−0.669**		1.183**			0.010	0.002	0.9128	6,490
		(0.124)	(0.036)	(0.021)		(0.038)			(0.006)	(0.002)		
	Cardiothoracic	0.733**	0.645**	−0.523**		−1.065**	−0.037**	0.053**	−0.093**	−0.005**	0.5844	255,014
		(0.030)	(0.004)	(0.002)		(0.049)	(0.012)	(0.013)	(0.002)	(0.001)		
	Gen-Trauma	0.437*	0.905**	−0.131**		−0.583**		−0.090**	−0.030**	0.004	0.8862	5,010
		(0.202)	(0.024)	(0.021)		(0.032)		(0.025)	(0.007)	(0.003)		
	General	0.388**	0.574**	−0.462**	0.163**	−0.195**	0.044**	−0.042**	−0.053**	0.001**	0.2381	850,563
		(0.016)	(0.002)	(0.002)	(0.002)	(0.003)	(0.005)	(0.002)	(0.001)	(0.000)		
	Oncology	−0.076	0.612**	−0.660**	0.266**	−0.177**	−0.008	−0.088**	−0.033**	0.007**	0.4798	77,406
		(0.053)	(0.006)	(0.005)	(0.005)	(0.006)	(0.011)	(0.005)	(0.003)	(0.001)		
	Ortho-Hand	0.387**	0.356**	−0.265**		−0.399**	0.106**	−0.203**	0.015*	−0.012**	0.5557	14,942
		(0.117)	(0.018)	(0.010)		(0.023)	(0.021)	(0.027)	(0.008)	(0.003)		
	Ortho-Oncology	1.493**	0.303**	−0.506**	0.226**	−0.294**	0.053**	0.050**	−0.003	−0.003*	0.6398	22,010
		(0.149)	(0.016)	(0.007)	(0.008)	(0.019)	(0.018)	(0.010)	(0.005)	(0.002)		
	Ortho-Spine	0.623**	0.437**	−0.590**	0.033**	−0.139**	0.016*	0.065**	−0.049**	0.009**	0.5362	79,004
		(0.050)	(0.005)	(0.004)	(0.004)	(0.014)	(0.007)	(0.004)	(0.002)	(0.001)		
	Ortho-Trauma	0.403**	0.692**	−0.590**	0.410**	−0.200**	0.031**	−0.118**	−0.071**	0.006**	0.4323	156,036
		(0.058)	(0.005)	(0.003)	(0.003)	(0.009)	(0.009)	(0.005)	(0.002)	(0.001)		
	Orthopedic	0.242**	0.588**	−0.562**	0.370**	−0.303**	0.024**	−0.113**	−0.039**	0.005**	0.324	442,040
		(0.022)	(0.003)	(0.002)	(0.002)	(0.004)	(0.006)	(0.003)	(0.001)	(0.001)		
	Plastic	0.415**	0.812**	−0.715**	0.310**	−0.403**	0.024**	−0.074**	−0.020**	0.005**	0.6739	116,446
		(0.022)	(0.004)	(0.003)	(0.004)	(0.006)	(0.009)	(0.004)	(0.003)	(0.001)		
	Thoracic	2.535**	0.554**	−0.592**	0.336**	−0.184**	0.048**	−0.073**	−0.025**	−0.010**	0.4134	209,152
		(0.053)	(0.004)	(0.003)	(0.004)	(0.005)	(0.010)	(0.004)	(0.002)	(0.001)		
Low PV	Colorectal	0.496**	0.581**	−0.456**	0.064**	−0.231**	0.024**	−0.072**	−0.016**	0.006**	0.393	207,206
		(0.044)	(0.005)	(0.003)	(0.004)	(0.005)	(0.008)	(0.004)	(0.002)	(0.001)		
	Edocrinolgoy	0.020	0.601**	−0.735**	0.417**	−0.325**	−0.026	−0.106**	−0.032**	0.000	0.4957	34,292
		(0.095)	(0.010)	(0.008)	(0.009)	(0.013)	(0.017)	(0.009)	(0.005)	(0.002)		
	Gastrointestinal	2.495**	0.270**	−0.480**	0.670**	0.090**		−0.146**	−0.014*	0.002	0.8889	4,758
		(0.151)	(0.038)	(0.025)	(0.073)	(0.027)		(0.028)	(0.007)	(0.002)		
	Gynecology	0.767**	0.603**	−0.493**	0.157**	−0.162**	0.025**	−0.071**	−0.031**	0.002**	0.3436	396,928
		(0.019)	(0.003)	(0.002)	(0.003)	(0.003)	(0.007)	(0.002)	(0.001)	(0.001)		
	Neurology	1.300**	0.506**	−0.535**	0.132**	−0.074**	0.025**	−0.034**	−0.029**	0.000	0.3968	277,172
		(0.045)	(0.003)	(0.002)	(0.003)	(0.005)	(0.007)	(0.003)	(0.002)	(0.001)		
	Ophthalmology	0.142	0.286**	−0.536**	0.043	0.029	0.102**	−0.136**	−0.021*	0.005	0.5423	10,828
		(0.113)	(0.021)	(0.017)	(0.038)	(0.030)	(0.028)	(0.027)	(0.009)	(0.004)		
	Oral	−0.061	0.623**	−0.323**	0.119**	−0.262**	−0.027	−0.066**	−0.012*	0.012**	0.3839	28,284
		(0.137)	(0.013)	(0.009)	(0.024)	(0.017)	(0.021)	(0.009)	(0.006)	(0.002)		
	Ortho-FootAnkle	3.532**	1.059**	0.013	−1.033**	0.126**		0.014	0.016*	−0.002	0.9352	3,914
		(0.147)	(0.036)	(0.026)	(0.035)	(0.039)		(0.019)	(0.007)	(0.002)		
	Ortho-Ped	0.806**	0.721**	−0.389**	0.433**	−0.065**	0.057**	−0.015	−0.049**	0.005*	0.4512	30,040
		(0.063)	(0.010)	(0.008)	(0.012)	(0.012)	(0.021)	(0.009)	(0.006)	(0.002)		
	Other	3.193**	−0.536**	−2.191**	0.523**	0.086*	0.005		−0.005*	0.001	0.9933	4,028
		(0.077)	(0.035)	(0.015)	(0.037)	(0.039)	(0.006)		(0.002)	(0.001)		
	Otolaryngology	0.517**	0.913**	−0.613**	0.409**	−0.417**	0.043**	−0.187**	−0.008**	0.007**	0.5327	263,540
		(0.027)	(0.003)	(0.003)	(0.003)	(0.004)	(0.009)	(0.003)	(0.002)	(0.001)		
	Pediatric	0.426**	0.706**	−0.476**	0.430**	−0.229**	0.031**	−0.129**	−0.010**	0.011**	0.3454	258,823
		(0.035)	(0.004)	(0.003)	(0.004)	(0.004)	(0.010)	(0.003)	(0.002)	(0.001)		
	Radiology	−6.729**	0.808**	−2.260**	5.196**	0.824**			0.002	−0.001	0.9908	1,886
		(0.153)	(0.025)	(0.043)	(0.132)	(0.031)			(0.003)	(0.001)		
	Urology	0.790**	0.615**	−0.381**	0.067**	−0.070**	0.040**	−0.050**	−0.050**	0.008**	0.342	424,444
		(0.020)	(0.003)	(0.002)	(0.003)	(0.003)	(0.007)	(0.002)	(0.001)	(0.001)		
	Vascular	0.020	0.851**	−0.566**	0.422**	−0.406**	0.019*	−0.129**	0.009**	0.003**	0.4376	299,674
		(0.065)	(0.004)	(0.003)	(0.003)	(0.005)	(0.010)	(0.004)	(0.002)	(0.001)		

This table shows our main results for the dyad-level ex post team performance model for patient times. For the process variability (PV) group and the respective specialties specified in the first two columns, we estimate the ex post patient time with dyad-level post-team-selection inputs—back-to-back (B2B), workload, and experience—specified in the first two rows. As with Table 1.12, we drop variables highly correlated with other variables in our sample to avoid a multi-colinearity problem and, for brevity, list only selective dyad shared (Λ) variables for post-team-selection inputs. Also, while not shown, each analysis includes additional controls for pre-team-selection inputs and previous durations before the start of each dependent variable duration. Also, each estimation includes extended controls for interactions between dyadic role-pair and experience variables. We use log minutes for surgical duration variables. Two stars (**) indicates significance at 1%, and one star (*) indicates significance at 5%.

Table 1.14: Dyadic ex post OR time performance with alternative outcome variables

		Procedure Time	Post-Turnover Time	LOS
Baseline	Intercept	-0.191^{**}	0.676^{**}	-1.166^{**}
Behavioral	SM over-allocation	0.638^{**}	1.144^{**}	
	SM under-allocation	-0.528^{**}	-1.151^{**}	
	TD over-allocation	0.217^{**}	0.343^{**}	
	TD under-allocation	-0.326^{**}	-0.268^{**}	
Dyad-Λ	B2B different setup	0.013	-0.023	-0.013
	B2B same setup	-0.096^{**}	-0.050^{**}	-0.000
	Workload scheduled	0.008^{**}	-0.060^{**}	0.022^{**}
	Experience non-specialty	0.001	-0.002	-0.000
Controls	Post-team controls	Yes	Yes	Yes
	Pre-team controls	Yes	Yes	Yes
	Precedent durations	Yes	Yes	Yes
Statistics	R-squared	0.3552	0.8906	0.2187
	Observations (K)	4,479.930	4,479.930	4,479.930

This tables presents extensions of our main dyadic patient time model with alternative outcome variables—procedure time, post-turnover time, and length of stay (LOS)—specified in the first two rows. As with Table 1.13, we show only the weighted average of estimates of all specialties (based on the number of observations within each specialty) for each outcome variable and the significance level determined using the weighted average standard error. For brevity, we list only selective dyad shared (Λ) variables for post-team-selection inputs—back-to-back (B2B), workload, and experience, after removing variables highly correlated with other variables in our sample. We use log minutes for surgical duration variables. Two stars (**) indicates significance at 1%, and one star (*) indicates significance at 5%.

Table 1.15: Expected average task-time estimations using post-team-selection inputs

		Patient Time		Procedure Time		Post-Turnover Time	
		Estimate	SE	Estimate	SE	Estimate	SE
Baseline	Intercept	3.708^{**}	0.010	2.799^{**}	0.014	1.333^{**}	0.026
Dyad-Λ	B2B different setup	0.792^{**}	0.095	0.968^{**}	0.128	-5.210^{**}	0.240
	B2B same setup	-0.670^{**}	0.009	-0.830^{**}	0.012	-0.025	0.023
	Workload scheduled	0.733^{**}	0.005	0.997^{**}	0.007	0.928^{**}	0.013
	Experience non-specialty	0.011^{*}	0.006	0.061^{**}	0.008	0.000	0.016
Statistics	R-squared	0.2891		0.2876		0.1106	
	Observations (K)	77,877		77,877		77,877	

The table reports our task diversity analysis using the predicted ex ante OR times as dependent variables—patient, procedure, and post-turnover times—and ex post team selection dyad shared (Λ) variables—back-to-back (B2B), workload, and experience—as independent variables. We use log minutes for surgical duration variables. Two stars (**) indicates significance at 1%, and one star (*) indicates significance at 5%.

Table 1.16: Newsvendor scheduling mismatch cost comparisons (in \$M)

		Absolute Mismatch (minutes)	Over-allocation (minutes)	Under-allocation (minutes)	Total Cost (\$M)
$C_u = 1 \times C_o$	Current Policy	34.55	26.44	40.18	38.65
	Ex Ante	33.73	27.24	40.06	37.73
	Team Ex Post	29.64	24.27	34.97	33.15
	Dyad Ex Post	27.81	22.88	32.89	31.11
$C_u = 2 \times C_o$	Current Policy	34.55	26.44	40.18	65.18
	Ex Ante	34.47	31.99	37.94	56.23
	Team Ex Post	30.43	28.45	33.26	49.31
	Dyad Ex Post	28.65	26.95	31.19	45.98
$C_u = 3 \times C_o$	Current Policy	34.55	26.44	40.18	91.71
	Ex Ante	35.65	34.89	36.96	70.39
	Team Ex Post	31.53	31.03	32.40	61.62
	Dyad Ex Post	29.76	29.51	30.23	57.13

This table tabulates an overview of scheduling mismatches and the total costs of the reserved patient times under the newsvendor model. We compute the reserved OR times of four different prediction models under the newsvendor model—with three different cost structures of over-allocation cost (C_o) and under-allocation cost (C_u) for each model, specified in the first two columns. Under each newsvendor cost scenario, the next three columns present the average absolute mismatch, over-allocation, and under-allocation levels in minutes for four models: the current policy (moving average and ad hoc adjustment, the default OR scheduling policy for our partner hospital), ex ante, team ex post, and dyad ex post models. The last column compares the total mismatch costs. We compute the total costs assuming that the over-allocation cost (idle time cost) is \$14.4 per minute (40% indirect cost of average total OR cost per minute in Childers and Maggard-Gibbons 2018).

Table 1.17: Total newsvendor cost breakdowns for surgical specialties (in $M)

		$C_u = 1 \times C_o$				$C_u = 2 \times C_o$				$C_u = 3 \times C_o$			
		Current Policy	Ex Ante	Team Ex Post	Dyad Ex Post	Current Policy	Ex Ante	Team Ex Post	Dyad Ex Post	Current Policy	Ex Ante	Team Ex Post	Dyad Ex Post
High PV	General	7.03	6.52	5.91	5.82	11.88	9.69	8.80	8.67	16.73	12.10	11.00	10.85
	Orthopedic	4.32	3.80	3.37	3.31	7.04	5.64	5.01	4.90	9.76	7.03	6.27	6.09
	Thoracic	1.84	1.86	1.65	1.56	2.94	2.81	2.47	2.31	4.03	3.54	3.12	2.88
	Ortho-Trauma	1.48	1.39	1.21	1.15	2.45	2.08	1.80	1.69	3.42	2.61	2.26	2.10
	Cardiothoracic	2.60	2.19	1.83	1.39	4.80	3.10	2.61	1.97	6.99	3.76	3.18	2.39
	Plastic	1.03	1.33	1.07	0.99	1.68	1.97	1.60	1.45	2.34	2.45	2.02	1.78
	Ortho-Spine	0.69	0.65	0.54	0.48	1.20	0.94	0.77	0.68	1.71	1.16	0.94	0.83
	Oncology	0.49	0.50	0.45	0.40	0.78	0.74	0.67	0.59	1.07	0.92	0.83	0.72
	Ortho-Oncology	0.22	0.19	0.17	0.13	0.39	0.27	0.26	0.18	0.57	0.32	0.32	0.23
	Ortho-Hand	0.19	0.16	0.15	0.11	0.33	0.25	0.23	0.17	0.47	0.33	0.30	0.20
	Bariatric	0.08	0.06	0.05	0.02	0.14	0.08	0.07	0.02	0.21	0.10	0.09	0.02
	Gen-Trauma	0.05	0.03	0.03	0.01	0.07	0.05	0.05	0.02	0.10	0.07	0.06	0.02
Low PV	Urology	3.52	3.29	2.98	2.83	6.11	4.90	4.40	4.20	8.70	6.14	5.48	5.24
	Gynecology	2.84	2.72	2.44	2.33	4.67	4.06	3.61	3.44	6.51	5.08	4.50	4.28
	Pediatric	2.14	2.08	1.88	1.82	3.71	3.18	2.83	2.71	5.27	4.05	3.58	3.39
	Neurology	2.54	2.47	2.19	2.02	4.31	3.63	3.22	2.97	6.08	4.51	3.98	3.67
	Vascular	2.76	3.27	2.70	2.59	4.66	4.88	4.06	3.86	6.56	6.11	5.10	4.81
	Otolaryngology	2.16	2.59	2.19	2.12	3.60	4.03	3.37	3.17	5.03	5.20	4.29	3.98
	Colorectal	1.52	1.55	1.36	1.27	2.59	2.32	2.02	1.87	3.66	2.90	2.51	2.31
	Ortho-Ped	0.30	0.28	0.25	0.21	0.49	0.43	0.37	0.31	0.68	0.54	0.46	0.39
	Edocrinolgoy	0.28	0.26	0.23	0.21	0.45	0.39	0.34	0.31	0.61	0.48	0.43	0.38
	Oral	0.29	0.27	0.26	0.22	0.47	0.41	0.39	0.33	0.65	0.51	0.48	0.40
	Ophthalmology	0.13	0.11	0.11	0.08	0.20	0.17	0.16	0.12	0.27	0.22	0.20	0.15
	Gastrointestinal	0.04	0.03	0.03	0.01	0.06	0.05	0.05	0.01	0.09	0.06	0.06	0.02
	Other	0.04	0.04	0.04	0.00	0.06	0.06	0.06	0.01	0.09	0.08	0.07	0.01
	Ortho-FootAnkle	0.03	0.04	0.03	0.01	0.05	0.05	0.04	0.01	0.07	0.07	0.05	0.01
	Radiology	0.03	0.02	0.03	0.00	0.05	0.03	0.04	0.01	0.06	0.04	0.04	0.01

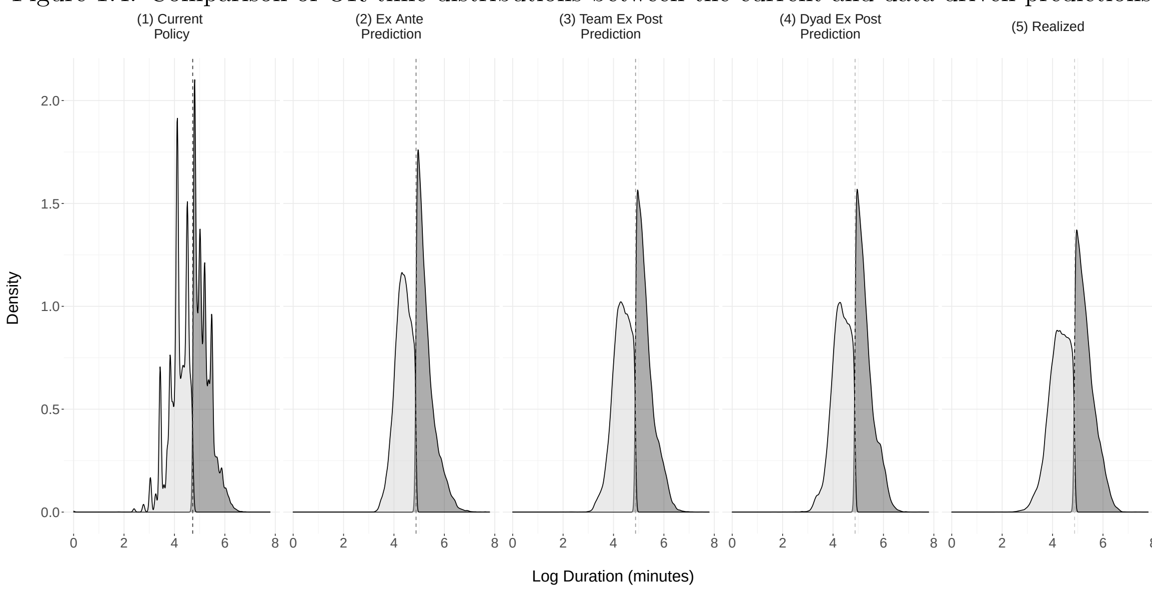
The table presents scheduling mismatches and the total costs of the reserved patient times under the newsvendor model across different specialties. We group specialties by the respective process variability (PV) group, specified in the first two columns. The first two rows indicate the three cost structures of over-allocation cost (C_o) and under-allocation cost (C_u). Under each newsvendor cost scenario, we compare the total mismatch costs (in $M) from four models: the current policy (moving average and ad hoc adjustment, the default OR scheduling policy for our partner hospital), ex ante, team ex post, and dyad ex post models. We compute the total costs assuming that the over-allocation cost (idle time cost) is $14.4 per minute (40% indirect cost of average total OR cost per minute in Childers and Maggard-Gibbons 2018).

Table 1.18: Scheduling mismatch overviews for under and over-allocations

	Absolute (minutes)		Over-allocation (minutes)		Under-allocation (minutes)	
	Current Policy	Ex Post Dyad	Current Policy	Ex Post Dyad	Current Policy	Ex Post Dyad
Ratio (%)	100.00	100.00	34.24	49.72	65.76	50.28
Mean	49.61	33.26	31.19	28.43	59.20	38.03
SD	63.41	39.54	40.79	34.37	70.56	43.54

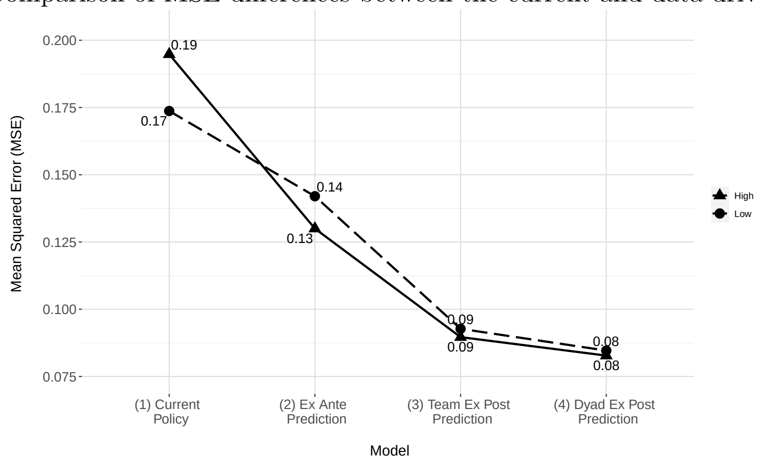
The table tabulates an overview of the actual and dyadic mismatch levels from the current policy and our dyad-level ex post prediction. We show the ratio, mean, and standard deviation of absolute mismatch (sum of over-allocation and under-allocation), over-allocation, and under-allocation levels with respect to the patient time in minutes, computed based on the current policy (moving average and ad hoc adjustment) and ex post dyad-level (predicted) models.

Figure 1.4: Comparison of OR time distributions between the current and data-driven predictions



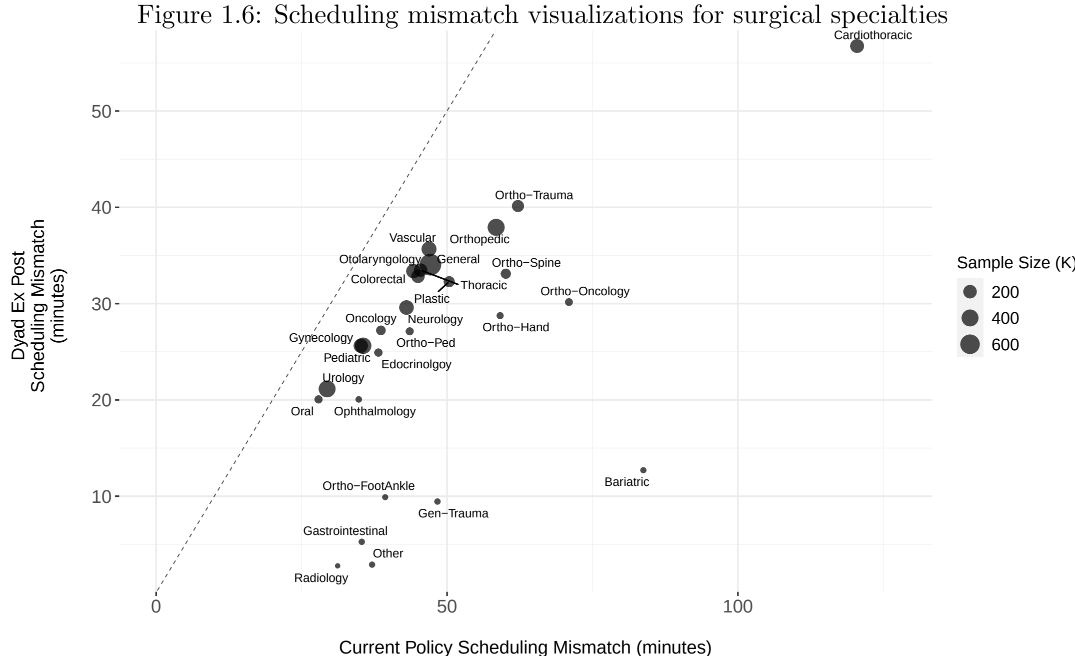
This figure compares the logged patient time distributions across five benchmarks in our analysis: (1) current policy (moving average and ad hoc adjustment), (2) ex ante, (3) team-level ex post, (4) dyad-level ex post, and (5) realized log patient times. Moving from (1) to (4), the amount and the granularity of data inputs increase: the first model involves only the ad hoc adjustment and moving average of the same surgeons' ten historical durations for the same procedure, the second model involves pre-team-selection inputs unconditional on teams, the third model extends to post-team-selection inputs conditional on teams at the team level, and the fourth model involves the most granular inputs, including post-team-selection inputs conditional on teams at the dyad level. Also note that the vertical dashed lines represent the mean value in each patient time distribution. The light-colored distribution is the distribution of OR times below the mean value; the dark-colored distribution, above the mean value.

Figure 1.5: Comparison of MSE differences between the current and data-driven predictions



This tables reports the mean squared errors (MSE) for the current policy and our data-driven models: current policy (moving average and ad hoc adjustment, the default OR time allocation policy for our partner hospital), ex ante, team-level ex post, and dyad-level ex post schedules. The dashed line indicates low process variability (PV) specialty group, whereas the solid line indicates high PV specialty group. As with Figure 1.4, moving from (1) to (4), the amount and granularity of data inputs increase: the first model involves only the ad hoc adjustment and moving average of the same surgeons' ten historical durations for the same procedure, the second model involves pre-team-selection inputs unconditional on teams, the third model extends to post-team-selection inputs conditional on teams at the team level, and the fourth model involves the most granular inputs, including post-team-selection inputs conditional on teams at the dyad level.

Figure 1.6: Scheduling mismatch visualizations for surgical specialties



This figure compares OR time scheduling mismatches with respect to patient time between the current policy (x-axis) and our dyad ex post prediction (y-axis) outcomes across different specialties. Specifically, we plot each specialty as dots, the size of which increases with sample size. The 45-degree dashed line indicates the benchmark where the current and data-driven models are equal to each other. Observations below the 45-degree reference line indicate the specialties that can benefit from our data-driven OR time allocation policies to reduce scheduling mismatch.

Chapter 2

Do New Partner and Procedure Exposure Influence Operating Room Nurse Turnover?

2.1 Introduction

Nurse turnover is a major concern for the U.S. healthcare system, threatening the provision of reliable healthcare services. High nurse turnover presents significant financial and operational challenges for hospitals. For an average hospital, the cost of turnover for a bedside registered nurse is $46,100, adding up to $5.2m-$9.0m in costs per year (NSI Nursing Solutions 2022). One way in which hospitals tackle this problem is by hiring travel nurses.[1] Yet, hiring one travel nurse, compared to a full-time nurse, costs an average hospital an additional cost of $210,184 per year (NSI Nursing Solutions 2022). More importantly, the nationwide nursing shortage is projected to be intensified with decreasing supply of the nursing workforce and increasing demand for surgical procedures (ResearchAndMarkets.com 2019, HRSA 2022). In addition, increased worker turnover is negatively associated with organizations' operational performance by decreasing remaining workers' productivity and work quality (Moon et al. 2023). In healthcare, prior research finds that the

[1] Travel nurses are registered nurses who take short-term contracts in hospitals, clinics, and other facilities to meet high healthcare demand in different parts of the country (Yang and Mason 2022, Faller et al. 2017).

turnover of clinicians results in reduced productivity, accounting for 42-66% of total turnover costs (Waldman et al. 2010).

While the reasons why nurses quit their job are diverse (Hayes et al. 2012), the factors that could create an unfavorable work environment for nurses include excessive workload, lack of teamwork, lack of growth opportunities, and unfavorable task assignments (ITA Group 2023, Bame 1993, Zaniboni et al. 2013). When facing an unfavorable work environment, nurses experience a high level of stress, leading to increased burnout and turnover (Shader et al. 2001). What contributes to this work environment of nurses is work scheduling—i.e., assigning workload, tasks, and teams in a nurse's schedule. To examine work scheduling impact on nurse turnover, a number of work scheduling studies in medical and nursing literature primarily use perceptual measures of nurse scheduling, such as perceived workload and teamwork measured by nurse surveys. While these prior studies provide important directions on how nurses' schedules can be configured, there are limited empirical insights between operational work scheduling characteristics and nurse turnover.

By contrast, a recent study in OM by Bergman et al. (2023) uses an operationally driven approach to quantify nurse scheduling. They use retrospective data on nurses' schedules and quantify schedule volatility of nurse schedule. Their study suggests that the schedule volatility of nurses increases the likelihood of voluntary nurse turnover. Nevertheless, empirical evidence about the impact of many other dimensions of nurse scheduling, such as task and partner assignments, is still limited despite the importance of nurse turnover. In this study, we focus on the operating room (OR) nurse scheduling problem of hospitals to manage OR nurse turnover. Our study is the first to document the work scheduling effect on OR nurse turnover that has been understudied in OM and nursing literature.

Examining work scheduling implications on OR nurse turnover is significant due to the unique characteristics of OR nurse schedules that are characterized by high levels of task and partner variety. Prior research finds that an increased variety of tasks and partners increases worker performance by reducing the boredom of repetitive tasks and cultivating organization-wise best practices in the long term (Staats and Gino 2012, Avgerinos and Gokpinar 2018, Akşin et al. 2021, Kim et al. 2018). However, a key question that has limited empirical evidence is whether the increased exposure to new tasks and partners, as a result of the increased variety, negatively affects worker turnover. In a complex knowledge work environment such as operating rooms, increased exposure and variety can lead to an excessive amount of information to process in a particular period (Laker et al. 2018) and increased worker stress (Misra and Stokols 2012). In turn, the increased stress can lead to a higher turnover intention of nurses (Shader et al. 2001).

To address the work scheduling implications on OR nurse turnover, we examine the impact of task and partner exposure on nurse turnover. Using our data containing granular information about task and team assignments of OR nurses and nurse departure, we aim to shed light on the following research questions: (1) How do procedure assignments influence OR nurse departure? and (2) How do partner assignments influence OR nurse departure?

To address these questions, we leverage the granularity of our OR nurse work scheduling data. Collected from a large South Carolina hospital's electronic health record (EHR) system, our data contain procedure logs and descriptions, staff assignments and logs, and surgery characteristics of 81,967 surgeries. The procedure data set includes unique identifiers classified by current procedure terminology (CPT) and the corresponding textual descriptions that differentiate a variety of procedures. The staff assignment data set contains staff identifier, staff role (e.g., physician, anesthesiologist, CRNA, surgical tech, circulator, etc.), and entry and exit logs associated with each surgery for every staff member. Lastly, our surgery data set contains surgery identifiers, perioperative timestamps, lead surgeon identifiers, and surgery characteristics.

To quantify the procedure exposure operationally, we leverage the procedure descriptions using textual analysis to compare the past and focal procedures, which is novel to the literature. Our textual analysis identifies granular differences between two sets of procedures by identifying the similarities and differences in words describing the procedures. To quantify the partner assignments, we build up on the broad team familiarity literature in OM but add a dimension of the partner exposure that has been understudied in prior research on team scheduling. Lastly, we quantify workload characteristics, motivated by an emerging stream of OM literature quantifying the scheduling characteristic of nurses. Our operationally driven data construction approach effectively captures nurse scheduling characteristics relating to diverse procedures, partners, and workload assignments.

For our estimation, we group our constructed variables into five categories. Each category is summarized by an index. These indices represent exposure, diversity, familiarity, workload volatility, and workload in nurses' schedules. These indices help us to evaluate how different of OR nurses' work environment, taken together, affect nurse departure. We use a linear probability model and estimate the impact of these five experience indices on nurse departure. One main empirical challenge for the estimation is a potential endogeneity of schedule experience indices to nurse departure. To address this endogeneity, we instrument for each of the five indices using the average indices of peer nurses working with the same lead surgeon in the same nurse role and tenure level.

Our results suggest that there are significant connections between nurse departure probability and how procedures, partners, and workload are configured in nurses' schedules. Our findings suggest that nurses' propensity to quit increases with high exposure and diversity to new procedures and partners and with high workload volatility, while decreasing with high workload in their schedules. Furthermore, these effects are significantly moderated by the seniority of nurses in the hospital. The positive impact of exposure on nurse departure probability is more pronounced with increasing seniority of nurses, while the positive effect of diversity and workload volatility gets less pronounced with increasing seniority. The negative workload effect becomes less pronounced with the increasing seniority of nurses.

Our findings provide strategic reasoning for why hospitals must pay increased attention to designing nurses' work schedules to mitigate the impact of the ongoing nursing shortage and high nurse turnover in the U.S. OR nurse schedulers should focus on balancing diverse work scheduling dimensions depending on the nurses' seniority. Also, the schedulers have to make sure that newly hired nurses maintain a certain workload that allows them to accumulate their human capital and skills and to ensure that they make enough labor income.

This paper contributes to the literature by documenting the empirical evidence of work scheduling effect on OR nurse turnover. To the best of our knowledge, this study is among the first to document the relationships between OR nurse turnover and diverse dimensions of work schedules. Our textual analysis that compares different procedures using procedure descriptions offers a novel way to operationally quantify task-related variables in a broad range of applications in OM. Furthermore, our results contribute to the literature by providing more insights about how diverse dimensions of work schedules can have different effects, across different institutions and stages of nurses' careers.

The rest of the paper proceeds as follows. Section 2.2 provides a review of the relevant literature. Section 2.3 describes our data and the variable construction to estimate the impact of procedure and partner exposure on nurse turnover. Section 2.4 presents our estimation strategy, followed by the findings in section 2.5. Section 3.5 concludes and discusses the managerial implications of our findings.

2.2 Literature Review

This paper contributes to the literature on work scheduling and its implications for worker turnover.

2.2.1 Turnover Drivers

While work scheduling is one of the classic themes of operations management (OM) literature, only a few studies examine the implications of work scheduling for worker turnover. For example, Bergman et al. (2023) operationally quantify the work schedule volatility (i.e., schedule consistency) of nursing home nurses. They find that high schedule volatility significantly increases voluntary nurse turnover and suggest several mitigation policies for scheduling volatility to reduce nurse turnover. Another related work is Ibanez (2022) which examines work scheduling implications for workforce management. Ibanez (2022) shows that offering employee-driven schedule flexibility (i.e., workers configure their own work schedule) in job descriptions attracts more job applicants. The positive impact of offering schedule flexibility is larger for temporary jobs than for permanent jobs. Furthermore, employer-driven schedule flexibility (i.e., workers are required to work on flexible schedules) rather reduces job applications. In sum, while these recent studies indicate that work scheduling significantly affects worker turnover and workforce management, only a few empirical studies in OM examine the work scheduling implications for worker turnover.

Our focus is work scheduling effect on OR nurse turnover. While there are limited insights about nurse turnover in OM literature, medical and nursing literature provides some insights about various schedule characteristics and nurse turnover, which we can leverage to operationalize and empirically investigate the turnover drivers. A few studies in nursing literature investigate how the work scheduling of nurses affects nurses' well-being and their turnover. For example, nurses are likely to leave the job if scheduling is inflexible (Leineweber et al. 2016, Wright et al. 2017); if the workload is excessive (Park et al. 2016, Bame 1993); if nurses are assigned to their preferred level of variety for skills and tasks required for the job (Zaniboni et al. 2013); if teamwork of healthcare teams is more collaborative (Al Sabei et al. 2022); if nurses have personal reasons to leave—such as family relocation, pregnancy, or career advancement (Dewanto and Wardhani 2018, Hayes et al. 2012). While these studies operationalize work schedule characteristics primarily based on the perception of nurses using surveys, interviews, and other qualitative data, they indicate that task assignments, team (and partners within teams) assignments, and how workload is distributed significantly affect nurses' intention to quit. These insights from nursing in practice require more empirical insights to be discovered, using the operational knowledge building upon OM work scheduling literature.

2.2.2 Impact of Turnover

In the literature on worker turnover, recent studies suggest that worker turnover affects the productivity and financial performance of organizations due to the resulting staffing instability and loss of organizational knowledge possessed by the leaving workers. For example, in a manufacturing context, Moon et al. (2023, 2022) show that worker turnover in a consumer electronics production system negatively affects the manufacturer's quality control outcome. Li et al. (2022) illustrate that high worker turnover results in decreased operational and financial performance of firms, measured in return on assets and future sales growth. The impact of turnover is more salient in service and knowledge-intensive operations, such as healthcare. For example, Waldman et al. (2010) show that the turnover of clinicians in a medical center results in reduced productivity and estimate the cost of reduced productivity as 42-66% of the total turnover costs. In sum, these studies collectively suggest that managing worker turnover and retention is crucial to managing worker productivity and operational performance.

Despite the significance of worker turnover on organizational performance, there are still limited empirical insights on how organizations can effectively manage worker schedules to reduce turnover. This paper contributes to the literature by addressing how organizations can strategically schedule the nurses' task and partner exposure to reduce worker turnover and mitigate the impact of worker turnover on remaining workers' productivity and organizational outcome.

2.2.3 Work Scheduling and Worker Productivity

Our study relates to work scheduling literature in OM, but we leverage existing insights in work scheduling literature to examine the work scheduling implications for OR nurse turnover. OM work scheduling literature concerns the allocation of resources—e.g., workers and machines—to meet the system demand including production and service provision. In this line of research, a broad range of analytical and empirical studies address how worker productivity changes by different schedule characteristics, such as allocating workers into a variety of tasks, groups of workers (i.e., teams), or workload.

Several studies suggest that task scheduling of workers affects worker performance. For example, Staats and Gino (2012) and Avgerinos and Gokpinar (2018) use empirical analyses to demonstrate that the degree of work specialization (i.e., working on a limited set of similar tasks), as opposed to variety, affects worker productivity. Staats and Gino (2012) show that intra-day work variety is less beneficial than work specialization for worker productivity, while inter-day work variety improves worker productivity.[2] In healthcare, Avgerinos and Gokpinar (2018) find that exposure to variety in parallel with the focal task (i.e., *concurrent variety*) benefits the surgeon's future productivity—measured by the duration of surgical procedure, while exposure to a variety independently from the focal task (i.e., *non-concurrent variety*) negatively affects the surgeon's productivity.

Other studies examine how scheduling a group of workers considering diverse worker characteristics can affect their collective productivity (e.g., team productivity), such as scheduling workers based on team's shared experience (i.e., prior experience of team members working together as a team) (Reagans et al. 2005, Avgerinos and Gokpinar 2017), worker's partner exposure (Akşin et al. 2021), and team diversity (Huckman and Staats 2011). For example, Reagans et al. (2005) and Avgerinos and Gokpinar (2017) show that increased shared experience of team members in surgical teams improves the team productivity measured in surgery duration. In a later study, Akşin et al. (2021) show that increased worker's prior partner exposure (i.e., the dispersion of shared experience across all prior partners) in ambulance transports can improve operational performance by allowing the workers to learn the best practices from interacting with various partners. Lastly, Huckman and Staats (2011) find that the diversity across team members in customer experience within software development teams decreases team productivity when tasks frequently change.

Lastly, a number of studies investigated how designing work schedule (i.e., distributing workload) affects worker productivity, including workload (KC et al. 2020a, KC and Terwiesch 2009), schedule volatility (i.e., schedule inconsistency) (Lu et al. 2022), intra-day scheduling (Ibanez and Toffel 2020), and responsible scheduling practice (i.e., schedule consistency, predictability, adequacy, and control) (Kesavan et al. 2022). Below, we discuss the workload and schedule volatility in detail, which are more relevant to our study. Studies that focus on the impact of workload on worker productivity indicate that workload and schedule volatility in a worker's schedule affects the

[2] In a Japanese bank's home loan application-processing line.

worker's productivity. For example, KC and Terwiesch (2009) show that a high workload temporarily improves hospital service workers' productivity, but the productivity gain does not last in the continuing high workload. KC et al. (2020a) show that physicians tend to select easier tasks as workload increases, which worsens the throughput and decreases physician productivity in the long term. Furthermore, Lu et al. (2022) show that low schedule volatility (i.e., schedule consistency) improves worker productivity. These studies focus on the relationship between work scheduling and work productivity and suggest that work scheduling significantly affects operational performance.

Our study adds to the work scheduling literature by examining the work scheduling implications for OR nurse turnover, leveraging the existing insights to operationalize the work schedule characteristics of OR nurses. This study makes several contributions to OM work scheduling and nurse turnover literature. First, we fill the gap in the OM work scheduling literature by operationalizing diverse work schedule characteristics of nurses with high levels of task and partner variety. Using novel textual analysis and operationally driven approaches, we quantify OR nurses' schedule characteristics regarding the diverse tasks, partners, and workload assigned to nurses' schedules. Second, we propose how OR nurse schedule characteristics can be operationalized using the retrospective schedule data of OR nurses to aid practitioners' decision-making. Motivated by an emerging stream of literature in OM work scheduling that operationalized the nurse schedule characteristics using the retrospective data (Bergman et al. 2023, 2022), we expand the scope of schedule characteristics to reflect the diverse tasks and partners given to OR nurses. In quantifying the task and partner assignments, we consider the level of exposure to diverse tasks and partners, which has been understudied in the current work scheduling literature. Lastly, by considering diverse dimensions of tasks, partners, and workload assignments, we show how these work schedule characteristics distinctively affect nurse turnover, which can provide practical insights for nurse scheduling and workforce management of hospitals.

2.3 Data and Variables

2.3.1 Data

We collected our primary data from a large South Carolina hospital's electronic health records (EHR).[3] The data show granular dimensions of nurse schedule characteristics assigned to

[3]Our research project is under an institutional review board (IRB)'s exempt approval for human subject research.

OR nurses, including surgical procedures, surgical team partners, and workload. This granularity of our data allows us to operationalize diverse dimensions of nurse schedule characteristics used for our primary independent variables of interest as well as control variables that we use to estimate the impact on OR nurse turnover.

More specifically, our raw data sets include information about procedures, staff assignments, and surgery characteristics of 81,967 surgeries across 27 surgical specialties from March 2016 to June 2019. Table 2.1 gives an overview of our raw and reconstructed data sets, including the number of observations and measurement level, followed by our major variable categories relating to how the information contained in each data set allows us to measure the nurse schedule characteristics. The procedure data set includes unique surgical procedure identifiers classified by current procedure terminology (CPT) of the American Medical Association[4] and the corresponding textual descriptions that differentiate distinct procedures. The staff assignment data set contains staff identifier, staff role (e.g., physician, anesthesiologist, CRNA, surgical tech, circulator, etc.), and each staff's entry and exit logs of all surgical team members associated with each surgery. Lastly, our surgery data set contains surgery identifiers, perioperative timestamps, lead surgeon identifiers, and surgery characteristics (e.g., add-on,[5] robotic, specialty) associated with each surgery. In addition, we complement our procedure data set and gain more accurate medical descriptions of surgical procedures by the official procedure description data from the CPT Codes Lookup of the American Academy of Professional Coders (AAPC).

We construct our sample by merging four sources of data: AAPC, surgery, procedure, and staff data sets. Each staff scheduling and procedure record corresponds to a unique surgery. Merged by unique surgery identifiers, these nurse-surgery level records capture the entry and exit logs of each staff in the OR and the procedures and partners (e.g., surgical team members) involved in each surgery. We use the complementary AAPC data set to replace the original procedure description records with the official description of CPT, which allows us to correct any typos and inconsistencies within similar types of surgeries in the original procedure data set.

In our data sets, we do not fully observe all nurses' true tenure (i.e., the number of months from the initial month of a nurse's appearance in the data set) because we do not have the job

[4] CPT code is the standard American Medical Association classification for surgical procedures, used in reporting medical, surgical, and diagnostic procedures (Dotson 2013)

[5] Add-on surgeries are surgeries scheduled on the day of surgery with a surge of demand after the scheduling job is finished the day before the surgery (Zhou and Dexter 1998).

appointment dates. To address this challenge, we make an assumption that we can fully censor the tenure of nurses whose work-start month is beyond the first two months since the beginning of the observation period of data (i.e., later than May 1, 2016). Hereafter, we refer to these nurses with fully censored tenures as *new nurses.* We refer to all other nurses as *incumbent nurses.* The assumption is sensible because it is unlikely that a nurse would take a break from their work for more than two months. Thus, in our empirical analysis, we focus on the new nurses only and exclude the nurse-surgery observations for which the nurses' tenure is not fully censored. After dropping the observations of incumbent nurses, we retain 103,608 nurse-surgery observations for our estimation.

Using our granular data, we construct our outcome variable, nurse departure, and main predictor variables about procedure, partner, and workload assignments of nurses. Below, we define and describe our major variables including our outcome and main predictor variables in more detail.

2.3.2 Outcome Variable

Our outcome variable, *Departure*, is a binary variable indicating whether the nurse departed from the OR within six months from the month of focal surgery. While we do not observe human resources records indicating whether a nurse quit, we observe complete records of nurses' time spent in the OR during the data set period. Hence, we assume that nurses departed from their OR position if one's last observation is before the last two months of the data set. We capture the six-month leading outcome because the timing of a nurse's decision to quit is not observable and because the impact of past schedules is likely to accumulate over time, thus influencing a nurse's decision to quit. Our specification of the outcome variable is similar to that of Bergman et al. (2023). They specify the turnover outcome as positive from the day of observed quitting to the prior 14 days of observed quitting. In contrast, we take a more extended period of lead time (six months) and take into account the long-term accumulation of past schedule experience of nurses that affects the decision to quit.

Table 2.2 summarizes the yearly nurse departures in our data set and the yearly surgery volume, including the number of departures, the number of days, the number of surgeries, and the staffing level of new and incumbent nurses. Figure 2.1 plots the nurse-level departure rate against their tenure for both incumbent and new nurses. This figure shows that almost 80% of new nurses depart within twelve months of their employment. In addition, we observe that new nurses in our data set are more likely to depart earlier in their tenure. Table 2.3 compares the count, ratio, and average nurse tenure for the last observation in the data set of departing and staying nurses within each OR nurse role (e.g., circulator, surgical tech, CRNA). From this table, we observe that about 40% of new nurses depart from the OR when their tenure is between 5 and 9 months, showing a significant turnover rate in the early tenure of nurses.

2.3.3 Main Predictor Variables

Next, we proceed by constructing our primary independent variables of interest to measure nurses' schedule characteristics regarding the nurses' procedures, partners, and workload. In measuring the schedule characteristics, we build up on the existing insights in work scheduling literature and propose methods to quantify nurses' procedure, partner, and workload schedule characteristics using retrospective data collected through the hospitals' EHR. After constructing schedule characteristic variables described in detail below, we cluster them into five indices of scheduling experiences that are likely to have different impacts on the nurse's departure intention: (1) exposure to procedures and partners, (2) diversity of procedures and partners, (3) familiarity with procedures and partners, (4) workload volatility, and (5) workload. Below, we describe our variable construction for schedule characteristic variables and how we cluster the schedule characteristics into five experience indices.

2.3.3.1 Procedure assignment characteristics

We measure three distinct procedure assignment characteristics every nurse is exposed to: procedure exposure, procedure familiarity, and procedure diversity. We measure these characteristics by comparing the procedures that a nurse performs in a focal surgery to the procedures in the nurse's recent surgeries within three months from the month of the focal surgery. The procedure exposure measures how new the focal procedures are to the nurse based on the similarities with the past procedures the nurse had performed. The procedure familiarity measures how familiar the focal procedures are to a nurse based on the number of past surgeries involving identical procedures. Lastly, procedure diversity measures how diverse the past procedures performed by the nurse were on average, compared to the focal procedures. A simple way to measure these three procedure assignment characteristics might be to compare a nurse's past and focal procedures by counting the unique identifiers of new procedures the nurse performs in a surgery. However, a major limitation of this method is that it neglects granular differences between two sets of procedures. Even two procedures with two distinct identifiers may have similarities if they belong to a broader category of procedures. For example, two procedures classified by two different CPT codes, 14000 and 14001, have almost identical textual descriptions with a minor difference in words that describe the size of tissue defects removed by the procedures.[6] To accurately account for the similarity and differences between the two sets of procedures, we leverage our granular procedure description records using a textual analysis in a novel way.

To specify our approach of comparing two sets of procedures, consider a nurse involved in a surgery. Each surgery has one or more procedures with textual descriptions classified by CPT. The goal is to compare the procedures in a surgery to the procedures a nurse had performed within the previous three months, in terms of how different their textual descriptions are from each other. The first step in achieving this goal is to extract all words describing the procedures of each surgery and all corresponding past surgeries the nurse had performed before the focal surgery. Before the comparison, we clean these words in four distinct stages. In stage 1, we remove the words that are not meaningful in natural language processing, such as determiners, coordinating conjunctions, and propositions. In stage 2, we classify words into three word families: "general", "rare", and "specialty" words, depending on their total appearances and the concentration of appearances across different specialties (see subsection A.1 in our appendix for details). We classified a word as "rare"—if a word appeared in the data set less than 50 times; as "specialty"—if a word is not "rare" and the word appears across at most two surgical specialties; as "general"—if a word is neither a "rare" nor a "specialty" word. Next, in stage 3, we only keep medical words found in the Unified Medical Language System (UMLS)—"an online meta-thesaurus of controlled vocabularies of biomedical terminologies developed by the U.S. National Library of Medicine (NLM) (Huang et al. 2019)." Lastly, in stage 4, we remove redundant words in each surgery's procedure descriptions. Table 2.4 shows the surgical specialty level statistics of word appearances, including the unique word count after each cleaning stage and the proportion of unique words and total words after stage 4 by each word family.

Using the cleaned words describing each surgery's procedures, we compare words of a surgery a nurse performed, to words of each past surgery the nurse has performed within the past three months from the month of focal surgery. The comparison begins with computing the differences in words describing two sets of procedures in two surgeries by counting the number of words not shared

[6]The AMA official textual description of the CPT code 14000 is "Adjacent tissue transfer or rearrangement, trunk; defect 10 sq cm or less" whereas that of the CPT code 14001 is "Adjacent tissue transfer or rearrangement, trunk; defect 10.1 sq cm to 30.0 sq cm".

between the two sets of words for each word family (e.g., rare, specialty, general). To formalize our variable construction for procedure assignments, consider a nurse n in a surgery s. Then, for each surgery s and for each word family w, we have the list of differences between the focal surgery and the past N_{nsw} surgeries the nurse had performed, $\{d_{nsw1}, d_{nsw2}, ..., d_{wsN_{nsw}}\}$. Figure A.1 in our appendix presents the example computation of word difference between a past surgery and a focal surgery. Using these differences between the focal surgery and past surgeries, we compute our three procedure assignment characteristics—procedure exposure, familiarity, and diversity—as follows:

$$E_{nsw}^{Procedure} = \min_i(d_{nswi}), \tag{2.1}$$

$$F_{nsw}^{Procedure} = \sum_{i=1}^{N_{nsw}} \mathbb{1}[d_{nswi} = 0], \tag{2.2}$$

$$D_{nsw}^{Procedure} = \frac{\sum_{i=1}^{N_{nsw}} d_{nswi}}{N_{nsw}}, \tag{2.3}$$

where $\mathbb{1}[\cdot]$ is the indicator function.

For each word family, the procedure exposure indicates how close the most similar recent surgery is to the focal surgery in terms of procedure words. A low value of procedure exposure indicates that the nurse had performed at least one similar surgery recently within three months prior to the focal surgery, and a zero value of procedure exposure indicates the nurse had performed an identical set of procedures in at least one past surgery. The procedure familiarity measures the degree of familiarity with the focal set of procedures by counting the number of past surgeries with the identical set of procedures if there were any. Lastly, procedure diversity measures the overall diversity of past procedures compared to the procedures in the focal surgery.

2.3.3.2 Partner assignment characteristics

For partner assignment characteristics, we measure nurses' three distinct partner assignment characteristics: partner exposure, partner diversity, and partner familiarity. We measure these characteristics based on a nurse's past three-month joint experiences with partners (i.e., team members) in a surgery. The partner exposure measures whether the nurse was exposed to each of the current partners during the recent three-month surgeries. The partner familiarity measures how familiar the partners were to the nurse if they worked with the nurse at least once in the recent three months. Lastly, partner diversity measures the diversity of partners the nurse worked with in all recent three-month surgeries.

To formalize our variable constructions for partner exposure and partner familiarity, we first compute the joint experiences of a nurse with her partners in a given surgery. In a typical surgical team, a nurse works with partners in diverse professionals: a surgeon, an anesthesiologist, a certified registered nurse anesthetist (CRNA), a surgical tech (ST), and a circulating nurse (CN). Surgical teams are fluid, meaning that the team members are drawn differently across teams rather than having one fixed team with the same partners. We consider a nurse n, in surgery s, working with a set of partners $\{1, 2, ..., N_{ns}\}$. We define the joint experience for a nurse n and a partner i as the number of prior surgeries where the nurse and the partner worked together in the past three months from the month of surgery s, [7] denoted by J_{nis}. In case there are multiple individuals in one role assigned to the surgery, we get the average joint experience between the focal nurse and partners in each role r: $J_{nrs} = \frac{1}{N_{nrs}} \sum_{i=1}^{N_{nrs}} J_{nis}$, where N_{nrs} is the number of the nurse n's partners in role r in the focal surgery s. Note that we differentiate the joint experience across partners in different roles. This way allows us to account for different partner experiences that a nurse could have due to the dynamics in their interprofessional relationships (Lingard et al. 2012).

To measure partner exposure and partner familiarity, we use this joint experience. For each nurse n in surgery s for the partners' role r, the partner exposure and familiarity are given by:

$$E_{nrs}^{Partner} = \mathbb{1}[J_{nrs} = 0], \tag{2.4}$$

$$F_{nrs}^{Partner} = J_{nrs}, \tag{2.5}$$

where $\mathbb{1}[\cdot]$ is the indicator function. Zero value of partner exposure indicates that the nurse had no joint experience with partners in role r within three months prior to the focal surgery. The partner familiarity measures the average familiarity with partners in role i in the number of recent three-month surgeries a nurse performed with the partners if there were any. Lastly, to measure the partner diversity, we compute the number of unique partners nurse n worked with in the past three

[7] The procedure is similar to Reagans et al. (2005) before the aggregation of dyadic shared experience at the team level.

months from the month of surgery s, for each role r. We denote this partner diversity as $D_{nrs}^{Partner}$. A high value of partner diversity indicates that the nurse had worked with diverse partners in role i within the past three months.

Furthermore, we construct an additional measure of partner familiarity that measures how familiar each nurse was with other nurses in terms of their independent joint experiences with different lead surgeons in the prior three months. Lead surgeons play a significant role in setting up the OR culture and coordination norms. For example, a lead surgeon formally sets up the OR norms by using a physician preference card in advance or informally by communicating with the rest of the team members as a leader of the entire surgery. We assume that nurses learn from prior experiences under these distinct OR norms set by different lead surgeons. If two nurses have had similar experiences working with a similar set of lead surgeons, these two nurses's expectations about OR norms are likely to be similar to each other, allowing them to coordinate seamlessly under a shared understanding of OR norms.

We call this additional partner familiarity *network familiarity* of partner nurses. To measure the network familiarity, we first obtain the prerequisite variable representing the difference between two nurses in their joint experiences with all lead surgeons who appeared in all past three-month surgeries. To formalize it, consider a surgery s and an absolute difference between a focal nurse n and a partner nurse j of their joint experience with surgeon i, $||J_{nis} - J_{jis}||$, where $||\cdot||$ is the distance measure. We define the total difference by summing up the absolute difference with all surgeons, d_{nsij}, by:

$$d_{njs} = \sum_{i=1}^{N_s^{surgeons}} ||J_{nis} - J_{jis}||, \text{ for all } i \in \{1, 2, ..., N_{ns}^{nurses}\}, \tag{2.6}$$

where $N_s^{surgeons}$ is the number of all lead surgeons found in all three-month past surgeries from the month of surgery s, and N_{ns}^{nurses} is the number of partner nurses for the nurse n in surgery s. Then, we take an average of d_{njs}, across all partner nurses in surgery s. We denote this average nurse-nurse distance measure as SND_{ns} and refer to it as *surgeon network distance.* Formally, the surgeon network distance is given by:

$$SND_{ns} = \frac{\sum_{j=1}^{N_{ns}^{nurses}} d_{njs}}{N_{ns}^{nurses}}, \tag{2.7}$$

To measure the additional partner familiarity, we reverse the value of the surgeon network distance, SND_{ns}. The network familiarity of nurse partners for nurse n in surgery s, $F_{ns}^{Network}$, is given by:

$$F_{ns}^{Network} = \max_{ns}(SND_{ns}) - SND_{ns}. \tag{2.8}$$

A low value of network familiarity indicates that the expectation of OR norms of nurse n is likely to be dissimilar to those expectations of partner nurses in surgery s.

2.3.3.3 Workload and Workload Volatility

To measure the workload assignment characteristics of nurses, we consider two distinct workload characteristics of nurses: (1) total workload and (2) workload volatility. To measure the total workload, we employ three variables that represent the weekly workload characteristics of nurses: the weekly number of shifts, the weekly total work hours of all shifts (i.e., the weekly sum of daily work hours between the first surgery start time and the last surgery end time of a workday), and weekly overtime.[8] For nurse n in week w, we denote these three weekly workload variables as W_{nw}^{Shifts} (weekly number of shifts), W_{nw}^{Hours} (weekly total work hours), and $W_{nw}^{Overtime}$ (weekly overtime).

To measure workload volatility, we construct three distinct workload volatility variables: daily work hour volatility, daily shift start volatility, and weekly overtime volatility. Volatility in work schedules can increase work-family conflict, elevating worker's stress from their job(Bergman et al. 2023). To measure the daily work hour volatility and weekly overtime volatility within a month, we adapt Bergman et al. (2023)'s approach that quantifies schedule volatility by the coefficient of variation (COV) of a workload measure.[9] The daily work hour volatility in month m for nurse n is given by $V_{nm}^{Workload}$, for a nurse n in month m by

$$V_{nm}^{Workhours} = \frac{\sqrt{\frac{1}{N_{nm}^D}\sum_{d=1}^{N_{nm}^D}(DW_{nmd} - \overline{DW}_{nm})^2}}{\overline{DW}_{nm}} \tag{2.9}$$

[8] To define the number of hours in overtime, we use the definition of overtime contained in the Fair Labor Standards Act (FLSA) described in the U.S. Department of Labor's webpage on overtime pay: https://www.dol.gov/agencies/whd/overtime. While 40 hours is the threshold to determine the overtime, we use 35 hours as a threshold because we do not observe the times between the shift start and the first surgery start and between the last surgery end and the shift end.

[9] Bergman et al. (2023) examined the schedule volatility of nurses in nursing homes, whereas our study focuses on operating room nurses.

where DW_{nmd} is the daily work hours on workday d in month m for nurse n, $\overline{DW}_{nm}$ is the monthly mean of DW_{nmi}, and N^D_{nm} is the total number of workdays the nurse worked on in month m. Similarly, the weekly overtime volatility in month m for nurse n is given by:

$$V_{nm}^{Overtime} = \frac{\sqrt{\frac{1}{N^D_{nm}} \sum_{d=1}^{N^D_{nm}} (WO_{nmd} - \overline{WO}_{nm})^2}}{\overline{WO}_{nm}} \tag{2.10}$$

where N^D_{nm} is the total number of weeks nurse n worked for in month m, and WO_{nmd} is weekly overtime on day d that belongs to a week in month m for nurse n. Weekly overtimes for all days that belong to a week are the same across the days. To account for the different number of days in a week that belongs to different months, we define the mean weekly overtime, $\overline{WO}_{nm}$, as the monthly weighted mean of WO_{nmd} with weights as the number of workdays that belong to month m for nurse n, $\frac{1}{N^D_{nm}} \sum_{w=1}^{N^D_{nwm}} (WO_{nwm} * N^D_{nwm})$, where N^D_{nwm} denotes the number of workdays for nurse n in week w that belongs to month m.

Lastly, we measure the shift start volatility by the standard deviation of shift start hours of workdays in month m for nurse s. Formally, we define the shift start volatility for nurse n in month m, V_{nm}^{Start}, by:

$$V_{nm}^{Start} = \sqrt{\frac{1}{N^D_{nm}} \sum_{d=1}^{N^D_{nm}} (H_{nmd} - \overline{H}_{nm})^2} \tag{2.11}$$

where H_{nmd} is the hour of first surgery start time on workday d in month m for nurse n, $\overline{H}_{nm}$ is the monthly mean of H_{nmi}, and N^D_{nm} is the total number of workdays the nurse worked on in month m. If the value is close to 0, it means that the start hours of the nurse's shifts in a given month were relatively consistent throughout the month. On the other hand, if the value is far from 0, it means that the shift start hours were relatively inconsistent.

2.3.3.4 Experience indices

In our empirical strategy, we cluster our constructed partner, procedure, and workload assignment variables by five experience indices: (1) exposure to procedures and partners, (2) diversity of procedures and partners, (3) familiarity with procedures and partners, (4) workload volatility, and (5) workload. We assume that the clustered variables in each index are likely to have similar directions of effects on nurse departure intention. Exposure, familiarity, and diversity measure the degree of information processing requirement of nurses; and workload and workload volatility measure the degree of difficulty a nurse has to experience due to their workload distribution. To construct the index variables, we get a weighted average of the variables within each experience index. Table 2.5 describes the granularity, measurement time horizon, the number of variables aggregated as a weighted average for each index, and the description of variables. Table 2.6 shows summary statistics of independent variables following the column specifying the within-index weights used to construct a weighted average experience index (e.g., exposure, diversity, familiarity, volatility, workload).

In determining within-index weights of partner, procedure, and workload assignment variables, we make several assumptions about the relative impact of these variables to represent each experience index. Broadly, if an index includes both procedure and partner assignment variables, we set equal weights to two sets of variables—procedure and partner sets (i.e., each set is assigned 50% of the weight out of 100%, the within-index total weight). For procedure assignment variables, we assume that the difficulty of processing procedure information is greatest for rare words, followed by specialty words and general words. Thus, we put the highest weight on rare words. For partner assignment variables, we assume that the exposure to the surgeon is most impactful, followed by the anesthesiologist and nurses. It is because a surgeon serves as an explicit and implicit leader of surgical teams, exercising the greatest influence on teamwork. For the rest of the non-surgeon team members, we assume that the relative impacts follow the traditional OR hierarchy structure (Avgerinos et al. 2020, Lingard et al. 2012). We denote these weighted averages for five experience indices for nurse n in surgery s as E_{ns}^{wa}, D_{ns}^{wa}, F_{ns}^{wa}, V_{ns}^{wa}, and W_{ns}^{wa}. In our empirical strategy, we use these index variables to examine the impact of work scheduling on nurse departure.

To prevent extreme values from affecting our estimation and match the scales of index variables, we clean and unit-scale the workload, procedure, and partner assignment variables before computing these experience indices. To clean the schedule characteristics, we winsorize the extreme values in order to prevent outliers (mostly due to data entry errors). Then, we unit-scale our continuous variables by dividing each observed value by the maximum value of all observations within each variable.

2.3.4 Control Variables

Our control variables include a range of variables to account for surgery and nurse characteristics including role, tenure, time/location fixed effects, and gaps in work history. In our appendix,

Table A.1 and Table A.2 detail the control variables included in our analysis.

2.4 Estimation Strategy

In this section, we describe our empirical approach to (1) estimate the effect of work scheduling on nurse departure, (2) discuss the potential endogeneity of the experience indices, and (3) describe our instrumental variable (IV) approach to address the potential endogeneity.

2.4.1 Model

To estimate the effect of work scheduling on nurse departure within the 6 months from the focal month m, we use the following specification of the linear probability model (LPM):

$$Departure_{n(m+1:m+6)} = \alpha + \beta_i I^{wa}_{ins} + \omega_i I^{wa}_{ins} \times Tenure_{ns} + Controls_{nsdm} + \epsilon_{insdm}, \tag{2.12}$$

where $Departure_{n(m+1:m+6)}$ indicates whether nurse n departed the OR within the previous six months from the focal month m (i.e., in any month from $m+1$ to $m+6$). I^{wa}_{ins} is experience index i for nurse n in surgery s, among indices for exposure, diversity, familiarity, workload volatility, and workload. $Controls_{nsdm}$ denotes control variables specified in Table A.1 and Table A.2 for nurse n in surgery s. We include the subscripts d and month m in the equation to indicate different granularity levels of our control variables including nurse-day and nurse-month levels. For example, nurse-day level control variables—such as holiday and weekend indicators—and nurse-month level control variables—such as the nurse's tenure group and work history gap—correspond to values on the day and in the month of a focal surgery. In addition, we include an interaction term between each index variable and monthly tenure of observation, $Tenure_{ns}$, as we expect that the effect of each experience index will vary their tenure that represents different stages of nurses' careers. We opt for an LPM instead of a logit or a probit model. It is because the interpretation of estimates is more intuitive. In addition, previous studies suggest that probit and logit models often make similar predictions when estimating marginal effects (Bergman et al. 2023, Angrist and Pischke 2009).

The potentially more important issue is the endogeneity of schedule characteristics: nurses' personal reasons affecting both their schedule characteristics and departure probability. We focus on it next.

2.4.2 Personal Reasons for Departure

Nurses' schedules are potentially endogenous if nurses have personal reasons to influence their own schedules due to an anticipated departure in the near future. Personal reasons—such as marriage, pregnancy, relocation following a spouse or family, or career advancement—are the top reasons for registered nurse resignations (Dewanto and Wardhani 2018, NSI Nursing Solutions 2023).

In our data, we do not observe these personal reasons for resignation. These personal reasons can drive both nurses' schedules and their departure in several ways. For example, when nurses have personal reasons for departure, they may attempt to reduce their workload and avoid the burden to learn new procedures or work with new partners. Alternatively, nursing schedulers who got a notice of resignation from these nurses may reduce workload or the number of new procedures in the nurses' schedule. Either of these two could also limit the variety of the nurse's partners. On the other hand, if the nurse expects to leave due to a career advancement opportunity, its effect on procedure, partner, and workload assignments is less clear. For example, whether these nurses would focus on specialization or generalization of their skills for their advanced career opportunities will change the nurses' selection behavior for the degree of procedure and partner variety. Furthermore, if the career opportunities are for management-level positions, they may attempt to reduce the workload for management-related skills rather than skills for performing surgical procedures. Thus, these unobserved personal reasons can drive both changes in the nurses' schedules and departure outcomes, but how these personal reasons will bias the coefficients estimated via ordinary least squares (OLS) is not clear.

To address the potential endogeneity of nurse schedules due to personal reasons, we adopt an instrumental variable (IV) approach. We instrument for each of the five experience indices at the nurse-surgery level using an average index value of peer nurses in similar surgeries. Our approach builds upon a prior study that used peer effects to instrument for schedule volatility in order to estimate the impact of schedule volatility on nurse departure (Bergman et al. 2023). We define peer observations of a nurse n in surgery s as all nurse-surgery observations matched with the nurse's role and tenure and the surgery's lead surgeon. These matching variables reflect the similarity of nurse's employment status over their employment window and the similarity of OR context set by the lead surgeon.[12] We define the peer average experience index, $\overline{I}^{p}_{ins}$, for nurse n in a surgery s for

[12]Defining the OR context by the lead surgeon also reflects the context of the surgical specialty, as most lead surgeons specialize in a particular surgical specialty.

each index i as follows:

$$\overline{I}^p_{ins} = \sum_{j=1}^{N^p_{ns}} \frac{I_{insj}}{N^p_{ns}} \tag{2.13}$$

where N^p_{ns} is the number of total matched nurse-surgery peer observations, and I_{insj} is the value of experience i for peer observation j matched with the characteristics of nurse n in surgery s.

Next, we discuss whether the proposed instruments—peer average indices—satisfy two assumptions for an IV estimation: instruments should (1) predict each of own experience index of the focal nurse (i.e., relevance) and (2) have no causal effect on nurse departure (i.e., validity). To address the relevance of our instruments, we estimate the following OLS regression for each measure of own experience index:

$$I_{ins} = \zeta + \boldsymbol{\gamma}^T \overline{\boldsymbol{I}}^p_{ns} + \theta_i \overline{I}^p_{ins} \times (N^{Peers}_{ns}) + \kappa(N^{Peers}_{ns}) + \upsilon_{ins}, \tag{2.14}$$

where I_{ins} is the own index level of type i (e.g., exposure, diversity, familiarity, volatility, and workload) for nurse n in surgery s, and the $\overline{\boldsymbol{I}}^p_{ns}$ is a vector of five peer average indices for nurse n in surgery s. N^{Peers}_{ns} is the number of peers used to compute the peer average indices for nurse n in surgery s. Note that we include the interaction terms between the peer average indices and the number of peers. It is to capture the accurate peer average effect on an own experience index because a larger peer group would result in a more accurate prediction of an own experience index.

For the estimation sample, we log-transform a variable if the variable's distribution is skewed and if the log transformation lessens the skewness. Appendix A.4 presents the scaling of all our independent variables and continuous control variables. In addition, we filter our data depending on whether we can capture nurses' prior three-month experience and leading six-month departure. To fully capture the nurses' prior three-month experience and leading six-month departure, in our empirical analysis, we only use nurse-surgery observations of which the nurses' tenure is greater than three months and of which the month is prior to the last six months of the data set (i.e., months before January 1, 2019). In addition, we drop the nurse-surgery observations if there is no peer to compute peer average,[13] no procedure descriptions to construct the procedure assignment variables, and no partners to construct the partner assignment variables. After dropping these observations, we retain 61,669 nurse-surgery observations for our estimation. Appendix A.5 presents the sampling

[13]Figure 2.2 illustrates the average number of peers observed at each tenure in months. From this table, we observe that the average number of peers in nurse-surgery observations decreases in the tenure of the focal nurse.

of our nurse-surgery observations.

Table 2.7 presents the first-stage estimation results of each own index on peer average instruments. From this table, the large F-statistic associated with the first-stage estimations for all models indicates that the instruments, $\overline{\boldsymbol{I}}^{p}_{ns}$, together with interaction terms of the number of peers, explain each endogenous index variable, I_{ins}, well. Furthermore, we conclude that our constructed instruments have no causal association with nurse departure because our matching procedure matches all observations in the entire data set rather than the observations at a particular time. Because each peer average index is a peer average throughout the entire period of data, it is not likely that any of the peer average indices affects the nurse departure in a particular month.

2.4.3 Instrumental variable (IV) approach

In sum, our two-stage least squares (2SLS) model is defined by:

$$I_{ins} \;=\; \theta \;+\; \boldsymbol{\gamma}^T \overline{\boldsymbol{I}}^{p}_{ns} \;+\; \boldsymbol{\theta}^T \overline{\boldsymbol{I}}^{p}_{ns} \times (N^{peers}_{ns}) \;+\; \zeta(N^{peers}_{ns}) \;+\; \upsilon_{ns}, \tag{2.15}$$

$$Departure_{n(m+1:m+6)} \;=\; \alpha \;+\; \beta_i \widehat{I}_{ins} \;+\; \omega_i \widehat{I}_{ins} \times Tenure_{ns} \;+\; Controls_{nsdm} \;+\; \epsilon_{nsdm} \tag{2.16}$$

where I_{ins} is the own index value of index i for nurse n in surgery s, and $\overline{\boldsymbol{I}}^{p}_{ns}$ is a vector of peer average instruments. N^{peers}_{ns} is the number of peers used to compute the peer average for nurse n in surgery s. $\widehat{I}_{ins}$ is the predicted own index value of index i for nurse n in surgery s. $\widehat{I}_{ins} \times Tenure_{ns}$ is an interaction term between each index i and monthly tenure of observation. $Controls_{nsdm}$ denotes control variables for nurse n in surgery s associated with day d and month m.

2.5 Results

Table 2.8 summarizes our instrumental variable estimation results on nurse departure described in Equation 2.16. First, from the results, we find that the coefficient of the interaction term between exposure index and nurse tenure is positive and significant, suggesting that the probability of nurse departure increases as the nurse tenure in the hospital increases. We find these results sensible due to the increased information a nurse has to process when the exposure to new procedures and partners increases during a particular period. Especially, the increasing positive effect of exposure by increasing nurse tenure is plausible if these nurses with low tenure are in their early careers. Compared to experienced nurses, early-career nurses tend to seek growth opportunities in their careers to develop their skills in performing various procedures and learning from diverse partners. While increased exposure to new procedures and partners may be stressful, it also allows early-career nurses to learn the procedures and how to work with the partners. Thus, early-career nurses are more likely to accept the responsibilities to work on new procedures and with new partners while the nurse's willingness to process new information decreases with increasing tenure.

Second, the coefficient of the diversity index is significant and positive, while the coefficient of its interaction term with nurse tenure is significant and negative. These results suggest that the probability of nurse departure increases with the diversity in a nurse's procedures and partners, but the positive effect of diversity becomes less pronounced when the nurse tenure increases. The diversity in tasks may result in a higher level of worker motivation due to reduced boredom from repeating similar tasks. However, for operating room nurses, the amount of information they have to process is high by default due to the assignments of various procedures and partners. Thus, any increase in diversity beyond the average can significantly decrease their intention to stay in their current positions. For the interaction effect between the diversity index and tenure, one explanation is that diversity in an early career represents a high level of exposure to new procedures and new partners, but the diversity in later stages of tenure represents the diversity of both familiar and new procedures and partners. Thus, the diversity effect may become less pronounced due to different levels of the newness of procedures and partners—that determines the level of diversity—throughout the nurses' tenures. Third, the coefficients of the familiarity index and the interaction term between the familiarity index and nurse tenure are not significant, suggesting that the familiarity with procedures and partners might not be a critical determinant of nurse departure.

Next, we find that the coefficient of workload volatility is positive and statistically significant. This means that nurses are more likely to depart if their workload is more volatile across days. The results are consistent with prior findings about the workload volatility effect of Bergman et al. (2023). What is different in our study is that we observe that the positive impact of workload volatility becomes less pronounced when the nurse tenure increases. One possibility is that the workload volatility might represent schedule flexibility allowed to the nurses when they get seniority in the hospital (i.e., employee-driven schedule flexibility), rather than workload volatility that is negatively associated with nurse turnover (Leineweber et al. 2016, Wright et al. 2017). Developing an empirical strategy to effectively separate the workload volatility effect from the employee-driven schedule flexibility is a promising avenue for future research. Lastly, the coefficient of the workload index is significant and negative, while the coefficient of its interaction term with nurse's tenure is significant and positive. Interestingly, these results suggest that the probability of nurse departure is lower when the workload is higher, but the negative effect of workload on nurse departure probability becomes less pronounced when a nurse's tenure increases. We find two explanations for how a high level of workload drives nurses to stay in their current positions. One explanation is that, because OR nurses' labor income is determined at an hourly rate, increased labor income from a high workload might compensate for the high workload and drive a nurse to decide to stay. Another explanation is that a high workload in nurses' early tenure may provide opportunities to gain the necessary skills for growth in their careers. Similar to the effect of the workload volatility index, examining the underlying mechanism of these results is a future research agenda. In sum, our results suggest that the probability of nurse departure increases with exposure, diversity, and volatility but decreases with familiarity and workload. The effects of these four experience indices are significantly moderated by the tenure of nurses.

In sum, our findings suggest that nurses' propensity to quit increases with high exposure and diversity to new procedures and partners and with high workload volatility, while decreasing with the workload in their schedules. Furthermore, these effects are significantly moderated by the seniority of nurses in the hospital. The positive impact of exposure on nurse departure probability is more pronounced with increasing seniority of nurses, while the positive effect of diversity and workload volatility gets less pronounced with increasing seniority. The negative workload effect becomes less pronounced with increasing seniority of nurses.

2.6 Concluding Remarks and Discussion

High nurse turnover presents significant financial and operational challenges for hospitals. The nationwide nursing shortage in the U.S. healthcare system makes it more important to retain existing nurses to provide reliable patient care with stable staffing. Work scheduling is a major factor that drives OR nurses' decision to quit, determining procedure, partner, and workload assignments in nurses' schedules. Especially, the knowledge-intensive work environment of OR nurses can lead to difficulty for nurses to learn a significant number of different procedures and different ways to work with diverse partners.

In this paper, we looked at the impact of work scheduling on the turnover of OR nurses who need to process a significant amount of information about different procedures and working with diverse partners. We quantify the work environment experience of nurses in terms of exposure, diversity, familiarity, workload volatility, and workload. Based on a detailed empirical analysis that tracks the procedure, partner, and workload assignments of nurses and addresses the potential endogeneity of these assignments due to nurses' personal reasons, we find that nurses' propensity to quit increases with high exposure and diversity to new procedures and partners and with high workload volatility, while decreasing with the workload in their schedules. Furthermore, these effects are significantly moderated by the seniority of nurses in the hospital.

OR nurse schedulers should focus on balancing diverse work scheduling dimensions depending on the nurses' seniority. The schedulers have to make sure that newly hired nurses maintain a certain workload that allows them to accumulate their human capital and skills and to ensure that they make enough labor income. For example, the schedulers may limit the senior nurses' exposure to new procedures and new partners during a particular time period to increase the probability of nurse retention. Also, the schedulers have to ensure the workload is enough for newly hired nurses to accumulate experience and make a certain amount of labor income.

There are several avenues for future research. In our data, we do not observe nurses' experience prior to their current position or their payroll. Also, our data come from a single hospital. Future studies can formally examine our explanations of the observed effects of nurses' diverse work scheduling dimensions to examine our assumptions and understand the underlying mechanisms of our results.

Table 2.1: Overview of data sets

		N	Level	Employment Data		Schedule Data			Surgery Data		Description
				Departure	Tenure	Procedure Experience	Partner Experience	Nurse Workload	Time and Location	Surgery Type	
Raw Data Sets	AAPC	2,959	Procedure	–	–	▲	–	–	–	–	AAPC official procedure codes and descriptions*
	Surgery	81,967	Surgery	▲	▲	▲	–	▲	○	○	Surgery-specific information and timestamps
	Procedure	155,418	Surgery-procedure	–	–	▲	–	▲	–	–	Procedure codes and descriptions in EHR
	Staff	471,825	Surgery-staff	▲	▲	▲	▲	▲○	○	○	Staff timestamps associated with each case
Reconstructed Data Set	Sample	241,819	Surgery-nurse	▲	▲	▲	▲	▲○	○	○	Combined study variables at surgery-nurse level

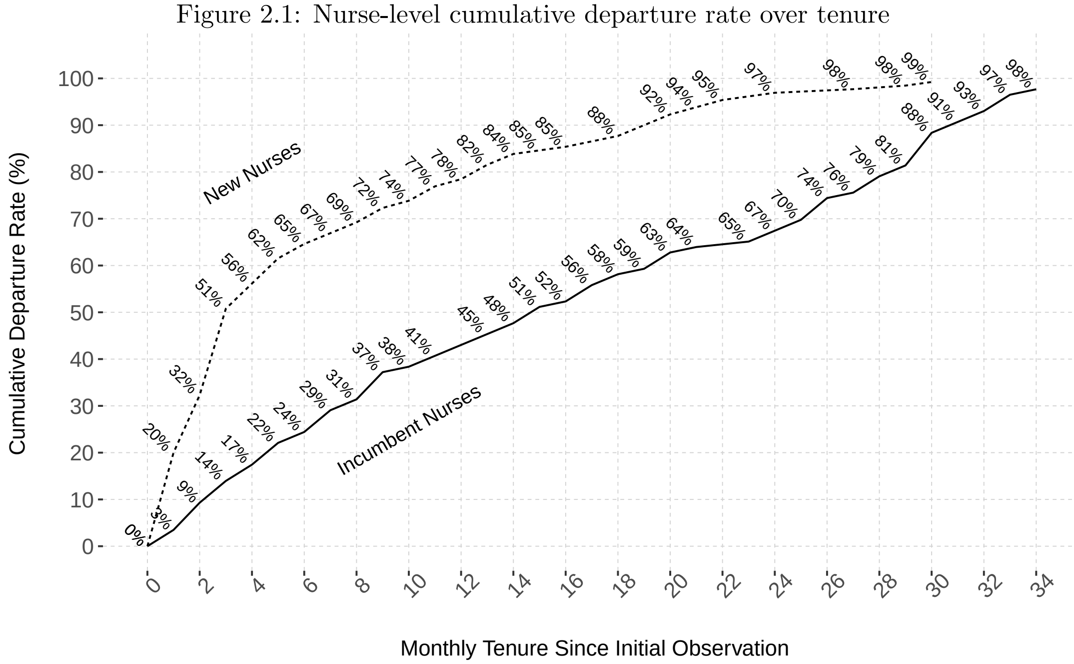
This table provides an overview of our raw data sources and reconstructed sample data set. The surgery, procedure, and staff data sets contain data relating to nurse employment and scheduling data as well as surgery data associated with the scheduling data, from March 2016 to June 2019. '○' indicates that we extracted the given variables from the raw data set without modification whereas '▲' indicates that the variables were constructed through modification of information in the raw data set and − is an indication of no data extracted from the raw data set for the variable group. In each row, we display how each raw data set relates to the major variable categories in our study (departure, procedure experience, partner experience, nurse workload, location and time, and surgery type) with a brief description of each data set.

Table 2.2: Yearly statistics of surgeries and employment

		2016	2017	2018	2019
Nurse Observations	Departing nurses	49	72	82	13
	New nurses	81	175	195	150
	Incumbent nurses	244	190	159	116
	All nurses	325	365	354	266
Calendar Observations	Calendar days	306	365	365	151
Surgery Observations	Surgeries	19,463	23,660	23,849	10,074

This table shows how many departures are observed in our data set each year, the number of days, the number of surgeries, and the staffing level of new and incumbent nurses in the dataset. The sample includes only operating room nurses such as CRNAs, circulators (CNs), and surgical techs (STs) and excludes physicians and other staff. New nurses are the nurses whose work-start month is beyond the first two months since the beginning of the observation period of data (i.e., later than May 1, 2016), otherwise, incumbent nurses.

Figure 2.1: Nurse-level cumulative departure rate over tenure



This figure illustrates the cumulative departure rate of incumbent and new nurses over their tenure (the number of months since the first month of the observation). New nurses are the nurses whose work-start month is beyond the first two months since the beginning of the observation period of data (i.e., later than May 1, 2016), otherwise, incumbent nurses.

Table 2.3: Nurse-level statistics of departing and staying nurses

		Incumbent Nurse			New Nurse		
		Circulator	Surgical Tech	CRNA	Circulator	Surgical Tech	CRNA
Nurse Count	Stayed	45	43	70	68	64	41
	Departed	40	28	18	67	44	19
	Total	85	71	88	135	108	60
Nurse Ratio	Stayed	52.9%	60.6%	79.5%	50.4%	59.3%	68.3%
	Departed	47.1%	39.4%	20.5%	49.6%	40.7%	31.7%
	Total	100%	100%	100%	100%	100%	100%
Nurse Tenure (months)	Stayed	33.22	35.84	28.73	15.54	16.2	15.12
	Departed	15.07	17.96	12.44	5.12	6.3	9.05
	Total	24.68	28.79	25.4	10.37	12.17	13.2

This table summarizes the total count of departing and staying nurses, their ratio to the total number of nurses in each role, and their average employment length in their last observed month. New nurses are the nurses whose work-start month is beyond the first two months since the beginning of the observation period of data (i.e., later than May 1, 2016), otherwise, incumbent nurses.

Table 2.4: Word processing overview by surgical specialty

	Unique Word Cleaning					Unique Word Clustering (Stage 4)			Total Word Appearances (Stage 4)			
	Raw	Stage 1	Stage 2	Stage 3	Stage 4	General (%)	Rare (%)	Specialty (%)	Total	General (%)	Rare (%)	Specialty (%)
None	1,640	1,545	1,545	1,545	1,545	45.57%	21.10%	33.33%	18,052	77.80%	1.87%	20.34%
General	1,573	1,471	1,471	1,471	1,471	42.62%	31.41%	25.97%	156,070	83.13%	1.11%	15.75%
Otolaryngology	1,181	1,103	1,103	1,103	1,103	42.43%	36.54%	21.03%	81,000	70.37%	2.50%	27.14%
Orthopedic	1,164	1,076	1,076	1,076	1,076	50.74%	32.62%	16.64%	97,524	92.36%	1.75%	5.89%
Pediatric	1,126	1,049	1,049	1,049	1,049	48.24%	28.22%	23.55%	50,448	85.44%	2.21%	12.35%
Urology	1,083	1,004	1,004	1,004	1,004	48.90%	26.10%	25.00%	154,295	41.30%	0.80%	57.90%
Thoracic	954	882	882	882	882	55.44%	24.94%	19.61%	52,597	84.04%	1.86%	14.10%
Plastic	895	825	825	825	825	52.97%	22.79%	24.24%	28,887	86.80%	2.45%	10.75%
Neurology	886	818	818	818	818	56.36%	26.65%	16.99%	105,480	81.95%	1.34%	16.70%
Vascular	883	803	803	803	803	50.93%	26.65%	22.42%	68,217	66.71%	1.49%	31.80%
Gynecology	875	804	804	804	804	52.86%	19.40%	27.74%	75,267	63.98%	0.84%	35.18%
Colorectal	799	734	734	734	734	55.86%	18.39%	25.75%	36,615	67.97%	1.26%	30.76%
Oncology	695	627	627	627	627	60.29%	16.27%	23.44%	15,886	83.66%	1.57%	14.76%
Ortho-Trauma	679	612	612	612	612	60.29%	26.80%	12.91%	30,755	95.75%	1.08%	3.17%
Ortho-Ped	549	498	498	498	498	61.04%	26.31%	12.65%	6,591	90.05%	5.52%	4.43%
Cardiothoracic	538	487	487	487	487	57.29%	24.23%	18.48%	38,289	62.93%	1.48%	35.58%
Ophthalmology	447	394	394	394	394	46.45%	38.32%	15.23%	3,659	56.90%	28.78%	14.32%
Ortho-Spine	440	390	390	390	390	72.56%	10.51%	16.92%	35,477	92.63%	0.28%	7.09%
Ortho-Oncology	419	376	376	376	376	74.73%	10.64%	14.63%	3,444	94.95%	2.15%	2.90%
Ortho-Hand	411	363	363	363	363	76.03%	17.36%	6.61%	4,764	94.31%	3.21%	2.48%
Ortho-FootAnkle	302	269	269	269	269	66.54%	20.45%	13.01%	856	86.68%	6.19%	7.13%
Edocrinolgoy	273	241	241	241	241	62.66%	10.37%	26.97%	8,463	78.66%	0.65%	20.69%
Bariatric	269	237	237	237	237	64.98%	10.13%	24.89%	957	78.68%	8.36%	12.96%
Oral	264	229	229	229	229	71.62%	13.97%	14.41%	2,728	89.08%	6.85%	4.07%
Other	240	211	211	211	211	72.99%	6.16%	20.85%	1,233	88.56%	1.62%	9.81%
Gastrointestinal	230	196	196	196	196	70.41%	14.29%	15.31%	1,162	90.62%	5.68%	3.70%
Gen-Trauma	227	192	192	192	192	83.85%	3.65%	12.50%	787	92.63%	1.14%	6.23%
Radiology	167	144	144	144	144	56.94%	22.22%	20.83%	762	42.52%	8.40%	49.08%

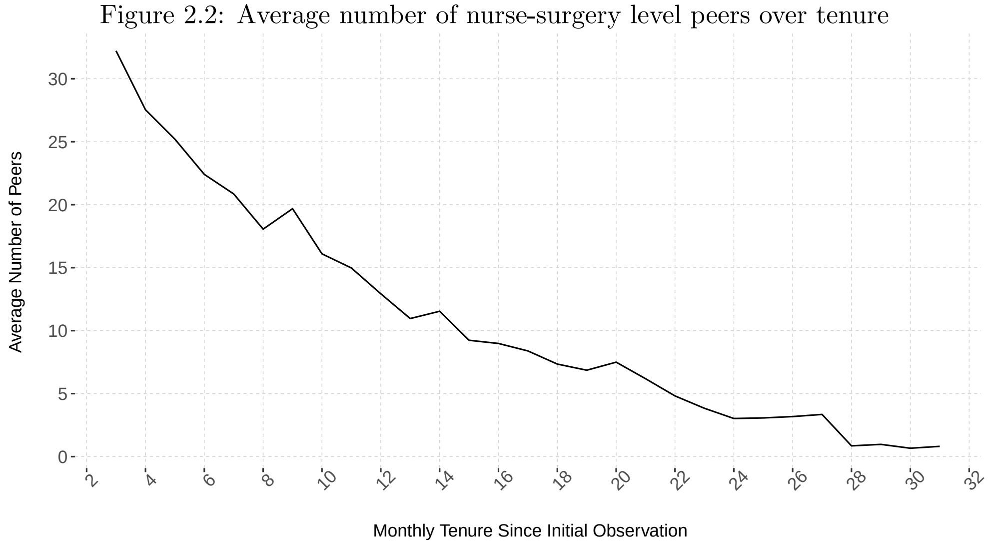
Our cleaning stages are as follows. Stage 1–remove non-meaningful words in natural language processing; stage 2–classify words into three word families: general, rare, and specialty words; stage 3–remove non-medical words that are not found in the Unified Medical Language System (UMLS); and stage 4–remove redundant words in each surgery.

Table 2.5: Description of independent variables

		Granularity	Time Horizon	N (variables)	Description
Exposure	Procedures	Nurse-surgery	3 months	3	Minimum word distance between the focal and the most similar recent surgery in each word family
	Partners	Nurse-surgery	3 months	5	Average new partner ratio in each partner role pair with no recent joint experiences
Diversity	Procedures	Nurse-surgery	3 months	3	Average word distance between the focal and recent surgeries in each word cluster category
	Partners	Nurse-surgery	3 months	5	Unique number of recent partners in each role pair with recent joint experiences
Familiarity	Procedures	Nurse-surgery	3 months	3	Total number of recent surgeries with the same (exact) word list in each word cluster category
	Partners	Nurse-surgery	3 months	5	Average number of joint recent surgery experiences with each role-pair partner
	Partner's network	Nurse-surgery	3 months	1	Average recent surgeon experience network distance with other nurses in the focal surgery
Volatility	Durations	Nurse-month	1 month	2	Monthly variation in workload duration variables—total and overtime deviations
	Shifts	Nurse-month	1 month	1	Monthly variation in the workload shift count variable
Workload	Durations	Nurse-week	1 week	2	Total amount of weekly work hours and overtime
	Shifts	Nurse-week	1 week	1	Total number of weekly shifts

This table describes the granularity, measurement time horizon, the number of variables aggregated as a weighted average for each index (e.g., exposure, diversity, familiarity, volatility, workload), and the description of variables.

Figure 2.2: Average number of nurse-surgery level peers over tenure



This figure presents the average number of nurse-surgery peers (y-axis) against tenure in months (x-axis).

Table 2.6: Summary statistics of independent variables

		Within-Index Weight (%)	Mean	SD	Min	Q1	Q2	Q3	Max
Procedure Exposure	Rare words	25.00	1.09	0.65	0.00	1.00	1.00	1.00	23.00
	Specialty words	16.67	1.30	1.89	0.00	0.00	1.00	1.00	29.00
	General words	8.33	4.31	6.27	0.00	0.00	2.00	6.00	70.00
Partner Exposure	Surgeon	18.75	0.60	0.49	0.00	0.00	1.00	1.00	1.00
	Anesthesiologist	12.50	0.98	0.13	0.00	1.00	1.00	1.00	1.00
	CRNA	6.25	0.98	0.14	0.00	1.00	1.00	1.00	1.00
	Circulator	6.25	0.23	0.42	0.00	0.00	0.00	0.00	1.00
	Surgical tech	6.25	0.24	0.43	0.00	0.00	0.00	0.00	1.00
Procedure Diversity	Rare words	25.00	2.07	0.83	0.00	1.65	1.96	2.29	23.21
	Specialty words	16.67	6.06	3.26	0.00	3.89	5.22	7.25	32.64
	General words	8.33	18.29	8.83	2.75	12.09	16.13	21.75	81.18
Partner Diversity	Surgeon	18.75	34.00	17.18	0.00	22.00	32.00	43.00	89.00
	Anesthesiologist	12.50	4.04	5.23	0.00	0.00	2.00	6.00	29.00
	CRNA	6.25	7.39	8.25	0.00	1.00	5.00	10.00	54.00
	Circulator	6.25	39.01	11.56	1.00	31.00	38.00	46.00	84.00
	Surgical tech	6.25	33.61	10.55	0.00	26.00	33.00	40.00	68.00
Procedure Familiarity	Rare words	0.00	0.02	0.17	0.00	0.00	0.00	0.00	5.00
	Specialty words	33.33	2.30	6.30	0.00	0.00	0.00	2.00	107.00
	General words	16.67	2.42	6.48	0.00	0.00	0.00	2.00	107.00
Partner Familiarity	Surgeon	9.38	3.56	7.54	0.00	0.00	0.00	3.00	54.00
	Anesthesiologist	6.25	0.03	0.34	0.00	0.00	0.00	0.00	13.00
	CRNA	3.12	0.03	0.27	0.00	0.00	0.00	0.00	9.00
	Circulator	3.12	9.62	12.72	0.00	1.00	5.50	14.00	154.00
	Surgical tech	3.12	10.99	12.92	0.00	0.50	7.00	17.00	92.00
	Network familiarity	25.00	329.33	58.09	61.00	291.00	337.67	374.75	444.67
Volatility	Work hours	33.33	0.35	0.25	0.00	0.20	0.29	0.39	2.02
	Shift start time	33.33	1.79	1.74	0.00	0.82	1.38	1.95	12.73
	Overtime	33.33	1.06	0.83	0.00	0.00	1.13	1.64	4.24
Workload	Shifts	33.33	3.78	0.89	1.00	3.00	4.00	4.00	7.00
	Work hours	33.33	31.54	9.92	0.13	27.10	32.68	36.95	180.63
	Overtime	33.33	1.81	5.13	0.00	0.00	0.00	1.95	145.63

This table shows summary statistics of independent variables following the column specifying the within-index weights used to construct each weighted average experience index (e.g., exposure, diversity, familiarity, volatility, workload). Note that the weight for procedure familiarity in rare words is set to zero. It is because we drop the variable due to very little variation. The nurse-surgery observations include only new nurses' nurse-surgery level observations if the tenure of the observation is greater than three months and the month of the observation's month is prior to January 1, 2019.

Table 2.7: First-stage regression results: effect of peer average experience indices

		(1) Exposure	(2) Diversity	(3) Familiarity	(4) Volatility	(5) Workload
	Intercept	0.400^{**}	0.144^{**}	0.051^{**}	0.002	-0.124^{**}
Peer Schedules	Exposure index	0.232^{**}	-0.021	0.040^{**}	0.024	0.131^{**}
	Exposure index × Peers	0.062^{**}	0.012^{**}	-0.021^{**}	-0.045^{**}	-0.033^{**}
	Diversity index	-0.135^{**}	0.222^{**}	0.049^{**}	0.092^{*}	0.327^{**}
	Diversity index × Peers	0.041^{**}	0.070^{**}	-0.024^{**}	-0.080^{**}	-0.067^{**}
	Familiarity index	-0.289^{**}	-0.207^{**}	0.587^{**}	0.520^{**}	0.414^{**}
	Familiarity index × Peers	0.055^{**}	0.020^{*}	0.016^{*}	-0.126^{**}	-0.088^{**}
	Volatility index	-0.030	0.079^{**}	0.012	-0.122^{**}	0.153^{**}
	Volatility index × Peers	-0.011^{*}	-0.042^{**}	0.004	0.106^{**}	-0.050^{**}
	Workload index	-0.031	0.225^{**}	0.081^{**}	0.313^{**}	0.519^{**}
	Workload index × Peers	0.019	-0.039^{**}	-0.036^{**}	-0.121^{**}	-0.071^{**}
Peer Count	Number of peers	-0.052^{**}	-0.003	0.015^{**}	0.069^{**}	0.073^{**}
Adjusted R^2 (%)		5.72	10.50	15.02	2.72	4.04
F-statistic		341.13	658.41	992.06	157.79	236.98
Observations		61,669	61,669	61,669	61,669	61,669

This table presents our first-stage regression results in our 2SLS model. The nurse-surgery observations include only new nurses' nurse-surgery level observations if the tenure of the observation is greater than three months and the month of the observation's month is prior to January 1, 2019. Out of 62,467 observations after removing observations that have no procedure words and no partners in the surgery to compare, we removed 798 observations (1.28%) that do not have any peer observations dropped (61,669 observations remaining). $^{**}p < 0.01$, $^{*}p < 0.05$.

Table 2.8: Instrumental variable estimation results: effect of experience indices on nurse departure

		IV F-stat	(1)	(2)	(3)	(4)	(5)
	Intercept		29.418	-6.800	39.261*	29.114	137.929**
Individual Schedule	Exposure index	341.13	41.219				
	Exposure index × Tenure	3881.17	3.911**				
	Diversity index	658.41		130.933**			
	Diversity index × Tenure	3006.43		-18.336**			
	Familiarity index	992.06			-5.617		
	Familiarity index × Tenure	5713.28			-2.701		
	Volatility index	157.79				82.629*	
	Volatility index × Tenure	1882.01				-7.508**	
	Workload index	236.98					-440.884**
	Workload index × Tenure	3800.71					20.947**
Individual Tenure			-1.801**	3.460**	0.145	0.997**	-5.605**
Adjusted R^2 (%)			0.0272	0.0624	0.0984	0.0784	-0.0689
Observations			61,669	61,669	61,669	61,669	61,669

This table presents our instrumental variable estimation results. We multiplied the binary departure outcome by 100 for ease of interpretation. Controls include variables in Table A.1 and Table A.2. The IV F-stat indicates the first stage F statistics. The nurse-surgery observations include only new nurses' nurse-surgery level observations if the tenure of the observation is greater than three months and the month of the observation's month is prior to January 1, 2019. $^{**}p < 0.01$, $^{*}p < 0.05$.

Appendix A Appendix

A.1 Classification of Procedure Words

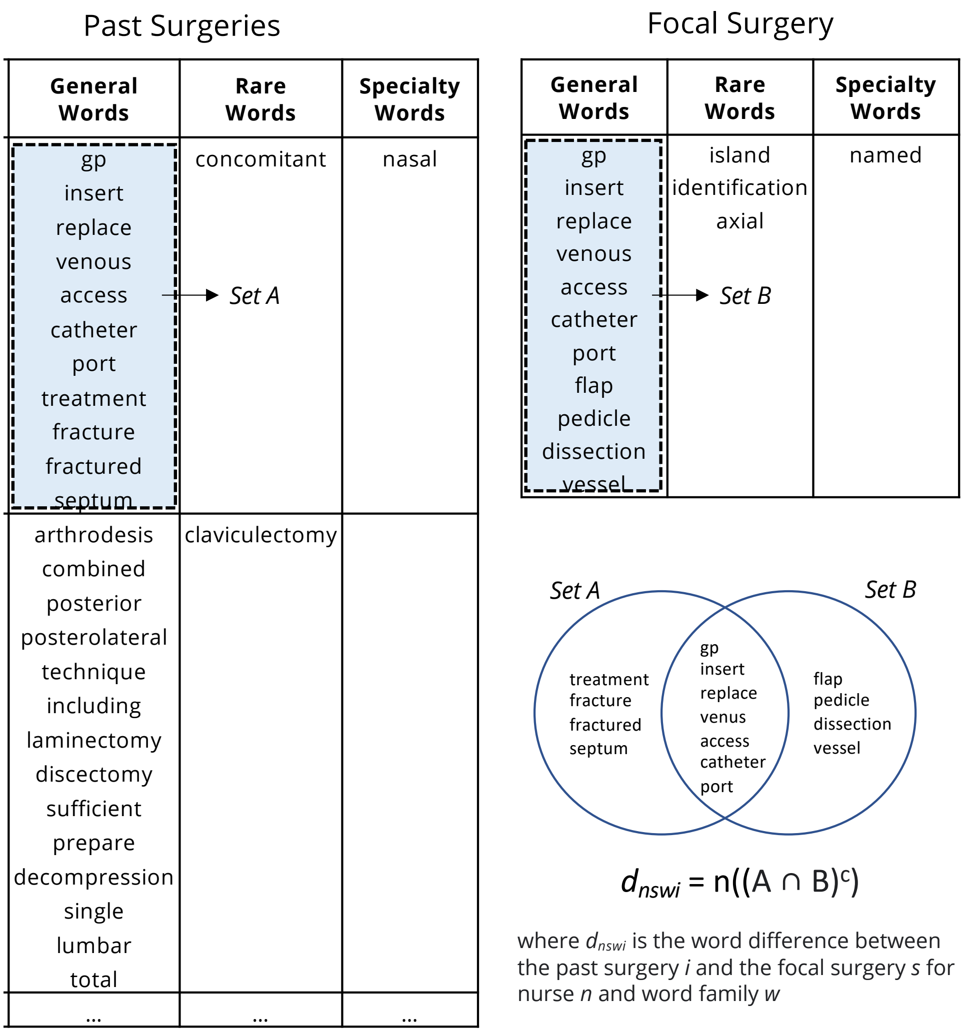
Our classification of procedure words is based on a word's total frequency and the frequency across different specialties. We define three word families each word can belong to: (1) rare words, (2) specialty words, and (3) general words. First, we count the total appearances in the data set and the appearances across different specialties for each word. Then, we follow the below criteria to classify the words into three word families. Rare words if the word's total frequency in the data set is less than 50 times. Next, if the word's total frequency is at least 50 times and the word's appearances are significantly dispersed in more than two surgical specialties. We evaluate the dispersion of a word's appearances across different specialties by using the Herfindahl Index, a statistical measure of concentration (Rhoades 1993). Using Herfindahl Index, we measure the concentration of word appearances of word w in different specialties by:

$$DI_w = \sum_{i=1}^{n} (\frac{Frequency_{iw}}{\sum_{i=1}^{n} Frequency_{iw}})^2 \tag{2.17}$$

where $Frequency_i$ is the number of appearances of word w in specialty i. The measure DI_w for word w indicates how concentrated the word's appearances are across different specialties. A word w is classified as a specialty word if the word's appearances are at least 50 times and if DI_w is at least 0.4. Lastly, if the remaining words after excluding the rare and specialty words are classified as general words as these words are not rare and their appearances in the procedure descriptions are significantly dispersed in multiple specialties.

A.2 Example of Word Difference Computation

Figure A.1: Example of word difference computation



A.3 Control Variables

Table A.1: Overview of continuous and binary control variables

		Granularity	Mean	SD	Description
Binary	Add-on surgery	Surgery	0.26	0.44	Add-on surgery indicator (yes or no)
	Robotic surgery	Surgery	0.05	0.21	Robotic surgery indicator (yes or no)
	Holiday	Day	0.00	0.06	Holiday indicator (yes or no)
	Returning nurse	Nurse	0.08	0.27	Indicator for whether the nurse has had at least one month of work gap
Continuous	Cumulative work gap	Nurse-month	0.27	1.25	The cumulative number of work gap in months prior to the focal month
	Tenure (months)	Nurse-month	10.94	6.59	The number of months since the month of initial observation

This table presents the granularity, the summary statistics, and the descriptions of the continuous and indicator control variables. The nurse-surgery observations include only new nurses' nurse-surgery level observations if the tenure of the observation is greater than three months and the month of the observation's month is prior to January 1, 2019.

Table A.2: Overview of categorical control variables

	Granularity	Levels	Description	Top Three Categories
Surgery start hour	Surgery	5	Surgery start hour	Noon (45%), Morning (39%), Afternoon (11%)
Lead surgeon	Surgery	278	Lead surgeon	Top 1 (3%), Top 2 (3%), Top 3 (2%)
Specialty	Surgery	27	Surgical specialty	General (19%), Vascular (10%), Urology (9%)
Room	Surgery	33	Operating room number	27 (5%), 33 (4%), 29 (4%)
Shift type	Nurse-day	4	Shift type	8 hrs (51%), 10 hrs (33%), 12 hrs (14%)
Shift start hour	Nurse-day	24	Shift start hour	6 (37%), 7 (33%), 8 (12%)
Day of week	Day	7	Day of week	Thu (21%), Tue (20%), Wed (19%)
Role	Nurse	3	Nurse roles	Circulator (57%), Surgical Tech (42%), CRNA (1%)
Year-month	Month	29	Year-month	2018-12 (5%), 2018-11 (5%), 2018-05 (5%)

This table presents the granularity, distribution, and descriptions of our categorical control variables. The nurse-surgery observations include only new nurses' nurse-surgery level observations if the tenure of the observation is greater than three months and the month of the observation's month is prior to January 1, 2019.

A.4 Scaling of continuous variables

We clean and scale our major variables in the following steps.

Step 0. Raw values

Step 1. Wisorizing extreme values (upper and lower bounds of wisorization shown in Table A.3)

Step 2. Unit-scaling by dividing a value by the maximum value of the variable

Step 3. Summarizing variables into experience indices

Step 4. Computing peer average experience indices

Step 5. Apply log-transformation if the skewness get lessened after log-transformation (logged variables shown in Table A.3)

Table A.3: Scaling of continuous variables

		Lower Bound	Upper Bound	Log-Transformation
Procedure Exposure	Rare words	0	0.95	Y
	Specialty words	0	0.95	Y
	General words	0	0.95	Y
Procedure Diversity	Rare words	0	0.95	Y
	Specialty words	0	0.95	Y
	General words	0	0.95	Y
Partner Diversity	Surgeon	0	1	N
	Anesthesiologist	0	1	N
	CRNA	0	1	N
	Circulator	0	1	N
	Surgical tech	0	1	N
Procedure Familiarity	Rare words	0	0.95	N
	Specialty words	0	0.95	N
	General words	0	0.95	N
Partner Familiarity	Surgeon	0	1	Y
	Anesthesiologist	0	1	Y
	CRNA	0	1	Y
	Circulator	0	1	Y
	Surgical tech	0	1	Y
	Network familiarity	0	1	Y
Peer Control	Number of peers	none	none	Y
Experience Indices	Exposure index	none	none	N
	Diversity index	none	none	N
	Familiarity index	none	none	N
	Volatility index	none	none	N
	Workload index	none	none	N
Peer Average Experience Indices	Exposure index	none	none	N
	Diversity index	none	none	N
	Familiarity index	none	none	N
	Volatility index	none	none	N
	Workload index	none	none	N
Controls	Cumulative work history gap	0	0.95	Y
	Tenure (months)	0	1	Y

This table presents the upper and lower bounds of winsorized variables and the log-transformation of variables.

A.5 Sampling

Table A.4: Sampling steps and outcomes

	Step	N (removed)	N (remaining)
Step 1	All nurse-surgery observations		255,068
Step 2	Keep only new nurses	151,460	103,608
Step 3	Keep observations before the last six months (i.e., before January 1, 2029)	20,288	83,320
Step 4	Keep observations with nurse tenure greater than 3 months	19,775	63,545
Step 5	Remove observations with no procedure descriptions	154	63,391
Step 6	Remove observations with only one nurse role	406	62,985
Step 7	Remove observations with immediate prior work gap (i.e., idle month)	518	62,467
Step 8	Remove observations with no peer observations	798	61,669

Chapter 3

Value of Cross-departmental Information-Sharing in Operating Room Scheduling: Review and Future Research Directions

3.1 Introduction

Research in operating room (OR) scheduling—allocating time to surgical procedures—is entering a new phase of research direction. Recent studies indicate that utilizing team information in OR scheduling can significantly improve the prediction accuracy of OR time (i.e., patient-in to patient-out duration in the OR) (Kim et al. 2022), which can smooth the daily operations of OR and reduce the idle time and overtime cost (Bartek et al. 2019, Zhou et al. 2016). At the same time, team configuration has been found to affect team productivity and operational outcomes (?Reagans et al. 2005, Avgerinos and Gokpinar 2017), providing opportunities to specify diverse team information for OR time prediction. In the most recent development, Kim et al. (2022) show that considering granular dyadic team information (i.e., information about each dyad of team members within a surgical team) can significantly improve the OR time prediction. The studies in

OR scheduling and surgical team scheduling suggest that there is a significant relationship between team configuration and operational outcomes (e.g., OR time) and that utilizing granular information about team members in surgical team scheduling is important to improve OR planning.

What is unknown in both literature and practice, however, is how hospitals can incorporate such granular team information in OR time prediction under the unique environment of surgical team scheduling. There are three practical challenges in using granular team information in OR scheduling. First, surgical team scheduling takes place through multiple decision-makers independent of each other. Each dyad of team members consists of two distinct professionals (e.g., surgeon, anesthesiologist, CRNA, etc.). Most of these dyads are assigned by two independent decision-makers in multiple decision-making stages. Second, the information about different team members becomes sequentially available throughout these multiple decision-making stages at different time points. Lastly, several studies suggest that there is an explicit and implicit hierarchical power structure within surgical teams, which can influence teamwork and team performance (Lingard et al. 2012, Kennedy et al. 2020, Avgerinos et al. 2020). Table 3.1 describes the multi-stage decisions regarding surgical team scheduling and the scheduling managers on the decision-making timeline. This table also specifies each dyad of scheduled and paired members at each decision, each dyad's hierarchical status, and whether a team member dyad is paired by two different departments (i.e., cross-departmental dyad) or within the departments. Hereafter, we call the information-sharing across different departments cross-departmental information sharing. From the table, we find that most of the team member dyads are cross-departmental dyads, meaning that the granular dyadic team information may not be utilized without cross-departmental information-sharing.

This uniqueness of surgical team scheduling poses an important question on whether incorporating cross-departmental information-sharing to utilize granular dyadic team information is valuable to improve OR time prediction. Motivated by recent insights in the literature and practical challenges in OR scheduling, we systematically review the team scheduling literature to provide an integrated review of the value of cross-departmental information-sharing for OR time prediction. Through our systematic literature review, we aim to evaluate the value of utilizing the dyadic team information across multiple department schedulers in OR time prediction by understanding the relative influence of multiple-decision makers and how the hierarchical power structure of dyads drives OR time. More specifically, our systematic literature review (SLR) aims to answer the following research question: (1) How much influence on OR time prediction each decision-maker has when updating the OR time by utilizing granular dynamic team information available to them?; (2) what is the value of cross-departmental information-sharing to utilize granular dyadic team information for OR time prediction?; and (3) is the traditional hierarchical power of each team member dyad consistent with the dyad's influence on OR time?

Because there are limited insights and research studies in the prior literature about the effect of using dyadic team information on OR time prediction, a systematic literature review (SLR) provides a perfect opportunity to gain insights to answer these research questions. In our SLR, we electronically search published articles in PubMed, JSTOR, ProQuest, and Web of Science. We use the combinations of context keywords (e.g., operating room or surgical team) and team scheduling keywords (e.g., team scheduling or team familiarity), which help us to search the most relevant articles that examine team scheduling effect on OR outcomes. We screen the identified articles using several eligibility criteria concerning the article's relevance to our research questions and their empirical methodologies. Lastly, we offer our SLR findings regarding our research questions and propose a future direction to empirically examine the research questions.

Our SLR findings suggest that there are still limited insights to answer our research questions. First, our reviewed studies do not address the multiplicity of decision-makers throughout the multiple stages of team scheduling. While the studies give some insights on the relative importance of team members scheduled by different decision-makers, they provide mixed results on which team member or dyad has a relatively larger impact on OR time. In turn, we find that evaluating the relative influence of each decision-maker on OR time prediction requires further research explicitly addressing the multiple decision stages. Second, the value of cross-departmental information-sharing is also unclear. Because of the limited insights on the relative influence of decision-makers from across different departments, whether a cross-departmental information-sharing of a particular combination of two departments would result in a better OR time prediction than the information-sharing of other department combinations is less clear. In addition, most studies focus on either individual experience or the team's shared experience independently, resulting in a partial evaluation of the significance of different information that can be shared across different departments on OR time prediction. Thus, we find that further research is required in this area. Third, only a few studies consider the hierarchical structure of team members to construct the team-level and dyad-level team experience variables. By default, these studies assume that the surgeon, the member with the highest hierarchical power in surgical teams, would affect the team performance most. As a result, the studies do not differentiate different non-surgeon members or non-surgeon dyads, providing limited insights to compare the hierarchical structure within a surgical team and the relative impacts of different dyads. In addition, we also find that the relationship between team experience, perceived hierarchical power, and team performance is more complex than previously thought. Perceived hierarchical power may be a potential mediator of team experience effect on team performance. Lastly, of the final 18 studies we reviewed, only three studies were in the operations management and management science (OM/MS) publication domain. This result indicates that research on OR predictions using granular team information is scant, calling for more development of research in this area.

In section 3.2, we provide an overview of our SLR procedures, followed by its findings in section 3.3. In section 3.4, we propose a strategy to empirically examine our research questions. Lastly, we conclude our paper with concluding remarks and managerial insights in section 3.5.

3.2 The Systematic Literature Review

3.2.1 Overview of systematic review

For the systematic review procedures, our study adapts the Preferred Reporting Items for Systematic reviews and Meta-Analyse (PRISMA),[1] but we exclude the meta-analysis step. In addition, while a meta-analysis can be useful in conducting an SLR, it is beyond the scope of our review because a meta-analysis of effect estimates is not possible due to limited empirical studies to answer our research questions. Instead, we provide an integrative summary of existing findings on our research questions and offer our suggestions for empirical strategies and managerial insights for practitioners.

The adapted systematic literature review steps used in this study are similar to the six steps of the review process of Wemmerlöv (2021). The difference is the fourth step (i.e., separating studies based on their methodologies in Wemmerlöv 2021), where we separate the reviewed studies based on team scheduling considerations (e.g., coverage of team members and dyadic team information). Because there is a limited range of empirical studies in OR team scheduling, this modification allows us to more effectively capture the differences between reviewed articles to answer our research

[1] "The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement, published in 2009, was designed to help systematic reviewers transparently report why the review was done, what the authors did, and what they found."(Page et al. 2021)

questions.

1. Identify purpose and frame the research questions
2. Select literature sources
3. Screen the literature
4. Separate studies based on team scheduling consideration
5. Analyze and synthesize the results
6. Document the review

3.2.2 Search Strategy

In our SLR, we electronically searched articles in PubMed, JSTOR, ProQuest, and Web of Science. In our search, we filtered: (1) only full-text articles published in peer-reviewed journals; (2) articles that contain our keywords (described below) in the title and abstract. To specify keywords, we used combinations of context keywords (e.g., operating room or surgical team) and team scheduling keywords (e.g., team scheduling or team familiarity), which helped us to search the most relevant team scheduling articles in the OR context. Table 3.2 shows our preliminary Google Scholar search outcomes in the number of search results[2] for each of our keyword combinations by topic and context. Our preliminary search shows the approximate size of existing research for each keyword combination. From this table, note that we use two sets of topic keywords: 'team scheduling' and 'team familiarity.' It is because team familiarity is the most widely-considered topic in surgical team scheduling literature (see Stucky and De Jong 2021 and Witmer et al. 2022 for the most recent reviews). In addition, we use variations of topic keywords—identified by our preliminary literature review, by which we capture a broad range of relevant articles that use similar team scheduling concepts.

In total, we found 492 articles from our initial search in PubMed, JSTOR, ProQuest, and Web of Science. Table 3.3 shows our initial search outcomes by topic keyword group and database. After removing redundant articles within and across these databases, we retained 425 articles for inclusion and exclusion steps.

[2] Google Scholar searches in the full text of all publications, including working papers and dissertations.

3.2.3 Screening

Our screening steps aim to identify articles that provide empirical insights relevant to our research questions. To achieve our purpose, we initially screened the articles based on our three main eligibility criteria:

- Relevance: whether the article addressed surgical team scheduling (i.e., team configuration) and whether their team performance outcomes include OR time (e.g., surgical time) or patient outcomes (e.g., readmission rate, mortality)
- Data: whether the article used empirical data including surveys, interviews, observational data, or retrospective data.
- Methodology: whether the article used empirical methodology with statistical analyses. Theory development and qualitative descriptive studies without statistical analyses are excluded.

Using our eligibility criteria, we excluded articles that explicitly state the ineligibility in the title, abstract, and full texts sequentially. In addition, we excluded articles in history, literacy, and law journals because the articles in these journals are not relevant to team scheduling. Table 3.4 shows our screening stages and the outcomes. Lastly, articles that did not meet the eligibility criteria but contained qualitative insights on our research questions were separately reviewed for discussion. After the screening, we retained 18 studies for our integrated review.

3.3 Findings of Systematic Literature Review

In this section, we describe our findings of systematic review concerning the methodologies used in the final 18 studies and the studies' relevance to our research questions. In addition, we adopt a benchmark study, Kim et al. (2022), in describing our systematic review findings. Because this study presents the most recent development on OR time prediction using granular team information, we evaluate the reviewed studies based on the benchmark study's approaches for OR time prediction.

3.3.1 Methodologies used in the final 18 studies

Table 3.5 shows the final 18 studies and our benchmark article from our literature search and their primary data analysis methodologies. Among the final 18 studies, only three studies are in the operations management and management science (OM/MS) journals, including *Management Science*, *European Journal of Industrial Engineering*, and *Computers & Industrial Engineering*. Of the remaining 15 studies, 14 studies are in medical and nursing journals whereas the remaining one article is in a management journal (*Human Relations*). The limited number of studies in OM/MS journals indicates that there are insufficient empirical insights in OM/MS about OR time prediction using granular team information.

For data analysis methodologies, most studies utilized causal inference with staff panel data and econometric analysis (panel causal inference) for their primary methodology. Most of these studies examine the causal effect of team scheduling on team outcomes, while some studies used mixed methods including panel causal inference, predictive analysis, and analytical modeling (Bayram and Chen 2020, Harmanli et al. 2021, Bayram et al. 2023). Only one paper adopts a predictive analysis of OR time to complement their analytical modeling approach (Bayram and Chen 2020), suggesting that there are limited insights about OR time prediction using team scheduling information. Furthermore, most studies utilized data collected for one or more procedures in a single specialty, suggesting the insights from these studies may have different implications when applied to an OR time prediction for multiple specialties. Lastly, most studies considered individual or team experience of team members as a primary predictor of team performance where OR time is the most common primary team performance outcome.

To discuss the types of data collected in each article, we leverage the data classification of our benchmark study. The benchmark study classifies the predictors of OR time prediction into two sets of information: pre-team selection information (i.e., information used for OR time prediction that becomes available before team selection) and post-team selection information (i.e., information used for OR time prediction that becomes available after team selection). Following the classification, we specify the pre-team selection information—including procedure and patient information—and post-team selection information—including granular information about staff who are assigned to surgical teams—as well as OR time information as an outcome measure. Table 3.6 summarizes the number of studies in our review that collect granular data categorized by OR outcome, pre-team selection, and post-team selection data.

For OR time outcome data, most studies collected OR patient duration (i.e., wheels-out time, incision-closure, or any part of duration within wheels-in to wheels-out time) while only a few papers collected OR turnover time (e.g., turnover time or wheels-out to wheels-in time) and detailed OR timestamps to compute the OR duration. For pre-team selection information, procedure codes, patient demographics, and patient clinical conditions are the most common pre-team selection information considered in their analysis (e.g., main predictors and controls). However, most of the data collection is limited to a few types of procedures or procedures within a specialty. The cross-sectional data of mixed surgical specialties were uncommon even though OR time prediction commonly takes place in large hospitals for diverse procedures from multiple specialties. In addition, the procedure data collected from these studies were limited to the procedure codes. While these codes can be used to identify the procedure complexity linked to the procedure code, granular procedure texts that can effectively compare different procedures were not utilized in any of the reviewed studies. For post-team selection information, most studies obtained staff panel data from the EHR of hospitals or staff surveys, which are associated with each surgery the staff performed in the data set period. These staff panel data were commonly used to compute team members' individual-, dyad-, or team-level experience. However, more granular staff scheduling information on the day of surgery, such as workload and pre-surgery schedule (e.g., back-to-back scheduling of staff in consecutive surgeries or whether the previous surgery of a team member was an add-on surgery that can affect their performance in the next surgery), were not considered in any of these studies.

The results suggest that there is limited use of granular data despite its potential to improve OR time prediction significantly for both pre-team selection information and post-team selection information. Most studies focused on empirical examination of team composition, rather than improving OR time prediction using the information together. On the other hand, our benchmark study collects the EHR data at the most granular level. To gain further insights building upon the benchmark study, there need to be more future studies to understand how the granular level of pre-team selection and post-team selection information can be used in practice for team scheduling and OR time prediction.

3.3.2 Separation of studies based on team scheduling consideration

In this section, we separate the final 18 studies based on their team scheduling considerations. First, we excluded studies if the studies discuss only team size as their team scheduling consideration (e.g., increasing or decreasing the team members within a surgical team, He et al. 2014); discuss only the hospital-level team familiarity of clinicians with other members, rather than within-team characteristics of surgical teams (Stucky et al. 2020); and discuss the intro-operative staff turnover and presence of different surgical team member roles (Cahan et al. 2021). The team scheduling consideration of these studies is not specific to using team information for OR time prediction prior to the surgery, which is less relevant to our research questions. Next, among the remaining 15 studies we excluded studies that cover only part of a surgical team to utilize the team information. Answering our research questions requires insights concerning the entire surgical team members and their dyadic relationships. The team coverage category in Table 3.7 describes how many studies in the 15 studies define their team using the information about a full team (i.e., a team of entire surgical team members) or partial team (i.e., a team of selected roles, such as a team of surgeons). There has been a shift of attention from considering teams of only surgeons to considering teams of entire surgical team members, acknowledging the importance of teamwork among all members of surgical teams in a complex OR environment. Earlier studies in team scheduling research in the OR context prior to 2016 examined the team member's experience on OR time primarily for teams of surgeons (Reagans et al. 2005, Kurmann et al. 2014, Özdemir-van Brunschot et al. 2015, Maruthappu et al. 2016). These studies with partial team coverage provide limited empirical insights about our research questions because they only examine teams assigned within a single department (i.e., surgery department for surgeons). Thus, we focus on the 11 studies that covered the full surgical team members and the full team members' dyads in their team scheduling consideration.

3.3.3 RQ1: Relative influence of multiple decision-makers

The second and third category in Table 3.7 describes how many studies in the remaining 11 studies utilized staff data at different granularity level: individual, team, and dyad, as well as how many studies considered the most granular dyadic data with consideration of multiple decision-makers and hierarchical structure of surgical teams. Among the 11 studies, three focused on the individual experience, three on only team-level experience, and five on both team-level experience and dyad-level experience of surgical team members in their surgical team scheduling considerations. However, among the studies that discuss dyad-level team scheduling, we find that no study addresses the multiple decision-makers in practice. Because there are no insights about multiple-decision makers, we discuss how these studies address the relative importance of individual team members and different team member dyads instead. Knowing the relative importance helps us infer the relative influence of multiple decision-makers when using granular team information available in a particular decision-making stage assigning key team members or key member dyads.

First, three studies examined the impact of team members' individual experience on team performance measured in OR time (Bayram and Chen 2020, Harmanli et al. 2021, Bayram et al. 2023). For the relative importance of team members' individual experience, these three studies provide mixed results. Bayram and Chen (2020) suggest that the high experience of surgeons is the most important determinant of OR time, followed by the experience of surgical techs and circulators. On the other hand, Harmanli et al. (2021) and Bayram et al. (2023) indicate that surgical tech's individual experience has the largest impact on OR time when compared to the experience of other members. The latter two studies' results contradict the consensus in a broad range of literature and practice that the surgeon is the most important team member as a leader of the surgical team Avgerinos et al. (2020). We find these results may be due to the uniqueness of robotic surgeries these three studies focused on. Compared to traditional surgeries, robotic surgeries involve new technologies and complex changes in the surgical process, requiring every surgical team member to learn and adapt to the non-traditional OR environment (Bayram and Chen 2020). In addition, the surgeon is not immediately positioned with the rest of the team members during the robotic surgeries (Bayram et al. 2023). As a result, it might be possible that the individual experience of non-surgeon team members becomes more important than that of surgeons for team coordination in the robotic OR environment. Hence, there is a lack of evidence on the relative importance of team members' individual experience on OR time in a traditional OR environment.

Next, eight studies discuss the team-level shared experience or dyad-level shared experience of surgical team members. Among these eight studies, we distinguish the team-level and dyad-level team scheduling studies: (1) team-level studies that aggregated all dyads' information at the team level (Sexton et al. 2018, Stucky et al. 2022, Zhang and Zheng 2022) and (2) dyad-level studies that differentiated unique team member dyads within teams (Frasier et al. 2019, Avgerinos et al. 2020, Parker et al. 2020, Mathis et al. 2021, Vaulont et al. 2021). The team-level studies give us limited insights on the relative importance of different dyads because they do not show the relative impacts of different dyads on team performance. In addition, two studies that use only the team-level aggregation assumed the impacts of different dyads are equally important, putting uniform weights on all dyads when aggregating the dyadic information into team-level information. While these studies collectively suggest that team-level familiarity is significantly associated with the change in team performance, the mechanism of how different dyads affect team performance differently is unclear because of team-level aggregation of team information. To gain more insights on granular dyadic impact, we discuss how the dyad-level studies discuss the relative importance of dyads.

The remaining five studies, Frasier et al. (2019), Avgerinos et al. (2020), Parker et al. (2020), Mathis et al. (2021), and Vaulont et al. (2021), examine how different dyad-level team composition affects the OR time. All of these five studies consider the shared experience within different dyads. We find several insights that we can leverage from these studies to answer our research questions. First, dyads' shared experiences within a team have unequal impacts on team performance (e.g., OR time or postoperative length of stay). Second, the difference between subteams of dyads within a team may drive different impacts of subteams on team performance. For example, specifying surgeon/non-surgeon dyads is one subteam specification of single dyads within a team. Lastly, the hierarchy of team members within a team affects the relative importance of dyads or dyad subteams, which will be discussed in detail in subsection 3.3.5.

More specifically, Parker et al. (2020) specified the team's shared experience at the dyad level in their empirical model and suggest that dyads that consist of a surgeon affect OR time more than other dyads. However, the relative importance of other dyads (e.g., anesthesia-ST) shows mixed results across different hospitals in their study. Mathis et al. (2021) suggest that surgeon/anesthesiologist or surgeon/physician assistant dyads were the only significant surgeon/non-surgeon dyads to explain OR time.[3] Yet, the study examined only surgeon/non-surgeon dyads, excluding non-surgeon/non-surgeon dyads.

The remaining three studies aggregated dyadic shared experience within subgroups, depending on the presence of a core member (i.e., surgeon) in the dyads (Vaulont et al. 2021, Avgerinos et al. 2020) or the cross-departmental status of dyads (Frasier et al. 2019). These studies provide mixed results about the relative influence of different dyads. Avgerinos et al. (2020) suggest that the dyadic-level shared experience of two surgeons is the most important determinant of OR time, followed by the dyadic experience of non-surgeon/non-surgeon dyads and that of surgeon/non-surgeon dyads. In contrast, Vaulont et al. (2021) find surgeon/non-surgeon dyads are more important than non-surgeon/non-surgeon dyads for team performance measured by the postoperative time. Lastly, Frasier et al. (2019) find no effect of the dyad-level shared experience on the team members' perceived

[3]Surgeon/non-surgeon dyads in Mathis et al. (2021) include surgeon/anesthesiologist and surgeon/surgery resident dyads among surgeon/anesthesiologist, surgeon/perfusionist, surgeon/scrub nurse, surgeon/physician assistant, and surgeon/surgery resident dyads. While their primary focus was team-level shared experience aggregating the dyad-level shared experience of surgeon/non-surgeon dyads, they examine the impacts of different dyads in their sensitivity analysis.

communication effectiveness.

In summary, we find that the reviewed studies do not address the multiplicity of decision-making in surgical team scheduling. Furthermore, there is a lack of evidence on the relative importance of the individual experience of team members in a traditional OR environment. For the relative importance of team member dyads, studies using the dyads' shared experience provide mixed results for the relative importance of surgeon/non-surgeon and non-surgeon/non-surgeon dyads and do not compare the relative importance of different non-surgeon/non-surgeon dyads. Our findings shed light on the need for future studies to provide a more comprehensive understanding of the relative influence on OR time prediction of multiple decision-makers—who determine the configuration of different dyads.

3.3.4 RQ2: Value of cross-departmental information-sharing

Because of the limited insights on the relative influence of decision-makers from across different departments, whether a cross-departmental information-sharing between any particular two departments would result in a better OR time prediction than other departments is less clear. In addition, besides the procedure and patient characteristics, which team information must be shared across different departments is not clear. The reviewed studies primarily considered two types of team information, the individual experience and the team's shared experience, for OR time prediction. However, most studies do not combine both individual experience and the team's shared experience for their predictors for OR time. As a result, the combined effect of two kinds of experience shared across different departments is less clear. Furthermore, three studies that address the importance of combined individual experiences on team performance focused on robotic surgeries in their research setting, and how knowing the individual experiences of surgical team members would improve the OR time prediction for traditional ORs is less clear.

3.3.5 RQ3: Hierarchical influence of team member dyads

Among the five studies that differentiate dyad-level shared experience, four studies consider the hierarchical structure of team members to construct the dyad-level team variables. Yet, the consideration primarily focuses on differentiating non-surgeon team members from the surgeon who is the surgical team member with the highest hierarchical power. Specifically, Mathis et al. (2021)

considered only team members dyads that include a surgeon (i.e., surgeon/non-surgeon dyads). Similarly, Vaulont et al. (2021), Mathis et al. (2021), and Avgerinos et al. (2020) subgroup the dyads based on the presence of surgeon(s) withing the dyads. Of these studies, Avgerinos et al. (2020) provide a detailed theoretical discussion and empirical results on how hierarchical status can affect the impact of different surgical dyads on team outcomes. They find that team's shared experience between high-power status members (surgeons in their study) has the largest impact on OR time than other dyads.[6] However, while the surgeon's importance in pairing with other team members is consistent through these studies, the impact of surgeon/non-surgeon dyads shows mixed results as described in subsection 3.3.3, making it difficult to evaluate whether the relative impact of dyads is consistent with the hierarchical structure.

In addition, the reviewed studies suggest that the relationships between a team's shared experience, team performance, and hierarchical power are more complex than previously thought. One relationship is that the perceived hierarchical power difference might mediate the effect of team's shared experience on OR time. It is because increased shared experience between two members of different hierarchical power might reduce reduced perceived hierarchical power difference, improving team coordination and reducing OR time. Frasier et al. (2019) provide anecdotal evidence on why this could be true. Frasier et al. (2019) observe the communication events of surgical teams in surgeries and explain how the perceived hierarchical power difference between a surgeon and a surgical tech can be reduced with a high level of shared experience.

> *"The surgery began with a scrub (i.e., surgical tech) who was experienced in the OR but had minimal familiarity with the attending surgeon. The surgeon made frequent, often repeated, requests for instruments, in part because the scrub was unable to anticipate the surgeons' pending needs. Partway through the operation, a second scrub entered and gave the first a temporary break. This scrub appeared more familiar with the attending surgeon and anticipated many of the surgeon's requests, having instruments ready. However, there was also a qualitative difference in how scrub responded to the surgeon: She appeared less intimidated, and called the attending by first name. In one instance, the surgery attending requested longer instruments, and the scrub observed that the surgeon risked contaminating instruments on her mask due to their length; the surgeon elected to keep*

[6] Avgerinos et al. (2020) suggest that the surgeon's hierarchical power is highest, followed by that of surgical tech or anesthesiologist.

the current instruments." (Frasier et al. 2019)

Also, another relationship is that the characteristics of OR environments might serve as a moderator of the individual experience effect on OR time. For example, our results suggest that the individual experience of a surgical tech can have a larger impact on OR time than that of a surgeon in robotic surgeries. In a robotic OR environment, the non-surgeon members' individual experience is critical to maintaining team performance, in which the relative impact of different surgical team members does not follow the traditional hierarchical power structure.

In summary, we find a lack of evidence to evaluate the consistency between dyads' impact on OR time and their hierarchical structure, calling for future studies in this area. Also, OR prediction may improve with a more comprehensive understanding of complex relationships among team familiarity, team performance, and hierarchical power structure. In the next section, we propose empirical strategies relevant to each of our research questions, which can be fruitful research directions for future studies.

3.4 Proposed Empirical Strategy

Our proposal for empirical strategies to answer our research questions begins with differentiating the impact of team information on OR time prediction from the impact of information about other OR time drivers. We build upon our benchmark study, Kim et al. (2022), for the empirical methodology to differentiate two types of impacts. Kim et al. (2022) use two-stage regression models that first regress OR time on *pre-team selection inputs*—i.e., information that becomes available before team selection, such as procedure and procedure characteristics (described in subsection 3.3.1). Then, they define the deviation between the first stage's predicted OR time and actual OR time as the measure of team performance. In the second stage regression, the study regresses the team performance measure on post-team selection inputs—i.e., team information that becomes available after team selection, to examine the impact of team information on OR time through team performance. More specifically, the first stage OR time prediction before team selection (i.e., ex-ante OR time) is made by the following ordinary least squares (OLS) model:

$$\underbrace{T_{sdt}^{Realized}}_{\text{Realized OR Time}} = \underbrace{\alpha_0 + \boldsymbol{\alpha}^T \cdot \boldsymbol{I}_{sd}^{Pre-team}}_{\text{Ex-ante OR Time}} + \underbrace{\Delta T_{sdt}^{Ex-Post}}_{\text{Ex-post OR Time Residual}}, \tag{3.1}$$

where s denotes a specific surgery, d denotes a specific surgery day, t denotes a specific surgery team, $T_{sdt}^{Realized}$ is the realized surgery time for surgery s, $\boldsymbol{I}_{sd}^{Pre-team}$ denotes the pre-team-selection information before team selections (see the full list in Kim et al. 2022), and $\boldsymbol{\alpha}$ is the ex ante OR time parameter vector. $\Delta T_{sdt}^{Ex-post}$ is the ex ante OR time residual that represents team performance. The second stage model regresses the team performance on the post-team selection inputs, given by:

$$\underbrace{\Delta T_{sdt}^{Ex-post}}_{\text{Ex-post OR Time Performance}} = \beta_0 + \underbrace{\boldsymbol{\beta}^T \cdot \boldsymbol{I}_{sdt}^{Post-team}}_{\text{Ex-post Team Selection Inputs}} + \nu_{sdt}, \tag{3.2}$$

where $\boldsymbol{\beta}$ is a vector of coefficients for post-team selection inputs $\boldsymbol{I}_{sdt}^{Post-team}$. For our proposal, we leverage and adapt the above second-stage model to compare the prediction accuracy of different post-team selection input specifications. In our proposed empirical strategy, for simplicity, we limit the post-team selection inputs to individual experience and dyads' shared experiences (i.e., the number of prior surgeries a dyad in a surgical team performed together in a particular period).

3.4.1 RQ1: Relative influence of multiple decision-makers

Our first research question examines the relative influence of multiple decision-makers on OR time prediction. Knowing which decision-maker has the most influence on OR time prediction accuracy when using granular team information can help hospitals to determine the key decision-maker and the timing to update the OR time prediction before the day of surgery.

What differentiates our proposed empirical strategy from Kim et al. (2022) is a sequential introduction of post-team selection inputs that become available throughout multiple decision-making stages. Specifically, future studies can divide the post-team selection inputs into four sets of team information that become available to four decision makers: (1) block committee, (2) anesthesia manager, (3) anesthesia nursing manager, and (4) nursing manager. For example, a block committee would have information about surgeons' individual experiences and shared experiences of surgeons. Assuming that the anesthesiologist is assigned prior to the CRNA, an anesthesia manager would have information about the individual experience of the surgeon and anesthesiologist and the shared experience between surgeons and anesthesiologists. If the information is shared across departments, the amount of information accumulates moving toward the last decision-maker, the nursing manager, who would have information about the individual experience and shared experience of all surgical members. To examine the relative influence of each decision-maker, one can specify Equation 3.2 into four models with increasing amounts of information toward the last stage of decision-making. The reduction in prediction error from a previous model would represent the influence of the corresponding decision maker on OR time prediction.

3.4.2 RQ2: Value of cross-departmental information-sharing

Our second research question examines the value of information-sharing across different departments. Depending on the scope of information shared across different departments, one can specify Equation 3.2 into three models: (1) no information-sharing, (2) information-sharing about individual experience, and (3) information-sharing about both individual and team's shared experience. The reduction in prediction error from a previous model with additionally shared information would represent the degree of benefits from sharing different scopes of information across departments.

3.4.3 RQ3: Hierarchical influence of team member dyads

Our last research question examines whether the hierarchical structure within a surgical team is consistent with the ranked impacts of dyads. One avenue to answer this research question is to re-examine the dyad impact on OR time by specifying both individual and dyadic shared experiences of team members in Equation 3.2. Yet, the data must control for the differences in OR environments across surgeries, such as the differences between robotic and traditional surgeries. Controlling for the OR environments would provide a more comprehensive understanding of why prior studies provided mixed results on the relative importance of different individuals and dyads. Another avenue for future studies is to explore the complex relationship among dyadic shared experience, hierarchical structure, and OR time. For example, examining the mediator role of perceived hierarchical power difference on the relationship between dyad shared experience and OR time is one way. Identifying the moderator role of OR environments (e.g., robotic surgeries) between hierarchical power difference and team performance is another interesting area.

3.5 Concluding Remarks and Discussion

We used a systematic review to gain insights to evaluate the value of cross-departmental information-sharing on OR time prediction, considering the multiplicity of team scheduling decisions and the hierarchical structure of surgical teams. Our review findings confirm that there is a lack of empirical evidence on the benefits of using granular team information and updating OR time prediction in multiple stages of OR team scheduling on OR time prediction accuracy. Much less is known about cross-departmental information-sharing as many studies do not consider the multiplicity of team scheduling decisions by different departments. In addition, which information between individual experience and team's shared experience should be shared across different departments is not clear as most studies do not show the interaction effect between two types of experience on OR time. Furthermore, our review findings indicate that more comprehensive future research on the relationships among the team's shared experience, hierarchical power, and team performance would help understand whether the traditional hierarchical power of team members is closely aligned with the dyadic influence on OR time. Lastly, our SLR findings call for additional empirical insights on OR time prediction using granular team information in OM/MS literature.

Taking together the insights from the reviewed studies and our benchmark study, we propose several empirical strategies that would help evaluate the relative influence of multiple-decision makers, the value of cross-departmental information-sharing, and the relative impact of dyads with different hierarchical power on OR time prediction. The resulting empirical insights through these strategies would help hospitals focus their resources and effort to improve their OR time prediction by selecting key decision-makers and key information about team members to be shared across different departments.

Table 3.1: Multi-stage team scheduling decisions and available team information by timeline

	Scheduling Manager	Scheduled Member	Paired Member	Hierarchical Status	Cross-Department	Available Team Information: Unique Scheduled Individuals	Unique Paired Individuals	Unique Dyads
Within 1–2 Month(s)	Block committee	Surgeon	Surgeon	High	No	315	314	98,910
Within 1 Week	Anesthesia	Anesthesiologist	Surgeon	High	Yes	56	315	3,865
	Anesthesia nursing	CRNA	Surgeon	High	Yes	163	315	7,048
	Anesthesia nursing	CRNA	Anesthesiologist	Medium	Yes	163	56	3,914
Within 24 Hours	Nursing	Circulator	Surgeon	High	Yes	253	315	15,714
	Nursing	Circulator	Anesthesiologist	Medium	Yes	253	56	4,231
	Nursing	Circulator	CRNA	Low	Yes	253	163	7,788
	Nursing	Surgical Tech	Surgeon	High	Yes	214	315	11,831
	Nursing	Surgical Tech	Anesthesiologist	Medium	Yes	214	56	3,711
	Nursing	Surgical Tech	CRNA	Low	Yes	214	163	7,068
	Nursing	Surgical Tech	Circulator	Low	No	214	253	14,405

This table describes the multi-stage decisions in surgical team scheduling. We assume that the anesthesia manager's decision occurs earlier than the anesthesia nursing manager's decision by which information regarding the anesthesiologist and surgeon/anesthesiologist dyad becomes available first. The first column after the timeline indicate the scheduling manager at each timeline, who schedules members indicated in the second column (e.g., block committee assigns surgeon). The third column indicates the members assigned in the previous stage and paired with team members assigned at each timeline. The fourth column indicates the hierarchical status of a dyad (i.e., the pair of a scheduled member and the paired member) in low, medium, and high, based on the combined hierarchical power of two members consisted in the dyad. The fifth column indicates whether the dyad consists of two members from across different departments. The last three columns show the available team information in terms of the number of unique individuals that each scheduling manager can schedule, the number of unique individuals that can be paired with the scheduled members, and the unique dyads as a combination of scheduled and paired members (based on the Kim et al. (2022)'s data of 81,967 surgeries between March 2016 and June 2019).

Table 3.2: Search keywords and preliminary Google Scholar search outcome

	Topic	Context: Operating room	Surgical team	Surgeon
Team Scheduling	Team scheduling	51	27	46
	Team assignment	142	42	147
	Team configuration	107	53	120
	Team allocation	35	20	57
	Team composition	2,150	893	2,380
Team Familiarity	Team familiarity	497	446	550
	Shared experience	1,470	311	8,790
	Team experience	1,240	1,430	2,620
	Familiar team	63	47	88
	Group familiarity	17	8	45

This table shows our preliminary Google Scholar search outcomes in the number of searches in the full text of publications, working papers, and dissertations (search date: May 28, 2023). The two left columns indicate the broad topic areas and variations of topic keywords in the areas. The three right columns show the number of Google Scholar search results for each combination of topic keywords and context keywords (e.g., Operating room, Surgical team, Surgeon).

Table 3.3: Initial search outcome by database and topic

	Team Scheduling Studies	Team Familiarity Studies	All Studies
JSTOR	55	220	275
ProQuest	15	22	37
Web of Science	35	79	114
PubMed	15	51	66
Total	120	372	492

This table shows our initial search outcome by the database (the first column), by our topic keyword areas (e.g., team scheduling and team familiarity), and the number of all searched studies (the last column) by the database.

Table 3.4: Screening stages and the results

		Candidate Studies	Removed Studies	Remaining Studies
Screening Stage 1	Article redundancy	492	73	419
Screening Stage 2	Title and journal relevance	419	357	62
Screening Stage 3	Abstract relevance	62	31	31
Screening Stage 4	Full text relevance	31	13	18

This table shows our screening stages and the outcomes, in the number of candidate studies before each screening stage, removed studies by each screening stage, and remaining studies after each screening stage.

Table 3.5: The final 18 studies from the literature search and their data analysis methodologies

	Study	Journal	Primary Methodology	Panel Data Period	Specialty Coverage	Primary Team Predictor	Primary Outcome
1	Reagans et al. (2005)	Manage Sci	Panel causal inference	5 years	1	Individual, team, organizational exp.	OR time
2	Kurmann et al. (2014)	World J. Surg.	Experiment	6 months	1	Team experience	Patient outcome
3	He et al. (2014)	Surg. Endosc.	Panel causal inference	1 year	1	Team size	OR time
4	Özdemir-van Brunschot et al. (2015)	World J Urol	Experiment	3 years	1	Team experience	OR time
5	Maruthappu et al. (2016)	J R Soc Med	Panel causal inference	14 years	1	Team experience	OR time
6	Sexton et al. (2018)	BMJ Qual Saf	Descriptive analysis	-	1	Team experience	Communication effectiveness
7	Frasier et al. (2019)	J. Surg. Res.	Panel causal inference	12 surgeries	-	Team experience	Communication effectiveness
8	Avgerinos et al. (2020)	Hum Relat	Panel causal inference	8 years	1	Team experience	OR time
9	Bayram and Chen (2020)	Eur. J. Ind. Eng.	Analytical modeling	4 years	1	Individual experience	OR time
10	Parker et al. (2020)	World J. Surg.	Panel causal inference	5 years	1	Team experience	OR time
11	Stucky et al. (2020)	AORN J	Panel causal inference	3 months	1	Degree centrality	Communication effectiveness
12	Cahan et al. (2021)	Bone Joint J	Panel causal inference	1 year	1	Intraoperative staff presence	OR time
13	Mathis et al. (2021)	Surgery	Panel causal inference	4.5 years	1	Team experience	OR time
14	Harmanli et al. (2021)	Int Urogynecol J.	Analytical modeling	5.5 years	1	Individual experience	OR time
15	Vaulont et al. (2021)	J Appl Psychol	Panel causal inference	2 years	> 1	Team experience	Post-operative time
16	Stucky et al. (2022)	J. Perianesth. Nurs.	Panel causal inference	6 months	1	Team experience	OR time
17	Zhang and Zheng (2022)	Am. J. Surg.	Panel causal inference	2 years	1	Team experience	OR time
18	Bayram et al. (2023)	Comput Ind Eng	Analytical modeling	4 years	1	Individual experience	OR time
Benchmark	Kim et al. (2022)	SSRN	Predictive analysis	3 years	27	Diverse team characteristics	OR time

This table shows the final 18 studies from our literature search, the publication domain (i.e., journal), and their data analysis methodologies including their primary methodology, panel data period, specialty coverage of their surgery data, primary team predictor of interest, and primary outcome of interest. The studies are sorted by year-last name (first author). We classified two studies that utilized natural experimental data (e.g., teams in different time periods have different team compositions by the institutional setting) as experiment/field studies (Kurmann et al. 2014, Özdemir-van Brunschot et al. 2015).

Table 3.6: Data categories collected by the final 18 studies

Category	Variable Category	SLR Studies	Example Reference	Kim et al. (2022)
OR Outcome	OR duration	16	Reagans et al. (2005)	Yes
Data Categories	OR turnover time	2	Stucky et al. (2022)	Yes
	OR timestamps	5	Harmanli et al. (2021)	Yes
Pre-team	Procedural texts	0	Not available	Yes
Selection	Procedures codes	15	Reagans et al. (2005)	Yes
Data Categories	Patient demographics	15	Reagans et al. (2005)	Yes
	Patient condition	15	Reagans et al. (2005)	Yes
Post-team	Staff workload	0	Not available	Yes
Selection	Staff pre-surgery schedule	0	Not available	Yes
Data Categories	Staff panel data	16	Reagans et al. (2005)	Yes
	Staff surveys	4	Kurmann et al. (2014)	No

This table summarizes the number of studies in our review that collect granular data categorized by OR outcome, pre-team selection, and post-team selection data. In classifying the types of data collected in each study, we leverage the data classification of our benchmark study. The benchmark study classifies the predictors of OR time prediction into two sets of information: pre-team selection information (i.e., information used for OR time prediction that becomes available before team selection) and post-team selection information (i.e., information used for OR time prediction that becomes available after team selection). Following the classification, we specify the pre-team selection information—including procedure and patient information—and post-team selection information—including granular information about staff who are assigned to surgical teams—as well as OR time information as an outcome measure. The last column specifies whether our benchmark study, Kim et al. (2022), collects the information specified in the variable category.

Table 3.7: Team scheduling information in the final 18 studies

Category	Team Consideration	SLR Studies	Example Reference	Kim et al. (2022)
Team	Partial team	4	Reagans et al. (2005)	No
Coverage	Full team	11	Avgerinos et al. (2020)	Yes
Staff Data	Individual level	3	Bayram and Chen (2020)	Yes
Granularity	Team level	6	Avgerinos et al. (2020)	Yes
	– Uniform team weighting	3	Mathis et al. (2021)	No
	– Non-uniform team weighting	0	Not available	Yes
	Dyad level	5	Avgerinos et al. (2020)	Yes
Dyadic	Distinct dyadic consideration	5	Avgerinos et al. (2020)	Yes
Coverage	Decision-maker consideration	0	Not available	Yes
	Hierarchy consideration	4	Avgerinos et al. (2020)	Yes

This table summarizes the number of studies in our review that considered different types and granularity of team consideration. The team coverage category in this table describes how many studies define their team using the information about a full team (i.e., a team of entire surgical team members) or partial team (i.e., a team of selected roles, such as a team of surgeons). Following the team coverage category, this table describes how many studies in the 11 studies that considered full team coverage utilized staff data at different granularity levels: individual, team, and dyad, as well as how many studies considered the most granular dyadic data with consideration of multiple decision-makers and hierarchical structure of surgical teams. The last column specifies whether our benchmark study, Kim et al. (2022), considered each team consideration component.

Bibliography

Zeynep Akşin, Sarang Deo, Jónas Oddur Jónasson, and Kamalini Ramdas. Learning from many: Partner exposure and team familiarity in fluid teams. *Management Science*, 67(2):854–874, 2021.

Sulaiman D Al Sabei, Leodoro J Labrague, Omar Al-Rawajfah, Raeda AbuAlRub, Ikram A Burney, and Sathish K Jayapal. Relationship between interprofessional teamwork and nurses' intent to leave work: The mediating role of job satisfaction and burnout. In *Nursing Forum*, volume 57, pages 568–576. Wiley Online Library, 2022.

Tammy D Allen, Ryan C Johnson, Kaitlin M Kiburz, and Kristen M Shockley. Work–family conflict and flexible work arrangements: Deconstructing flexibility. *Personnel psychology*, 66(2):345–376, 2013.

David Anderson, Guodong Gao, and Bruce Golden. Life is all about timing: An examination of differences in treatment quality for trauma patients based on hospital arrival time. *Production and operations management*, 23(12):2178–2190, 2014.

Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2009.

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

Emmanouil Avgerinos and Bilal Gokpinar. Task variety in professional service work: When it helps and when it hurts. *Production and Operations Management*, 27(7):1368–1389, 2018.

Emmanouil Avgerinos, Ioannis Fragkos, and Yufei Huang. Team familiarity in cardiac surgery operations: The effects of hierarchy and failure on team productivity. *Human Relations*, 73(9):1278–1307, 2020.

Sherry I Bame. Organizational characteristics and administrative strategies associated with staff turnover. *Health Care Management Review*, pages 70–86, 1993.

Matthew A Bartek, Rajeev C Saxena, Stuart Solomon, Christine T Fong, Lakshmana D Behara, Ravitheja Venigandla, Kalyani Velagapudi, John D Lang, and Bala G Nair. Improving operating room efficiency: machine learning approach to predict case-time duration. *Journal of the American College of Surgeons*, 229(4):346–354, 2019.

Armagan Bayram and Xi Chen. Optimising teams and the outcomes of surgery. *European Journal of Industrial Engineering*, 14(1):127–145, 2020.

Armagan Bayram, Senay Solak, Oz Harmanli, and Bahriye Cesaret. Selecting a winning team: Management of surgical team composition in robotic surgery. *Computers & Industrial Engineering*, 175:108819, 2023.

Alon Bergman, Hummy Song, Guy David, Joanne Spetz, and Molly Candon. The role of schedule volatility in home health nursing turnover. *Medical Care Research and Review*, 79(3):382–393, 2022.

Alon Bergman, Guy David, and Hummy Song. "i quit": Schedule volatility as a driver of voluntary employee turnover. *Manufacturing & Service Operations Management*, 2023.

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

Johan Braeken and Marcel ALM Van Assen. An empirical kaiser criterion. *Psychological Methods*, 22(3): 450, 2017.

Gerard Cachon and Christian Terwiesch. *Matching supply with demand: An introduction to operations management*. McGraw-Hill, 3rd edition, 2012. ISBN 978-0-07-352520-4.

Eli M Cahan, Henry C Cousins, Joshua T Steere, Nicole A Segovia, Matthew D Miller, and Derek F Amanatullah. Influence of team composition on turnover and efficiency of total hip and knee arthroplasty. *The bone & joint journal*, 103(2):347–352, 2021.

Brecht Cardoen, Erik Demeulemeester, and Jeroen Beliën. Operating room planning and scheduling: A literature review. *European journal of operational research*, 201(3):921–932, 2010.

Lei Chen. Curse of dimensionality. In *Encyclopedia of Database Systems*. 2009.

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

Hojun Choi, Ahmet Colak, Sina Golara, and Achal Bassamboo. Understanding the value of feature information sharing. *Working Paper*, 2022.

Joshua Choper, Daniel Schneider, and Kristen Harknett. Uncertain time: Precarious schedules and job turnover in the us service sector. *ILR Review*, page 00197939211048484, 2019.

Aryo Dewanto and Viera Wardhani. Nurse turnover and perceived causes and consequences: a preliminary study at private hospitals in indonesia. *BMC nursing*, 17:1–7, 2018.

P Dotson. Cpt (®) codes: What are they, why are they necessary, and how are they developed? *Advances in Wound Care*, 2(10):583–587, 2013.

Marinus JC Eijkemans, Mark Van Houdenhoven, Tien Nguyen, Eric Boersma, Ewout W Steyerberg, and Geert Kazemier. Predicting the unpredictable: a new prediction model for operating room times using individual characteristics and the surgeon's estimate. *The Journal of the American Society of Anesthesiologists*, 112(1):41–49, 2010.

Marcia Faller, Bob Dent, and Jim Gogek. A single-hospital study of travel nurses and quality: what is their impact on the patient experience? *Nurse Leader*, 15(4):271–275, 2017.

Samer Faraj and Yan Xiao. Coordination in fast-response organizations. *Management science*, 52(8):1155–1169, 2006.

Louis Ferré. Selection of components in principal component analysis: a comparison of methods. *Computational Statistics & Data Analysis*, 19(6):669–682, 1995.

Lane L Frasier, Sudha R Pavuluri Quamme, Yue Ma, Douglas Wiegmann, Glen Leverson, Eva H DuGoff, and Caprice C Greenberg. Familiarity and communication in the operating room. *Journal of Surgical Research*, 235:395–403, 2019.

Jody Hoffer Gittell et al. Relational coordination: Coordinating work through relationships of shared goals, shared knowledge and mutual respect. *Relational perspectives in organizational studies: A research companion*, pages 74–94, 2006.

Peter Groves, Basel Kayyali, David Knott, and Steve Van Kuiken. The 'big data' revolution in healthcare: Accelerating value and innovation. Technical report, 2013. URL `https://www.mckinsey.com/~/media/mckinsey/industries/healthcaresystemsandservices/ourinsights/thebigdatarevolutioninushealthcare/the_big_data_revolution_in_healthcare.pdf`.

Oz Harmanli, Senay Solak, Armagan Bayram, Beril Yuksel, and Keisha Jones. Optimizing the robotic surgery team: an operations management perspective. *International Urogynecology Journal*, 32:1379–1385, 2021.

Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

Laureen J Hayes, Linda O'Brien-Pallas, Christine Duffield, Judith Shamian, James Buchan, Frances Hughes, Heather K Spence Laschinger, and Nicola North. Nurse turnover: a literature review–an update. *International journal of nursing studies*, 49(7):887–905, 2012.

Wenjing He, Shenghao Ni, Gengzhen Chen, Xuewu Jiang, and Bin Zheng. The composition of surgical teams in the operating room and its impact on surgical team performance in china. *Surgical endoscopy*, 28:1473–1478, 2014.

Narges Hosseini, Mustafa Y Sir, CJ Jankowski, and Kalyan S Pasupathy. Surgical duration estimation via data mining and predictive modeling: a case study. In *AMIA annual symposium proceedings*, volume 2015, page 640. American Medical Informatics Association, 2015.

HRSA. Nurse workforce projections, 2020-2035. Technical report, 2022.

Kuang-Yuan Huang, Indushobha Chengalur-Smith, and Alain Pinsonneault. Sharing is caring: Social support provision and companionship activities in healthcare virtual support communities. *MIS quarterly*, 43(2):395–424, 2019.

Robert S Huckman and Bradley R Staats. Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management*, 13(3):310–328, 2011.

Maria Ibanez. How schedule flexibility affects job applications. *Available at SSRN*, 2022.

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

ITA Group. 7 things that cause nurse turnover (and 8 things that stop it). Web, 2023. Retrieved June 9, 2023, from https://www.itagroup.com/insights/employee-experience/7-things-cause-nurse-turnover-8-things-to-stop-it.

George J Jiang, Bing Liang, and Huacheng Zhang. Hedge fund manager skill and style-shifting. *Management Science*, 2021.

Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065): 20150202, 2016.

JH Jung, Christoph Schneider, and Joseph Valacich. Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management science*, 56(4):724–742, 2010.

Henry F Kaiser. The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1):141–151, 1960.

Diwas S KC and Christian Terwiesch. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management science*, 55(9):1486–1498, 2009.

Diwas S KC, Bradley R Staats, Maryam Kouchaki, and Francesca Gino. Task selection and workload: A focus on completing easy tasks hurts performance. *Management Science*, 66(10):4397–4416, 2020a.

Diwas Singh KC, Stefan Scholtes, and Christian Terwiesch. Empirical research in healthcare operations: past research, present understanding, and future opportunities. *Manufacturing & Service Operations Management*, 22(1):73–83, 2020b.

Rachel R Kelz, Timothy T Tran, Patrick Hosokawa, William Henderson, E Carter Paulson, Francis Spitz, Barton H Hamilton, and Bruce L Hall. Time-of-day effects on surgical outcomes in the private sector: a retrospective cohort study. *Journal of the American College of Surgeons*, 209(4):434–445, 2009.

Erin Kennedy, Lorelei Lingard, Christopher J Watling, Roberto Hernandez Alejandro, Jeanna Parsons Leigh, and Sayra M Cristancho. Understanding helping behaviors in an interprofessional surgical team: How do members engage? *The American Journal of Surgery*, 219(2):372–378, 2020.

Saravanan Kesavan, Susan J Lambert, Joan C Williams, and Pradeep K Pendem. Doing well by doing good: Improving retail store performance with responsible scheduling practices at the gap, inc. *Management Science*, 68(11):7818–7836, 2022.

Jaeyoung Kim, Ahmet Colak, Lawrence Fredendall, and Robert Allen. Reducing or time mismatches using data-driven operations scheduling. *Available at SSRN 4175660*, 2022.

Song-Hee Kim, Hummy Song, and Melissa Valentine. Cross-disciplinary temporary teams in health care organizations: The role of partner variety and shared experience. *SSRN Electronic Journal*, 2018. doi: 10.2139/ssrn.3176306.

Anita Kurmann, S Keller, F Tschan-Semmer, J Seelandt, NK Semmer, Daniel Candinas, and Guido Beldi. Impact of team familiarity in the operating room on surgical complications. *World journal of surgery*, 38: 3047–3052, 2014.

Lauren F Laker, Craig M Froehle, Jaime B Windeler, and Christopher John Lindsell. Quality and efficiency of the clinical decision-making process: Information overload and emphasis framing. *Production and Operations Management*, 27(12):2213–2225, 2018.

Juha-Matti Lehtonen, Paulus Torkki, Antti Peltokorpi, and Teemu Moilanen. Increasing operating room productivity by duration categories and a newsvendor model. *International Journal of Health Care Quality Assurance*, 2013.

Constanze Leineweber, Holendro Singh Chungkham, Rikard Lindqvist, Hugo Westerlund, Sara Runesdotter, Lisa Smeds Alenius, Carol Tishelman, et al. Nurses' practice environment and satisfaction with schedule flexibility is related to intention to leave due to dissatisfaction: A multi-country, multilevel study. *International journal of nursing studies*, 58:47–58, 2016.

Qin Li, Ben Lourie, Alexander Nekrasov, and Terry Shevlin. Employee turnover and firm performance: Large-sample archival evidence. *Management Science*, 68(8):5667–5683, 2022.

Lorelei Lingard, Meredith Vanstone, Michele Durrant, Bonnie Fleming-Carroll, Mandy Lowe, Judy Rashotte, Lynne Sinclair, and Susan Tallett. Conflicting messages: examining the dynamics of leadership on interprofessional teams. *Academic Medicine*, 87(12):1762–1767, 2012.

Guanyi Lu, Rex Yuxing Du, and Xiaosong Peng. The impact of schedule consistency on shift worker productivity: An empirical investigation. *Manufacturing & Service Operations Management*, 24(5):2780–2796, 2022.

Mahiben Maruthappu, Antoine Duclos, Charlie D Zhou, Stuart R Lipsitz, John Wright, Dennis Orgill, and Matthew J Carty. The impact of team familiarity and surgical experience on operative efficiency: a retrospective analysis. *Journal of the Royal Society of Medicine*, 109(4):147–153, 2016.

Michael R Mathis, Steven Yule, Xiaoting Wu, Roger D Dias, Allison M Janda, Sarah L Krein, Milisa Manojlovich, Matthew D Caldwell, Korana Stakich-Alpirez, Min Zhang, et al. The impact of team familiarity on intra and postoperative cardiac surgical outcomes. *Surgery*, 170(4):1031–1038, 2021.

Jerrold H May, William E Spangler, David P Strum, and Luis G Vargas. The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management*, 20(3):392–405, 2011.

Velibor V Mišić and Georgia Perakis. Data analytics in operations management: A review. *Manufacturing & Service Operations Management*, 22(1):158–169, 2020.

Shalini Misra and Daniel Stokols. Psychological and health outcomes of perceived information overload. *Environment and behavior*, 44(6):737–759, 2012.

Ken Moon, Prashant Loyalka, Patrick Bergemann, and Joshua Cohen. The hidden cost of worker turnover: Attributing product reliability to the turnover of factory workers. *Management Science*, 68(5):3755–3767, 2022.

Ken Moon, Patrick Bergemann, Daniel Brown, Andrew Chen, James Chu, Ellen A Eisen, Gregory M Fischer, Prashant Loyalka, Sungmin Rho, and Joshua Cohen. Manufacturing productivity with worker turnover. *Management Science*, 69(4):1995–2015, 2023.

NSI Nursing Solutions. 2022 NSI national health care retention & rn staffing report. Technical report, 2022.

NSI Nursing Solutions. 2023 NSI national health care retention & rn staffing report. Technical report, 2023.

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

William D Owens, James A Felts, and EL Spitznagel Jr. Asa physical status classifications: a study of consistency of ratings. *Anesthesiology*, 49(4):239–243, 1978.

Denise MD Özdemir-van Brunschot, Michiel C Warlé, Michel F van der Jagt, Janneke PC Grutters, Sharon BCE van Horne, Heinrich J Kloke, Johannes A van der Vliet, Johan F Langenhuijsen, and Frank C d'Ancona. Surgical team composition has a major impact on effectiveness and costs in laparoscopic donor nephrectomy. *World Journal of Urology*, 33:733–741, 2015.

Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, et al. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery*, 88: 105906, 2021.

Sonal Parasrampuria. Hospitals' use of electronic health records data, 2015-2017. Technical report, 2019. URL `https://www.healthit.gov/sites/default/files/page/2019-04/AHAEHRUseDataBrief.pdf`.

Shin Hye Park, Stephanie Gass, and Diane K Boyle. Comparison of reasons for nurse turnover in magnet® and non-magnet hospitals. *The Journal of nursing administration*, 46(5):284–290, 2016.

Sarah Henrickson Parker, Xue Lei, Shimae Fitzgibbons, Thomas Metzger, Shawn Safford, and Seth Kaplan. The impact of surgical team familiarity on length of procedure and length of stay: inconsistent relationships across procedures, team members, and sites. *World journal of surgery*, 44:3658–3667, 2020.

Yuqi Peng, Yan Dong, Ahmet Colak, and Sriram Venkataraman. Bilateral r&d productivityand supply chain networks. *SSRN*, 2022.

Gary P Pisano, Richard MJ Bohmer, and Amy C Edmondson. Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Science*, 47(6):752–768, 2001.

Ray Reagans, Linda Argote, and Daria Brooks. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management science*, 51(6):869–881, 2005.

ResearchAndMarkets.com. Us outpatient surgical procedures market by surgical procedure type, patient care setting - us forecast to 2023 report. Technical report, 2019.

Stephen A Rhoades. The herfindahl-hirschman index. *Fed. Res. Bull.*, 79:188, 1993.

Aleda Roth, Anita L Tucker, Sriram Venkataraman, and Jon Chilingerian. Being on the productivity frontier: Identifying "triple aim performance" hospitals. *Production and Operations Management*, 28(9):2165–2183, 2019.

Daniel Schneider and Kristen Harknett. Consequences of routine work-schedule instability for worker health and well-being. *American Sociological Review*, 84(1):82–114, 2019.

Kevin Sexton, Amanda Johnson, Amanda Gotsch, Ahmed A Hussein, Lora Cavuoto, and Khurshid A Guru. Anticipation, teamwork and cognitive load: chasing efficiency during robot-assisted surgery. *BMJ quality & safety*, 27(2):148–154, 2018.

Karen Shader, Marion E Broome, Carroll D Broome, Mary Ellen West, and Mary Nash. Factors influencing satisfaction and anticipated turnover for nurses in an academic medical center. *JONA: The Journal of Nursing Administration*, 31(4):210–216, 2001.

Hummy Song and Robert S Huckman. Cohort turnover and operational performance: The july phenomenon in teaching hospitals. *Available at SSRN 3037753*, 2018.

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

Pieter S Stepaniak, Christiaan Heij, Guido HH Mannaerts, Marcel de Quelerij, and Guus de Vries. Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study. *Anesthesia & Analgesia*, 109(4):1232–1245, 2009.

Pieter S Stepaniak, Wietske W Vrijland, Marcel de Quelerij, Guus de Vries, and Christiaan Heij. Working with a fixed operating room team on consecutive similar cases and the effect on case duration and turnover time. *Archives of surgery*, 145(12):1165–1170, 2010.

Conor Stewart. Projected growth in global healthcare data volume 2020. In Statista - The Statistics Portal. Retrieved March 21, 2022, Web, 2020. URL `https://www.statista.com/statistics/1037970/global-healthcare-data-volume/`.

David P Strum, Allan R Sampson, Jerrold H May, and Luis G Vargas. Surgeon and type of anesthesia predict variability in surgical procedure times. *The Journal of the American Society of Anesthesiologists*, 92(5):1454–1466, 2000.

DP Strum, LG Vargas, AR Sampson, and JH May. Individual surgeon variability is a multiplicative function of surgical time. *Anesthesia & Analgesia*, 88(2S):48S, 1999.

Christopher H Stucky and Marla J De Jong. Surgical team familiarity: an integrative review. *AORN journal*, 113(1):64–75, 2021.

Christopher H Stucky, Marla J De Jong, Felichism W Kabo, and Christine E Kasper. A network analysis of perioperative communication patterns. *AORN journal*, 111(6):627–641, 2020.

Christopher H Stucky, Marla J De Jong, and Yuliang Liu. Military surgical team performance: the impact of familiarity, team size, and nurse anesthesia students. *Journal of PeriAnesthesia Nursing*, 37(1):86–93, 2022.

Christian Terwiesch, Marcelo Olivares, Bradley R Staats, and Vishal Gaur. Om forum—a review of empirical operations management over the last two decades. *Manufacturing & Service Operations Management*, 22 (4):656–668, 2020.

Manuel J Vaulont, Jennifer D Nahrgang, Margaret M Luciano, Lauren D'Innocenzo, and Carolyn T Lofgren. The room where it happens: The impact of core and non-core roles on surgical team performance. *Journal of applied psychology*, 106(11):1767, 2021.

Sriram Venkataraman, Lawrence D Fredendall, Kevin M Taaffe, Nathan Huynh, and Gilbert Ritchie. An empirical examination of surgeon experience, surgeon rating, and costs in perioperative services. *Journal of Operations Management*, 61(1):68–81, 2018.

J Deane Waldman, Frank Kelly, Sanjeev Arora, and Howard L Smith. The shocking cost of turnover in health care. *Health care management review*, 35(3):206–211, 2010.

Urban Wemmerlöv. The retrospective determination of process improvement's economic value at the individual manufacturing firm level: Literature review and proposed measurement framework. *Journal of Operations Management*, 67(2):182–211, 2021.

Hunter DD Witmer, Ankit Dhiman, Alonzo D Jones, Alison M Laffan, Dan Adelman, and Kiran K Turaga. A systematic review of operative team familiarity on metrics of efficiency, patient outcomes, cost, and team satisfaction. *Annals of Surgery*, 276(6):e674–e681, 2022.

Christina Wright, Peggy McCartt, Diane Raines, and Marilyn H Oermann. Implementation and evaluation of self-scheduling in a hospital system. *Journal for Nurses in Professional Development*, 33(1):19–24, 2017.

Yan Xiao, Alan Jones, Beilei Belinda Zhang, Monica Bennett, Simon C Mears, Jay D Mabrey, and Donald Kennerly. Team consistency and occurrences of prolonged operative time, prolonged hospital stay, and hospital readmission: a retrospective analysis. *World journal of surgery*, 39(4):890–896, 2015.

Yuqian Xu, Mor Armony, and Anindya Ghose. The interplay between online reviews and physician demand: An empirical investigation. *Management Science*, 67(12):7344–7361, 2021.

Jie Yang, Christina WY Wong, Kee-hung Lai, and Alfred Ngome Ntoko. The antecedents of dyadic quality performance and its effect on buyer–supplier relationship improvement. *International Journal of Production Economics*, 120(1):243–251, 2009.

Y Tony Yang and Diana J Mason. Covid-19's impact on nursing shortages, the rise of travel nurses, and price gouging. *Health Affairs Forefront*, 2022.

Sara Zaniboni, Donald M Truxillo, and Franco Fraccaroli. Differential effects of task variety and skill variety on burnout and turnover intentions for older and younger workers. *European Journal of Work and Organizational Psychology*, 22(3):306–317, 2013.

Yao Zhang and Bin Zheng. Familiarity of surgical teams: Impact on laparoscopic procedure time. *The American Journal of Surgery*, 224(5):1280–1284, 2022.

Jinshi Zhou and Franklin Dexter. Method to assist in the scheduling of add-on surgical cases-upper prediction bounds for surgical case durations based on the log-normal distribution. *The Journal of the American Society of Anesthesiologists*, 89(5):1228–1232, 1998.

Zhengyuan Zhou, Daniel Miller, Neal Master, David Scheinker, Nicholas Bambos, and Peter Glynn. Detecting inaccurate predictions of pediatric surgical durations. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 452–457. IEEE, 2016.