Mechanisms Driving Service Duration: A Large-Scale Empirical Analysis

by

Seyyed Mohammad Hossein Eshraghi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Management Sciences

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Using large-scale MRI services data, a multi-type multi-priority scheduling system, we measure service duration and show that a number of covariates, including the shift during which the procedure is performed, patients' priority, case mix workload of the proceeding patients and batching of similar procedures affect service duration. We find that the effects of various mechanisms on service duration may depend on the hospital type (teaching or community hospitals) and customer type (high priority versus low priority patient). For instance, adjusting for other factors, we find that MRI scan duration for emergent patients during the night shifts is significantly longer than the day shifts (around 4%) and the decrease is even higher in teaching hospital settings (around 8%), but low priority patients undergo shorter procedure duration during night shifts. We also show that the inverted-U-shaped relationship observed between the service duration and workload in the literature is also evident in the MRI services. We also find that sequencing consecutive procedures of similar body types is a significant mechanism that reduces patients' MRI scan time. As a result, adding an extra job of a similar scan type reduces the service time by 4%. We find that the effects of workload and sequencing are both endogenous, thus the OLS estimator might fail to determine the true effects. Thereby, we constructed a simultaneous equations model and used a three-stage least square (3SLS) estimator to correct the endogeneity and simultaneity biases of workload and sequencing factors, respectively.

Acknowledgements

First, I would like to thank my supervisor Dr. Hossein Abouee-Mehrizi for his dedicated support, guidance and insightful feedback throughout this research. I also express my thanks to Dr. Somayeh Sadat for her thoughtful helps and recommendations.

Second, I would like to thank Dr. Selcuk Onay and Dr. Jangho Yang, who accepted to be the readers of this thesis and helped me with their comments.

Lastly, I express my sincere gratitude to my family and Reyhaneh for all the support they have shown me through this research, the culmination of two years of distance learning.

Dedication

This thesis is dedicated to my parents. For their endless love and encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

COVID-19 has made resource management in health systems very critical since many hospitals have reduced the resources allocated to elective medical procedures to ensure that they have enough capacity for patients who are infected with COVID-19. Due to this adjustment in resource allocation, many elective medical procedures have been delayed. For instance, in Ontario, Canada, pandemic has left backlog of more than 16 million of medical procedures with MRI services having the highest number of cancelled or delayed procedures (around 500,000 cases) (Ireland, 2021).

Queueing theory is one of the main tools utilized in the operations management literature to analyze service systems and get insights on different aspects of the system such as staffing, scheduling, and pricing. Early analytical models mainly ignore the behavioral factors affecting the performance of the system (see e.g., Shortle et al. 2018) due to the lack of empirical evidence revealing the factors that should be considered in analyzing service systems. Recent empirical papers on the impact of behavioral factors affecting the performance of service systems shed light on how queueing models utilized to study service providers should be modified (Delasay et al., 2019). As a result, there have recently been papers in the literature that consider insights released by empirical papers in their modeling of service systems (see e.g., Delasay et al. 2016a, Abouee-Mehrizi and Barjesteh 2019). Furthermore, empirical studies in the literature find an inverted U-shaped relationship between workload and staff performances in different settings (see e.g., Tan and Netessine 2014) However, there are still important hidden aspects of service systems that should be revealed by empirical studies to provide insights for analyzing and improving the performance of service systems. For example, does an inverted U-shaped relationship between workload and procedure duration hold in a Magnetic Resonance Imaging (MRI) setting? Does the procedure starting time during a day affect the service time? If so, does the patients' priority level matter? Furthermore, how does sequencing the MRI scan appointments affect the service performance? Moreover, do behavioral mechanisms observed in the whole network of hospitals hold for individual sites as well? Having access to a large dataset with granular patient-level data, the focus of this thesis is to answer these questions and provide detailed insights on behavioral factors affecting the performance of service systems.

The extant literature on behavioral queueing science has been on the increase with the recent availability of data that allows exploring behavioral relationships. Most recently, Delasay et al. (2019) develop a general framework that conceptualizes mechanisms through which load characteristics (changeover, load, and extended load) affect service time determinants (work content, service speed, and in-process delay) at the level of each system component (server, network, and customer). They identify a total of 25 mechanisms, out of which 15 impact service time components through the server (namely physical setup, forgetting, loss of rhythm, task reduction, engagement, server early task initiation, multitasking-cognitive sharing, social speedup pressure, social loafing, multitasking- time sharing, multitasking-interruptions, service cancellation, learning by doing), and four impact service time components through the customer (namely customer early task initiation, return, abandonment, and deterioration).

We were able to get access to a large-scale data over a long period of time that allows a multi-mechanisms exploration at the aggregate level (e.g. all servers pooled together) as well as more granular levels (e.g. by different server types or at the level of individual servers). In fact, one of the contributions of this thesis is demonstrating how various mechanisms that impact system performance may differ in magnitude at different levels of the system, as well as for different types of customers (e.g. high priority versus low priority patients).

We examine these mechanisms in the setting of MRI scheduling and delivery in the province of Ontario, Canada. MRI is a medical imaging technique that generates images of the organs in the body. In addition to exploring what behavioral mechanisms affect service times, estimating a predictive model to obtain an accurate forecast for the service duration of patients is crucial for bringing down unnecessary delays. Excessive wait times are emotionally draining for patients and families and can worsen health outcomes for patients and incur costs to patients, caregivers, and health systems (Sutherland et al., 2019). It is estimated that excessive MRI wait times in Canada directly impacted the Canadian economy by imposing 0.7 billion dollars in costs (271.1 million dollars in the province of Ontario), and their aggregate (sum of direct, indirect, and induced) impact on the Canadian economy was at 1.4 billion dollars (535.3 million dollars in Ontario) in 2017 (Sutherland et al., 2019). The direct impact includes the effect of reduced employment due to excessive wait times, while the indirect and induced effects include the costs of other supply chain impacts and consumer spending resulting from the reduced output (Sutherland et al., 2019). Moreover, the COVID-19 pandemic exhausted the healthcare system all over the world and caused huge backlogs of elective surgeries. Wang et al. (2020) estimated that between March 15 and June 13, 2020, the incremental backlog in the province of Ontario is roughly more than 145,000 surgeries.

MRI scheduling and delivery is an interesting context from both the queueing and scheduling modeling perspective. In this context, customers (patients) and servers (providers) co-produce services. The system involves multi-type multi-priority customers with priority-specific service level constraints (i.e. wait time targets), and the behavioral factors on the side of servers and customers affect service duration.

In this thesis, we conduct econometric analyses that look into determinants of service (procedure) duration. We explore roughly 2 million records of service (procedure) time duration 66 hospitals respectively, including both teaching and community hospitals in the province of Ontario, Canada. We measure service (procedure) duration and show that a number of covariates, including the shift during which the procedure was performed, case-mix workload of the proceeding patients and batching of similar procedures affect service duration, likely through the following mechanisms: social speedup pressure, task reduction, early task initiation, forgetting, learning by doing, engagement, and physical setup. Due to the large-scale data that we have access to, we are able to investigate the model for teaching and community hospitals separately, as well as at the level of individual hospitals. In fact, one contribution of this thesis is showing how the magnitude of the effect of various mechanisms on service duration may depend on the server type (i.e. teaching or community hospital), as well as on the type of the customers (e.g. high priority versus low priority patient). For example, we show that the inverted-U-shaped relationship observed between the service duration and workload depends on the hospital type. Interestingly, while the inverted U-shaped relationship is observed at both hospital types, the semielasticity of workload on procedure duration for scans with average workload level is steeper at teaching hospitals, i.e., the inverted U-shaped behavior is more narrow. Moreover, we find that MRI scan duration for emergent patients during the night shifts is significantly longer than that in the day shifts (around 4%) and the decrease is even higher in teaching hospital settings (around 8%) than community hospitals, but low priority patients undergo shorter procedures duration during night shifts than high priority patients.

The remainder of this thesis is organized as follows. In Chapter 2, we present relevant literature to this thesis. Chapter 3 includes the empirical setting in this thesis along with the proposed hypotheses. In Section 3.1 we present our research setting and in Section 3.2 we encompass the hypotheses developments for our research. In Chapter 4 we discuss the dataset and the variable definition for the empirical model along with some descriptive analysis to support our model. Chapter 5 delivers the econometric specification along with the endogeneity challenges of our model. Chapter 6 includes the results and the robustness checks for our empirical models that incorporate varying data settings and modification to our econometric model and the estimating procedure that we adopted for our main results. Lastly, Chapter 7 presents our discussion, limitations, and proposed areas for future research which concludes this thesis.

Chapter 2

Related Literature

This thesis is mainly related to two streams of papers in the literature, namely papers that empirically analyze service systems and papers that model scheduling of service delivery.

While there are many papers in the literature that study patient scheduling (see, e.g., Ahmadi-Javid et al. 2017, Dantas et al. 2018 and Zhu et al. 2019 for a comprehensive review of this literature), there are only a few papers that empirically investigate scheduling practices. For instance, Hockenberry et al. (2008) examine the effects of temporal distance and task repetition on the productivity of cardiac surgeons. They find that surgeon productivity decreases as the length of break between procedures increases; however, task repetition reduces that effect. Norris et al. (2014) empirically investigate how attributes such as the appointment time/date affect the patients' no shows or cancellations. They show that longer "lead time" is likely to result in higher cancellation rates. They further observe that as the appointment times distant from the midnight to the early morning, the cancellations and no shows rates decrease. Liu et al. (2018) study the connection between patient choice behavior and operational attributes in appointment scheduling and show that delays and in-clinic waiting times have a negative impact on patients' utility. Ibanez et al. (2018) empirically show that physicians deviate from FCFS discipline manner by selecting similar tasks (i.e. batching) and also prioritizing shortest expected processing time patients. Further, they show that "deviating" from the assigned order decreases the productivity. Nonetheless, in our framework, radiologists are not able to deviate from the scheduled appointments order and the main question lies within whether batching contributes to radiologists' performance or not.

Putting similar tasks together to improve the productivity in short term has been studied in other areas as well. For example, Staats and Gino (2012) empirically show that "same-day, same-task experience" has a positive effect on staff performance in a bank loan application processing. It has also been observed that the productivity of servers may increase as the number of services that they perform increases due to their learning by doing. In particular, they find for short term practice (i.e., during a single day), task specialization (based on the same-day volume) results in improvements in workers' productivity. However, are interested to show that how the task repetition (based on the past experience of a worker) improves the productivity. Contrarily, it has also been observed that when servers do not focus on one particular type of service and switch between different service types, the service duration increases due to physical setup required for different service types (see, e.g., Schultz et al. 2003), and forgetting the information required for the procedure (see, e.g., Delasay et al. 2016b).

There are also empirical papers in the literature that study the impact of workload on the performance of a system. For example, empirical papers show that as the workload in a system increases, due to social speedup pressure, servers may speed up to avoid delaying the service of people waiting. In a serial production system, Schultz et al. (1998) observe that adjacent workers may adjust their work speed as the workload (measured by inventory level between them) changes. Staats and Gino (2012) analyze the loan application-processing line in a bank and observe that loan application processors speed up as the load of applications increases. Speeding up to avoid delaying the service of people waiting has been observed in other settings as well, including supermarket cashiers and in-hospital patient transporters (see, e.g., Delasay et al. 2019 for a comprehensive review of the related literature).

Some empirical papers also demonstrate that servers may terminate service before completion or eliminate discretionary service steps as the workload in the system increases, which is defined as task reduction in Batt and Terwiesch (2017). For instance, Delasay et al. (2016b) analyze a dataset of emergency medical service (EMS) responses and realize that when the EMS load is very high, the emergency department staff admit patients coming through ambulances faster. They also argue that when EMS load is not critically high, the paramedics increase their engagement on the scene (i.e., spend more time with patients on the scene) to avoid transportation to a hospital. Moreover, as the workload increases, servers may be unable to keep a high level of effort for a long time and their speed decreases after sometime. This has been observed, for example, Kc and Terwiesch (2009) empirically demonstrate that higher workload decreases the service duration of cardiothoracic surgeries as the surgeons are initially motivated; however, this mechanism would not hold for longer periods of time and increase the service time eventually. Tan and Netessine (2014) estimate an inverted U-shaped relationship between the workload, which is defined as the average number of tables that waiters handle at the same time, and the meal duration in a restaurant chain setting. There are other empirical papers in the healthcare literature that focus on the effects of clinical or operational features on service quality (see, e.g., Lu and Lu 2017 and Lu et al. 2021).

To the best of our knowledge, this is the first empirical work that examines the effects of patient sequencing, procedure shifts, and workload on procedure duration in an advance scheduling framework. Using a large patient level data set, we construct an empirical framework to answer how batching the sequence of procedures, shifts that procedures are performed in, and short term workload level prior to performing an MRI scan, affect the system productivity. Furthermore, we acknowledge the feedback effects of the radiologists' performance on operational covariates that drive the service duration of other procedures.

While there are a few papers in the literature that empirically study patient scheduling, there are many papers that analyze the problem analytically. A stream of papers in this literature focus on determining the sequencing and timing of appointments within a period (a day for instance) so that the direct wait time of patients are minimized. For example, Denton et al. 2007, Mancilla and Storer 2012, and Begen and Queyranne (2011) study job sequencing in operations rooms. The same day allocation, which seeks to correctly use resources to schedule appointments for patients on the same day while facing different uncertainties including patient arrivals and no-shows, has been also studies in the literature see e.g. Green et al. (2006), Muthuraman and Lawley (2008), Kolisch and Sickinger (2008), Liu and Ziya (2014).

Another stream of papers that analytically investigate patient scheduling focus on advance scheduling of patients to the future periods. For example, Patrick et al. (2008) study advance scheduling of multi-priority patients considering patients' wait time targets. Gocgun and Puterman (2014) consider multi-priority advance scheduling in chemotherapy with due date and time windows. Truong (2015) studies advance scheduling of urgent and regular patients and characterizes the optimal scheduling policy. Diamant et al. (2018) also investigated the advance scheduling process of multi-type outpatients in a multi-stations. Sauré et al. (2020) extended the former work into accounting multi-priority, multi-class patient scheduling, while incorporating random appointment duration, nonetheless, patients' required service time depends merely on their class. In this thesis we empirically analyze advance patient scheduling to determine characteristics that should be considered in analyzing advance patient scheduling.

Chapter 3

Empirical Setting and Hypotheses

3.1 Empirical Setting

This research is based on patient-level data on MRI scheduling and delivery in Ontario, Canada. Canada has a universal, publicly funded health care system. Each provincial or territorial insurance plan covers a basket of services, free at the point of care, to provide access to health care that is based on need rather than the ability to pay. However, long waits for elective care and inequities still exist (Martin et al., 2018).

In Ontario, to receive MRI, patients have to be formally put on a waitlist once a decision to treat by their physician is made. Physicians specify the service type, which refers to the body part that needs to go through an MRI scan. Percent of cases by service type, as well as the mean, median, and standard deviation of the MRI scan procedure duration of each service type, based on our data from 2014 to 2017, are presented in Table 3.1.

Table 3.1: Descriptive summary for procedure duration

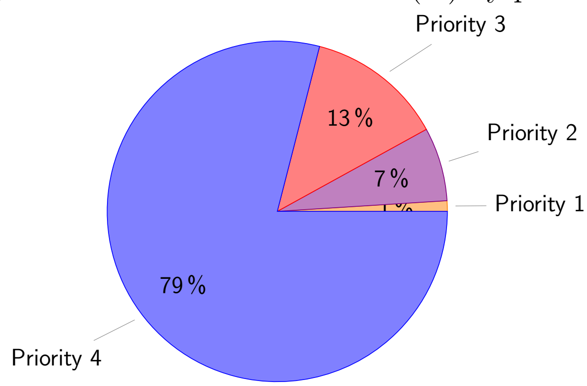
Service Types	% of Cases	Mean	Standard Deviation	Median
Abdomen	7.50%	36.25	13.96	34
Breast	2.49%	35.02	11.22	33
Cardiac	0.97%	59.37	21.12	59
Extremities	26.80%	27.65	11.26	25
Head (Brain)	28.7%	29.76	13.39	27
Head and Neck	4.10%	38.71	16.42	35
Pelvis	3.86%	38.90	14.04	37
Peripheral Vascular	0.08%	37.57	22.35	27
Spine	25.20%	27.27	13.33	23
Thorax	0.38%	45.01	19.09	42

Notes. The unit of reported numbers for procedure duration is by minutes. Total numbers of observations for descriptive summary of procedure duration is 1,954,088.

From now on, we may use the terms service duration/time and procedure duration/time, interchangeably.

Furthermore, physicians attribute a priority level to patients, based on clinical evidence. There are four priority levels: emergent (priority 1), inpatient or urgent (priority 2), semi-urgent (priority 3), and non-urgent (priority 4). By definition, failure to diagnose/treat would result in serious morbidity/mortality for priority 1 patients, significant deterioration/deficit for priority 2 patients, moderate deterioration/deficit for priority 3 patients, and minimal deterioration/deficit for priority 4 patients. Depending on these priority ratings, the Ontario Ministry of Health and Long-Term Care developed wait time

Figure 3.1: Patients distribution (%) by priority



targets under which the patients should be served; emergent levels: less than 24 hours, inpatient/urgent: within 48 hours, semi-urgent: within 2-10 days and non-urgent: within 4 weeks. Figure 3.1 shows patient distribution (percent) by priority in our dataset from the year 2014 to 2017.

Shortly after patients are put on the waitlist, the schedulers review their records and, depending on patient priority and available capacity, create an appointment (for a future procedure date) for the patient. On the procedure day, the scheduler assigns an MRI scanner for each procedure based on the type of patient receiving the procedure; *outpatients*: which is referred to individuals who arrive and depart on the same day of the scheduled procedure day, *inpatients*: who were admitted to the hospital before the procedure day and may remain afterward, *emergency*: which is referred to those who registered through the emergency department (ED) and *research*: which is referred to the rare number of participants in a research study. The scheduled MRI procedure might be cancelled and rescheduled into the future for several reasons (e.g., lack of facility resources, change in medical status, patients choosing to defer, no shows, etc.), hence, in case of procedure cancellations the scheduler may re-assign the scanner to accommodate another patient to avoid system's under-utilization.

3.2 Hypotheses Development

Service duration is the duration of time the actual procedure (MRI scan) takes from the start to finish, including the setup time. It is expected that patient clinical characteristics such as priority and service type affect service duration: Patients of higher priority may demand more diligence in treatment and as such take longer, and different service types, as described earlier in Table 3.1, vary in average service duration, since preforming MRI scans on different body parts requires distinct sequence types and orientations (Lauenstein and Semelka, 2006). However, we hypothesize that certain behavioral and mechanistic factors might also affect the relationship between patients' procedure duration and a number of characteristics not related to patient's clinical status, namely the shift during which the service was provided, case-mix workload of proceeding patients, clinic congestion and batching of services, even after controlling for patient's clinical characteristics (such as patient service type, and general anesthesia requirements), time trends, and hospital fixed effects.

We first explore the effect of workload on MRI procedure duration. Providers may respond differently to (extended) workload. On the one hand, service duration may increase due to a mechanism known as engagement, through which increased workload increases attention to all tasks (Delasay et al., 2019). In line with studies ergonomically showing that as engagement increases, labor performance decreases (Barker and Nussbaum, 2011), we hypothesize that performing back-to-back procedures on MRI machines would initially increase workload in the system, thus increasing service duration through the engagement mechanism, as providers are mentally fatigued (see e.g., Cakir et al. 1982, Fan and Smith 2017, Setyawati 1995). Furthermore, heavy workload may lead to exhaustion (Bentzen et al., 2016) and stress (Glaser et al. 1999, Rabe et al. 2012), which can decrease system's outcome. On the other hand, social speedup pressure exerts pressure on providers to speed up in order to avoid delaying the service of others (Delasay et al., 2019). Moreover, performing consecutive MRI scans may stimulate physicians' motivations (Locke, 1968), induce creativity (Shao et al., 2019) or improve cognitive performance (Lupien et al., 2007) which could reduce procedures' duration. These conflicting mechanisms would lead to the following hypothesis:

Hypothesis 1 *As the prior workload for procedures increases, the service duration of the next patient first increases and then decreases: i.e., there is an inverted U-shaped relationship between workload and service time.*

Hypothesis 1 is aligned with empirical findings in the literature that workload initially increases service time and then has a decreasing effect (e.g., Tan and Netessine 2014).

Another mechanism that may affect the MRI procedure duration is that performing repetitive tasks which require similar skills and methods (similar type) increases the productivity (e.g., Staats and Gino 2012), thus we hypothesize that performing a batch of similar service types decreases service duration through one or more of the following mechanisms: i) reducing physical setup since "additional tasks required when changing to service a different customer class" (Staats and Gino, 2012), ii) lack of forgetting which is "loss of required information from immediate memory" (see e.g., Bendoly et al. 2014, Froehle and White 2014), iii) learning by doing ("productivity gains through learning over short horizons") (e.g., see e.g., Delasay et al. 2019). Furthermore, healthcare providers may batch similar procedure to improve efficiency and save time (Ibanez et al., 2018). Thus we form our next hypothesis as:

Hypothesis 2 *Service duration decreases as the batches of procedures of similar type increase in size.*

We next consider the impact of time of night shifts on the MRI procedure duration. We hypothesize that after adjusting for procedures' type and severity and the level of congestion, the procedures performed overnight shifts are faster than the procedures that took place during day shifts. We believe that inconvenience hours at night shifts might drive physicians to perform non-discretionary tasks quickly (especially for less difficult procedures) (Oliva and Sterman, 2001) (i.e. task reduction or terminating service before completion or eliminating one or more discretionary service steps) due to the inconvenience of the night shift. Thus, we present our hypothesis regarding the impact of night shifts as follows:

Hypothesis 3A *Service duration of MRI procedures decreases in the night shifts.*

However, another stream of conflicting mechanisms can drive the physicians to perform MRI scans slower in the night shifts. Physicians' sleep quality may be affected by night shifts and thus the mechanism of "deteriorated cognitive efficiency" which may deteriorate the physicians' performance (Di Muzio et al. 2020, Amirian 2014) might kick in. Moreover, reduction in visual memory capacity at night shifts (Rollinson et al., 2003) may also decrease the physicians' performance, which would result in an increase in the service time (especially for more difficult procedures). Hence, we form the following competing hypothesis regarding the impact of night shifts:

Hypothesis 3B *Service duration of MRI procedures increases in the night shifts.*

In the next chapter, we introduce the dataset that we use to test our hypotheses and provide some descriptive statistics.

Chapter 4

Data and Variables

4.1 Data Collection

This research is based on empirical data gathered in the Wait Time Information System (WTIS) and MRI Efficiency Data. WTIS is a web-based tool that is used to collect and to report wait time information in the province of Ontario, Canada, for over 190 procedures in 13 key surgical areas and diagnostic imaging cases (Garay et al., 2015). In particular, we use the data related to the scheduling and delivery of MRI services in Ontario. The data includes the most major time stamps of the patient journey, including the decision to treat date and time (order received date and time), appointment created date and time, scheduled procedure date and time, rescheduled procedure date and time (if applicable), actual service date with start and finish time or alternatively (if applicable) date procedure no longer required, as well as facility and site name. For each patient, service type, priority level (1 to 4), whether the exam requires a general anesthetic and the assigned MRI scanner ID are recorded which account for heterogeneity across patients' medical conditions and enable us to capture these effects. The MRI efficiency dataset provides procedures' starting time (feet-in-time) and finishing time (feet-out-time). The MRI Efficiency dataset also includes the hospital type (teaching vs community) at which the patients' procedures were performed.

We received data for over 66 hospitals, incorporating more than 2.3 million patient time stamp entries from the year 2013 to 2017 captured in WITS. Due to the high volume of missing data before 2014, we further limited the data to the period starting from January 2014 to the end of March 2017. We also excluded records of patients who had been removed from the waitlist due to not requiring the procedure anymore or who had been scheduled for an appointment but missed their procedure ($\approx$ 7%). As a result, our records were reduced to a total of 1,954,088 service durations.

4.2 Measures and Controls

To test our hypotheses, we consider $ServDur_i$ as the dependent variable which includes the service time (setup time inclusive) of MRI scan procedure for patient i. We define this variable by calculating the difference between the actual starting and finishing time of procedures, measured in minutes.

We define the independent variables $\boldsymbol{Shift}_i$ as a categorical variable denoting the shift in which the procedure of patient i starts. The physicians' shifts changes at 7 A.M., 3 P.M. and 11 P.M., hence we use this variable to measure the effect of night shifts, in addition, to control for the effect of the shift itself for other hypotheses. We refer to the shift starts at 11 P.M., 7 A.M., and 3 P.M. as the first, second, and third shift, respectively. While most of the procedures were performed in the day shifts; ($\approx$ 46.8%) in the shifts starting from 7 A.M. and ($\approx$ 42.7%) in the shifts starting from 3 P.M., ($\approx$ 10.5%) of procedures were performed in the night shifts which starts from 11 P.M. until 7 A.M. in the next day. Retrospectively, we observe that the service time reduces in the night shifts. Specifically, the average service time during the first shift is 27.0 minutes and will increase to 32.8 and 28.4 minutes in the second and third shifts, respectively. Nonetheless, the increase in the procedure duration in the day shifts may be due to the following reasons: 1) less congestion in the hospital in the night shifts may reduce the stress level of physicians (Kc and Terwiesch, 2009) which increase their performance, 2) since less than 3% of procedures in the night shifts are emergent/urgent cases (versus 88.6% non-urgent cases), reduction in the service time may stem from the fact that procedures performed during these shifts are predominantly non-urgent cases which require less diligence to perform than the urgent cases, and 3) the scheduler may only offer night shifts appointments to patients with specific body type scans which are easier to perform than other parts (e.g., head versus extremities). All in all, heterogeneity in patients' medical conditions may also affect their procedure time in the night shifts, yet we employ certain control levels in the proposed specifications to properly test the hypothesis regarding the physicians' performance in night shifts.

The next set of variables that we define relates to the impact of past procedures on the upcoming MRI scan. Because our dataset does not track the physicians' identifier for each procedure, we consider procedures on each MRI machine within batches with at most τ minutes of idle time between two consecutive procedures. Figure 4.1 illustrates an arbitrary batch of procedures performed on a given scanner. The two cut-off points imply that there is at least τ minutes of idle time between finishing time of procedure a_{k_1} and starting time of procedure b_1 on the same MRI machine. This is also true for procedures b_{k_2} and c_1. Based on this batching criterion, we define Seq_i as a count variable indicating the total number of *consecutive* procedures of similar type (body part) right before the procedure of patient i on the same MRI machine and within the same batch. For instance, $Seq_i = 3$ implies that prior to the procedure of patient i, two other procedures with similar service types were performed consecutively on the same machine with at most τ minutes of idle time between them. This measure helps us to test whether as the number of consecutive procedures with a similar type increases, the learning by doing mechanism kicks in and reduces the service time (Hypothesis 2).

Table 4.1: Variables definition

Variable	Description
$ServDur_i$	Actual procedure (MRI scan) duration (minutes) for patient i (including setup time)
Seq_i	Count variable indicating the total number of consecutive procedures of similar type as of patient i's procedure (inclusive) on the same shift and scanner
$Shift_i$	Categorical variables indicating the shift of day in which the patient i's procedure took place
$Workload_i$	Total service duration of consecutive procedures prior to patient i in the same shift and scanner
$LWorkload_i$	Total service duration of procedures from previous week but within a 2-hours difference of starting hour from that of procedure i at the same hospital
$NProcShift_{h(i)}$	Total number of scheduled procedures on the same date and shift as patient i at the same hospital site h
$FirstJob_i$	Binary variable, equal to 1 if patient i's procedure was the first scan in the batch
$LastJob_i$	Binary variable, equal to 1 if patient i's procedure was the last scan in the batch
$UrgentRecent_i$	Count variable indicating the total number of urgent (first or second priority) procedures prior to patient i's procedure in the same batch and scanner
$Priority_i$	Clinical priority of patient i (1=emergent; 2=inpatient/urgent; 3=semi-urgent; 4=non-urgent)
$ServType_i$	Service type of patient i (body part being scanned)
$RepType_i$	Binary variable, equal to 1 if patient i's procedure body part is the same as patient $(i-1)$
$PatType_i$	Categorical variable indicating the entry type of patient i (Inpatient, Outpatient, Emergency, Research participant)
$GenAnesth_i$	Binary variable, equal to 1 if patient i's procedure requires general anesthesiology
$Teaching_{h(i)}$	Binary variable, equal to 1 if hospital h where the patient i was served is a teaching (vs. community) hospital
$OptHour_{h(i)}$	The hospital's total operating hour at which the patient i was served (8 hrs/16hrs/24 hrs)
$ScannerID_i$	The Scanner ID which is assigned to perform procedure of patient i
$Year_i$	Categorical variables indicating the year in which the patient i's procedure took place
$Month_i$	Categorical variables indicating the month in which the patient i's procedure took place
$Week_i$	Categorical variables indicating the week number (1 to 52) in which the patient i's procedure took place
$Weekday_i$	Categorical variables indicating the weekday (Mon. to Sun.) in which the patient i's procedure took place

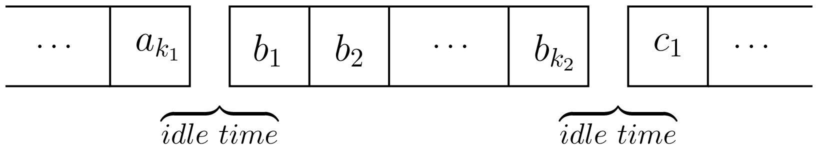


Figure 4.1: Batches of MRI procedures performed on a MRI machine

$Workload_i$ denotes the total service duration of procedures right before procedure i performed on the same MRI machine and within the same batch of procedures. Specifically,

$$Workload_i = \sum_{j \in \mathcal{B}_i^\tau} ServDur_j \cdot \mathbb{1}_{\{s_j < s_i\}},$$

where $\mathcal{B}_i^\tau$ denotes the set of procedures performed within the same batch and on the same machine as procedure i with at most τ minutes of idle time between two consecutive procedures. The starting time of procedures j and i are denoted by s_j and s_i, respectively, and $\mathbb{1}_{\{\cdot\}}$ is an indicator function. The more physicians perform back-to-back procedures or procedures that require more diligence (and naturally takes more time to perform), the more workload increases prior to the procedure of patient i. This variable helps us to capture the effect of extended workload on the upcoming MRI scan.

$FirstJob_i$ is an indicator variable for procedures performed at the beginning of each batch (after at least τ minutes of idle time from the previous procedure that was performed on the same machine). We use this variable to control the effect of sequence-dependent setup time (Mousakhani, 2013) since performing the first job in a batch may require more

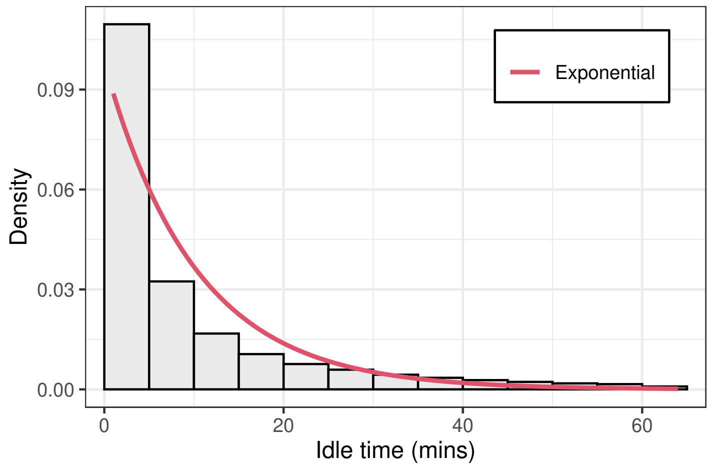


Figure 4.2: Distribution of idle time between procedures on the same machine

time to setup. Moreover, the procedure duration of the first job may include a warm-up time which affects the performance of the physician on the upcoming jobs (see e.g., Abdalla et al. 2015).

In order to conduct the hypothesis testing in the next sections, we consider $\tau = 15$ minutes as the cut-off threshold for determining a batch of back-to-back procedures. Figure 4.2 illustrates the empirical and fitted distribution for the idle time between procedures on an MRI scanner. We will use different cut-off thresholds (e.g., τ=5, 10, and 20 minutes) in the robustness checks provided in Section 6.2 to examine whether the results are consistent in different settings. Table 3.1 summaries the definition of variables.

We also control for several variables that could affect procedures' service duration. $NProcShift_{h(i)}$ accounts for the number of scheduled procedures in the same shift and

Table 4.2: Descriptive statistics of variables (n = 1,954,088)

Variable	Median	Mean	Std. dev.
(1) $ServDur$	27	29.98	13.72
(2) Seq	1	1.57	1.15
(3) $Workload$	31	63.04	86.68
(4) $LWorkload$	219	193.71	90.06
(5) $NProcShift$	12	12.11	3.93
(6) $FirstJob$	0	0.35	0.48
(7) $LastJob$	0	0.38	0.48
(8) $UrgentRecent$	0	0.16	0.52
(9) $RepType$	0	0.47	0.50
(10) $GenAnesth$	0	0.01	0.09
(11) $Teaching$	0	0.48	0.50
(12) $OptHour$	24	21.67	2.13

hospital as patient i's procedure, which controls the congestion level in each shift. We also control for patients' medical characteristics, namely, $Priority_i$ and $ServType_i$ which indicate the patient's severity level and the body part being scanned, respectively.

In addition to patients' characteristics, we control for the effect of hospital operational capacity using $OptHour_{h(i)}$ as a proxy that captures the daily operating hours of hospitals, which are either 8 hrs, 16 hrs or 24 hrs. Lastly, we include time trends as a set of categorical vectors that identify the year, week of the year, and weekday on which the procedure takes place. Weeks are numbered consecutively from 1 to 52, starting from the first week of January to the last week in December. Weekdays are numbered consecutively from 1 to 7 from Sundays to Saturdays. Including $Week_i$ and $Weekday_i$ in the specification reduces the unobserved effect of staff's skill, since the schedule of physicians and nurses remains almost the same throughout the weekdays. Figure 4.5 shows how the service duration pattern changes across all hospitals by date, month, day of the week, and hours of the day.

4.3 Descriptive Statistics

Table 4.2 shows the descriptive summary statistics of the patient-level variables considering $\tau = 15$ minutes as the cut-off point for the idle time between procedures for determining the batches (see Figure 4.1). Based on the historical data, the average service duration of MRI scans is 29.98 minutes. We observe that the average service duration in community hospitals is approximately 16.5% shorter than in teaching hospitals, as illustrated in Figure 4.3. This discrepancy motivated us to test the hypotheses in teaching and community hospitals separately, to examine whether the intrinsic behavior of physicians in teaching and community hospitals leads to dissimilar mechanisms driving procedure duration.

As Figure 4.3 illustrates, $ServDur$ seems to be right-skewed; hence, using the log transformation on the dependent variable increases the normality of the error term (Velleman and Hoaglin, 1981). Furthermore, using the log transformation on the dependent variable makes the interpretation of covariates easier, as changing the independent variable (while other variables are adjusted) translates into the percentage change in the dependent

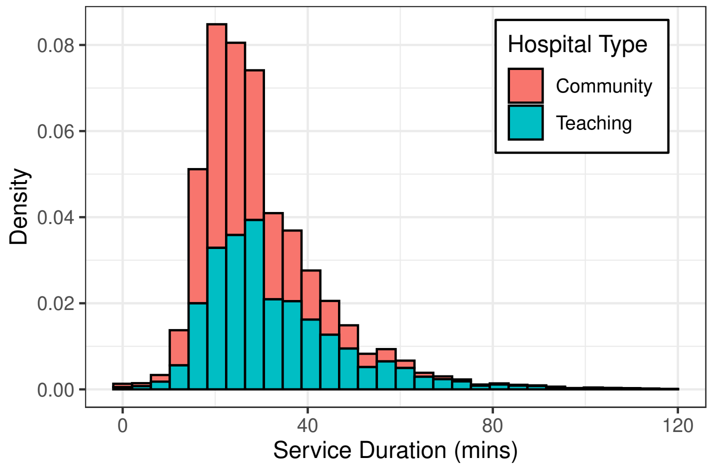


Figure 4.3: Histogram of service duration across teaching and community hospitals

variable (see e.g., Wooldridge 2016, Kennedy 2008).

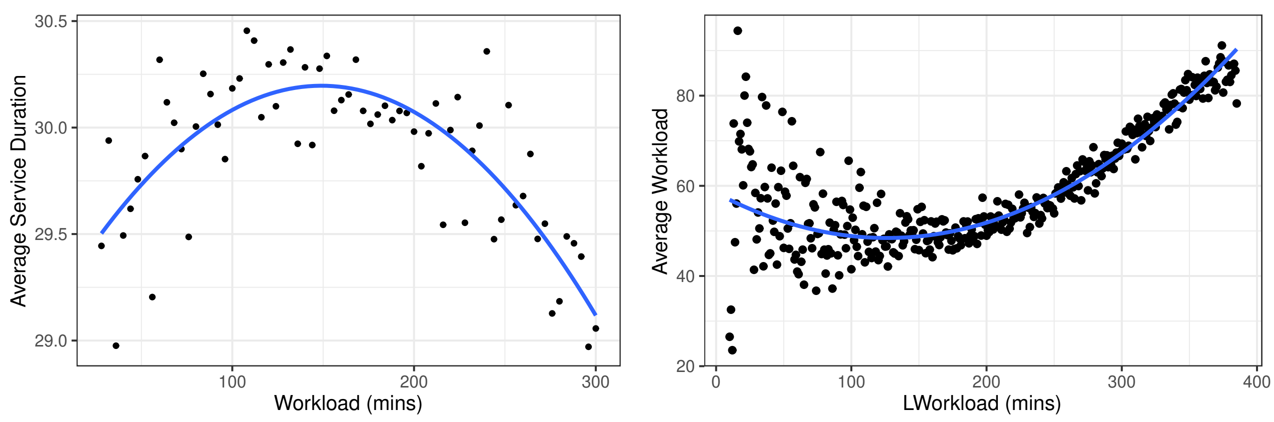
Table 4.3 presents the correlation matrix of continuous variables. We find that the correlation between $ServDur$ and Seq is approximately -0.13, implying that (before controlling for other factors) the batch size of consecutive procedures of similar type (i.e., body part) is associated with shorter procedure duration for the upcoming scan; nonetheless, an appropriate identification strategy is required to test whether or not this correlation implies causation.

The correlations between variables are relatively small, which confirms that there is no considerable multicollinearity problem in our model. We also obtain the variance inflation factor (VIF) of covariates to ensure all covariates have a VIF less than the commonly accepted threshold of 5 (Hair et al., 1998).

Table 4.3: Correlation matrix of continuous variables

Measures	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) $ServDur_i$	1.00											
(2) Seq_i	−0.13	1.00										
(3) $Workload_i$	−0.03	0.34	1.00									
(4) $LWorkload_i$	0.23	−0.07	0.00^a	1.00								
(5) $NProcShift_{h(i)}$	−0.30	0.14	0.19	−0.09	1.00							
(6) $FirstJob_i$	0.08	−0.37	−0.54	0.00^a	−0.21	1.00						
(7) $LastJob_i$	0.08	−0.08	−0.16	0.01	0.15	−0.07	1.00					
(8) $UrgentRecent_i$	0.03	0.04	0.46	0.06	0.07	−0.23	−0.07	1.00				
(9) $RepType_i$	−0.12	0.52	0.02	−0.07	0.04	−0.06	-0.00^a	−0.05	1.00			
(10) $GenAnesth_i$	0.18	−0.04	−0.05	0.08	−0.13	0.08	0.07	−0.02	0.01	1.00		
(11) $Teaching_{h(i)}$	0.20	−0.05	−0.05	0.28	−0.22	0.05	0.06	0.03	−0.03	0.09	1.00	
(12) $OptHour_{h(i)}$	0.05	−0.01	0.05	0.18	0.04	−0.02	−0.01	0.06	−0.03	0.00^a	0.11	1.00

Notes. [a] The absolute values are less than 0.005. All other correlation coefficients are significant at the 0.01 level.



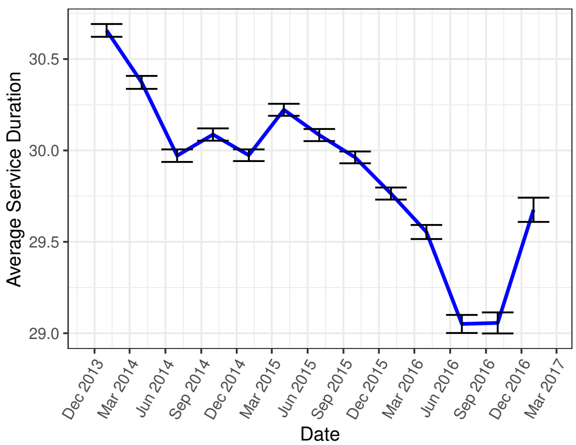
(a) Impact of workload on service duration

(b) Impact of lagged workload on workload

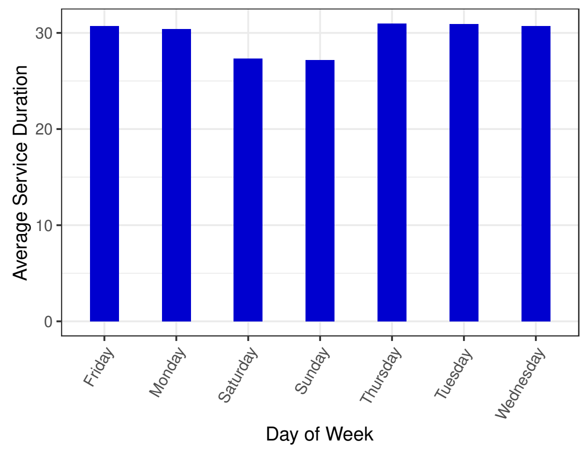
Figure 4.4: Descriptive of Workload and LWorkload

We also observe that the correlation between $Workload$ and $ServDur$ is not economically significant (-0.03). However, the Pearson correlation coefficient merely measures the strength of *linear* relationship between variables rather than *non-linear* relationships. Figure 4.4a depicts the relationship between workloads of prior procedures and the service duration of the current procedure. We observe that as the workload increases, the average of service duration first increases and then (roughly after 150 minutes) it begins to decrease.

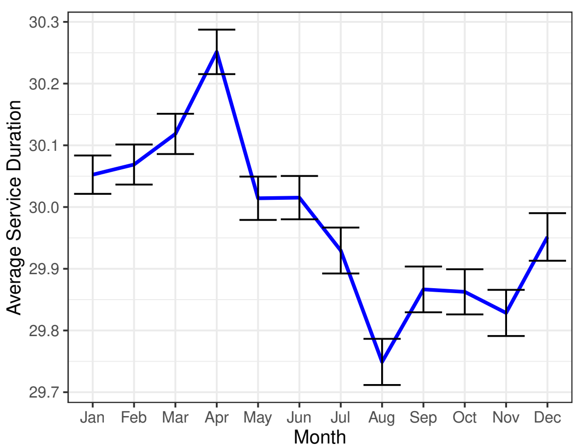
Figure 4.5a illustrates how the average procedure duration varies through time different time trends. As it can be observed, the average service duration generally has a decreasing trend from starting from the year 2014 until the starting year of 2017. This downtrend might be indicative of improvement in performing MRI scans throughout all hospitals. Hence, we should control for these trends in our model. Furthermore, Figure 4.5b demonstrates that the procedure duration of MRI scans is, in general, slower over the workdays, possibly due to more pressure from congestion levels on radiologists that affect their performance. Thereby, we control for the effects of the day of the week in our specification. Moreover, dependency of procedure duration across different months and different hours of the day is also illustrated in Figures 4.5c and 4.5d, respectively. We will control for these effects in our model to increase the strength of our results.



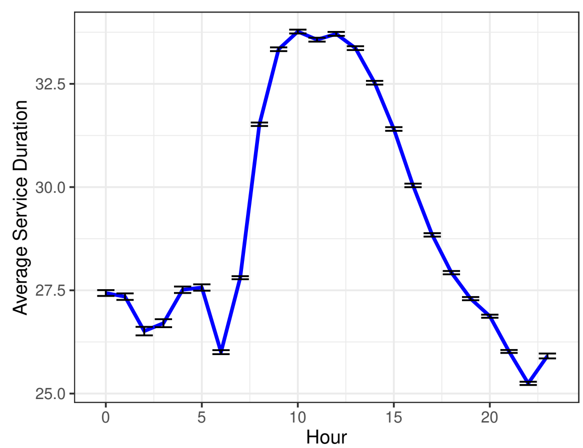
(a) The average service duration by date



(b) The average service duration by day of the week



(c) The average service duration by month



(d) The average service duration by hours of the day

Figure 4.5: The average service duration by date, month, day of the week, and hours of the day

Chapter 5

Econometric Specifications

In this chapter, we present a rigorous econometric specification that measures the impact of night shifts, workload, and sequencing on the MRI scan duration.

5.1 Econometric Model

To test the hypotheses regarding the drivers of MRI procedure duration, we estimate the following regression model:

$$\begin{aligned} log(ServDur_i) =& \beta_0 + \beta_{11} Workload_i + \beta_{12} Workload_i^2 + \boldsymbol{Shift}_i \boldsymbol{\beta}_s \\ &+ \beta_2 Seq_i + \boldsymbol{X}_i \boldsymbol{\beta}_p + \delta_{h(i)} + \epsilon_i, \end{aligned} \tag{5.1}$$

$Workload_i$ and $Workload_i^2$ denote the linear and quadratic effects of prior workload, which are included to account for their non-linear effects. Similar to Freeman et al. (2017), we standardized both linear and quadratic variables of workload for interpretation purposes that can capture the quadratic effects between workload and service time as well. In addition, $\boldsymbol{Shift}_i$ stands for the set of binary vectors for the shift in which the procedure i is performed. We consider the first shift denoted by $Shift_i^{(1)}$ (i.e., from 11 P.M. to 7 A.M.), which captures the night shifts, as the reference group. Therefore, $\boldsymbol{Shift}_i$ comprises two binary variables for the other shifts, i.e., the second shift, $Shift_i^{(2)}$ (from 7 A.M. to 3 P.M.) and the third shift, $Shift_i^{(3)}$ (from 3 P.M. to 11 P.M.) when MRI scan of patient i took place.

We also control for several factors that may affect the dependent variable. Specifically, $\boldsymbol{X}_i$ is a vector of control variables to control for time-invariant heterogeneity in attributes of patients and hospitals. The control vector includes the body part being scanned ($ServType_i$), the priority level of patients ($Priority_i$), the indicator capturing whether patient i required general anesthesiology prior to the MRI scan ($GenAnesth_i$), and patient entry type ($PatType_i$). Seq_i accounts for the effect of sequencing, which is a count variable for the total number of consecutive procedures of similar types as of patient i's procedure (inclusive). Note that, unlike the Ibanez et al. (2018) framework, since the procedures are scheduled in advance, radiologists may not be able to "deviate" from the procedure calendar. Hence, after controlling for the procedures' difficulty level (through the proxies $Priority$ and $ServType$) and the radiologists' engagement level (through the proxies $Workload$ and $Workload^2$), batches are assumed to be exogenous.

We also control for the congestion level in the shift ($NProcShift_i$), the scanner used to perform the procedure of patient i ($ScannerID_i$), and whether the patient i procedure started at the beginning of a chain ($FirstJob_i$), or at the end of a chain ($LastJob_i$). Moreover, as Figure 4.5 illustrates, time factors significantly affect the service duration, thus we include several temporal factors such as ($Year_i$), ($Month_i$), ($Week_i$) and ($Weekday_i$) to control for the seasonal trends which may affect procedure duration. We also incorporate the hospitals' operating hours per day ($OptHour_{h(i)}$) as a proxy to control for hospital operational capacity. The fixed effect of a hospital-scanner is included in $\delta_{h(i)}$, where $h(i)$ is the hospital-scanner at which the patient i is treated. This indicates that we consider the joint interaction between the effects of hospitals and MRI scanners as the fixed effect in our model. The reason is that scanners do not move between hospitals; therefore, including the effect of $ScannerID$ as an additional fixed effect in our model makes the effect of hospitals unidentified. Lastly, ϵ_i is the mean zero error term (see Table 4.1 for the detailed description of control variables).

Although specification (5.1) controls for the hospitals' heterogeneity, unobservable staffs' training level and skill may be correlated with both $Workload$ and $ServDur$. As the skill of the staff who performed procedure i increases, not only the service duration of patient i is likely to decrease (see e.g., Hancock 1986), but also the service duration of patients within the same batch and prior to patient i is likely to decrease, which is reflected in the workload. Disregarding this endogeneity may lead to omitted variable bias in the estimation of $Workload$. Furthermore, the impact of sequencing may also be endogenous. The reason is that as Ibanez et al. (2018) suggest, the scheduler may consider giving appointments to procedures with "shortest-expected processing time", hence procedures with short (expected) processing time may mechanistically be served at a longer sequence of procedures of similar types. In other words, procedures with short (expected) service duration might have higher sequencing orders than other MRI jobs. Neglecting this reverse causality may lead us to draw erroneous conclusions regarding the true impact of sequencing on the procedures' duration. Moreover, the emphasis on the "expected" procedure duration was because, at the time that appointments are assigned to patients, schedulers have merely an "approximation" of the service time. However, since their expectation is highly correlated with the actual service time that the patient will experience, we expect that the sequencing depends on the procedure duration.

We expect that the effect of night shifts ($\boldsymbol{Shift_i}$) is uncorrelated with radiologists' skill level and therefore, exogenous. In addition, specification (5.1) controls for the impact of congestion level and adjusts for the effect of procedures' difficulty (through the procedures' emergency level and service types) and radiologists' engagement level (through the workload). Therefore, changes in the service time during night shifts would be due to staff's behavioral factors than the attribute of patients or hospitals. Furthermore, although patients are not exogenously *assigned* to the night shifts but *choose* a night appointment, we expect that the impact of ($\boldsymbol{Shift_i}$) is exogenous. The reason is that patients may *choose* night shifts to reduce their waiting time than their procedure duration. Hence, their decision would not endogenously affect their service duration. We next propose an identification strategy to alleviate the endogeneity bias of OLS for estimating the impacts of workload and sequencing.

To overcome those potential endogeneity and simultaneity biases, we use the simultaneous equations models (SEM) to simultaneously specify the endogeneity of workload and sequencing through a system of equations. We then use a three-stage least square (3SLS) estimation method (Zellner, 1962) to consistently estimate the impacts of the endogenous covariates.

5.2 Endogeneity Challenge

We address the endogeneity challenge by introducing a set of instrumental variables to assess the endogenous treatments and estimate their true impact on service duration.

A valid instrument variable must satisfy two conditions (1) relevance, that is, it should be correlated with the endogenous variable, and (2) exclusion restriction, that is, it must be uncorrelated with the unobserved error term (Wooldridge, 2016). In other words, the impact of the instrument variable on service duration should be merely through the workload.

We utilize two types of instruments in our identification. First, we exploit $UrgentRecent_i$ as the count variable for the total number of urgent/emergent (i.e., priority 1 or 2) patients who have been served before patient i within the same batch and on the same MRI machine, i.e.:

$$UrgentRecent_i = \sum_{j \in \mathcal{B}_i^\tau} \mathbb{1}_{\{s_j < s_i | Priority_j \leq 2\}},$$

where $Priority_j$ and s_j are the priority level and starting time of patient j service, respectively. We include patients of priority 1 and 2 in the definition of $UrgentRecent$ as those patients should be served within 24 and 48 hours of their arrival time, respectively, compared to 10 and 28 days of priority 3 and 4, respectively. Thus, the scheduler treat them similarly.

The more urgent/emergency patients are served, the more workload increases for procedures in the same batch. Furthermore, performing urgent/emergency procedures leads to a significant increase in workload. Based on our descriptive analysis, the average procedure duration for patients with a high level of severity, takes 23.5% longer than that for semi-urgent or non-urgent procedures. In addition, $UrgentRecent$ is highly correlated with $Workload$ (≈ 0.46), thus satisfying the relevance condition.

We also expect that the total number of urgent/emergency procedures prior to each procedure affects the service duration of the upcoming procedure only through the workload and thus is uncorrelated with the staff's skills. The reason is that urgent/emergency procedures should usually get served under 48 hours. Thus, it is more likely that they accept the first available (earliest) appointment. The median of the waiting time from receiving an MRI order until its corresponding procedure starts is 3 hours for emergent patients and 23 hours for urgent patients, respectively. In comparison, the median waiting time to get the service for semi-urgent and non-urgent patients is 11.6 and 41.25 days, respectively. This difference shows that there are considerably fewer opportunities for emergent/urgent patients to book their appointment with preferred hospitals rather than semi and non-urgent procedures, resulting in an uncorrelated relationship between the staff's skill and schedule of emergent/urgent patients. Therefore, satisfying the exclusion restriction condition.

Similar to Tan and Netessine (2014), we also include the week lagged values of workload both in the linear and quadratic forms, namely, $LWorkload$ and $LWorkload^2$, as additional instrumental variables for both linear and quadratic endogenous variables of workload. Furthermore, since the number of procedures (along with their starting time) is not fixed for each batch, instead of calculating the lag of workload from the previous week at the *same hour*, we compute the lag of workload by taking the summation of service times for procedures which were performed at the same hospital from previous week within a 2-hours interval of the current procedure starting hour. For example, the lagged workload of a procedure taking place at 9 A.M. would be the summation of service time of procedures with starting times from 7 A.M. until 11 A.M., from the previous week at the same hospital. In Section 6.2, we will also use other interval thresholds (e.g., 30-minutes, 1-hour and 3-hours) to compute those instrumental variables to demonstrate the robustness of our results. We used a weekly delay to calculate the instrumental variables because hospital operational managers are likely to adjust their service capacity and radiologists' schedule based on the congestion level and workload of the previous week. Hence, we expect that the weekly lag of workload and workload are correlated, and satisfy the relevance condition. Furthermore, as Table 4.3 shows, the Pearson correlation between $Workload$ and $LWorkload$ is statistically zero; however, when we examine the graphical relationship between those two variables (see Figure 4.4b), a U-shaped pattern is evident. We also include the quadratic value of this lagged variable to account for the non-linear relationship. Therefore, $LWorkload_i = \sum_{j \in \mathcal{W}_i} ServDur_j \cdot \mathbb{1}_{\{|H(s_j) - H(s_i)| < 2\}}$, where $\mathcal{W}_i$ is the set of procedures performed at the same hospital as procedure i in the previous week, and $H(s_k)$ denotes to the hour at which the procedure k is performed. We again standardize both linear and quadratic variables of lagged workload for interpretation purposes. It is also sensible to expect that after controlling the time trends, workload level from the previous week is exogenous and would not be correlated with the unobserved factors for procedure duration of MRI scans for the current week (e.g., staffs' serving skill), as the managers' adjustment to staff schedule is likely to diminish the relationship between workload level of the past week and staff skill. Hence, we expect that those instruments satisfy the exclusion restriction condition. Hence, similar to Tan and Netessine (2014) for two endogenous variables (i.e., $Workload$ and $Workload^2$) we use three instrumental variables (i.e., $LWorkload$, $LWorkload^2$ and $UrgRecent$), which makes the effects of workload identifiable.

Based on the above discussion, we include the following specifications to describe the instrumental variables estimation:

$$\begin{aligned} Workload_i =& \pi_0 + \pi_1 UrgentRecent_i + \pi_{21} LWorkload_i + \pi_{22} LWorkload_i^2 + \\ & \boldsymbol{X}_i^{'} \boldsymbol{\pi}_p + \omega_{h(i)}^{'} + e_i^{'}, \end{aligned} \tag{5.2}$$

and,

$$\begin{aligned} Workload_i^2 =& \gamma_0 + \gamma_1 UrgentRecent_i + \gamma_{21} LWorkload_i + \gamma_{22} LWorkload_i^2 + \\ & \boldsymbol{X}_i^{'} \boldsymbol{\gamma}_p + \omega_{h(i)}^{''} + e_i^{''}, \end{aligned} \tag{5.3}$$

where π_1 and γ_1 are the coefficients that capture the impact of $UrgentRecent$ on linear and quadratic terms of workload, respectively. Similarly, π_{21}, π_{22}, γ_{21}, and γ_{22} capture the non-linear impact of lagged workload on the linear and quadratic terms of the endogenous treatment. Similar to specification (5.1), we control for the fixed effects of shift, hospital, and batch of procedure i by including the vector of covariates $\boldsymbol{X}_i^{'}$. Heterogeneity in the intrinsic workload level of hospital-scanner have been incorporated using the fixed effects $\omega_{h(i)}^{'}$ and $\omega_{h(i)}^{''}$, and the error in the estimation of the linear and quadratic terms are captured in $e_i^{'}$ and $e_i^{''}$, respectively.

We next introduce an instrumental variable to address the simultaneity bias of the impact of sequencing in equation (5.1). As we discussed earlier in this section, the schedulers may be prone to fit "shortest expected procedures" between MRI scans to fill the gaps between MRI slots in the procedures calendar. Further, Ibanez et al. (2018) empirically demonstrated that radiologists are likely to "deviate" from the pre-determined sequence of jobs. Therefore, the batching sequence of procedures may be affected by the "expected" procedure duration, which is assessed by the scheduler prior to giving the appointment, and thus, makes Seq_i endogenous. We utilize the instrumental variable $RepType_i$ as an indicator variable, which equals 1 if the body part of procedure i is the same as that of procedure $(i-1)$, and equals zero otherwise. The idea for using $RepType_i$ as an instrumental variable for sequencing is that the scheduler may try to schedule procedures with similar (repeated) body types close to each other in a batch in order to diminish the setup time between procedures. Also, we expect that scheduling procedures avoid reducing the radiologists' cognitive sense by changing their routine, repetitively. Furthermore, following the results of Ibanez et al. (2018), radiologists' are likely to deviate from their calendar to serve jobs with repeated service types. In addition, the correlation between Seq and $RepType$ is approximately (0.52), and thus, it satisfies the relevance condition. Also, after controlling for effects of task difficulty and radiologists' engagement level in the service time equation, we expect that $RepType_i$ would be uncorrelated with the unobserved factors that determine the service time of procedure i, and thus, the impact of repetitive procedure type on the service duration is merely through the endogenous treatment, Seq. Hence, satisfying the exclusion restriction assumption. Now that we have an exogenous variable that can help us to consistently identify the impact of sequencing, we use the following equation to model the sequencing:

$$\begin{aligned} Seq_i =& \theta_0 + \theta_1 log(ServDur_i) + \theta_{21} Workload_i + \theta_{22} Workload_i^2 \\ & + \theta_3 RepType_i + \boldsymbol{X}_i^{'} \boldsymbol{\theta}_p + \omega_{h(i)}^{'''} + e_i^{'''}, \end{aligned} \tag{5.4}$$

where θ_1 captures the effect of "expected" service duration of procedure i. Since the expected service duration (at the time of advance scheduling) is not captured in the dataset, we use the actual service time as a proxy. We also include $Workload$ and $Workload^2$ to control for the effect of radiologists' engagement level. Also, $RepType$ would be the instrumental variable for Seq. Similar to other equations, we control for the fixed effects of shift, hospital, and batch of procedure i by including the vector of covariates $\boldsymbol{X}_i^{'}$. And lastly, we include the hospital-scanner fixed effects through $\omega_{h(i)}^{'''}$ and other unobserved factors is capture through the error term $e_i^{'''}$. We simultaneously estimate the system of equations given in (5.1) - (5.4) while assuming that the error terms of all four equations are correlated and follow a multivariate standard normal distribution, i.e., $(e_i, e_i^{'}, e_i^{''}, e_i^{'''}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is called the contemporaneous covariance matrix.

Our structural model satisfies the rank condition since for each equation, the number of instruments is as large as the number of RHS endogenous variables, i.e., we use $LWorkload$, $LWorkload^2$ and $UrgRecent$ as instruments for $Workload$ and $Workload^2$ and also use $RepType$ as an instrument for Seq, and thus making our model identified (Wooldridge, 2016). We exploit an instrumental variable three-stage least square (3SLS) estimation method to jointly estimate all four equations to account for potential correlation between their error terms.

First, we use OLS to estimate specifications (5.2)-(5.4) and obtain the predicted endogenous independent variables, i.e., $\widehat{Workload}_i$, $\widehat{Workload}_i^2$ and $\widehat{Seq}_i$. Second, we estimate the coefficients of each equation and predict the residuals and estimate the contemporaneous covariance matrix $\mathbf{\Sigma}$. Then, we use the generalized least square (GLS) to estimate and obtain the unbiased and consistent estimation for coefficients (Davidson et al., 2004). 3SLS estimation method is asymptotically more efficient than 2SLS in estimating simultaneous equation models. We also perform the Wu-Hausman endogeneity test with the F statistic of 1617.49, which strongly rejects the null hypothesis ($p < 0.001$) that impacts of workload and sequencing are exogenous on service duration, and thus, supports the ideas of the simultaneous equation model.

A significant non-zero estimation for instrumental variables in the first-stage, i.e., π_1, π_{21}, π_{22}, γ_1, γ_{21}, γ_{22} and θ_3 implies that the instrumental variables are not "weak", and thus satisfy the relevance condition. In Section 6.1, we verify the relevance condition of the instrumental variables by comparing the F-statistics of the first-stage estimations with the suggested rule of thumb for weak instruments (Staiger and Stock, 1997).

Chapter 6

Results and Robustness Checks

6.1 Results

Table 6.1 summarizes the results with $log(ServDur)$ as the dependent variable. We estimate the parameters of specification (5.1) at three sample levels, namely, the pool of all hospitals, teaching hospitals, and community hospitals. The R-squared for 3SLS estimations are computed to be negative, hence are not reported. The IV estimations help us provide consistent estimations of ceteris paribus; hence, goodness-of-fit is not a factor and thus R-squared has no meaningful interpretations (Wooldridge, 2016).

We find that the estimations of the quadratic specification showcase an inverted U-shaped behavior of prior $Workload$ on the upcoming procedure duration. The estimation for linear and quadratic terms are significantly positive and negative, respectively. This finding is consistent in the proposed models across different sample levels and different

Table 6.1: Estimation results of models on $log(ServDur)$ as the dependent variable

	All Hospitals		Teaching Hospitals		Community Hospitals	
Coefficient	OLS	3SLS	OLS	3SLS	OLS	3SLS
$Workload$	0.0613^{***}	1.2696^{***}	0.0468^{***}	1.4027^{***}	0.0731^{***}	1.2572^{***}
	(0.0010)	(0.0311)	(0.0015)	(0.0638)	(0.0014)	(0.0344)
$Workload^2$	-0.0380^{***}	-1.1031^{***}	-0.0288^{***}	-1.2257^{***}	-0.0462^{***}	-1.0714^{***}
	(0.0008)	(0.0277)	(0.0012)	(0.0566)	(0.0012)	(0.0303)
$Shift^{(2)}$	0.2025^{***}	0.1493^{***}	0.1910^{***}	0.1253^{***}	0.2154^{***}	0.1708^{***}
	(0.0011)	(0.0019)	(0.0016)	(0.0037)	(0.0016)	(0.0024)
$Shift^{(3)}$	0.1029^{***}	0.0729^{***}	0.1155^{***}	0.0817^{***}	0.0952^{***}	0.0670^{***}
	(0.0011)	(0.0016)	(0.0015)	(0.0027)	(0.0016)	(0.0022)
Seq	-0.0261^{***}	-0.0431^{***}	-0.0234^{***}	-0.0400^{***}	-0.0261^{***}	-0.0425^{***}
	(0.0002)	(0.0006)	(0.0004)	(0.0009)	(0.0003)	(0.0008)
Controls	Included	Included	Included	Included	Included	Included
Observations	$1,954,088$	$1,954,088$	$944,080$	$944,080$	$1,010,008$	$1,010,008$
Prob $> \chi^2$	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Notes. Standard errors are in parentheses. Control vectors include fix effects for patients' characteristics, hospital-scanner fixed effect, and time effects such as year, week, and weekday.
$^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$.

estimation methods, which supports Hypothesis 1. The quadratic function specification allows us to capture the critical point at which the service time marginally alters, which is equals to $-b/(2a)$ in a quadratic function in the form of $f(x) = ax^2 + bx + c$ (Tan and Netessine, 2014). Although the OLS estimation for the impact of workload is statistically significant, yet it lacks economic significance. Nonetheless, after addressing the endogeneity bias using the 3SLS estimation, the inverted U-shaped behavior of workload becomes more evident. Furthermore, after addressing the endogeneity bias, the estimation for the magnitude of linear and quadratic terms of workload has been increased in the 3SLS estimation, demonstrating the considerable presence of endogeneity. The 3SLS estimation at "all-hospital level analysis" for $Workload$ and $Workload^2$ shows that the critical workload is about $(1.2696/(2 \times 1.1031) \approx 0.575)$ of the standard deviation of workload. The critical point of 0.575 implies that serving patients consecutively for up to nearly half the standard deviation is associated with an increased service time for the next patient, likely due to the physician's mental engagement level. However, approximately after the critical workload level, the service time decreases, likely due to the physician realizing how behind the schedule s/he is and speeding up under the social speedup pressure (supporting Hypothesis H1).

In addition, the critical workload is roughly half the standard deviation above the sample mean of all hospitals (see Table 4.2). In other words, (assuming the normality of workload) approximately 71.7% of procedures are performed below the critical workload level, and thus, the marginal effect for these procedures is positive. In particular, for the procedures with an average value of $Workload$ (which is 0 due to normalizing), the marginal effect would be approximately 1.2696. This estimation implies that on average, one unit of Z-score increase in the accumulative service duration of past procedures in a batch would increase service times by approximately $(e^{1.2696} - 1 \approx 2.559)$ 255.9%. The critical workload can be calculated similarly at teaching and community hospitals, separately. Specifically, using the 3SLS estimation on teaching hospitals sample, we observe that the critical workload is around $(1.4027/(2 \times 1.2257) \approx 0.572)$ of workload Z-score while for community hospitals it is approximately equal to $(1.2572/(2 \times 1.0714) \approx 0.586)$ of workload Z-score. This result shows that (even after adjusting the different workload levels) radiologists at teaching hospitals reach the critical workload level earlier than their peers at community hospitals.

The coefficient of Seq is estimated at -0.0261 as per OLS estimation and its magnitude increased to -0.0431 after overcoming the simultaneity bias using the 3SLS method, which supports Hypothesis 2. This estimation implies that performing procedures of a similar type consecutively (in the same batch) is associated with a decrease in service duration. Specifically, adding an extra procedure of a similar type to a batch of procedures approximately reduces the procedure duration of the upcoming procedure by $(e^{-0.0431} - 1 \approx -0.0421)$ 4.21%. The underlying mechanism could be due to the elimination of physical setup, lack of forgetting or learning by doing.

As Table 6.1 suggests, the estimation for the corresponding coefficient of (Seq) at both teaching and community hospitals is negative, that is, -0.0400 and -0.0425, respectively. These estimations are consistent with Hypothesis 2 which means that adding an extra procedure of a similar type results in a 3.92% and 4.16% decrease in service duration of the next patient at teaching and community hospitals, respectively. The difference in the marginal effect between the teaching and community hospitals is likely because of fewer interruptions at community hospitals (compared to the interruptions caused by trainees at teaching hospitals), reinforcing the lack of forgetting and learning by doing mechanisms at community hospitals. An interesting observation based on our descriptive analysis of the historical data is that the average length of a sequence of procedures with similar body types at community hospitals is approximately 9.3% longer than those at teaching hospitals. This difference indicates that not only the marginal effect of sequencing the procedures of a similar type is higher (more negative) at community hospitals than that at teaching hospitals but also community hospitals adopt the sequencing of procedures more often than their counterpart.

We next explore the impact of night shifts on the service duration. Table 6.1 shows that the OLS estimations of $Shift^{(2)}$ (7A.M.-3 P.M.) and $Shift^{(3)}$ (3P.M.-11 P.M.) are both statistically significant. Also, after correcting the endogeneity and simultaneity biases of workload and sequencing effects using 3SLS estimation, the corresponding impacts of shifts remain positive and significant, which again supports Hypothesis 3A. However, we expect that patients' priorities affect the radiologist's behavior at night shifts. Figure 6.1 illustrates that the average service duration increases at first and third shifts for semi-urgent and non-urgent patients rather than for all patients, which indicates that H3A is supported for low priority patients. On the other hand, Figure 6.1 shows that for urgent/emergency patients, the service duration increases at first and third shifts, which is aligned with Hypothesis 3B for high priority patients.

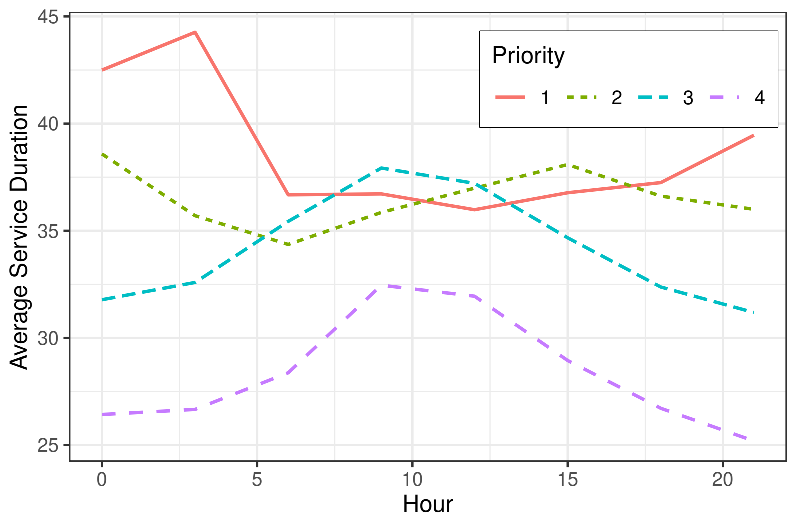


Figure 6.1: Average service duration based on procedure starting hour and patient priority

To investigate whether the impact of night shifts are different with respect to patients' priorities, we again use the 3SLS estimation method to simultaneously estimate equations (5.1)-(5.4) at a more granular level by incorporating interaction terms between shifts and patients' priorities. These interacting coefficients allow us to test for marginal changes in service duration of procedures at different shifts across different priority groups (Wooldridge, 2016). Note that, we incorporate the interaction terms to determine the causal relationship between nightshift and procedure duration as per each priority separately, and not necessarily for increasing the models' goodness-of-fit.

We re-estimate our SEM using the 3SLS method while adding the interaction terms between $\boldsymbol{Shift}_i$ and $\boldsymbol{Priority}_i$ to capture the impact of shifts on service duration across patients with different priorities. We revise equation (5.1) as follows:

$$\begin{aligned} log(ServDur_i) =& \beta_0 + \beta_{11} Workload_i + \beta_{12} Workload_i^2 + \boldsymbol{Shift_i \beta_s} \\ & + \boldsymbol{Shift}_i \times \boldsymbol{Priority}_i \boldsymbol{\beta}_I + \beta_2 Seq_i + \boldsymbol{X}_i \boldsymbol{\beta}_p + \delta_{h(i)} + \epsilon_i, \end{aligned} \tag{6.1}$$

Table 6.2: 3SLS estimation for interaction terms of the effect of Shifts on $log(ServDur)$ across different priorities

Coefficient	All Hospitals	Teaching Hospitals	Community Hospitals
$Shift^{(2)}$	-0.0390^{***}	-0.0841^{***}	0.0309
	(0.0094)	(0.0111)	(0.0340)
$Shift^{(3)}$	-0.0305^{***}	-0.0361^{***}	0.0022
	(0.0095)	(0.0011)	(0.0344)
$Shift^{(2)} \times Priority^{(2)}$	0.1069^{***}	0.1285^{***}	0.0622
	(0.0108)	(0.0131)	(0.0353)
$Shift^{(2)} \times Priority^{(3)}$	0.2083^{***}	0.2336^{***}	0.1954^{***}
	(0.0098)	(0.0111)	(0.0347)
$Shift^{(2)} \times Priority^{(4)}$	0.1940^{***}	0.2165^{***}	0.1418^{***}
	(0.0093)	(0.0111)	(0.0340)
$Shift^{(3)} \times Priority^{(2)}$	0.0866^{***}	0.1008^{***}	0.0481
	(0.0111)	(0.0130)	(0.0359)
$Shift^{(3)} \times Priority^{(3)}$	0.1304^{***}	0.1460^{***}	0.1258^{***}
	(0.0100)	(0.0118)	(0.0352)
$Shift^{(3)} \times Priority^{(4)}$	0.1015^{***}	0.1139^{***}	0.0634^{*}
	(0.0096)	(0.0112)	(0.0343)
Controls	Included	Included	Included
Observations	$1,954,088$	$944,080$	$1,010,008$
Prob $> \chi^2$	< 0.0001	< 0.0001	< 0.0001

Notes. Standard errors are in parentheses. Control vectors include fix effects for patients' characteristics, hospital-scanner fixed effect, and time effects such as year, week, and weekday. $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$.

where $\boldsymbol{\beta}_I$ includes the interacting terms between patients' $Priority$ and the $Shift$ at which the procedure took place. Table 6.2 summarizes the estimation of interaction terms using three samples of the data, namely, aggregation of all hospitals, teaching hospitals, and community hospitals. Note that, since both shifts and priorities are categorical vectors, we consider $Shift^{(1)}$ (i.e., night shifts) and $Priority^{(1)}$ (i.e., emergent patients) as the reference groups for procedure shifts and patients' priorities, respectively.

We find that in comparison with the first shifts (i.e., night shifts), performing procedures in the second shifts results in approximately ($e^{-0.0390} - 1 \approx -0.0382$) 3.82% reduction in the procedure duration of patients listed as priority 1. This result shows that emergent patients undergo longer procedure duration in the night shifts which support Hypothesis 3B. On the other hand, Table 6.2 shows that in comparison with the first shifts (i.e., night shifts), performing procedures of urgent patients (i.e., priority 2) in the second shifts leads to ($e^{-0.0390+0.1069} - 1 \approx 0.0692$) 6.92% increase in the service duration. Similarly, the procedure duration of semi-urgent patients (i.e., priority 3) and non-urgent patients (i.e., priority 4) would also increase by ($e^{-0.0390+0.2083} - 1 \approx 0.1841$) 18.41% and ($e^{-0.0390+0.1940} - 1 \approx 0.1676$) 16.76%, respectively. Therefore, unlike emergent patients, for urgent, semi-urgent and non-urgent procedures, the service duration is in fact shorter at night shifts, which support Hypothesis 3A.

These results are consistent with our initial insights based on Figure 6.1, and demonstrate that for emergent procedures, which require more diligence than their counterpart cohort, night shifts may "deteriorate cognitive efficiency" which increase the service duration of urgent patients. Conversely, performing low priority procedures at night shifts may stimulate physicians to perform non-discretionary tasks quickly and finish their working shift, which is also illustrated in Figure 6.1. We observe the same pattern for procedures that took place at the third shift. That is, in comparison with first shifts, the service duration of emergent procedures is 3.00% shorter at third shifts, which supports Hypothesis 3B, but other priorities (2, 3 and 4) experience longer procedure duration at third shifts. In particular, urgent procedures took 5.77% longer during the third shifts, semi-urgent 10.50% longer, and non-urgent took 7.30% longer during the third shifts.

We now restrict our analyses to estimate the interaction effect between shifts and patients' priorities at the teaching and community hospitals. Columns 2 and 3 of Table 6.2 show the 3SLS estimation of interaction terms at teaching and community hospitals, respectively. Considering night shifts as the reference group, we find that performing emergent (i.e., priority 1) procedures in the second shift at teaching hospitals is associated with ($e^{-0.0841} - 1 \approx -0.0806$) 8.06% decrease in the duration of the procedure. This result is consistent with that for all hospitals from column 1 and shows that service duration of emergent procedures in night shifts is longer than that during day shifts, and thus, supports Hypothesis 3B. Nonetheless, for other priority levels (i.e., 2, 3 and 4), service duration in the second shifts is slower than that at first shifts. In particular, the service duration of patients with priority 2 increases by ($e^{-0.0841+0.1285} - 1 \approx 0.0454$) 4.54%, priority 3 by ($e^{-0.0841+0.2336} - 1 \approx 0.1612$) 16.12% and priority 4 by ($e^{-0.0841+0.2165} - 1 \approx 0.1445$) 14.15%. These results provide evidence to support Hypothesis 3A which indicates that procedures in night shifts undergo a faster service duration. Additionally, we find the very same pattern for procedure duration at the third shift in comparison with the base shift (night shifts) at the teaching hospitals. The procedure duration performed at the third shifts is associated with ($e^{-0.0361} - 1 \approx -0.0354$) 3.54% decrease for patients of priority 1, ($e^{-0.0361+0.1008} - 1 \approx 0.0668$) 6.68% increase for patients of priority 2, ($e^{-0.0361+0.1460} - 1 \approx 0.1161$) 11.61% increase for patients of priority 3, and ($e^{-0.0361+0.1139} - 1 \approx 0.0809$) 8.09% increase for patients of priority 4.

The third column of Table 6.2 shows the 3SLS estimation for interaction between shifts and patients' priorities at the community hospitals. We now find that there is no statistically significant difference between procedure duration of emergent or urgent patients during any of the three shifts. However, for low priority patients (i.e., semi-urgent and non-urgent), radiologists at community hospitals perform the MRI scans at first and third shifts much slower than those at the night shifts, which supports Hypothesis 3A. Specifically, we find that for semi-urgent procedures, in comparison with night shifts, the procedure duration took roughly ($e^{0+0.1954} - 1 \approx 0.2157$) 21.57% longer during the second shifts and ($e^{0+0.1258} - 1 \approx 0.1340$) 13.40% during the third shifts. Similarly, for non-urgent patients, we find that the procedures took ($e^{0+0.1418} - 1 \approx 0.1523$) 15.23% longer during the second

Table 6.3: Estimations of instrumental variables on endogenous treatments using different hospital sample level

	All Hospitals			Teaching Hospitals			Community Hospitals		
Coefficient	$Workload$	$Workload^2$	Seq	$Workload$	$Workload^2$	Seq	$Workload$	$Workload^2$	Seq
$UrgentRecent$	0.655***	0.738***		0.642***	0.725***		0.681***	0.776***	
	(0.001)	(0.001)		(0.001)	(0.001)		(0.002)	(0.002)	
$LWorkload$	0.053***	0.035***		0.008**	−0.008**		0.074***	0.057***	
	(0.002)	(0.002)		(0.002)	(0.003)		(0.002)	(0.003)	
$LWorkload^2$	−0.025***	−0.024***		0.019***	0.023***		−0.042***	−0.047***	
	(0.002)	(0.002)		(0.002)	(0.003)		(0.002)	(0.003)	
$RepType$			1.048***			0.979***			1.100***
			(0.001)			(0.001)			(0.003)
Controls	Included	Included	Included	Included	Included	Included	Included	Included	Included
Observations	1,954,088	1,954,088	1,954,088	944,080	944,080	944,080	1,010,008	1,010,008	1,010,008
R^2	0.480	0.308	0.337	0.484	0.287	0.419	0.479	0.325	0.254
Prob > χ^2	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Notes. Standard errors are in parentheses. Control vectors include fix effects for patients' characteristics, hospital-scanner fixed effect, and time effects such as year, week, and weekday.
$^*p \le 0.05$, $^{**}p \le 0.01$, $^{***}p \le 0.001$.

shifts and ($e^{0+0.0634} - 1 \approx 0.0654$) 6.54% during the third shifts.

The discrepancy between the interacting effect of shifts and patients' priorities at teaching and community hospitals is interesting. While the urgent, semi-urgent and non-urgent patients at teaching hospitals experience shorter service duration in night shifts (supporting H3A), only semi- and non-urgent patients at community hospitals experience shorter service duration in night shifts. Furthermore, while emergent procedures at teaching hospitals require more time to perform at night shifts, there is no significant evidence that radiologists at community hospitals perform high priority procedures (e.g., emergent or urgent) slower at night shifts.

Since the 3SLS estimation results depend on the validity of instrumental variables, we examine the relevance condition of the instrumental variables. Table 6.3 shows the first-stage regression of instrumental variables (i.e., $UrgentRecent$, $LWorkload$, $LWorkload^2$ and $RepType$) on linear and quadratic terms of $Workload$ and also on Seq, respectively. The corresponding coefficient of $UrgentRecent$, $LWorkload$ and $LWorkload^2$ are all statistically significant in the first-stage regressions for $Workload$ and $Workload^2$ at different hospital-level data samples, including all-hospitals, teaching hospitals, and community hospitals. The coefficient of *Reptype* is also significant across all sample levels. Also, the F statistics for all of the endogenous treatments equations across all three sample levels are greater than 10, which is suggested as a rule of thumb criteria for testing the weak instruments (Staiger and Stock, 1997). However, since we have more than one endogenous regressor, we also apply the Cragg-Donald F statistic (see e.g., Cragg and Donald 1993) which again rejects the null hypothesis indicating that the instruments are weak (p-value < 0.001) at all sample levels. Thus, the relevance condition is met.

6.2 Robustness Checks

6.2.1 Alternative Variable Setting

In this section, we examine the robustness of the service duration model. First, we test the robustness of the model against alternative definitions for the shifts, by considering two distinct shifts, namely, day shift (from 7 A.M. to 3 P.M.) and night shift (from 3 P.M. to 7 A.M. the following day). We repeat the 3SLS estimation of the simultaneous equation model given in (5.2)-(6.1) which includes the interaction terms between shift and priorities. Similar to the previous section, we consider night shifts as the reference group. The results show significant negative coefficients in day shifts for patients with priority 1 ($-0.0379, p < 0.001$) and significant positive coefficients in day shifts for patients with and priority 2 ($0.0046, p < 0.001$), priority 3 ($0.0842, p < 0.001$) and priority 4 ($0.0899, p < 0.001$), which is consistent with the findings provided in the previous section. Using this setting for shifts, we again find a significant negative coefficient for *Seq* which is ($-0.0401, p < 0.001$) and also significant inverted U-shaped behavior of workload with statistically significant coefficients of (1.815, -1.589) for linear and quadratic terms, respectively.

Recall that we considered $\tau = 15$ minutes as the cut-off threshold for the idle time between batches of procedures in the results provided in the previous section. When we set $\tau = 5, 10$ and 20 minutes as the cut-off threshold, we obtain consistent results for the impacts of sequencing, workload and night shifts for high and low priority patients as before. We also change the definition of *Seq* from the number of consecutive procedures of similar body types to the number of non-consecutive of the same type but in the same batch. We find that the estimate for *Seq* is quite robust to this modification with an estimation of (-0.0603, p-value < 0.001). This estimation implies that adding one procedure of a similar type in the same batch (but not necessarily back-to-back) will decrease the service duration of the next procedure by ($e^{-0.0603} - 1 \approx -0.0585$) 5.85%.

We also narrow down the data into shorter time frames, that is from January 2015 to December 2016, and observe that the results are quite consistent with the ones provided in the previous section. We also exclude records from week 1 and week 52 throughout the data (i.e., first and last week of the year) since both scheduler and the physicians' behavior could be affected by the Christmas holidays. We find similar results to the previous settings. Moreover, we conduct our analyses by separating the patients served on weekdays (i.e., Mon. to Fri.) versus weekends (i.e., Sat. and Sun.). The descriptive analysis shows that the average service duration on weekdays is 31.09 (s.d. = 14.45) minutes whereas over the weekends it reduces to 27.58 (s.d. = 11.81) minutes, (see Figure 4.5). Performing the 3SLS estimation on the weekdays' sample, we again obtain an inverted U-shaped behavior for prior workload with a significant coefficient of (1.591, −1.394) and the critical Z-score of workload at 0.570, implying that approximately 71.56% of procedures of patients, served on weekdays, are performed below the critical workload level with a positive marginal effect. In addition, the impact of prior workload on the procedure of patients served over the weekends (while adjusting for patients priority) follows an inverted U-shaped behavior with a significant coefficient of (0.573, −0.500) and the critical Z-score of workload at 0.573. Therefore, by assuming the normality of workload over weekends (mean = 53.33, s.d. =

88.42), roughly 71.67% of those procedures are performed below the critical workload level; therefore, supporting the Hypothesis 1. We also find that our main results are robust in sign and statistical significance for procedures performed on weekdays and weekends. The impact of Seq is strongly significant and negative on both weekdays $(-0.046, p < 0.001)$ and weekends $(-0.027, p < 0.001)$. This estimation suggests that performing consecutive similar body type procedures reduces the service duration of the next procedure during weekdays by 4.49% and during weekends by 2.66%.

We also check the robustness of the main results by modifying the definition of instrumental variables. We vary the definition of $UrgentRecent$ to the total number of *consecutive* emergent or urgent procedures and also the total number of *non-consecutive* emergent procedures and we again observe robust results in sign, magnitude, and statistical significance. We also performed the Cragg-Donald F statistic for all instrumental variables and found no "weak instruments".

6.2.2 Heteroskedasticity and Serial Correlation

The main results that we obtained from Section 6.1 substantially depend on the fairly strong assumption that the error terms of the structural equations (5.2)-(6.1) are homoskedastic and serially uncorrelated. We used the 3SLS estimation method, which allows structural disturbances to be correlated across each equation. Nonetheless, the variance-covariance matrix within each equation is still assumed to be homoskedastic and serially uncorrelated, an assumption that may not be satisfied in our setting. Note that due to the statistically significant dependency between the conditional variance and the explanatory variables captured by the Breusch-Pagan test for heteroskedasticity (Breusch and Pagan, 1979), the assumption of homoskedasticity is violated. Furthermore, specification (5.1) is similar to a lagged-dependent variable (LDV) model, since we defined $Workload_i$ as the summation of service duration of procedures within the same batch but prior to procedure i. Therefore, the workload (independent variable) for the second procedure in a batch is equal to the service duration (dependent variable) of the previous procedure. The same story applies to the rest of the procedures in the batch. Therefore, it is sensible to anticipate that the error terms of procedures within the same batch are serially correlated. In order to address those two concerns, we propose to use heteroskedastic robust standard errors which are clustered by each batch. Clustering standard errors allow all procedures within a batch to be correlated with each other. In order to simultaneously estimate equations (5.2)-(6.1) while accounting for heteroskedasticity and autocorrelation between error terms of each equation, we employ the Generalized Method of Moments (GMM) estimator. The GMM estimator is strongly consistent and asymptotically normal for large samples (Davidson et al., 2004). We first use the generalized IV to obtain the "consistent but inefficient estimates", then we use the efficient feasible GMM estimator which will be used to calculate the residuals and estimate $\hat{\mathbf{\Sigma}}$ to address both heteroskedasticity and serial correlation (see Davidson et al. 2004, pp. 355-358). We use Hansen et al. (1996) approach to estimate the structural model in two steps. First, we use the identity weight matrix to simplify the criterion function. Then, in the second step, we use the optimal clustered robust (clustered by batch) weight matrix to obtain the consistent disturbance matrix.

The main challenge in applying the above procedure to our main results (using the sample from all hospitals) is the computational complexity. Nonetheless, using the GMM estimator on each hospital separately is computationally feasible and thus, we can demonstrate that the results from Section 6.1 are robust to heteroskedasticity and serial correlation. Also, since the impact of hospitals is captured as a fixed effect in the proposed model, the results presented above may be predominated by larger hospitals. In addition, the impacts of workload, sequencing and shifts can be different at each hospital. We explore the hypotheses at the individual hospital level by estimating the simultaneous equation model, given in equations (5.2)-(6.1) using the GMM estimator. We find that at 10% significance level, out of 66 hospitals, the quadratic coefficient for the impact of workload is significantly negative at 51 hospitals, positive at only 1 hospital, and the other 14 hospitals have an insignificant coefficient for $Workload^2$. This estimation demonstrates the inverted U-shaped relationship between workload and procedure duration at most hospitals. Furthermore, using the GMM estimator at the individual hospital level, we find negative coefficients for the impact of Seq at 48 hospitals (out of 66) at 10% significance level, 17 hospitals have an insignificant estimation for Seq, and only 1 hospital has significant positive estimation for Seq. This result supports Hypothesis 2 for most of the hospitals and indicates that after correcting the simultaneity bias, a longer sequence of procedures with a similar service type is associated with a shorter service duration of the next MRI scan. We also find consistent results for the impact of night shifts at individual hospital levels. We find that out of 55 hospitals, which operate at night shifts, in 49 hospitals the high priority patients undergo longer service duration while low priority patients experience shorter service duration at night shifts. This observation supports Hypothesis 3B and Hypothesis 3A. The impact of night shifts is insignificant for both high priority and low priority patients in the other 6 hospitals.

6.2.3 Inverted U-Shaped Robustness

In this section, we provide the statistical hypothesis testing regarding the inverted U-shaped behavior for ($Workload$), which may erroneously provide a positive extreme point while considering quadratic specification (Lind and Mehlum, 2010). Specifically, we provide robustness checks to examine whether the slopes of the service time model are significantly positive and then negative at lower and higher points of ($Workload$), respectively. To test the following hypothesis, we consider the interval $[Workload_l, Workload_h]$ as an interval for the observations of standardized ($Workload$), which is $[-0.727, 5.040]$ and construct the following two standard one-sided t-test:

$$H_0^l : \beta_{11} + 2\beta_{12}Workload_l \leq 0 \quad \text{versus}$$

$$H_1^l : \beta_{11} + 2\beta_{12}Workload_l > 0,$$

$$H_0^h : \beta_{11} + 2\beta_{12}Workload_h \geq 0 \quad \text{versus}$$

$$H_1^h : \beta_{11} + 2\beta_{12}Workload_h < 0.$$

Note that the rejection area for the above test is a convex cone:

$$\begin{aligned}R_\alpha = \{(\beta_{11}, \beta_{12}) : &(\beta_{11} + 2\beta_{12}Workload_l) \\ &\cdot \left(\sqrt{s_{11} + 4Workload_l s_{12} + 4Workload_l^2 s_{22}}\right)^{-1} > t_\alpha \quad \text{and} \\ &(\beta_{11} + 2\beta_{12}Workload_h) \\ &\cdot \left(\sqrt{s_{11} + 4Workload_h s_{12} + 4Workload_h^2 s_{22}}\right)^{-1} < -t_\alpha\},\end{aligned}$$

where s_{11}, s_{22}, and s_{12} are the estimated variances of β_{11}, β_{12} and the covariance between them from the variance-covariance matrix, respectively. Lastly, t_α is the t-value at the α significance level.

Table 6.4 provides hypothesis testing for alternative behavior of ($Workload$) for the service duration model. This table shows that the slope for ($Workload$) is statistically positive (2.873) at the lower bound and negative (-9.845) at the upper bound. Further-

Table 6.4: Alternative inverted U-shaped hypothesis testing for standardized $Workload$

	Lower Bound	Upper Bound
Interval	-0.727	5.040
Slope	2.873	-9.845
t-value	40.218	-39.547
p-value	0.000	0.000

more, the p-values for both standard tests are approximately zero. Thereby, we reject the null hypothesis that the effect of ($Workload$) on the service duration is U-shaped or monotone.

Chapter 7

Insights and Conclusions

Using more than three years of data (2014-2017), encompassing approximately 2 million records of service (MRI scan procedure) time duration from 66 hospitals in the province of Ontario (Canada), we conducted econometric analyses that look into drivers of service duration.

We showed that a number of behavioral covariates, including the shift during which the procedure was performed, case-mix workload of the proceeding patients and batching of similar procedures affect service duration. We showed that compared to the day shifts, while the service duration is longer during the night shifts for emergent procedures (around 4%), low priority procedures experience shorter MRI scan time. Possibly because providers may be tempted to perform non-discretionary tasks quickly (i.e. the task reduction mechanism) during the night shift for low priority procedures. On the other hand, it is sensible that since performing emergent procedures requires more dedication and thoroughness (compared to their counterparts), these procedures during night shifts take longer to perform. We also showed that the proceeding case-mix workload affects subsequent service duration. In particular, the case-mix workload of the proceeding procedure lengthens the service duration up to around 0.57 standard deviation from the workload mean (likely due to the engagement level mechanism) and shortens the duration beyond that point (likely due to the social speedup mechanism). Finally, we observed that sequencing similar service types together (i.e. batching) is associated with a shorter service time, likely due to any of the following mechanisms: reducing physical setup, lack of forgetting, and learning by doing. In particular, performing an additional procedure of a similar body type within a batch of the procedure is likely to reduce the next procedure duration by around 4.2%.

The main challenge in estimating those mechanisms was endogeneity and simultaneity biases of workload and sequencing in our model, respectively. As we were unable to capture the staff skill in our dataset, the omitted variable bias in our specification would have lead OLS estimates to be inconsistent. For instance, the decrease in handling skills of radiologists in a batch contributes to both increasing workload before a procedure and also increasing the MRI scan. Furthermore, as (Ibanez et al., 2018) show, the schedulers are likely to schedule jobs with the "shortest expected procedure time" between appointment slots and thus increasing their sequence order. Hence, simultaneity bias also could have hurt our conclusions. To tackle these issues, we proposed a simultaneous equations model

(SEM) to estimate the true impacts of workload and sequencing effects. We used lagged value of workload from the previous week and the number of urgent/emergent cases prior to a procedure as instrumental variables for workload and also used an indicator variable for a procedure with repeated scan body type as an instrumental variable for the impact of sequencing. We employed the 3SLS estimator to simultaneously estimate the equations. Another challenge in estimating our econometric model was heteroskedasticity and serial correlation in the model. As we anticipated, the error terms of service time specification of procedures that were performed closed to each other are correlated. To overcome this challenge we turned our attention to the Generalized Method of Moments (GMM) estimator that allowed us to simultaneously estimate equations while considering heteroskedastic robust standard errors which were clustered by each batch. However, due to the computational limitations of the GMM estimator on a large dataset, we were not able to use the GMM on the sample of all hospitals as a single model. Instead, we performed the estimation using the sub-sample analysis of each hospital separately. We obtained the same results at most of the hospitals.

Given the large-scale data that we have access to, we were able to investigate the model for teaching and community hospitals separately, as well as at the level of individual hospitals. In fact, one contribution of this thesis is showing how the magnitude of the effect of various mechanisms on service duration may depend on the server type (i.e. teaching or community hospital), as well as on the type of the customer (e.g. high priority versus low priority patient). We observed that the reduction of the service duration during the night shift is more intense at community hospitals, likely because providers at community hospitals are fully trained physicians and as such more dexterous at providing services. The magnitude of the effect of a sequence of similar body type procedures on the subsequent service duration is also different in teaching and community hospitals. The inverted U-shaped relationship between workload and procedure duration for scans is evident at both teaching and community hospitals. However, with average workload level, the curve is less steep at community hospitals, likely because the providers are fully trained physicians.

We also observed a higher magnitude of reduction in service duration as a result of batching services together in community hospitals, likely due to fewer interruptions occurring at community hospitals (compared to the interruptions caused by trainees at teaching hospitals), which reinforces the lack of forgetting and learning by doing mechanisms at community hospitals. We also investigated the effect of behavioral factors on service duration at the level of individual hospitals and found that all hypotheses are supported by the majority of hospitals and that the few exceptions are largely due to the existence of outlier cases.

Our findings have both theoretical and practical implications. On the theory side, the context of MRI scheduling and delivery in Ontario is interesting as it involves multi-type multi-priority customers with priority-specific service level constraints (i.e. wait time targets). The few recently developed theoretical multi-class queueing models in the literature with service level constraints typically assume a multi-class setting with a specific service level for each class (e.g. Sun and Whitt 2018).

However, our empirical analysis shows that both priority levels and service types affect service duration. Furthermore, a mix of various service types comprise each priority level and should collectively meet the priority-specific service levels. As such, results of multi-class queues cannot be easily extended to the multi-type multi-priority queueing systems with wait targets and new theoretical queueing models are needed to model such systems, and also take into account the various behavioral mechanisms revealed in this study.

On the practical side, an interesting observation is the opposing effects of batching (sequencing of procedures with similar service types) and the workload on the service duration. Batching similar procedures decreases service duration but also increases the workload which would increase the service time through the engagement level mechanism. This finding has implications for system administrators and policymakers, as they have to balance the trade-off between maximizing efficiency and controlling the workload by finding the appropriate level of batching. A high level of batching of similar services would indeed increase the system efficiency, as the average service duration becomes shorter. However, this increase in efficiency comes at the price of diminishing radiologists' performance as the engagement level effect kicks in. Finding the appropriate level of batching to balance this trade-off between efficiency and workload level is a policy question that would likely be best answered by future research in the field of social sciences. Furthermore, one of the most important implications of this research is to use the predictive model for MRI scan procedure duration to reduce the uncertainties and delays in the scheduling system for hospital administrators and aid them in clearing the long backlogs of elective surgeries as the COVID-19 pandemic significantly reduced the hospitals' service capacities and thus increased the waiting times. In particular, around 95% of batches observed in our data set have a size of at most 7 procedures with an average sequence length of similar procedures of 1.57. Therefore, considering the ample backlogs of procedures created during the COVID-19 pandemic, we can schedule procedures with similar types in the same batches so that the average service duration is reduced. To highlight the impact of proper sequencing in patient scheduling, we built a trace-driven simulation using the MRI data. We observed that, for a given hospital, the average capacity (i.e., the number of scheduled procedures per day) could increase from 31.35 (sd = 14.73), under the current setting to 45.23 (sd = 2.58) while fixing the workload, the average service duration, and all other covariates. These estimations demonstrate that by fixing the workload level and only considering the sequencing policy, the average daily capacity can be increased by roughly 44.27%.

Another interesting implication of this work on the practical side is using the service duration prediction model discussed in Section 6.1 to determine the appointment times. One of the main challenges in scheduling the procedures throughout the day is the uncertainty in the starting and finishing time of procedures. The accumulation of delays produced by the difference between scheduled and actual starting times of procedures may under-utilize the hospital resources. However, with the OLS estimated model that we introduced in Section 6.1, we can use the clinical and operational covariates to predict the service duration of procedures and schedule the starting time procedures, accordingly. We built another trace-driven simulation model which incorporates the prediction model that we estimated in Section 6.1. We randomly select 18 hospitals and simulate the MRI scheduling system using the *same order* of scheduled procedures (using current practice) and use predicted procedure duration to determine the starting time of procedures. The rolling simulation at all hospitals starts from 2014 until the end of 2016 and find that the average daily capacity can be improved by up to 11.16%.

Another area for future research might be looking at the quality of service vis-à-vis service duration. To the extent that reductions in service duration result from efficiency maximizing mechanisms (such as reducing setups, learning by doing, and lack of forgetting), likely, the quality of service is not negatively affected with a shorter duration. However, other mechanisms such as task reduction or social speedup pressure may harm the quality of service. Likewise, the mechanisms responsible for lengthening service duration may also affect service quality. Future research would likely need to explore the effects of various mechanisms not only on service duration but also on service quality.

Theoretical scheduling models, with a few exceptions such as Kopach et al. (2007), typically either consider same-day scheduling or advance scheduling. In many settings, such as ours, same-day scheduling is impractical as patients need advance notice to make it to the appointment. Developing theoretical scheduling models that find the optimum time window for scheduling patients in advance of the appointment is another area of proposed future research. Such theoretical scheduling models should also closely consider endogenous service time distributions as driven by factors uncovered in our study, such as shift, case-mix workload of the proceeding patients and batching.

It is important to consider the potential limitations of this study and interpret the results accordingly. Like most empirical research, we acknowledge the threat of omitted variable bias. It might have been helpful to add more variables such as the patient's place of residence, but this data was not available to us. For instance, it is likely that patients who are willing to travel a distance to undergo the procedure experience shorter wait times. Another limitation is the transferability of our findings to healthcare systems unlike that of Ontario with its universal insurance that allows patients to receive care in almost any location. Our findings may not be transferrable to other health systems, such as the USA, where patients may have limitations or preferences to be served in some locations over others due to insurance coverage or other considerations.

Improving queueing system performance requires an understanding of the behavior of customers and servers, who co-produce the services. Our study sheds light on a number of behavioral factors that affect service duration, wait times, and customer's tendency to defer services in the context of a multi-priority multi-type queueing system. Although results may differ from one setting to another, the mechanisms identified in this study may be likely candidates to investigate in similar settings, and could reveal behaviors that should be considered in future modeling endeavors, as well as addressed by system administrators and policymakers who work at improving the systems in practice.

References

Abdalla, G., Moran-Atkin, E., Chen, G., Schweitzer, M. A., Magnuson, T. H., and Steele, K. E. (2015). The effect of warm-up on surgical performance: a systematic review. *Surgical endoscopy*, 29(6):1259–1269.

Abouee-Mehrizi, H. and Barjesteh, N. (2019). Multi-class state-dependent service systems with returns. *Available at SSRN 3464764*.

Abouee-Mehrizi, H. and Baron, O. (2016). State-dependent m/g/1 queueing systems. *Queueing Systems*, 82(1-2):121–148.

Ahmadi-Javid, A., Jalali, Z., and Klassen, K. J. (2017). Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research*, 258(1):3–34.

Amirian, I. (2014). The impact of sleep deprivation on surgeons' performance during night shifts. *Dan Med J*, 61(9):B4912.

Anand, K. S., Paç, M. F., and Veeraraghavan, S. (2011). Quality–speed conundrum: Trade-offs in customer-intensive services. *Management Science*, 57(1):40–56.

Arlotto, A., Chick, S. E., and Gans, N. (2014). Optimal hiring and retention policies for heterogeneous workers who learn. *Management Science*, 60(1):110–129.

Ayvaz, N. and Huh, W. T. (2010). Allocation of hospital capacity to multiple types of patients. *Journal of Revenue and Pricing Management*, 9(5):386–398.

Barker, L. M. and Nussbaum, M. A. (2011). Fatigue, performance and the work environment: a survey of registered nurses. *Journal of advanced nursing*, 67(6):1370–1382.

Batt, R. J. and Terwiesch, C. (2015). Waiting patiently: An empirical study of queue abandonment in an emergency department. *Management Science*, 61(1):39–59.

Batt, R. J. and Terwiesch, C. (2017). Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Science*, 63(11):3531–3551.

Begen, M. A. and Queyranne, M. (2011). Appointment scheduling with discrete random durations. *Mathematics of Operations Research*, 36(2):240–257.

Bekker, R., Borst, S. C., Boxma, O. J., and Kella, O. (2004). Queues with workload-dependent arrival and service rates. *Queueing Systems*, 46(3-4):537–556.

Bendoly, E., Swink, M., and Simpson III, W. P. (2014). Prioritizing and monitoring concurrent project work: Effects on switching behavior. *Production and Operations Management*, 23(5):847–860.

Bentzen, M., Lemyre, P.-N., and Kenttä, G. (2016). Development of exhaustion for high-performance coaches in association with workload and motivation: A person-centered approach. *Psychology of Sport and Exercise*, 22:10–19.

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, pages 1287–1294.

Cakir, A., Hart, D. J., and Stewart, T. F. (1982). Visual display terminals. a manual covering ergonomics, workplace design, health and safety, task organization.

Cayirli, T. and Gunes, E. D. (2014). Outpatient appointment scheduling in presence of seasonal walk-ins. *Journal of the Operational Research Society*, 65(4):512–531.

Chakraborty, S., Muthuraman, K., and Lawley, M. (2010). Sequential clinical scheduling with patient no-shows and general service time distributions. *IIE Transactions*, 42(5):354–366.

Chakravarty, A. and Debnath, J. (2015). Life cycle costing as a decision making tool for technology acquisition in radio-diagnosis. *Medical Journal Armed Forces India*, 71(1):38–42.

Cragg, J. G. and Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, pages 222–240.

Dantas, L. F., Fleck, J. L., Oliveira, F. L. C., and Hamacher, S. (2018). No-shows in appointment scheduling–a systematic literature review. *Health Policy*, 122(4):412–421.

Davidson, R., MacKinnon, J. G., et al. (2004). *Econometric theory and methods*, volume 5. Oxford University Press New York.

Deber, R. B. (2018). *Treating Health Care: How the Canadian System Works and how it Could Work Better*. University of Toronto Press.

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

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

Delasay, M., Ingolfsson, A., Schultz, K., et al. (2016b). Inventory is people: how load affects service times in emergency response. *Cross-Functional Inventory Research*, 10:21.

Denton, B., Viapiano, J., and Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health care management science*, 10(1):13–24.

des Bordes, E. and Büyüktahtakın, İ. E. (2017). Optimizing capital investments under technological change and deterioration: A case study on mri machine replacement. *The Engineering Economist*, 62(2):105–131.

Di Muzio, M., Diella, G., Di Simone, E., Novelli, L., Alfonsi, V., Scarpelli, S., Annarumma, L., Salfi, F., Pazzaglia, M., Giannini, A. M., et al. (2020). Nurses and night shifts: Poor sleep quality exacerbates psychomotor performance. *Frontiers in neuroscience*, 14:1050.

Diamant, A., Milner, J., and Quereshy, F. (2018). Dynamic patient scheduling for multi-appointment health care programs. *Production and Operations Management*, 27(1):58–79.

Erdogan, S. A., Gose, A., and Denton, B. T. (2015). Online appointment sequencing and scheduling. *IIE Transactions*, 47(11):1267–1286.

Erlang, A. K. (1909). The theory of probabilities and telephone conversations. *Nyt. Tidsskr. Mat. Ser. B*, 20:33–39.

Fan, J. and Smith, A. P. (2017). The impact of workload and fatigue on performance.

In *International symposium on human mental workload: Models and applications*, pages 90–105. Springer.

Feldstein, M. (1977). Quality change and the demand for hospital care. *Econometrica: journal of the Econometric Society*, pages 1681–1702.

Fox, J. and Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87(417):178–183.

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

Froehle, C. M. and White, D. L. (2014). Interruption and forgetting in knowledge-intensive service environments. *Production and Operations Management*, 23(4):704–722.

Garay, J., Cartagena, R., Esensoy, A. V., Handa, K., Kane, E., Kaw, N., and Sadat, S. (2015). Strategic analytics: towards fully embedding evidence in healthcare decision-making. *Healthc Q*, 17:23–7.

Gerchak, Y., Gupta, D., and Henig, M. (1996). Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science*, 42(3):321–334.

Glaser, D. N., Tatum, B. C., Nebeker, D. M., Sorenson, R. C., and Aiello, J. R. (1999).

Workload and social support: Effects on performance and stress. *Human Performance*, 12(2):155–176.

Gocgun, Y. and Puterman, M. L. (2014). Dynamic scheduling with due dates and time windows: an application to chemotherapy patient appointment booking. *Health care management science*, 17(1):60–76.

Green, L. V. and Savin, S. (2008). Reducing delays for medical appointments: A queueing approach. *Operations Research*, 56(6):1526–1538.

Green, L. V., Savin, S., and Wang, B. (2006). Managing patient service in a diagnostic medical facility. *Operations Research*, 54(1):11–25.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., et al. (1998). *Multivariate data analysis*, volume 5. Prentice hall Upper Saddle River, NJ.

Hancock, P. A. (1986). The effect of skill on performance under an environmental stressor. *Aviation, space, and environmental medicine*.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280.

Heckman, J. J. (1977). Dummy endogenous variables in a simultaneous equation system. Technical report, National Bureau of Economic Research.

Hockenberry, J., Lien, H.-M., and Chou, S.-Y. (2008). The impacts of task repetition and temporal breaks in production on human capital and productivity. *Journal of Human Capital*, 2(3):303–335.

Huh, W. T., Liu, N., and Truong, V.-A. (2013). Multiresource allocation scheduling in dynamic environments. *Manufacturing & Service Operations Management*, 15(2):280–291.

Ibanez, M. R., Clark, J. R., Huckman, R. S., and Staats, B. R. (2018). Discretionary task ordering: Queue management in radiological services. *Management Science*, 64(9):4389–4407.

Ingolfsson, A., Almehdawe, E., Pedram, A., and Tran, M. (2020). Comparison of fluid approximations for service systems with state-dependent service rates and return probabilities. *European Journal of Operational Research*, 283(2):562–575.

Ireland, N. (2021). Pandemic has left backlog of almost 16 million medical procedures in ontario, doctors say. *CBC News*.

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

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

Kc, D. S. and Terwiesch, C. (2017). Benefits of surgical smoothing and spare capacity: an econometric analysis of patient flow. *Production and Operations Management*, 26(9):1663–1684.

Kennedy, P. (2008). *A guide to econometrics*. John Wiley & Sons.

Klassen, K. J. and Rohleder, T. R. (2002). Demand and capacity management decisions in services: How they impact on one another. *International Journal of Operations & Production Management*.

Kolisch, R. and Sickinger, S. (2008). Providing radiology health care services to stochastic demand of different customer classes. *OR spectrum*, 30(2):375–395.

Kong, Q., Li, S., Liu, N., Teo, C.-P., and Yan, Z. (2020). Appointment scheduling under time-dependent patient no-show behavior. *Management Science*.

Kopach, R., DeLaurentis, P.-C., Lawley, M., Muthuraman, K., Ozsen, L., Rardin, R., Wan, H., Intrevado, P., Qu, X., and Willis, D. (2007). Effects of clinical characteristics on successful open access scheduling. *Health care management science*, 10(2):111–124.

Lauenstein, T. C. and Semelka, R. C. (2006). Emerging techniques: whole-body screening and staging with mri. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 24(3):489–498.

Levine, A. and Finkel, D. (1990). Load balancing in a multi-server queuing system. *Computers & operations research*, 17(1):17–25.

Lind, J. T. and Mehlum, H. (2010). With or without u? the appropriate test for a u-shaped relationship. *Oxford bulletin of economics and statistics*, 72(1):109–118.

Liu, N. (2016). Optimal choice for appointment scheduling window under patient no-show behavior. *Production and Operations Management*, 25(1):128–142.

Liu, N., Finkelstein, S. R., Kruk, M. E., and Rosenthal, D. (2018). When waiting to see a doctor is less irritating: Understanding patient preferences and choice behavior in appointment scheduling. *Management Science*, 64(5):1975–1996.

Liu, N. and Ziya, S. (2014). Panel size and overbooking decisions for appointment-based services under patient no-shows. *Production and Operations Management*, 23(12):2209–2223.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational behavior and human performance*, 3(2):157–189.

Lu, L. X. and Lu, S. F. (2018). Distance, quality, or relationship? interhospital transfer of heart attack patients. *Production and operations management*, 27(12):2251–2269.

Lu, S. F. and Lu, L. X. (2017). Do mandatory overtime laws improve quality? staffing decisions and operational flexibility of nursing homes. *Management Science*, 63(11):3566–3585.

Lu, S. F., Serfes, K., Wedig, G., and Wu, B. (2021). Does competition improve service quality? the case of nursing homes where public and private payers coexist. *Management Science*.

Lu, Y., Musalem, A., Olivares, M., and Schilkrut, A. (2013). Measuring the effect of queues on customer purchases. *Management Science*, 59(8):1743–1763.

Luo, J., Kulkarni, V. G., and Ziya, S. (2012). Appointment scheduling under patient no-shows and service interruptions. *Manufacturing & Service Operations Management*, 14(4):670–684.

Lupien, S. J., Maheu, F., Tu, M., Fiocco, A., and Schramek, T. E. (2007). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and cognition*, 65(3):209–237.

Maddala, G. S. (1986). *Limited-dependent and qualitative variables in econometrics*. Number 3. Cambridge university press.

Mancilla, C. and Storer, R. (2012). A sample average approximation approach to stochastic appointment sequencing and scheduling. *IIE Transactions*, 44(8):655–670.

Martin, D., Miller, A. P., Quesnel-Vallée, A., Caron, N. R., Vissandjée, B., and Marchildon, G. P. (2018). Canada's universal health-care system: achieving its potential. *The Lancet*, 391(10131):1718–1735.

Mousakhani, M. (2013). Sequence-dependent setup time flexible job shop scheduling problem to minimise total tardiness. *International Journal of Production Research*, 51(12):3476–3487.

Muthuraman, K. and Lawley, M. (2008). A stochastic overbooking model for outpatient clinical scheduling with no-shows. *Iie Transactions*, 40(9):820–837.

Norris, J. B., Kumar, C., Chand, S., Moskowitz, H., Shade, S. A., and Willis, D. R. (2014). An empirical investigation into factors affecting patient cancellations and no-shows at outpatient clinics. *Decision Support Systems*, 57:428–443.

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

Patrick, J., Puterman, M. L., and Queyranne, M. (2008). Dynamic multipriority patient scheduling for a diagnostic resource. *Operations research*, 56(6):1507–1525.

Pavel, M. S., Chakrabarty, S., and Gow, J. (2015). Assessing willingness to pay for health care quality improvements. *BMC health services research*, 15(1):43.

Priestman, T., Dunn, J., Brada, M., Rampling, R., and Baker, P. (1996). Final results of the royal college of radiologists' trial comparing two different radiotherapy schedules in the treatment of cerebral metastases. *Clinical oncology*, 8(5):308–315.

Rabe, M., Giacomuzzi, S., and Nübling, M. (2012). Psychosocial workload and stress in the workers' representative. *BMC public health*, 12(1):1–11.

Rollinson, D. C., Rathlev, N. K., Moss, M., Killiany, R., Sassower, K. C., Auerbach, S., and Fish, S. S. (2003). The effects of consecutive night shifts on neuropsychological performance of interns in the emergency department: a pilot study. *Annals of emergency medicine*, 41(3):400–406.

Sauré, A., Begen, M. A., and Patrick, J. (2020). Dynamic multi-priority, multi-class patient scheduling with stochastic service times. *European Journal of Operational Research*, 280(1):254–265.

Schultz, G. S., Sibbald, R. G., Falanga, V., Ayello, E. A., Dowsett, C., Harding, K., Romanelli, M., Stacey, M. C., Teot, L., and Vanscheidt, W. (2003). Wound bed preparation: a systematic approach to wound management. *Wound repair and regeneration*, 11:S1–S28.

Schultz, K. L., Juran, D. C., Boudreau, J. W., McClain, J. O., and Thomas, L. J. (1998). Modeling and worker motivation in jit production systems. *Management Science*, 44(12-part-1):1595–1607.

Setyawati, L. (1995). Relation between feelings of fatigue, reaction time and work productivity. *Journal of human ergology*, 24(1):129–135.

Shao, Y., Nijstad, B. A., and Täuber, S. (2019). Creativity under workload pressure and integrative complexity: The double-edged sword of paradoxical leadership. *Organizational Behavior and Human Decision Processes*, 155:7–19.

Shortle, J. F., Thompson, J. M., Gross, D., and Harris, C. M. (2018). *Fundamentals of queueing theory*, volume 399. John Wiley & Sons.

Soltani, M., Batt, R., Bavafa, H., and Patterson, B. (2019). Does what happens in the ed stay in the ed? the effects of emergency department physician workload on post-ed care use. *The Effects of Emergency Department Physician Workload on Post-ED Care Use (November 28, 2019)*.

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

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica: journal of the Econometric Society*, pages 557–586.

Stein, C., Truong, V.-A., and Wang, X. (2020). Advance service reservations with heterogeneous customers. *Management Science*, 66(7):2929–2950.

Sun, X. and Whitt, W. (2018). Delay-based service differentiation with many servers and time-varying arrival rates. *Stochastic Systems*, 8(3):230–263.

Sutherland, G., Russell, N., Gibbard, R., and Dobrescu, A. (2019). The value of radiology, part ii. *The Conference Board of Canada*.

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

Tay, A. (2003). Assessing competition in hospital care markets: the importance of accounting for quality differentiation. *RAND Journal of Economics*, pages 786–814.

Truong, V.-A. (2015). Optimal advance scheduling. *Management Science*, 61(7):1584–1597.

Velleman, P. F. and Hoaglin, D. C. (1981). *Applications, basics, and computing of exploratory data analysis*. Duxbury Press.

Wang, J., Vahid, S., Eberg, M., Milroy, S., Milkovich, J., Wright, F. C., Hunter, A., Kalladeen, R., Zanchetta, C., Wijeysundera, H. C., et al. (2020). Clearing the surgical backlog caused by covid-19 in ontario: a time series modelling study. *CMAJ*, 192(44):E1347–E1356.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, pages 817–838.

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association*, 57(298):348–368.

Zhu, S., Fan, W., Yang, S., Pei, J., and Pardalos, P. M. (2019). Operating room planning and surgical case scheduling: a review of literature. *Journal of Combinatorial Optimization*, 37(3):757–805.