Drivers of Physician Productivity and Performance: The Role of Queuing Systems, Relative Performance Feedback, and Cohort Turnover

DRIVERS OF PHYSICIAN PRODUCTIVITY AND PERFORMANCE: THE ROLE OF QUEUING SYSTEMS, RELATIVE PERFORMANCE FEEDBACK, AND COHORT TURNOVER

A DISSERTATION PRESENTED

BY

HUMMY SONG

TO

THE COMMITTEE ON HIGHER DEGREES IN HEALTH POLICY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE SUBJECT OF

HEALTH POLICY (MANAGEMENT)

HARVARD UNIVERSITY

CAMBRIDGE, MASSACHUSETTS

JANUARY 2017

Dissertation Advisors: Professors Robert S. Huckman & Anita L. Tucker Hummy Song

DRIVERS OF PHYSICIAN PRODUCTIVITY AND PERFORMANCE: THE ROLE OF QUEUING SYSTEMS, RELATIVE PERFORMANCE FEEDBACK, AND COHORT TURNOVER

ABSTRACT

This dissertation investigates how operational choices about the design of work systems in health care settings impact the way physicians deliver care. Across three studies, I examine the following potential drivers of physician productivity and performance: queuing systems, relative performance feedback, and cohort turnover.

In Chapter 1 – coauthored with Anita L. Tucker and Karen L. Murrell – we analyze the impact of pooled versus dedicated queuing systems on wait times and processing times in an emergency department (ED) setting. We find that – in contrast to what traditional queuing theory would predict – patients' average wait times and lengths of stay (LOS) are *shorter* when physicians are assigned patients under a dedicated queuing system. Interviews and observations of physicians suggest that improved performance under a dedicated system stems from physicians' increased ownership over patients and resources, which enables them to more actively manage the flow of patients. In Chapter 2 – coauthored with Anita L. Tucker, Karen L. Murrell, and David R. Vinson – we examine whether and how publicly (as opposed to privately) disclosing relative performance feedback (RPF) affects the speed and quality with which physicians deliver care in the ED. We find that public disclosure of RPF enables the identification and diffusion of best practices around ED workflow, which

ultimately helps reduce physician processing times without sacrificing quality. In Chapter 3 – coauthored with Robert S. Huckman and Jason R. Barro – we explore when and how cohort turnover affects hospital operational performance in U.S. teaching hospitals. Despite the anticipated nature of the cohort turnover and the supervisory structures that exist in teaching hospitals, we find that the annual cohort turnover of resident physicians in July results in increased resource utilization; we find limited evidence of negative effects on quality. Particularly in major teaching hospitals, we find evidence of a strong anticipation effect in which hospitals exhibit a gradual trend of decreasing performance that begins several months before the actual cohort turnover.

Together, these studies advance our understanding of how certain operational choices enable physicians to deliver care more efficiently without adversely affecting the quality of care. In doing so, this dissertation contributes to the literatures on health care operations, behavioral operations, and the productivity of service and knowledge workers. In addition, it provides practical implications for hospital managers who make important operational decisions that affect the efficiency and quality of care delivery.

Contents

List of Tables

List of Figures

ACKNOWLEDGMENTS

Writing this dissertation was truly a journey, and one I am grateful to have had the privilege of experiencing. Although it is my name alone that appears on the cover of this dissertation, there are many people who have contributed to it. To them, and to everyone who has supported me throughout this adventure, I want to say: *Thank you*.

First and foremost, I am incredibly thankful to my family, which has been an unwavering source of support from day 1. My parents instilled in me the confidence that I can achieve whatever I set out to explore. As academics themselves, they have shown me first-hand the joys of pursuing one's passions and curiosities. My sister Humbi has blessed me with her boundless energy and creativity, and she continues to inspire me to think bigger and outside the box.

I am fortunate to have had four outstanding committee members, each of whom deserves tremendous credit for having shaped my thinking and experiences during the PhD program.

Since the day I was admitted, Rob Huckman has had an open door—always willing to answer my "quick" questions, providing thoughtful feedback on my work, and even reading through STATA code line by line. Rob taught me to be a more careful and critical empiricist. He also helped me learn when and how to say no to projects, for which I know I will continue to thank him as I progress in my academic career.

I consider one of my luckiest days to be the day I met with Anita Tucker to pitch my half-baked research ideas and ask about potential field sites. Anita took a chance on me when I was a first-year PhD student trying to find her way, and I am forever grateful for it. I have benefited immensely from the countless hours she spent with me—collecting data, traveling to field sites, working through paper revisions, navigating the job market, and so much more. I hope to be able to pay it forward to my future advisees. Also, I would be remiss not to say thanks to her family, who so generously shared her time and energy with me.

Sara Singer also welcomed me onto her research team while I was still in my first year, and I have learned so much from our collaborations. Despite her busy schedule and close to zero hours of sleep, Sara always made time to meet—whether to talk about my latest set of findings or to help me think through my research direction. Sara is living proof that multitasking and happiness can coexist, with the key ingredient being an underlying passion for research. I hope to be able to carry her energy and spirit with me as I embark on my career.

Last but not least, Ryan Buell has been a wonderful champion and a role model. It is not an exaggeration to say it must be in his DNA to be exceptionally kind, thoughtful, and generous. Ryan never missed an opportunity to provide me detailed feedback on my work, and his insights always helped me push the work forward. His dedication to and enthusiasm for both research and teaching is something I hope to embody throughout my career.

Several others contributed directly to this dissertation as well. Most importantly, I had the privilege of working with two remarkable collaborators from Kaiser Permanente, Drs. Karen Murrell and David Vinson. Karen's dedication and passion for improving operational efficiency in emergency departments has been infectious and invigorating. David has been instrumental to keeping our collaborations moving forward, and his humorous emails have been a joy to read.

Also at Kaiser Permanente, Shannon Beckwith, Kristen Gregory, Mark Kauffman, Brent Soon, and Derek Ursin have been gracious with their time in providing us access to the data used in Chapters 1 and 2. I am also grateful to the several dozen anonymous physicians who participated in interviews and allowed me to observe their work.

This work also benefitted greatly from the numerous questions I asked of Andrew Marder at HBS and Simo Goshev at IQSS. They taught me how to improve the efficiency of my statistical coding and more. For Chapter 3, Sarah Woolverton at HBS provided tremendous help in organizing and synthesizing several gigabytes of Nationwide Inpatient Sample data, and Jean Roth at NBER provided access to data from the American Hospital Association's Annual Survey of Hospitals. Lydia Kim provided expert research assistance for Chapter 1.

There are many who have made this PhD program one I am proud to have been a part of. Making it through was by no means easy, but Joe Newhouse and Kathy Swartz's insightful comments during Tuesday seminars, their words of wisdom (and reassurance) at orientation, and the annual welcome parties in Joe's backyard made this a program that was challenging yet enriching and also homey. Tom McGuire's advice to me to keep an open mind about research direction—rather than staying married to what I proposed in my application essay—ended up being one of the most valuable pieces of advice I received early on. Debbie Whitney, Ayres Heller, and Jessica Livingston worked tirelessly to make this program a true community and family. I will always look forward to visiting our "headquarters" at 14 Story Street. Together with John Korn, Dianne Le, Jen Mucciarone, Marais Young, and Karla Cohen, they helped me navigate the ins and outs of an interdisciplinary PhD program. I am also grateful for the financial support from the HBS Doctoral Programs and the Daniel A. Buckley Scholarship for providing funding for my graduate studies.

There are also those who have kept me balanced and sane over these past several years. Ghideon Ezaz read multiple drafts of each of these chapters, and he listened to me practice presenting Chapter 1 so many times he could give the presentation himself. He continues to answer my numerous questions about hospital operations and surveys his friends and colleagues to help pre-test my research ideas. His level-headed perspective and constant support have been essential to me surviving graduate school. Tomoko Harigaya answered countless text messages with questions about statistical methods and STATA code. I was also fortunate to be a beneficiary of her approach to de-stressing—cooking delicious multi-course meals and baking delectable desserts. Teddy Svoronos, Mark Shrime, Hannah Neprash, Joe Hero, Chantelle Boudreaux, and Loren Saulsberry made learning fun. It is with them I learned not only about health policy and econometrics, but also about picklebacks and terms from Cards Against Humanity that shall not be named. Carolyn Deller, Jin Hyung Kim, Lisa Kwan, Rajesh Vijayaraghavan, Akash Chattopadhyay, Adi Gupta, and Rachel Arnett opened up their homes and became a second family to me. They made me look forward to going into the office, and together we celebrated many personal and family milestones. My track-mates Melissa Valentine, Scott Lee, Chinny Offor, Ashley-Kay Basile, Olivia Jung, Bethany Gerstein, Michaela Kerrissey, Phil Saynisch, Maryaline Catillon, and Lauren Taylor were excellent companions in navigating the dual worlds of HBS and GSAS across multiple campuses. Beyond the academic bubble, Amrit Chaudhuri, Dina Guzovsky, Julia Kamin, Sheryl Kluberg, and Cynthia Maertz provided a listening ear, delightful company, and wonderful friendship over the years.

Finally, I thank the many physicians, nurses, hospital administrators, and other health care providers who participated in my research. They generously shared their time, allowing me to absorb as much as I could about health care operations as it unfolds in real time. I am grateful to have had the opportunity to learn, and look forward to continuing on with the next phase of this journey.

INTRODUCTION

Health care is an industry that touches each and every one of us. In the United States, more than 83% of adults and 92% of children have contact with a health care professional in a given year, contributing to 929 million physician office visits, 126 million hospital outpatient visits, 136 million emergency department (ED) visits, and 35 million hospital inpatient visits per year according to the latest available data from 2014 (CDC - National Center for Health Statistics 2016). As an industry, the health care sector employs more than 11.5 million Americans (U.S. Bureau of Labor Statistics 2016a, b). As a share of the nation's Gross Domestic Product, total health expenditures constitute 17.5% (Centers for Medicare & Medicaid Services 2015).

Despite its scale and significant reach, the health care sector suffers from numerous inefficiencies. Better understanding the barriers to and potential solutions for improving efficiency in health care delivery is of substantial interest to health care providers, consumers, and policymakers alike. Meaningful solutions may equip health care providers to deliver higher quality care at lower costs while simultaneously enabling more patients to access care.

The field of operations management has developed various theories and approaches for improving the efficiency of organizations across many industries. For example, several operations management scholars have studied the newsvendor model and its applications to manufacturing and service industries to help forecast demand when demand is stochastic (Cachon 2002, Porteus 1990). Others have examined pooling inventory, queues, tasks, and servers, respectively, as a way to mitigate the negative effects of variability (Corbett and Rajaram 2006, Eppen 1979, Mandelbaum and Reiman 1998). Standardized work has also been studied as an effective approach for reducing variability in a variety of industries (Andritsos and Tang 2014, Spear and Bowen 1999). Though some of these principles have been adopted by health care delivery systems and organizations (e.g., lean thinking (Holden 2011, Mazzocato et al. 2010)), broader adoption and progress has been slow for two reasons. First, many of these principles in earlier operations management research emerged from manufacturing settings and related industries, from which health care diverges in important ways. For example, variation in patient needs and physician practice patterns lead to high levels of variability and an inherent need for some degree of flexibility. Second, regulatory aspects that are fundamental to the organization of health care delivery systems impose external constraints within which managers must make their decisions. This necessitates more careful thinking around how and whether certain operational choices will lead to improved productivity and performance.

In this dissertation, I examine how certain operational choices may enable managers of health care organizations to improve productivity without eroding quality. Specifically, I examine three potential drivers of physician productivity and performance: queuing systems, relative performance feedback, and cohort turnover. In doing so, I seek to develop greater insight into how (a) behavioral responses to operational choices and (b) operational choices around coordination affect the productivity and performance of physicians, often in unexpected ways. The work in this dissertation contributes to the growing body of empirical research in health care operations, behavioral operations, and, more broadly, to the literature on the productivity of service and knowledge workers. This work also has substantial implications for practitioners, especially given the high cost of labor in service industries, and in particular, the need for efficiency gains in health care.

This dissertation is comprised of three chapters. In Chapter 1, "The Diseconomies of Queue Pooling: An Empirical Investigation of Emergency Department Length of Stay" – coauthored with Anita L. Tucker and Karen L. Murrell, we examine the impact of pooled versus dedicated queuing systems on wait times and processing times in the ED setting. In a context where physicians have discretion over their work, we find that patients' average wait times and lengths of stay are shorter when physicians are assigned patients under a *dedicated* queuing system. This is in contrast to what traditional queuing theory would predict. Interviews and observations of physicians suggest that improved performance under a dedicated system stems from physicians' increased ownership over patients and resources, which enables them to more actively manage the flow of patients. Our findings suggest that the benefits from this improved flow management in a dedicated queuing system can be large enough to overcome the longer wait times predicted to arise from non-pooled queues.

In Chapter 2, "Public Relative Performance Feedback in Complex Service Systems: Improving Productivity through the Adoption of Best Practices" – coauthored with Anita L. Tucker, Karen L. Murrell, and David R. Vinson, we explore whether and how publicly (as opposed to privately) disclosing relative performance feedback about individual workers' processing times may affect physician productivity and service quality. Using three years of data from two EDs, we find that publicly disclosing relative performance feedback among physicians within the department leads to an 8.6% reduction in physician processing times and no significant reduction in quality. This benefit is greater when workers are carrying out unstandardized, rather than standardized, tasks. Our analyses suggest these effects may primarily stem from the identification and diffusion of best practices around workflow that is enabled by public disclosure. In this paper, we move beyond the conventional focus on standardizing work tasks and instead examine a way for managers to foster improved management of *workflow* in complex service systems.

In Chapter 3, "Cohort Turnover and Operational Performance: The July Phenomenon in Teaching Hospitals" – coauthored with Robert S. Huckman and Jason R. Barro, we explore how the annual cohort turnover of resident physicians affects operational performance at the hospital level.

Despite the anticipated nature of the turnover and the supervisory structures that exist in teaching hospitals, we find increased resource utilization at all teaching hospitals following the cohort turnover. With regards to clinical quality, we find limited evidence of a decrease in performance in major teaching hospitals and no significant effect in minor teaching hospitals. In major teaching hospitals, we also find evidence of a strong anticipation effect in which hospitals exhibit a gradual trend of decreasing operational performance that begins several months *before* the actual cohort turnover. We find that hospitals may be able to manage these effects through the facilitation of knowledge transfer from departing to entering physicians by proactively improving their overall quality of nursing and increasing their intensity of quality assurance.

CHAPTER 1

The Diseconomies of Queue Pooling:

An Empirical Analysis of Emergency Department Length of Stay

1.1 Introduction

Improving efficiency and customer experience is a key objective for service organizations. Skillful application of operations management principles may help achieve these goals. In particular, queue management decisions—such as queue structure and job routing policies—may impact how long customers have to wait for service and their service times.

Prior work has demonstrated through analytical models that pooling separate streams of identical customers into a single queue served by a bank of identical servers is more efficient than having a set of dedicated queues because pooling results in shorter wait times for service (Eppen 1979, Kleinrock 1976). Having a pooled queue structure leads to a reduction in wait time because it enables customers to be processed by any available server from a bank of servers, rather than having to wait for a specific server to become available. That said, prior analytical research also suggests that pooling queues may not always yield the expected performance improvements (Debo et al. 2008, van Dijk and van der Sluis 2009, Hopp et al. 2007, Jouini et al. 2008, Loch 1998, Mandelbaum and Reiman 1998). For example, combining streams of customers who have different processing requirements can introduce inefficiencies that erode the benefits of pooling (Benjaafar 1995, Green and Nguyen 2001, Mandelbaum and Reiman 1998, Rothkopf and Rech 1987). In addition, the perceived unfairness of a pooled queue, in which faster servers are assigned more customers than their peers, may negatively impact the speed at which servers work (Doroudi et al. 2011). Thus, the overall impact of queue pooling in service settings is ambiguous.

To our knowledge, there have been few field-based, empirical studies on the impact of pooled versus dedicated queue management systems on the speed of service. This is an important omission because, in service settings, servers can adjust how they manage their work to increase or decrease their service rate (Doroudi et al. 2011, Hopp et al. 2009). Operations management scholars advocate for more studies that examine how human behavior can alter the dynamics between operational variables and performance (Boudreau et al. 2003, Jouini et al. 2008). Thus, empirical research that examines the impact of queue structure on servers' behaviors can provide new insights for operations management theory and increase the relevance of queuing theory and research to practice.

To address this gap, we leverage the introduction of a new policy that changed the queuing system in only one part of a hospital's Emergency Department (ED), but not the other, from a pooled system to a dedicated system. The parallel trend in performance of the two parts of the ED before the queuing system change, and the fact that the change affected only one part of the ED, allows us to use a difference-in-differences approach to empirically test the impact of a change in the structure of the queuing system on the average wait time to be seen by an ED physician and the average length of stay (LOS) in the ED. LOS is a measure of service time and starts with the time the physician begins delivering care to the patient and ends with either a bed request for admission to the hospital or the discharge of a patient to home or to an outside facility. We use the term LOS rather than service time to more clearly convey that this measure encompasses both (a) the value-added time when clinicians are providing care, as well as (b) the time that the patient is occupying an ED bed but is not receiving active care (e.g., when the physician is waiting for test results or treating other patients).

The ED under study switched from a pooled to a dedicated queuing system to be able to handle the larger volume of patients predicted to occur due to the closing of a nearby ED. For both the pooled and dedicated queuing systems, a fairness constraint in the form of a round robin (RR) routing policy was used to assign patients to physicians, in which patients were evenly distributed across physicians independent of physician speed or idle time. The ED had this policy because physicians were paid a fixed salary and did not receive additional compensation for treating more patients or working more hours than scheduled. As a result, there were few financial incentives available to increase physician productivity, and instead, work was allocated equally among physicians. Using a difference-in-differences approach, we find that, on average, the use of a dedicated queuing system with a RR routing policy as a fairness constraint—after controlling for individual patient, physician, time, and ED characteristics—is associated with a 17% decrease in patients' average LOS and a 9% decrease in their average wait time relative to the control group. This represents a 39-minute reduction in LOS and a four-minute reduction in wait time—a meaningful time savings for the ED.

Operations management theory suggests a possible reason why the pooled queuing system with a fairness constraint is associated with a longer average LOS than the dedicated queuing system with a fairness constraint. Similar to workers in other service settings (Debo et al. 2008, Hasija et al. 2010, Tan and Netessine 2014), physicians in the dedicated queuing system are strategic servers who change their behaviors in response to their assigned responsibilities and ownership over the work routines and resources needed to accomplish those responsibilities (Cachon and Zhang 2007, Gilbert and Weng 1998, Hopp et al. 2007, 2009). Interviews with physicians suggest that, in this context, the increased ownership that stems from a dedicated queuing system with a fairness constraint leads to a situation in which the improvements in service rates due to better flow management are greater than the variability-buffering benefits of a pooled queuing system with a fairness constraint.

This paper makes a contribution to the literature on queue pooling because prior research has emphasized customer behaviors that reduce the process losses of dedicated queues, but fewer papers have empirically examined the impact of employee behaviors on the performance of dedicated versus pooled queuing systems (Boudreau et al. 2003, Hopp et al. 2007, Jouini et al. 2008). Our work thus informs the debate over the benefit of a pooled queue, which enables flexibility in the routing of jobs to servers, and a dedicated queue, which enables improvement in wait and service times through better flow management.

1.2 Prior Research and Hypotheses

1.2.1 Prior Research on Queue Management and Service Times

Operations scholars have investigated at least two different contexts in which pooling may occur: inventory waiting to be processed (production-inventory systems) and customers waiting for service (queuing networks). Most closely related to our research context, studies of queuing networks focus on the effect of pooling queues of customers, servers, and tasks in service organizations (Mandelbaum and Reiman 1998). Much of this research has been conducted with call centers, and has shown that the benefits of flexible servers and pooled queues can outweigh potential drawbacks (Anupindi et al. 2005, Bassamboo et al. 2010, Gans et al. 2003, Jouini et al. 2008). Researchers have reached similar conclusions in other settings, such as mail delivery, finding that pooling can improve quality while concurrently reducing costs (Ata and Van Mieghem 2008). Furthermore, prior research has found that pooling is beneficial and wait time reductions are achieved even when work is allocated fairly among servers using a RR routing policy (Hyytiä and Aalto 2013, Raz et al. 2006). In fact, Armony and Ward (2010) find that pooling with a fairness constraint outperforms classical pooling when the arrival rate of customers is high because faster servers have an incentive to slow their service rate under systems in which work is allocated based on server availability instead of a fair distribution across servers.

On the other hand, some analytical models have shown that the behavioral responses of servers and customers can reduce the expected benefits of queue pooling (van Dijk and van der Sluis 2008, Hopp et al. 2007, Loch 1998, Mandelbaum and Reiman 1998, Rothkopf and Rech 1987). Most pertinent to our study, strategic servers may reduce the effectiveness of queue pooling (Cachon and Zhang 2007, Debo et al. 2008, Hopp et al. 2007, 2009, Jouini et al. 2008). First, they may manipulate customer service times to be higher or lower by managing their tasks differently when it benefits them to do so (Hopp et al. 2007, Link and Naveh 2006, Tan and Netessine 2014). For example, in the restaurant industry, Tan and Netessine (2013) find that wait staff adjust the services offered to customers so that customers spend less time in the restaurant when workload is high. Similarly, Oliva and Sterman (2001) find that bank employees reduce the steps they go through to approve loans when workload is high, even though this erodes bank profitability. Second, strategic servers can also slow down their work pace. Using analytical models, Debo et al. (2008) show that when workers are paid by the quantity of work completed, such as taxicab drivers and lawyers, they add unnecessary tasks when business is slow, thereby increasing service time for their customers. Similarly, Hasija et al. (2010) find that call center agents take more time to answer customers' queries when they have low workloads if their contract rewards them for keeping utilization above a minimum threshold. Collectively, these studies suggest that service time is impacted by strategic servers' responses to incentives and responsibilities.

Even when strategic servers do not have direct financial incentives to adjust their service rates, they may still manipulate their service times if they have a high degree of perceived ownership over their assigned jobs. Employees feel higher levels of ownership when they are given the resources and responsibility to manage the complete workflow of a meaningful task (Hackman and Oldham 1976). By design, dedicated queuing systems with a fairness constraint afford higher levels of ownership than do pooled queuing systems with the same fairness constraint because in the former, each server has been explicitly assigned the responsibility for efficiently completing the work waiting in his or her queue. In contrast, pooled queuing systems provide lower levels of ownership because the responsibility for depleting the queue is dispersed over multiple servers. Thus, strategic servers in dedicated queuing systems with a fairness constraint may be more motivated to efficiently manage their workload than those in pooled queuing systems with a fairness constraint (Doroudi et al. 2011, Gilbert and Weng 1998).

1.2.2 Queue Management and Strategic Physician Behavior in the ED

ED physicians are strategic servers, as defined by Cachon and Zhang (2007). To illustrate how physicians operate as strategic servers, consider an ED physician who has a patient with a headache. The physician can treat the patient using any combination of the following tasks: obtain a detailed medical history to generate possible causes of the headache, order a computed tomography scan, or prescribe an aspirin. The physician's choice can impact the patient's LOS because of variance in the time required for the different options. In addition, the physician can influence patient LOS by proactively pulling for information, such as x-ray results, rather than waiting for that information to be pushed. The physician can also control his or her own utilization because there are usually multiple patients under the care of an ED physician. Thus, physicians can reduce their own idle times and further increase the flow of patients through the system.

In this paper, we consider two different types of queuing systems in the context of an ED. In a pooled queuing system—which is typical for most EDs in the United States—a physician is assigned to a patient only once the patient is placed in an ED bed. This means patients in the waiting room remain in a pooled queue while waiting for an open bed. In a dedicated queuing system, physicians are assigned to patients at the point of triage. Here, patients in the waiting room are, in effect, waiting to be seen by a specific physician. In the dedicated queuing system, each physician thus has greater ownership over his or her workload even before the patient is placed in an ED bed.

In the ED that we study, each physician in the dedicated system also controls his or her own bank of resources (e.g., beds and nurses) necessary to facilitate the flow of his or her own patients. Physicians are assigned patients in a RR fashion that fairly allocates patients among all physicians independent of physicians' service rates. In addition, they can only go home when all of their assigned patients are discharged or nearly discharged (e.g., awaiting a test result), and are not paid extra for working past the scheduled end of their shift. Therefore, physicians have an incentive and the ability to manage their workload as efficiently as possible. For example, physicians can coordinate the care of their patients with their nurses to prioritize getting test results back for a patient so he can be discharged, and then quickly move a patient from the waiting room into that vacated bed. In contrast, in the pooled queuing system, physicians do not "own" patients in the waiting room, nurses and beds are shared among all physicians, and they rely on a triage nurse, called the "internal triage" nurse, to manage the flow of patients into available beds for the entire ED. Thus, in the pooled queuing system, physicians' have ownership over a much smaller portion of the patient flow process. Based on our interviews with physicians and observations of their practice patterns, we suspect that the higher level of ownership of one's workload and the resources necessary to manage that workload afforded by the dedicated queuing system increases physicians' perceived ownership over patient flow. This results in physicians having a faster rate of discharging patients throughout their entire shift than when in the pooled system.

Prior theoretical operations management research suggests that when strategic servers have ownership and responsibility for managing flow, it can lead to lower service times. Gilbert and Weng (1998) and Cachon and Zhang (2007) construct analytical models of a buyer's choice of queue structure for allocating demand among two suppliers. They find that suppliers in a dedicated system produce the goods faster than those in a pooled system because the dedicated system's suppliers have more incentive to invest in production capacity. The dedicated system provides certainty that they will benefit from their capacity investments, which can be thought of as having ownership over a demand stream in combination with the responsibility over production resources needed to meet that demand. Similarly, in the context of a hospital's inpatient department, Best et al. (2015) use a stylized model to show that a patient flow director with increased ownership and responsibility for managing flow is able to attain a significant decrease in patient LOS. The authors suggest that this decrease is attained from increased motivation to cut non-value-added time and better coordinate patient care among doctors, nurses, and case managers.

In the context of an ED, switching from a pooled to a dedicated queuing system should similarly affect the behavior of physicians by increasing the degree of ownership physicians have over their patients' flow through the ED. Specifically, we hypothesize that ED physicians may attain a shorter average LOS for their patients when they work in an ED with a dedicated queuing system with a fairness constraint. Prior research suggests that servers work slower at low workloads because there is no need to work fast due to the slack capacity (Tan and Netessine 2014). However, in our ED setting, workloads are typically at high levels due the ED's ability to staff according to historical demand and to send clinicians home early during periods of unexpectedly low demand. Therefore, we hypothesize a direct, positive effect of a dedicated queuing system on LOS.

Hypothesis 1*: LOS is shorter in the ED when physicians are working in a dedicated queuing system as opposed to a pooled queuing system.*

We further consider how dedicated queuing systems may affect patients' average wait times. A priori, it is unclear whether dedicated queues with strategic servers will result in shorter or longer wait times for customers. On the one hand, when under a dedicated queuing system, if patients currently being cared for spend less time in an ED bed and if a physician proactively places the next patient from his or her queue into the newly available bed, the next patient's wait time may decrease due to an indirect queuing effect. In other words, the benefits of a dedicated queue—fair assignment of work and ownership over patients, resources, and patient flow—may overcome the negative impact on wait time of using a dedicated rather than a pooled queue. Thus, we predict the following:

Hypothesis 2a*: Wait time is shorter in the ED when physicians are working in a dedicated queuing system as opposed to a pooled queuing system.*

On the other hand, switching from a pooled to a dedicated queue may result in an increase in wait time, due to the well-known inefficiency of forcing customers' whose server is busy to wait for that server to be free, even if another server is idle (Eppen 1979, Kleinrock 1976). The inefficiency of dedicated queues might overpower the possible reduction in wait times due to faster service times. Therefore, we test the following competing hypothesis.

Hypothesis 2b*: Wait time is longer in the ED when physicians are working in a dedicated queuing system as opposed to a pooled queuing system.*

To understand the behavioral mechanism through which different queuing systems may impact LOS, we explore the rates at which physicians discharge patients during different time periods throughout their shifts. We hypothesize that the higher level of ownership over patient flow afforded by a dedicated queuing system, as opposed to a pooled queuing system, motivates physicians to more efficiently manage patient flow throughout the duration of the entire shift. Physicians in the dedicated system may be able to efficiently manage patient flow—and thus achieve higher discharge rates—by proactively "pulling" for lab, x-ray, and consult results; improving coordination with nurses to prioritize tasks necessary for discharge; initiating the discharge process sooner for patients ready for discharge; and making sure that nurses place waiting patients into available beds as soon as possible. This hypothesized increase in discharge rate is in contrast to only speeding up towards the end of the shift, which would be predicted if physicians were only subject to a deadline effect and were not better managing patient flow (Deo et al. 2014).

Prior theoretical research suggests that physicians in the dedicated system will have a greater incentive to consistently work at a higher rate because they can reap the benefits that stem from achieving a faster rate of production (Gilbert and Weng 1998). In our setting, the benefits to physicians of obtaining a higher discharge rate are (a) more time to spend with current patients, which increases both patient and physician satisfaction (Hopp et al. 2007); (b) idle time if the physician has no additional patients currently in queue (Armony and Ward 2010); and (c) less work remaining for the physician to complete before he or she can go home. In a pooled queuing system, these benefits do not necessarily accrue to physicians who work at a higher rate because the misalignment of responsibility for patient flow and ownership over patients and resources prevents physicians from being able to reap these benefits. Thus, we hypothesize that physicians working in a dedicated queuing system will attain higher rates of discharging patients throughout the shift. Specifically, we hypothesize that this increase in discharge rate will emerge a few hours after the beginning of a shift because the average LOS is greater than two hours and, therefore, it would not be possible to discharge many patients in the first two hours of one's shift. However, after this initial two-hour period, the faster discharge rate will be present throughout the remainder of the shift, rather than only at the end of the shift.

Hypothesis 3: *A physician's discharge rate of patients is greater for each non-initial time period of the shift when physicians are working in a dedicated queuing system as opposed to a pooled queuing system.*

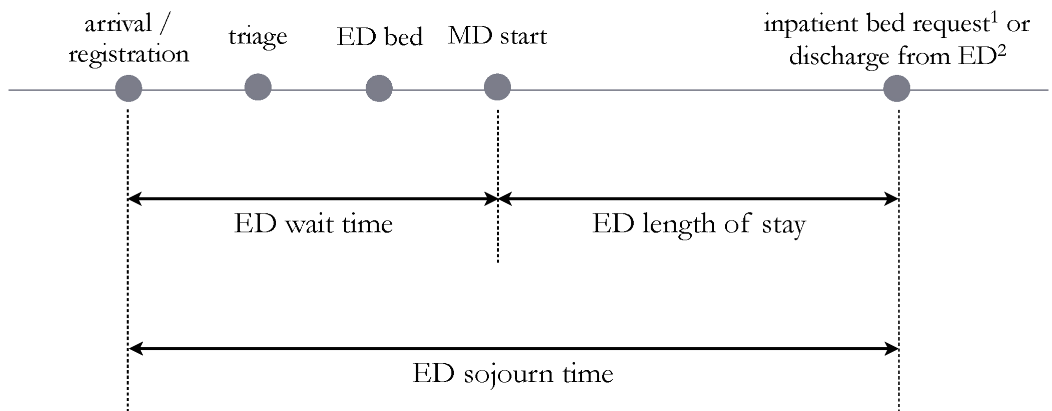
1.3 Setting, Data, and Empirical Methods

1.3.1 Research Setting

Our data come from the ED of a 162-bed hospital in northern California. We select this ED for study because in August 2008, it experienced an intervention—which we describe in more detail in chapter 1.3.2—that transformed a part of the ED from having a pooled queuing system to a dedicated queuing system for the patients waiting to be seen in the ED. We use data from a time span before and after the intervention (March 2007 to July 2010) to test our hypotheses about the impact of queuing systems on average LOS, wait time, and discharge rate in the ED.

Depending on the time of day, this ED had an average of two to five physicians staffing 41 ED beds and up to nine hallway gurneys. One bed was located in the resuscitation room and reserved for patients arriving without a pulse, three beds were in the trauma bay reserved for trauma intakes, four beds were in the Rapid Care Area (RCA) for low severity patients, and a minimum of two beds were reserved for psychiatric patients. This ED experienced an average 5% increase in patient volume each year, from approximately 65,000 patients in 2007 to 76,000 patients in 2010. The average daily patient volume was 178 patients in 2007 and 212 patients in 2010. This was a relatively large patient volume in comparison to other EDs in the surrounding areas.

This ED, like many others, had a standardized patient flow process (Figure 1.1). Upon a patient's arrival, a registration clerk conducted a brief registration process. A second triage nurse, called the "external triage" nurse, obtained vital signs, collected the chief complaint, and assigned an Emergency Severity Index (ESI) triage category—a commonly used, standard ranking of ED patient severity that ranges from levels 1 (highest acuteness) through 5 (lowest acuteness). This triage process accounted for a patient's expected level and type of resource utilization, and was used to route a patient to either the main area (main ED) or the RCA. The two areas of the ED each had its own equipment and staff to deliver care to patients (e.g., the RCA had its own computer terminals and vital sign monitors that were separate from the main ED's equipment). Ninety-eight percent of higher acuteness patients (ESI levels 1, 2, or 3) were treated in the main ED. Seventy-five percent of lower acuteness patients (ESI levels 4 or 5) were treated in the RCA. Lower acuteness patients were treated in the main ED when main ED beds were available and the waiting room census was low (15% of lower acuteness patients) or when they arrived between 11 p.m. and 7 a.m. when the RCA was closed (9% of lower acuteness patients).

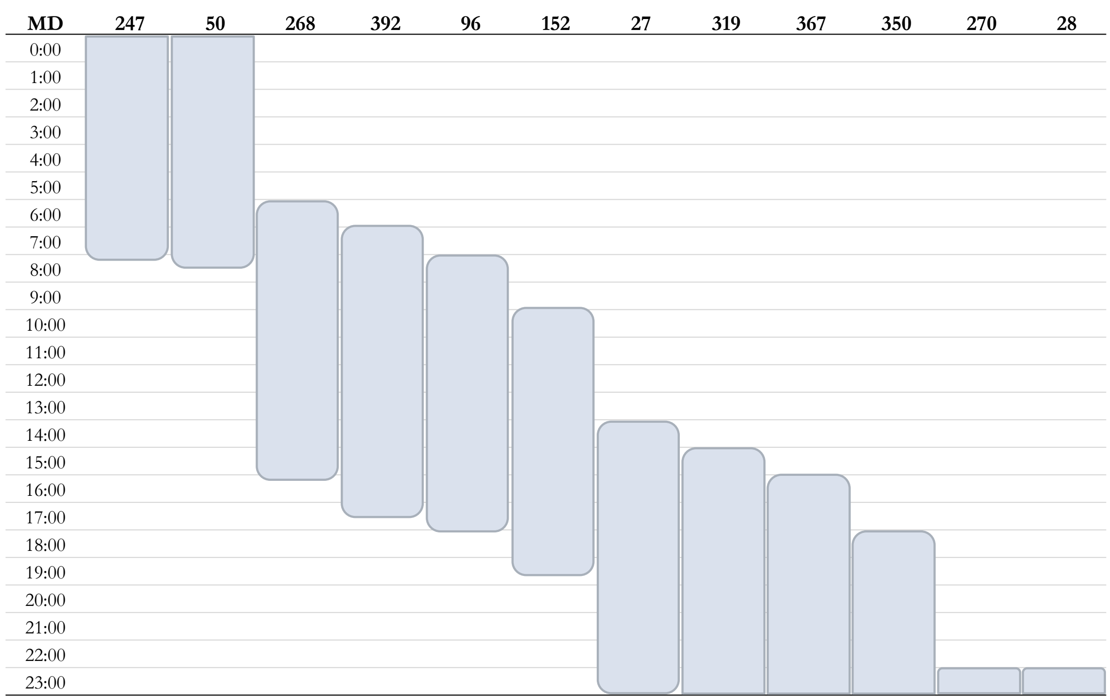


[1] For patients who were admitted to the hospital.
[2] For patients who were discharged home or to an outside facility.

Figure 1.1. Standard patient flow in the ED

In this ED, a computer system assigned each patient to a specific attending physician, either upon assignment to a bed (pooled queuing system) or at the point of triage (dedicated queuing system). The assigned physician assumed responsibility for completing the set of physician-related tasks for that patient during the patient's ED visit, such as taking the patient's history, prescribing medications, and ordering tests or treatments. This physician could consult other physicians concerning his or her patient's care, but this did not transfer the responsibility for patient care to the consulting physician. It was common for a physician to serve multiple patients simultaneously. In other words, a physician did not need to discharge one patient before starting work for the next patient.

Physicians arrived at staggered times throughout the day, such that there was not a certain time at which all physicians changed shifts (Figure 1.2). Physician shift times were determined in advance by the ED chief, and the ED scheduler assigned individual physicians to each of the pre-determined shift times. Physicians could change shifts on the hour between 5 a.m. and 11 a.m., between 2 p.m. and 5 p.m., and at 11 p.m. or midnight. Between 7 a.m. and 11 p.m., there was usually one physician working in the RCA and four physicians working in the main ED. During the overnight shift from 11 p.m. to 7 a.m., there were a minimum of two physicians and a maximum of four physicians working in the main ED.



Notes. MD numbers across the *x* axis are unique physician identifiers. Shaded bars indicate the duration of a physician's shift.

Figure 1.2. Example of physician shift distribution over a 24-hour period

Physicians were assigned to either the RCA or the main ED for the full duration of a shift by the ED scheduler. They were paid a flat rate for their shift without any additional compensation for the services provided or the number of hours worked. Thus, there were no incentives to stretch out treatment times by providing additional services. Prior to leaving the shift, physicians were expected to discharge or at least complete a care plan for the cohort of patients assigned to them (e.g., indicate what next steps should be taken if the lab test comes back positive versus negative), which incentivized physicians to get their patients through the system as efficiently as possible. Physicians were not required to stay if they had patients who were simply boarding in the ED, waiting to be transferred to an inpatient unit or to another facility. To allow physicians enough time to either complete a care plan or discharge the patients who had been assigned to them, they were assigned new patients only up until two hours before the scheduled end of their shifts. Patients arriving during the last two hours of a physician's shift were assigned to one of the other physicians on shift or, if it was close enough in time, to the oncoming physician. Because physician shifts were sufficiently staggered, there was always a physician available to take newly arriving patients and this did not induce greater variation in system productivity.

1.3.2 Intervention: Change in the Patient Assignment System

In August 2008, the main ED implemented an intervention called the Patient Assignment System (PAS). PAS restructured the main ED from having a pooled queuing system to a dedicated queuing system. Prior to PAS, higher severity patients due to be seen in the main ED returned to the waiting room after being triaged, with the exception of ESI level 1 patients who proceeded directly to the resuscitation room. When a bed became available in the main ED, the internal triage nurse placed the next patient of highest severity in this bed. Our interviews with ED physicians revealed that this process often resulted in a delay from the bed becoming available to a patient being placed in the bed because the internal triage nurse was not responsible for patient flow through the ED, and the physicians did not feel responsible for making sure that empty beds were filled quickly. Once a patient was placed in a bed, the computer system assigned each patient to a physician using a RR routing policy, which means that each patient was assigned to a physician in a set order that evenly distributed patients among physicians regardless of each physician's current workload. Once this assignment occurred, the physician could see the assigned patient listed under his or her panel when logged onto the patient management system on one of the ED computers. Thus, when a patient was waiting in the waiting room, he or she was in a pooled queue waiting to be assigned to any one of the, on average, four physicians on shift in the main ED.

Prior to the PAS intervention, the patient only entered a specific physician's queue after being placed in an available main ED bed by the internal triage nurse. It was at this point that the physician had ownership of the patient, not before. The only exception to the RR routing policy was made when a physician was currently involved in the resuscitation of an ESI level 1 patient, in which case another physician could voluntarily take on that physician's next patient. In addition, at the beginning of a physician's shift, the computer system assigned one, two, or three consecutive patients to the oncoming physician. The specific number of consecutive patients to whom a physician was assigned was automatically determined by the computer system based on the rate of patient arrivals. This RR routing policy was instituted to prevent physicians from unfairly selecting "easier" patients and to ensure that the faster physicians would not be unequally assigned more work simply because of their higher service rates. This simultaneously made patient routing to physicians both fair and nearly random rather than due to a physician's seniority or speed of discharging patients. It was feasible to implement because there were two organizational structures in place to minimize the variation in workload across the physicians staffing the main ED: (a) the hospital's trauma team assumed primary responsibility for incoming trauma patients and thus did not disproportionately increase the workload of an ED physician; (b) the RCA cared for lower severity patients. Thus, there was limited variation in patient intensity among the patients being assigned to the physicians staffing the main ED.

After PAS implementation, the computer system still used the RR routing policy but assigned each patient to a physician at the point of triage. This means that, when a physician logged onto the patient management system to view his or her panel of patients, the display showed not only those patients who were already placed in ED beds, but also those who were still in the waiting room. This increased physicians' perceived ownership of their patients because they were responsible for their patients' care and experience from triage onward—which included their time in the waiting room—rather than just from placement in an ED bed. In conjunction, it was now the physicians' responsibility to make sure their next patient from the waiting room was placed in an available main ED bed. To enable physicians to carry out this additional responsibility, six main ED beds and two hallway gurneys were allocated to each physician working in the main ED. In addition, two nurses were assigned to each physician to help care for patients, although each physician typically worked with other nurses outside of these two nurses during the course of the shift because (a) nurses' shift change times were not aligned with that of the physician and (b) nurses had designated break times during which a relief nurse substituted in for the duration of the break.

After PAS implementation, the computer system's RR routing policy was maintained and adhered to, even if there was a physician who had waiting patients while another physician had an available ED bed and no waiting patients. Hence, patient assignment remained independent of a physician's speed of discharging patients. Similarly, the incentive of having to stay until all patients had been cared for remained constant, though now physicians also had to care for the patients who had been assigned to them who were still in the waiting room.

In the RCA, the process used to assign patients to the physician working in the RCA did not change over the course of our study. A lower severity patient was assigned to a physician when he or she was called to be seen in the examination room, not while in the waiting room. Thus, the RCA physician was not responsible for any patient who was still waiting in the waiting room at the conclusion of his or her shift; any patient still waiting became the responsibility of the next physician coming on to the shift.

1.3.3 Data

This study uses approximately 3.5 years of de-identified electronic medical record (EMR) data of all 238,946 patients treated in the ED from March 1, 2007 to July 31, 2010. The dataset contains patient-level information including, but not limited to, the following: the patient's time of arrival and departure, LOS, ESI level, attending physician, and disposition. We exclude patients with no attending physician or ESI level listed on their record, patients who left without being seen by a physician, patients who had a LOS of zero minutes or less, patients whose records lacked a time stamp for when the physician began caring for the patient, and patients who were admitted to the hospital but whose records lacked a time stamp for when a bed request was made. In addition, we exclude patients whose LOS was greater than 48 hours; most of these patients presented with a psychological condition and were waiting to be discharged to an appropriate facility. We exclude these observations from our dataset because their extended LOS was typically driven by placement logistics rather than by physicians' levels of productivity. In addition, we exclude patients of ESI level 1 (i.e., patients needing resuscitation) and patients who died in the ED because their LOSs were likely to be driven by factors other than physician productivity. Finally, we exclude trauma patients because the hospital's trauma team, not a particular ED physician, primarily cared for these patients. Altogether, we exclude 12,817 patients or 5.4% of the overall sample.

Using this sample of 226,129 patients, we create a patient-level panel dataset that treats the physician as the panel variable. For our analyses, we exclude data from August 2008 to account for an acclimation period because the exact date of PAS implementation is unknown. In addition, we limit our sample to the patients seen by physicians who were full-time employees of this ED. Physicians who worked in this ED but were not full-time employees tended to be employees of other hospitals in the hospital's network who were brought in to cover small portions of shifts when the full-time ED physicians were not able to staff the ED (e.g., during physician staff meetings). This results in a final sample of 217,213 patients.

In addition to the EMR data, we also gathered qualitative data through 86 hours of observations of ED staff and unstructured interviews about workflow in the ED with ED physicians, nursing staff, and the ED unit leadership.

1.3.4 Dependent Variables

Our key dependent variables are ED wait time, ED LOS, and patient discharge rate. ED wait time is defined as the time from a patient's arrival to the ED to the time the physician began delivering care. ED LOS starts with the time the physician began delivering care to the patient and—for patients admitted to the hospital—ends with a bed request for admission to the hospital, thus excluding the time spent boarding in the ED and any time spent in an inpatient unit. For patients discharged to home or to an outside facility, ED LOS ends at the time of discharge. We log-transform ED wait time and LOS because each of their distributions are otherwise right-skewed. Patient discharge rate is defined as the number of patients discharged per hour by a given physician in a specified two-hour period of the shift, such as the first two hours, second two hours, or final two hours.

We employ a set of additional dependent variables for analyses that extend the main findings and consider possible alternate explanations. These include binary indicators for whether a lab was ordered, an x-ray was ordered, a patient was admitted to the hospital, a patient died in the ED, or a patient returned to the ED within 72 hours, respectively.

1.3.5 Independent and Control Variables

Patient assignment intervention in main ED. The implementation of PAS marks the time at which the main ED transitioned from having a pooled queuing system to a dedicated queuing system. We capture this transition with a binary interaction term, $PAS \times main$, which is equal to 1 in the main ED after the implementation of PAS and 0 otherwise (i.e., in the main ED before the implementation of PAS, or in the RCA at any time). To account for an acclimation period, we designate the pre-PAS period to include up to July 31, 2008 and the post-PAS period to begin with September 1, 2008.

Control variables. We account for several factors that may affect our dependent variables and may be correlated with our independent variables, *PAS* and *main*. These include factors related to the patient's condition, the state of the ED, the physician's practice experience, and time trends. To account for the variation in LOS due to the severity of a patient's condition, we control for the patient's acuteness and age. We account for patient acuteness using a series of dummy variables that reflect ESI levels 2, 3, 4, and 5, respectively. The combination of a patient's ESI level and age is the best approximation we have for patient condition and severity because our dataset does not include patients' specific diagnoses (e.g., diagnosis-related groups (DRGs)). It is important to control for patient acuteness because the patient mix in this ED changed over time, wherein more patients presenting to the main ED were of higher acuteness and more patients presenting to the RCA were of lower acuteness after PAS implementation.

To capture ED busyness and congestion, we control for the total number of physicians working during a given morning, afternoon, or overnight shift; the number of patients waiting to be seen by this physician at a given time; the number of patients being seen by this physician at a given time;

whether an ESI level 1 patient was present in the ED; and whether a trauma patient was present in the ED. Relatedly, to account for other systematic differences in patients' LOSs that would arise from differences in structural elements of the ED, we control for the general time frame of the physician's shift (morning, afternoon, or overnight) and the location of the shift (main ED or RCA).

To account for systematic differences arising from differences in physicians' experience working in this particular ED, we control for the number of shifts the physician has worked in this ED since the beginning of the dataset up until the point of each patient encounter. As we explain in more detail in chapter 1.3.6, we also include physician fixed effects to account for other unobserved differences by physician.

Finally, we account for time trends and related influences by including dummy variables for day of the week and by using month-year fixed effects.

1.3.6 Empirical Models

Our main analyses use a difference-in-differences framework to examine the relative changes in LOS and wait time for patients seen in the main ED and the RCA before and after PAS implementation. We use linear regression models with month-year and physician fixed effects and clustered standard errors. We cluster standard errors by physician to account for within-physician correlations of the error terms, both within and across shifts, rather than imposing the usual assumption that all error terms are independently and identically distributed. The fixed-effects models allow us to capture time trends and to control for unobservable individual physician effects that do not vary over time, such as level of motivation, innate ability, and practice routines. These are important to account for because they may significantly influence a physician's productivity level in ways that cannot be measured (McCarthy et al. 2012).

In addition to the standard assumptions of linear regression models, fixed-effects models make two key assumptions, both of which are satisfied in our study. First is the assumption of strict exogeneity, which means the observation-specific error term is uncorrelated with the covariates of the observation and all other observations belonging to the same cluster (Wooldridge 2010). This is a plausible assumption in our context because (a) there is a low likelihood that patients with multiple visits are treated by the same physician and (b) the patient error term is unlikely to be correlated with the covariates for other patients of the same physician. In addition, the unobservable random traits of physicians that affect their patients' average LOS are not likely to be associated with the key independent variable of interest. Specifically, the RR routing policy makes it unlikely that the fastest physicians receive the most complicated cases since patient assignment to physicians is random and is not driven by physician speed or physician preference.

We use fixed-effects models rather than random-effects models because we do not believe that the random effects assumption of zero correlation between the month-year effect or physician effect and the other covariates (such as the number of shifts worked by the physician) holds. By using fixed-effects models, we can account for the unobserved traits of each month-year and of each physician that are associated with a patient's LOS and also correlated with the independent variables of interest. Accordingly, we conduct the Durbin-Wu-Hausman test, which rejects the random-effects model in favor of the fixed-effects model ($\chi^2 > 169.45$, $p < 0.001$).

ED LOS. To test Hypothesis 1, we estimate the following difference-in-differences model at the patient level:

$$\ln LOS_{ijt} = \alpha_0 + \alpha_1 main_{ij} + \alpha_2 PAS_t \times main_{ij} + \delta \mathbf{X}_{ijt} + \theta_t + \gamma \mathrm{MD}_i + \varepsilon_{ijt} \tag{1.1}$$

Here, $\ln LOS_{ijt}$ represents the logged number of minutes that patient i of physician j stayed in the ED in month-year t; $main_{ij}$ indicates whether patient i of physician j was seen in the main ED; $PAS_t \times main_{ij}$ is an interaction term equal to 1 when the patient was seen in the main ED after the implementation of PAS; $\mathbf{X}_{ijt}$ is a vector of patient, physician, and day-of-week covariates; θ_t is a

Table 1.1. Summary definition of variables

Variable	Description	Level of Analysis
Main dependent variable		
ED wait time	Logged number of minutes elapsed between patient arrival to ED and MD start.	Patient
ED length of stay	Logged number of minutes elapsed between MD start and bed request (for patients admitted to hospital) or discharge from ED (for patients discharged home or to an outside facility).	Patient
Discharge rate	Number of patients discharged per hour by a given physician in a given 2-hour period of the shift (e.g., penultimate 2 hours, final 2 hours).	Physician-shift (2-hour period)
Independent and control variables		
ESI level	4 indicators for patient's ESI level (from highest to lowest: 2, 3, 4, 5).[1]	Patient
Age	Patient age in years.	Patient
MDs on shift	Number of all physicians working at any point during this shift.	Physician-shift
Current waiting count	Number of patients waiting to be seen by this physician at this time.	Patient
Current patient count	Number of patients being seen by this physician at this time.	Patient
Shift number	Indicator for what number shift this is for this physician in this dataset.	Physician-shift
ESI level 1 patient present	Indicator for presence of ESI level 1 patient (= 1 for present, = 0 for absent).	Patient
Trauma patient present	Indicator for presence of trauma patient (= 1 for present, = 0 for absent).	Patient
Arrival shift type	3 indicators for type of shift during which patient arrived (morning, afternoon, overnight).	Patient
Months since March 2007	Indicator for what number month this is in this dataset.	Patient
Day of week	7 indicators for day of week of shift.	Patient
Main ED	Shift location (= 1 for Main ED, = 0 for Rapid Care Area).	Physician-shift
PAS implemented	Indicator for whether PAS was implemented (= 1 for pre-implementation, = 0 for post-implementation).	Physician-shift
Interaction	PAS × Main ED.	Physician-shift
Additional dependent variables		
Lab ordered	Indicator for whether lab was ordered (= 1 for ordered, = 0 for not ordered).	Patient
X-ray ordered	Indicator for whether x-ray was ordered (= 1 for ordered, = 0 for not ordered).	Patient
Admitted to hospital	Indicator for whether patient was admitted to hospital upon discharge from ED (= 1 for admitted, = 0 for not admitted).	Patient
Died in ED	Indicator for whether patient died in ED (= 1 for died in ED, = 0 for did not die in ED).	Patient
Revisit within 72 hours	Indicator for whether patient returned to ED within 72 hours (= 1 for returned, = 0 for did not return).	Patient
Shift duration	Number of hours for which physician worked in ED during this shift.	Physician-shift
ED sojourn time	Logged number of minutes elapsed between arrival to ED and bed request (for patients admitted to hospital) or discharge from ED (for patients discharged home or to an outside facility).	Patient
ED boarding time	Logged number of minutes elapsed between bed request and discharge from ED (if admitted to hospital).	Patient

[1] Although the Emergency Severity Index (ESI) uses five categories, we have four indicators for patient ESI level because we exclude patients of ESI level 1 from our analysis.

vector of month-year fixed effects, MD_i is a vector of physician indicators; α's and δ's represent vectors of coefficients; γ represents a vector of physician fixed effects; and ε is the time-varying error term not already captured. Table 1.1 provides summary definitions for all variables included in our models.

In estimating Equation (1.1), we use the difference-in-differences estimator, $PAS_t \times main_{ij}$, to compare the difference in patients' average LOS in the main ED and the RCA before PAS implementation to the difference after PAS implementation. Because the queue structure did not change in the RCA, whereas the main ED moved from having a pooled to a dedicated queuing system, we consider the shifts worked in the RCA as comprising the untreated comparison group and those worked in the main ED as comprising the treatment group. By using a difference-in-differences approach, we are able to control for any bias caused by variables common to the main ED and the RCA, even when those variables are unobserved. Although the acuteness of patients seen in the two parts of the ED differed, thus implying differences in treatment processes and levels of patient LOS, the RCA serves as a reasonable control because, as our interviews with ED leadership and staff indicate, there were no changes besides PAS during the study period that affected only one part of the ED and not the other. Furthermore, we find that average LOS in the main ED and the RCA, respectively, exhibit parallel trends in the 17 months preceding the implementation of PAS.

After establishing the parallel trend assumption (Abadie 2005, Duflo 2001), we estimate the effect of transitioning from a pooled to a dedicated queuing system on patients' average LOS by examining the coefficient on the interaction term, $PAS_t \times main_{ij}$. We predict that this coefficient, α_2, is negative and statistically significant, suggesting that the dedicated queuing system is associated with a shorter average LOS than the pooled queuing system.

ED Wait Time. To test Hypotheses 2a and 2b, we estimate the following difference-in-differences model at the patient level:

$$\ln wait_{ijt} = \beta_0 + \beta_1 main_{ij} + \beta_2 PAS_t \times main_{ij} + \delta \mathbf{X}_{ijt} + \theta_t + \gamma \mathrm{MD}_i + \varepsilon_{ijt} \tag{1.2}$$

In this model, all variables remain the same as in Equation (1.1) with the exception of $\ln wait_{ijt}$, which represents the logged number of minutes that patient *i* of physician *j* in month-year *t* spent in the waiting room upon arrival to the ED. We use the same difference-in-differences approach as we do in testing Hypothesis 1. Here, we estimate the effect of PAS on patients' average ED wait time by examining the coefficient on the difference-in-differences estimator, $PAS_t \times main_{ij}$. Hypothesis 2a predicts that this coefficient, β_2, is negative and statistically significant due to an indirect queuing effect, suggesting that the dedicated queuing system is associated with a shorter average wait time than the pooled queuing system. Hypothesis 2b predicts that β_2 is positive and statistically significant due to the inefficiency of dedicated queues, suggesting that the dedicated queuing system is associated with a longer average wait time than the pooled queuing system.

Discharge Rate. To test Hypothesis 3, we estimate the following model at the physician-shift two-hour period level:

$$\ln discrate_{kj} = \varphi_0 + \varphi_1 PAS + \delta \mathbf{X}_{kj} + \gamma \mathrm{MD}_k + \varepsilon_{kj} \tag{1.3}$$

Here, $dischrate_{kj}$ represents the number of patients discharged per hour by physician *j* in a given two-hour shift period *k*; *PAS* indicates whether PAS had been implemented; φ's and δ's represent vectors of coefficients; and all other variables remain the same. For this analysis, we limit the sample to patients seen in the main ED and conduct a pre-post analysis. We do not employ a difference-in-differences approach because the different discharge processes in the main ED and the RCA make the comparison difficult, and because we are interested in the change in discharge rates during each of the two-hour periods over the course of a physician's shift rather than the change in the average discharge rate of a physician's shift. Therefore, we estimate Equation (1.3) separately for the first, second, penultimate, and final two-hour periods of a physician's shift. This allows us to examine whether and at what point during a physician's shift the implementation of PAS in the main ED affects the discharge rate of patients. Because the discharge rate is small and discrete and because the data are not over-dispersed, we employ a Poisson model with physician fixed effects.

If the dedicated queuing system results in a reduction in patients' average LOS, we would expect the discharge rate in each of the two-hour periods of a physician's main ED shift to increase after PAS implementation. This is because, after PAS implementation, physicians are more likely to engage in strategic behaviors throughout the shift to ensure that their patients' average LOS is as short as possible. However, because many of the preliminary tasks may be unaffected by the post-PAS increase in ownership, we expect that the discharge rate may be unaffected in the first two-hour period of a physician's shift. Accordingly, we predict that the coefficient on *PAS* will be positive and statistically significant for each of the second, penultimate, and final two-hour periods of a physician's shift, while it will not exhibit a statistically significant change for the first two-hour period of a shift.

Additional analyses. To better understand our main findings and consider possible alternate explanations, we conduct several additional analyses. We begin by considering two competing explanations that could account for the decrease in average LOS post-PAS. First, patients might have experienced shorter LOSs in the ED because physicians "cut corners" by stinting on care (Oliva and Sterman 2001). We assess this possibility by estimating Equation (1.1) with two different dependent variables, both measured at the patient level: whether labs are ordered for a patient and whether x-rays are ordered for a patient. Data on whether labs or x-rays are ordered for a patient are obtained directly from the hospital's EMR system. For each of these variables, we estimate Equation (1.1) as a logistic regression because both are binary indicator variables. Second, we consider whether the decrease in LOS stems from physicians shifting their work onto other clinicians. In the context of the ED, the most plausible scenario is ED physicians admitting more patients to the hospital, so that patients appear to stay in the ED for a shorter period of time. We examine this possibility by estimating Equation (1.1) with admission to the hospital as the dependent variable. Data for whether the patient is admitted to the hospital come from the EMR system and are measured at the patient level. Again, we estimate a logistic regression because admission to the hospital is a binary dependent variable. Next, we examine the possibility of the quality of care in the ED declining as an unintended consequence of PAS implementation in the main ED. As proxies for quality, we examine whether the patient returned to the ED within 72 hours after an initial visit and whether the patient died in the ED. We estimate Equation (1.1) as a logistic regression with each of these binary indicators as the dependent variable, respectively. For the analysis of ED revisits, we employ a 72-hour time period, which is the standard quality metric used to capture returning ED patients (Keith et al. 1989). For the analysis of patient mortality in the ED, we include a subset of previously excluded patient-level observations—specifically patients of ESI level 1, patients who died in the ED, and trauma patients.

In addition, we consider the potential impact of PAS on the duration of a physician's shift, which is measured as the number of hours for which a physician worked in the ED during a particular shift. Though this does not directly address why having a dedicated queuing system may decrease patients' average LOS, it is an important consideration if implementing a similar system at other EDs. If having a dedicated queuing system results in physicians staying longer to finish caring for their assigned patients, it may not be feasible to implement elsewhere for reasons of cost and physician burnout. To assess this possibility, we estimate a regression of a similar form as Equation (1.1) but with the shift duration as the dependent variable and at the physician-shift level. We use the shift duration and not the log of shift duration because the variable is normally distributed. We estimate this regression at the physician-shift level because the dependent variable (i.e., shift duration) is calculated at this level. If physicians are working longer hours as a result of PAS implementation, we would expect to see a positive and statistically significant coefficient on the interaction term, *PAS* × *main*.

Finally, we examine the impact of PAS implementation on sojourn time, which is the sum of ED wait time and LOS. We also examine the impact of the queue structure on ED boarding time to assess whether the change results in an admitted patient waiting longer for an inpatient bed. We estimate Equation (1.1) with logged ED sojourn time and logged ED boarding time, respectively, as the dependent variable.

1.4 Results

1.4.1 Descriptive Statistics

Table 1.2 presents means and standard deviations for all continuous variables included in the empirical models, stratified by location (main ED or RCA) and time period (pre-PAS or post-PAS). Table 1.3 presents the correlations between all continuous variables included in the empirical models. Table 1.4 presents percentages for all categorical or binary variables in the empirical models stratified by location and time period. As shown in Table 1.2, the average LOS for a patient seen in the main ED is approximately 3.5 hours, and it is about 50 minutes for a patient seen in the RCA. There are, on average, three or four physicians staffing the main ED during a given eight-hour period (i.e., morning shift, afternoon shift, overnight shift), and one physician staffing the RCA. None of the correlations between variables in the same regression model have levels close to or higher than 0.80, minimizing concerns about multicollinearity (see Table 1.3). We also check for multicollinearity by calculating variance inflation factors (VIF). The largest VIF is 5.45 and the mean VIF is 2.52 (not shown), both of which fall well below the conventional threshold of 10, providing additional evidence that multicollinearity is not a concern (Wooldridge 2012). As Table 1.4 shows,

Table 1.2. Summary statistics of continuous variables included in models

	Main ED				RCA			
	Pre-PAS		Post-PAS		Pre-PAS		Post-PAS	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. ED length of stay (minutes)	212.7	210.7	210.3	227.3	46.6	65.1	46.8	61.3
2. ED wait time (minutes)	43.9	42.9	33.6	30.4	54.8	43.6	46.0	33.4
3. Discharge rate	1.8	0.9	1.8	0.8	3.3	1.3	3.4	1.4
4. Age (years)	43.3	24.3	42.5	24.6	28.4	20.6	26.0	20.2
5. MDs on shift	3.4	0.9	3.7	1.0	1.0	0.2	1.0	0.3
6. Current waiting count	1.9	1.1	1.7	0.9	3.9	2.6	3.5	2.3
7. Current patient count	5.3	2.7	5.3	2.7	6.2	3.4	5.9	3.1
8. Shift number	115.1	72.0	335.5	151.5	135.7	78.4	373.8	126.8
9. Shift duration (hours)	9.7	1.5	9.2	1.3	10.2	1.2	10.0	1.0
10. ED sojourn time (minutes)	256.7	210.4	243.9	226.2	101.4	78.2	92.7	69.0
11. ED boarding time (minutes)	329.0	418.7	165.3	252.0	256.3	390.0	122.5	249.4

Notes. N = 217,213. Excludes all observations from August 2008 to account for an acclimation period.

Table 1.3. Correlations of continuous variables included in models

Variable	1	2	3	4	5	6	7	8	9	10	11
1. ED length of stay (minutes)	1										
2. ED wait time (minutes)	-0.13*	1									
3. Discharge rate	-0.22*	0.22*	1								
4. Age (years)	0.30*	-0.12*	-0.16*	1							
5. MDs on shift	-0.05*	0.07*	0.13*	-0.02*	1						
6. Current waiting count	-0.19*	0.50*	0.52*	-0.18*	0.13*	1					
7. Current patient count	-0.04*	0.33*	0.45*	-0.06*	0.08*	0.61*	1				
8. Shift number	-0.04*	-0.07*	0.08*	-0.05*	0.21*	0.01*	0.004*	1			
9. Shift duration (hours)	-0.11*	0.11*	0.19*	-0.05*	0.12*	0.16*	0.10*	-0.04*	1		
10. ED sojourn time (minutes)	0.98*	0.05*	-0.15*	0.28*	-0.04*	-0.10*	0.02*	-0.06*	-0.09*	1	
11. ED boarding time (minutes)	0.40*	0.07*	-0.02*	0.07*	-0.02*	0.06*	0.01	-0.17*	-0.003*	0.41*	1

Notes. N = 217,213. Excludes all observations from August 2008 to account for an acclimation period.
* $p < 0.05$

nearly 75% of main ED patients are of ESI level 3, with the remainder being predominantly split between ESI levels 2 and 4. About 65% of main ED patients had a lab ordered compared to less than 9% of RCA patients.

As expected, patients' average LOS differs significantly by their acuteness. Although, for brevity,

Table 1.4. Percent of sample by categorical and binary variables included in models

Variable	Main ED Pre-PAS	Main ED Post-PAS	RCA Pre-PAS	RCA Post-PAS
ESI level 2	7.88	14.05	--	--
ESI level 3	74.10	73.70	--	--
ESI level 4	17.68	11.85	96.23	96.53
ESI level 5	0.34	0.40	3.77	3.47
ESI level 1 patient present	8.78	9.16	9.69	9.92
Trauma patient present	7.36	27.68	7.62	29.25
Morning shift	34.21	35.55	40.58	37.73
Afternoon shift	44.59	43.63	53.87	55.99
Overnight shift	21.20	20.82	5.55	6.28
2007	57.92	--	56.23	--
2008§	42.08	14.39	43.77	16.16
2009§	--	52.47	--	54.80
2010	--	33.15	--	29.04
January	5.78	8.73	6.35	8.59
February	6.18	8.45	6.63	8.18
March§	12.38	9.63	11.81	9.17
April§	11.74	9.18	11.18	8.70
May§	12.07	9.85	12.24	9.36
June§	11.63	8.93	11.25	8.45
July§	12.12	9.19	11.44	8.87
August§	5.88	4.50	5.96	4.72
September	5.66	8.10	5.72	8.68
October	5.55	8.01	5.77	9.17
November	5.46	7.72	5.75	8.34
December	5.55	7.69	5.89	7.76
Sunday	15.15	14.75	15.01	15.09
Monday	14.89	15.19	15.00	15.22
Tuesday	14.08	14.11	14.50	14.07
Wednesday	13.93	13.40	13.72	13.58
Thursday	13.84	13.86	13.89	13.40
Friday	13.84	13.95	13.37	13.33
Saturday	14.27	14.73	14.51	15.31
Lab ordered	64.12	66.91	8.83	8.05
X-ray ordered	38.44	39.46	27.37	26.92
Admitted to hospital	14.11	12.42	0.38	0.30
Revisit within 72 hours	4.99	5.04	2.89	2.77

Note: N = 217,213. Excludes all observations from August 2008 to account for an acclimation period.

§ Because the study period begins on March 1, 2007 and ends on July 31, 2010, it is not surprising that a larger percentage of patients in our dataset presented to the ED in the months between March and July (inclusive) and in the years of 2008 and 2009, respectively. Because all observations from August 2008 have been excluded, it is also not surprising that this percentage is smaller for the month of August. When these summary statistics are produced with the inclusion of observations all from January 1, 2007 to December 31, 2010, we obtain an approximately uniform distribution of patients across all months of the year.

we do not display the numbers in a table, we find that for patients of ESI levels 2 to 5, the relationship between LOS and ESI level is a generally monotonically increasing one, with patients of a higher acuteness having a longer LOS. We account for the non-linearity of this relationship by adjusting for patient acuteness using a dummy variable for each ESI level.

1.4.2 Patient Assignment System Implementation in the Main ED

Both the qualitative and quantitative data suggest that PAS was implemented as described, though not without challenges. An ED physician remarked on one of the key challenges during implementation: "[PAS] was the hardest thing we have ever done. When we first started with the PAS system, it was a rocky road because sometimes there were patients in the waiting room when there was an open bed." This comment, in combination with the first author's observations of the ED workflow, suggests that physicians largely abided by PAS and the RR routing policy. In our EMR data, we find further support for the general adherence to the RR routing policy. In particular, patient demographics across physicians are well balanced and there is little variation in the average acuteness of patients assigned to each physician, suggesting it is not the case that certain physicians are being assigned particular types of patients. Furthermore, on average, there are only one or two ESI level 2 patients seen by a physician on a given main ED shift (*mean* = 1.4, *s.d.* = 0.5), suggesting the workload across physicians remains relatively balanced, thus allowing physicians to feasibly adhere to the RR routing policy.

However, there are rare situations when the RR routing policy is violated. Although the internal triage nurse cannot bypass the patient assignment generated by the computer system, physicians working in the main ED can bypass the RR assignment determined by the computer when another physician has an exceptionally time-consuming workload of ESI level 1 patients. One physician stated: "The expectation is that each physician sees the patients assigned to him or her. Ninety-nine percent of the time, this happens…[but] we help each other if someone gets slammed with a critical

[ESI level 1] patient… I remember one case last year where a physician got three critical patients in a row. That is extremely rare. He did not ask anyone, but two of his colleagues came and took two of the three patients [onto their panels].” This corroborates our understanding of the RR routing policy, in which other physicians can voluntarily take on the next patient assigned to a physician caring for an ESI level 1 patient.

1.4.3 Base Results

ED LOS. We estimate Equation (1.1) to assess the impact of having a pooled queuing system (versus a dedicated queuing system) on patients’ average LOS in the main ED. Table 1.5 column (1) presents a fixed-effects model that captures the effect of moving from a pooled to a dedicated queuing system. We find that the difference in patients’ average LOS between the main ED and the RCA is greater prior to PAS implementation. Once the main ED adopts a dedicated queuing system, this difference in patients’ average LOS is reduced. This difference-in-differences is captured by the coefficient on the interaction term, $PAS \times main$ (α_2 = -0.17, p < 0.001), and indicates that the transition from a pooled queuing system to a dedicated queuing system is associated with a highly significant reduction in the difference between the average LOS in the main ED and the RCA. This 17% decrease in the difference in average LOS in the main ED and the RCA after the implementation of PAS corresponds to a 39-minute decrease in LOS in the main ED relative to the RCA for an average patient of ESI level 3 seen by an average physician in the main ED. In other words, the average patient’s LOS in the main ED when compared to that in the RCA is significantly longer in the pooled queuing system than in the dedicated queuing system. This result offers strong support for Hypothesis 1, which predicts that, in our setting, pooled queuing systems are associated with a longer average LOS compared to dedicated queuing systems.

This finding is consistent with strategic changes in physicians’ behaviors to improve the management of their overall workflow. After PAS implementation, physicians change their practice behaviors because (a) they are aware of their full set of assigned patients, even those still in the waiting room, and (b) they have ownership over a designated bank of beds and nurses. In addition,

Table 1.5. Fixed-effects models at patient level

Variables	(1) Logged ED Length of Stay	(2) Logged ED Wait Time
Main ED	0.642***	0.377***
	(0.0307)	(0.0330)
PAS x Main ED	-0.174***	-0.0854**
	(0.0211)	(0.0265)
ESI level 3	-0.401***	0.415***
	(0.0159)	(0.0129)
ESI level 4	-1.211***	0.698***
	(0.0203)	(0.0241)
ESI level 5	-1.578***	0.617***
	(0.0252)	(0.0291)
Age	0.00773***	-0.00260***
	(0.000233)	(0.000188)
MDs on shift	-0.00559	0.0148
	(0.00302)	(0.00855)
Current waiting count	0.00184	0.189***
	(0.00171)	(0.00561)
Current patient count	0.000909	0.0201***
	(0.00167)	(0.00476)
Shift number	-0.000484*	-2.51e-05
	(0.000236)	(0.000359)
ESI level 1 patient present	0.0169**	0.0604***
	(0.00528)	(0.00806)
Trauma patient present	0.00844	0.0650***
	(0.00505)	(0.00562)
Afternoon shift	-0.0605***	0.0726***
	(0.00711)	(0.0161)
Overnight shift	-0.0731***	-0.157***
	(0.0131)	(0.0283)
Constant	4.546***	2.298***
	(0.0538)	(0.0611)
Observations	217,161	217,213
Number of ED physicians	40	40
Adjusted R^2	0.519	0.298

*** p <0.001, ** p <0.01, * p <0.05

Notes. All regressions are estimated at the patient level and include day of week controls, month-year fixed effects, and physician fixed effects. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician.

when one of their designated beds becomes available due to a patient discharge, physicians post-PAS are responsible for ensuring that their next patient from the waiting room is placed in that bed as quickly as possible. Specifically, according to interviews with physicians and observations of their practice patterns, physicians change their practice behaviors by (a) proactively "pulling" for lab results, x-ray results, and specialty consult results rather than waiting for this information to be "pushed"; (b) jointly managing their own workflow with that of the nurses with whom they are paired to better coordinate various tasks; (c) initiating the discharge process sooner for patients who are ready for discharge; and (d) making sure that patients are brought in from the waiting room as soon as one of their main ED beds becomes available rather than waiting for the internal triage nurse to place the next patient in an open bed. Collectively, these proactive actions lead to a shorter average LOS for patients in the main ED and result in a decrease in the difference in average LOS between the main ED and the RCA.

To confirm that the implementation of PAS only affected the main ED and not the RCA, a necessary condition for using the difference-in-differences framework, we conduct two analyses. First, using a pre-post analysis that is limited to the RCA, we examine whether there is a discontinuous jump in LOS in the RCA when PAS is implemented. We find no evidence of a significant increase or decrease in LOS in the RCA after PAS implementation ($\alpha_2 = 0.01, p \approx 0.84$). Second, we check for a change in the slope of LOS trends in the RCA before and after PAS implementation. A Wald test on the equality of coefficients also suggests no change in the trend of LOS in the RCA after PAS implementation ($p \approx 0.71$). Both of these findings indicate that the effects of PAS implementation were limited to the main ED and did not affect the RCA, thereby validating the use of the difference-in-differences model.

ED Wait Time. We estimate Equation (1.2) to examine the impact of having a pooled queuing system on patients' average wait time in the main ED. The results are summarized in column (2) of

Table 1.5. We find that the difference in patients' average wait time between the main ED and RCA decreases after PAS implementation (β_2 = -0.09, $p < 0.01$). This 9% decrease corresponds to a four-minute decrease in wait time in the main ED relative to the RCA for an average patient of ESI level 3 seen by an average physician in the main ED. In other words, the average patient's wait time in the main ED when compared to that in the RCA is significantly longer in the pooled system than in the dedicated system. This offers strong support for Hypothesis 2a, which predicts that, in our setting, dedicated queuing systems are associated with a shorter average wait time compared to pooled queuing systems. We do not find support for Hypothesis 2b, which relies on traditional queuing theory to predict that a pooled queue yields a shorter average wait time than do dedicated queues. In the dedicated system, the shorter wait times may be attained because, instead of waiting for the internal triage nurse to initiate placing the next patient in an open bed, physicians operating under PAS are able to initiate placement of the next patient from their queue into their newly available bed. Our findings are also consistent with the expectation of an indirect queuing effect, where patients experience shorter wait times because the patients who are receiving care have a shorter average LOS, which in turn makes beds in the main ED available sooner.

Discharge Rate. To better understand how dedicated queuing systems impact patients' average LOS, we estimate Equation (1.3). We examine whether, and at what point during a physician's shift, the implementation of PAS affects the discharge rate of patients in the main ED.

Columns (1) – (4) of Table 1.6 present fixed-effects models estimated at the physician-shift two-hour period level for each of the following four time periods: the first, second, penultimate, and final two-hour periods of a physician's shift. We find that in the second, penultimate, and final two-hour periods of a shift, the discharge rate in the main ED exhibits a significant increase after PAS implementation. Specifically, after PAS implementation, the discharge rate is 1.05 times greater (φ_1 = 1.05, $p < 0.05$) in the second two hours, 1.07 times greater (φ_1 = 1.07, $p < 0.001$) in the penultimate

Table 1.6. Fixed-effects models at physician-shift levels

Variables	(1) Discharge Rate in First 2 Hours of Shift	(2) Discharge Rate in Second 2 Hours of Shift	(3) Discharge Rate in Penultimate 2 Hours of Shift	(4) Discharge Rate in Last 2 Hours of Shift	(5) Shift Duration
Main ED	--	--	--	--	1.060***
					(0.166)
PAS	1.042	1.053*	1.069***	1.051**	--
	(0.0275)	(0.0245)	(0.0190)	(0.0204)	
PAS x Main ED	--	--	--	--	-0.0904
					(0.0855)
Percent of ESI level 3 patients	1.001	1.002***	1.002***	1.001*	0.0137***
	(0.000641)	(0.000430)	(0.000440)	(0.000404)	(0.00162)
Percent of ESI level 4 patients	1.007***	1.008***	1.005***	1.004***	0.0325***
	(0.000761)	(0.000535)	(0.000668)	(0.000553)	(0.00233)
Percent of ESI level 5 patients	1.012***	1.007*	1.001	1.006*	0.0311***
	(0.00368)	(0.00323)	(0.00299)	(0.00240)	(0.00537)
Average age of patients	1.000	1.000	1.001***	1.000	-0.0208***
	(0.000269)	(0.000267)	(0.000295)	(0.000386)	(0.00309)
MDs on shift	0.982*	0.981*	0.976**	0.968***	-0.331***
	(0.00736)	(0.00884)	(0.00730)	(0.00662)	(0.0239)
Average waiting count	1.015	1.014	0.999	1.001	-0.530***
	(0.0184)	(0.00961)	(0.00925)	(0.00694)	(0.0364)
Average patient count	1.094***	1.123***	1.111***	1.104***	0.661***
	(0.0172)	(0.00861)	(0.00510)	(0.00432)	(0.0260)
Shift number	1.000	1.000	1.000	1.000	7.06e-05
	(0.000231)	(0.000226)	(0.000184)	(0.000142)	(0.000805)
Percent of time ESI level 1 patient present	0.996	0.995	0.980	1.022	0.0495
	(0.0279)	(0.0186)	(0.0172)	(0.0161)	(0.0541)
Percent of time trauma patient present	0.998	0.965*	1.000	1.016	-0.0234
	(0.0185)	(0.0157)	(0.00994)	(0.0132)	(0.0479)
Afternoon shift	0.936***	1.064**	1.154***	1.156***	-0.134*
	(0.0133)	(0.0208)	(0.0199)	(0.0152)	(0.0541)
Overnight shift	0.806***	1.031	1.094***	0.980	-1.834***
	(0.0203)	(0.0382)	(0.0225)	(0.0235)	(0.130)
Constant	--	--	--	--	6.917***
					(0.301)
Observations	3,922	8,594	10,675	10,905	14,153
Number of ED physicians	38	39	38	40	40
Adjusted R^2	--	--	--	--	0.329

*** p <0.001, ** p <0.01, * p <0.05

Notes. Columns (1) – (4) are conditional fixed-effects Poisson models estimated at the physician-shift 2-hour period level with linear time trends by month, day of week controls, physician fixed effects, and heteroskedasticity robust standard errors. Discharge rate reflects the number of patients discharged per hour by a given physician in a given 2-hour period of the shift, and coefficients have been exponentiated to show incident rate ratios. Column (5) is a fixed-effects linear regression model estimated at the physician-shift level with day of week controls, month-year fixed effects, physician fixed effects, and heteroskedasticity robust standard errors clustered by physician. Shift duration is expressed in hours.

two hours, and 1.05 times greater (φ_1 = 1.05, $p < 0.01$) in the final two hours of a physician's main ED shift. We also find that this increase in discharge rate does not manifest in the first two hours of a physician's main ED shift (φ_1 = 1.04, $p \approx 0.12$). Based on observations in the ED and the fact that the average LOS of a patient seen in the main ED is 211 minutes (i.e., approximately 3.5 hours), the lack of significant difference in discharge rates in the first two hours of a shift may be due to the fact that the baseline amount of time necessary for patient care in the main ED is greater than two hours and, therefore, it is difficult for physicians to have a faster discharge rate during the first two hours of a shift.

Our findings are thus consistent with Hypothesis 3, which predicts that physicians in a dedicated queuing system exhibit a higher discharge rate that is sustained throughout the entire shift, which indicates that physicians are engaging in strategic behaviors over the entire course of the shift. This may be attributable to their greater ownership for patient flow and the resources needed to manage patient flow that comes with working in the ED's dedicated queuing system.

1.4.4 Consideration of Alternate Explanations and Unintended Consequences

Though our finding of a reduction in the difference between main ED and RCA patients' average LOS in a dedicated queuing system versus a pooled queuing system is consistent with an increase in physicians' strategic behavior to more efficiently manage patient flow, we consider alternate explanations that could also be consistent with our finding. We also explore the possibility of unintended consequences arising when implementing a dedicated queuing system.

Testing for changes in the provision of care. First, one possibility is that physicians stint on care after PAS implementation because of the increased pressure to care for all patients in their dedicated queues. If fewer services are provided to patients, they may stay in the ED for a shorter amount of time. For example, if a patient who would have otherwise received an x-ray does not, she

Table 1.7. Logistic regression models at patient level for alternate explanations and unintended consequences

Variables	(1) Lab Ordered	(2) X-ray Ordered	(3) Admitted to Hospital	(4) Revisit within 72 hours	(5) Died in ED
Main ED	1.451***	-0.103*	1.673***	0.167***	--
	(0.120)	(0.0503)	(0.102)	(0.0446)	
PAS	--	--	--	--	-0.669*
					(0.301)
PAS × Main ED	-0.0847	-0.0260	-0.188	0.00770	--
	(0.0468)	(0.0382)	(0.132)	(0.0506)	
ESI level 2	--	--	--	--	-5.270***
					(0.420)
ESI level 3	-0.693***	-0.508***	-1.007***	0.0381	-7.457***
	(0.0319)	(0.0334)	(0.0322)	(0.0375)	(0.538)
ESI level 4	-2.550***	-0.799***	-2.929***	-0.374***	-8.820***
	(0.0430)	(0.0476)	(0.0804)	(0.0633)	(1.008)
ESI level 5	-3.275***	-2.348***	-5.300***	-0.577***	--
	(0.0732)	(0.117)	(0.994)	(0.155)	
Age	0.0176***	0.0221***	0.0389***	0.00125*	0.0284***
	(0.000652)	(0.000937)	(0.000692)	(0.000515)	(0.00435)
MDs on shift	-0.0146	-0.0205	-0.00650	-0.0106	-0.0915
	(0.0180)	(0.0113)	(0.0127)	(0.0136)	(0.0974)
Current waiting count	0.0131	0.00726	-0.0187	-0.0232**	-0.0480
	(0.00723)	(0.00572)	(0.0111)	(0.00803)	(0.0750)
Current patient count	-0.0165***	0.00217	-9.74e-05	-0.00847	-0.00174
	(0.00462)	(0.00381)	(0.00435)	(0.00436)	(0.0307)
Shift number	-0.000511	-0.000515*	-0.000314	8.50e-05	-7.19e-05
	(0.000303)	(0.000260)	(0.000265)	(0.000122)	(0.000688)
ESI level 1 patient present	0.0316	-0.0101	-0.0432	0.00331	-0.0577
	(0.0276)	(0.0163)	(0.0288)	(0.0467)	(0.439)
Trauma patient present	0.0117	0.0241	-0.00760	0.0513	-0.204
	(0.0183)	(0.0148)	(0.0286)	(0.0369)	(0.174)
Afternoon shift	-0.0999	0.0286	0.0460	-0.00747	0.226
	(0.0552)	(0.0274)	(0.0338)	(0.0327)	(0.162)
Overnight shift	-0.175***	-0.0177	0.000590	0.0713	0.459
	(0.0532)	(0.0330)	(0.0472)	(0.0559)	(0.282)
Constant	-0.558***	-0.767***	-4.317***	-3.035***	-1.687*
	(0.160)	(0.0771)	(0.122)	(0.131)	(0.729)
Observations	193,807	193,807	193,807	193,807	132,952
Pseudo R^2	0.331	0.0679	0.257	0.0110	0.564

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Notes. All regressions are logistic regression models estimated at the patient level. Columns (1) – (4) include day of week controls, month-year fixed effects, and physician fixed effects. Column (5) includes linear time trends by month, day of week controls, and physician fixed effects. Column (5) also includes previously excluded observations – specifically patients of ESI level 1, patients who died in the ED, and trauma patients. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician.

would likely stay in the ED for a shorter duration because she would not need to wait for the x-ray machine to become available, have the x-ray taken, and wait for the radiologist to read the films. If physicians are stinting on care post-PAS, we would be mistaken to assume that the reduced LOS stems from an increase in physicians' strategic behaviors to more efficiently manage patient flow.

We do not find strong evidence of stinting on care after the transition to a dedicated queuing system in the main ED. In column (1) of Table 1.7, we examine the change in a patient's likelihood of having a lab test ordered. We find that the coefficient for $PAS \times main$ is not statistically significant (α_2 = -0.08, $p \approx 0.07$), suggesting that the difference in the likelihood of having a lab test ordered for a patient in the main ED and the RCA does not change significantly after PAS. Similarly, in column (2) of Table 1.7, we do not find a statistically significant change in a patient's likelihood of having an x-ray ordered (α_2 = -0.03, $p \approx 0.50$). This suggests that there is no meaningful change in the difference in a patient's likelihood of receiving an x-ray between the main ED and the RCA before and after the implementation of PAS. In combination, these results suggest that physicians are not systematically stinting on care in the main ED as compared to the RCA as a result of PAS implementation.

Testing for changes in the likelihood of a patient's admission to hospital. A second possibility is that ED physicians may be reducing patients' average LOS in the ED by passing them off to other hospital departments earlier. If an ED physician decides to have a patient admitted to the inpatient unit for further evaluation, rather than taking the time to conduct further evaluation while the patient is still in the ED, the patient's LOS in the ED may appear to be shorter than it would be otherwise.

We do not find evidence of main ED patients exhibiting a higher likelihood of admission to the hospital, relative to RCA patients, after PAS. As shown in column (3) of Table 1.7, we find that the difference in a patient's likelihood of being admitted to the hospital when in the main ED versus the

RCA does not change significantly after PAS implementation (α_2 = -0.19, $p \approx 0.16$).

Testing for changes in the quality of care. Next, we consider two potential unintended consequences of this transition from a pooled to a dedicated queuing system in the main ED. We assess whether patients are more likely to return to the ED within 72 hours of being seen, which could be an unintended consequence of physicians providing lower quality or insufficient care in order to decrease patient LOS. Similarly, if physicians are providing lower quality care such that more patients are dying in the ED, this truncating effect on LOS may result in a decrease in the average patient's LOS in the ED.

We do not find evidence of lower quality of care as measured by revisits to the ED within 72 hours. As is summarized in column (4) of Table 1.7, we find no statistically significant changes after PAS implementation in the difference in the likelihood of returning to the ED within 72 hours of an initial visit (α_2 = 0.01, $p \approx 0.88$). Even when using a more inclusive cutoff of 7 days (results not shown), we find no statistically significant changes in the difference in the likelihood of revisit (α_2 = 0.07, $p \approx 0.11$).

In addition, we do not find evidence of lower quality of care as measured by mortality in the ED. These results are presented in column (5) of Table 1.7. Due to the lack of variation in the dependent variable among patients of ESI level 5 and patients seen in the RCA, these two categories of patients are omitted from the analysis. In the resulting analysis, comparing patient mortality in the main ED before and after the implementation of PAS, we find that the likelihood of dying in the ED decreased after the transition to a dedicated queuing system (α_2 = -0.67, $p < 0.05$). This suggests that the quality of care, as measured by patient mortality in the ED, *improved* after PAS was implemented, thereby reducing concerns that the assignment of patients in the waiting room to a specific physician might adversely affect patients.

Testing for potential impact on the duration of a physician's shift. Finally, we consider the potential impact of PAS on the duration of a physician's shift. As summarized in column (5) of Table 1.6, we find no statistically significant change in the difference between the duration of a shift in the main ED and the RCA before and after PAS implementation (α_2 = -0.09, $p \approx 0.30$). This suggests that physicians are not working longer hours in the main ED as a result of the intervention.

1.4.5 Specification Tests

To examine the robustness of our main findings about LOS, we test a variety of other specifications in addition to the reported models. Due to space constraints, these results are not reported in tables.

First, we use a limited model specification that includes only patient ESI levels as control variables. We retain patient ESI levels because the average acuteness of patients arriving in the main ED increased over time while that of patients arriving in the RCA decreased over time. We find that the base result remains very robust to this limited model specification (α_2 = -0.16, $p < 0.001$), with the magnitude of the effect decreasing only slightly from 17% to 16%.

We then repeat our estimation of Equation (1.1) using non-logged LOS and bootstrapped standard errors. With this alternate model specification, we find that PAS implementation is associated with a 23-minute reduction in the difference in LOS between the main ED and the RCA (α_2 = -22.73, $p < 0.001$). Even when not using a log-level specification to account for the heavily skewed nature of the dependent variable, we obtain results that are robust to our base findings.

Although our interviews with ED staff suggest that there were no other interventions besides PAS that were applied to only the main ED or only the RCA during the study period (March 1, 2007 to July 31, 2010), we apply our analyses to shorter time frames around PAS implementation to nullify the possibility of other effects. When we limit the time frame to three months, seven months, 12 months, 15 months, and 18 months before and after the intervention, we find that our base results remain robust to these shorter time frames ($\alpha_2 < -0.10$, $p < 0.001$).

Next, we repeat our analyses using logged ED sojourn time to test for the impact of PAS on a more holistic measure of patient experience. We find that PAS is associated with a 10% decrease in the difference between main ED and RCA sojourn times before and after PAS implementation ($\alpha_2 = -0.10$, $p < 0.001$). This suggests that when taking both wait time and LOS into account, PAS is associated with a reduction in the average time that patients spend in the ED. In addition, we examine the impact of PAS on logged ED boarding time, which is the amount of time that patients being admitted to the hospital spend waiting for an inpatient bed. We find no statistically significant change in the difference in ED boarding times for patients in the main ED and the RCA before and after PAS implementation ($\alpha_2 = -0.25$, $p \approx 0.09$). This is consistent with our expectation because ED boarding time is primarily determined by the inpatient unit's capacity to admit a new patient rather than ED physicians' productivity levels.

Next, we limit our sample to those patients seen in the main ED and conduct a pre-post analysis, comparing the average LOS of patients before and after PAS. We find that our main findings are robust to this alternate specification that does not use a difference-in-differences approach, where PAS is associated with a 5% decrease in LOS in the main ED ($\alpha_2 = -0.05$, $p < 0.01$).

In addition, our results do not appear to be driven by differences in patient care delivered in the two areas of the ED. To examine this, we assess whether the transition from a pooled system to a dedicated system differentially affects LOS depending on the location of a patient's ED care. To conduct this analysis, we use the same empirical model as Equation (1.1), but limit the sample to patients of ESI levels 4 and 5, and with each independent variable of interest interacted with ESI level 5. We limit the sample to these patients because they constitute the group of patients who are potentially seen in both areas of the ED (because all ESI level 4 and 5 patients are seen in the main ED after 11 p.m.). This analysis suggests that there are no differential effects by the location of a patient's ED care ($p \approx 0.32$).

Furthermore, we examine whether the base results are sensitive to heterogeneity in patient acuteness. In other words, we examine whether the transition from having a pooled queuing system to a dedicated queuing system has a greater impact on patients with a higher ESI level as opposed to those with a lower ESI level. Using a similar approach as above, we explore this possibility by limiting the sample to patients of ESI levels 2 and 3, and interacting each independent variable of interest with ESI level 3. For this analysis, we limit the sample to patients of these two ESI levels because they exhibit two different groups with relatively longer LOS (for ESI level 2, *mean* = 332 minutes, *s.d.* = 330 minutes) and shorter LOS (for ESI level 3, *mean* = 212 minutes, *s.d.* = 202 minutes). This analysis suggests that patients of higher acuteness (ESI level 2) are likely to experience a greater decrease in LOS after the implementation of PAS compared to patients of a relatively lower acuteness (ESI level 3) ($\alpha = 0.42$, $p < 0.01$). While it is beyond the scope of this paper to examine why this heterogeneity arises, we speculate that it may be due to the prioritization of higher acuteness patients (ESI level 2) within each physician's dedicated queue.

We also repeat our analyses using several different exclusion criteria in constructing our sample and find that our results are robust in all of the following analyses. First, we include all observations that had previously been excluded as outliers (i.e., patients with a LOS greater than 48 hours). Then, to test our hypotheses on an even more homogeneous set of patients and ensure that our findings are not driven by outliers, we exclude observations with a LOS greater than one day (24 hours) and the average duration of one shift (9.4 hours), respectively. Next, we test our hypotheses on a sample that includes ESI level 1 patients, which were previously excluded. Finally, we test our hypotheses on a sample that excludes patients arriving by ambulance and patients presenting with a psychological condition, respectively, both of whom were previously included. All coefficients of interest and their corresponding significance levels remain robust to these alternate specifications (α_2

< -0.13, $p < 0.001$).

Finally, we use hierarchical linear models, which specify random effects rather than fixed effects at the physician level. We conduct this analysis to test each of our hypotheses with greater efficiency gains. We use three levels for our multilevel analyses: patient, physician-shift, and physician. The effect of transitioning from having a pooled queuing system to a dedicated queuing system remains robust to this model specification ($\alpha_2 = -0.18$, $p < 0.001$).

1.5 Discussion and Conclusions

Using 3.5 years of data from a hospital's ED, we find that patients experience shorter LOS when physicians work in a dedicated queuing system with a fairness constraint as opposed to a pooled queuing system with the same fairness constraint. Although we are unable to precisely test the mechanism for the shorter LOS in the dedicated system, we believe that the improved performance stems from strategic physician behaviors triggered by physicians' greater ownership over patient flow and the resources needed to smooth flow through the ED. This suggests that the flow management benefits associated with a dedicated queuing system with a fairness constraint may outweigh the variability-buffering benefits of a pooled queuing system. We consider, but find no empirical support for, alternate explanations for this reduction in LOS, such as changes in the provision of care or lower quality care.

We find evidence that physicians' strategic behaviors persist throughout the entire shift. In particular, examination of physicians' discharge rates in two-hour periods over the course of the shift shows that physicians exhibit a higher discharge rate when working in a dedicated queuing system as opposed to a pooled queuing system soon after beginning the shift. This increase in discharge rates is sustained throughout the remainder of the shift. In describing how the implementation of PAS increases physicians' ability to manage patient flow, one physician said, "Before PAS, the physician had no control or responsibility over getting the next patient into an empty bed. I often had idle time and had more than enough time to see more patients; I just couldn't get them to me from the waiting room. I wasn't in control so I didn't do much to get patient turnover to happen faster. Now, with PAS, I am responsible for getting my patients from the waiting room into my beds. I do this by making sure that tasks are being done so that I can discharge my current patients…. It changed the whole responsibility for patient flow from [the] one [internal triage] nurse onto me to manage my patients."

To quantify the impact of our findings, we calculate effect sizes. We find that moving from a pooled queuing system to a dedicated queuing system is associated with a 17% decrease in the difference in LOS between the main ED and RCA. For an average patient of ESI level 3 seen in the main ED by an average physician, this corresponds to a 39-minute decrease in LOS in the main ED relative to the RCA. This is a particularly meaningful difference in the context of a hospital's emergency room. With approximately 200 patients in the ED every day, this is roughly equivalent to an additional 130 patient-hours per day that are saved with the dedicated queuing system. Once we take into account the large costs associated with emergency room care, it becomes clear that the time and cost implications are substantial. If these findings are generalizable to other EDs, this would have significant practical implications for EDs across the country faced with large increases in patient volume accompanied by constrained budgets.

Nevertheless, it is important to consider the potential limitations of dedicated queuing systems. In systems with less homogenous patient populations, a dedicated queuing system with fairness constraints might result in imbalanced workloads among different care providers.

1.5.1 Theoretical Contributions

This paper contributes to the operations management literature on queue pooling in several ways.

Our paper is one of a few to use empirical data to examine the effect of queue management systems on wait times and service times. We find that when servers have ownership over patient flow and key resources, dedicated queuing systems with a fairness constraint are associated with *shorter* wait times and service times than pooled systems with a fairness constraint. Our findings illustrate the importance of accounting for the interaction between human behavior and queuing system design when predicting performance (Boudreau et al. 2003, Jouini et al. 2008). When queuing theory does not account for strategic server behavior, it suggests that pooling queues should result in shorter wait times even when fair routing policies are used (Armony and Ward 2010). In our study, we find that wait times are longer for the pooled system. Thus, our paper provides empirical support for prior analytical models that predict that human behaviors can reduce the benefits of using a common pool (Best et al. 2015, Cachon and Zhang 2007, Gilbert and Weng 1998, Hopp et al. 2007, Jouini et al. 2008, Wang et al. 2010). We are also able to add quantitative, empirical evidence to the debate that the benefits that arise from lean manufacturing's practice of assigning a specific person to service a specific stream of work outperforms the flexibility benefits from a pooled system (Spear and Bowen 1999). Our paper demonstrates how employees' willingness and ability to manage flow create an advantage for dedicated systems over pooled systems.

We speculate that queue pooling results in longer LOS because, in the pooled system, physicians do not feel completely responsible for patient flow because the internal triage nurse is responsible for moving patients from the waiting room to available beds. This result is similar to, but distinct from, Chan's finding (2015) that ED physicians work slower when they are assigned patients by a triage nurse than when physicians—collectively as a group—assign patients to physicians. Chan asserts that this "foot-dragging" behavior occurs in the nurse-managed system because physicians delay discharges to overstate their true workload to the nurse in hopes of avoiding being assigned another patient. The findings in the Debo et al. (2008) study are also driven by servers' misleading behaviors. Another mechanism in the literature that explains why dedicated queuing systems have faster service times than pooled systems is that managers can better supervise the smaller teams of workers that result from splitting up a large pooled system into a set of dedicated systems and a healthy competition emerges among the different dedicated systems (Jouini et al. 2008).

In contrast, we propose a different underlying mechanism for the improvement in throughput times: better flow management arising from strategic physician behaviors. In our study, a computer-automated RR routing policy fairly assigns patients to physicians both before and after the intervention. Thus, unlike physicians in Chan's study (2015), physicians in our study are not deliberately working slower to overstate their workloads. Furthermore, the fact that only a handful of physicians are working in this ED at any one time suggests that the Jouini et al. (2008) emphasis on the challenge of managing a large pool of employees is not what is driving our results. Also, physicians were not given any information about other physicians' average LOS, so competition is not the explanatory mechanism (Jouini et al. 2008). Instead, we propose that making a single physician—as opposed to a group of physicians—accountable for efficiently managing patient flow leads to a reduction in the wait time and LOS through better flow management practices.

Our findings build on the Schultz et al. (1998) study of the motivational impact of low inventory levels on production line workers' speeds. Schultz et al. (1998) find that low inventory motivates slower workers to speed up, enough to cancel the productivity loss due to the blocking and starving that occurs in low inventory production lines. We examine a different lever to increase workers' motivation: the queue structure of incoming jobs. We find that, when physicians work in a dedicated queuing system, they are able to attain shorter average LOS and wait times for their patients by managing their workloads more efficiently. We suggest that this may be because the dedicated system affords physicians a higher level of ownership over patient flow. We find that the motivational benefits of the dedicated queuing system outweigh the inefficiencies introduced by unpooling the queue. Thus, our study furthers the Schultz et al. (1998) finding by proposing that queue structure is another job design factor that interacts with human behavior in ways that can reverse predicted relationships between work system design and performance.

1.5.2 Implications for Practice

Our study has important implications for workplace managers and health care policy makers. Our findings suggest that managers of work settings with strategic servers should design work systems to mitigate behaviors that benefit the employee to the detriment of customers or the organization. One possible mechanism is to give strategic servers greater ownership and responsibility for managing their workflow and to route work evenly across all servers regardless of differences in work pace, which removes the benefit of working slower than one's peers. EDs may benefit from implementing dedicated, fair queuing systems in which patients are assigned to physicians immediately following triage. To our knowledge, this is not currently in place at most EDs; most EDs employ a pooled queuing system that assigns patients to physicians once placed in a bed. Thus, the potential for improvement is significant.

1.5.3 Limitations and Future Research

This study has limitations, and its results should be interpreted accordingly. First, we note the threat of omitted variable bias, common to many empirical models. While it would have been helpful to include more patient characteristics in our model, such as patient diagnoses or medical comorbidities, these data were protected information and not available for use. However, this is not an important threat to validity because patients are randomly assigned to physicians rather than by physician choice. This is supported by the fact that the average ESI level of patients seen by each physician is less than one standard deviation away from the average ESI level of all patients seen in the ED (*mean* = 3.33, *s.d.* = 0.64).

Second, our study is of a single hospital's ED and its response to a single intervention. The fact that our data come from a single ED makes it impossible for us to use another ED as the control in our difference-in-differences analysis. While we are confident that the RCA is a good control for our study, there would be advantages to using data from another ED with a similar patient population that did not implement the PAS system. We were unable to do this in our study because the PAS system was implemented in all EDs in the hospital's network. Though the generalizability of our findings is limited because we studied only one ED, we believe our findings have strong theoretical underpinnings. Nevertheless, future research could examine a larger sample of EDs to study a wider variety of routing policies and queue structures. Given that prior literature has found a variety of different mechanisms that may explain the shorter service times in dedicated systems, such research might enable greater clarity in which mechanisms are most powerful and under what conditions. In addition, these effects and suggested mechanisms could be studied in different empirical contexts for further theory development.

Third, our study raises the possibility that better flow management—arising from ownership over key resources—enables physicians in dedicated queuing systems to reduce their patients' average wait times and LOS. However, we are unable to precisely identify and test the mechanisms conclusively. Instead, we suggest these potential mechanisms based on interviews with physicians and observations of their practice patterns and leave it to future research to disentangle the mechanisms responsible for the reduced times.

Fourth, future research could consider how dedicated queuing systems affect patient and physician satisfaction, since changes in wait times and LOS may be associated with perceptions of fairness and the general satisfaction of both parties. These data are not available from the time period of our study, but have recently become more widely available.

Finally, implementing a dedicated queuing system is merely one way to try to attain the goal of shorter wait times and LOS in EDs. Future research should consider other mechanisms, such as financial incentives or interventions that leverage social pressure (Chan 2015). For example, do physicians increase their work rates when provided information about each other's average LOS? It may be possible to use a combination of interventions so that EDs can capture the benefits of pooling while simultaneously avoiding the slower service rates that seem to arise from queuing systems where responsibility for customers is shared across multiple servers.

1.5.4 Conclusions

Effectively using queue design to create both fairness and efficiency is an important opportunity for service organizations. While results may differ across different settings, the mechanisms through which changes in LOS occur may help shed light on improvement opportunities in other contexts. Our findings are especially timely and could have significant implications for health care delivery as EDs across the country contemplate ways to handle the anticipated increases in ED patient volume as a result of the recent health reform legislation (U.S. Senate 2010).

CHAPTER 2

Public Relative Performance Feedback in Complex Service Systems: Improving Productivity through the Adoption of Best Practices

2.1 Introduction

In service organizations, a key objective for managers is to create conditions that help workers improve their productivity without eroding service quality. One commonly employed approach for improving quality and productivity is standardized work (Spear and Bowen 1999). This approach has been shown to positively affect quality and productivity in a variety of industries, including health care (Andritsos and Tang 2014, Chandrasekaran et al. 2012, Shah et al. 2008), automotives (Ohno 1988), retail (DeHoratius and Raman 2008), banking (Frei et al. 1999, Staats and Gino 2012), and software (Narayanan et al. 2009, Staats et al. 2011). In each of these cases, standardized work leads to operational improvements by reducing variability in how tasks are executed. For example, checklists ensure that workers do not forget key steps or deviate from evidence-based treatments (Pronovost et al. 2006).

Though the benefits of standardized work have been documented, many processes in complex service organizations remain unstandardized, due to the challenges presented by inherent variability in customer needs and differences in worker behaviors that arise from worker discretion and imperfect monitoring (Klassen and Menor 2007, Tucker et al. 2007). In such settings, it may not be feasible to develop and implement a standardized protocol for each specific work task. Instead, managers may need to consider alternate ways to motivate and enable workers to better manage their *workflow* across all tasks. We draw on Georgakopoulos et al. (1995) to define workflow as a worker's selection and sequencing of a set of tasks necessary to accomplish the objectives for his or her customers. To illustrate, standardized work would specify that emergency department (ED) physicians order a beta blocker within five minutes of diagnosing a patient with a heart attack. In contrast, standardized workflow would specify that physicians order at the beginning of a patient's visit all pertinent tests and medications needed to treat the most probable diagnosis, rather than ordering them in a serial fashion.

One way to standardize workflow may be by publicly disclosing—among workers within an organization—relative performance feedback (RPF) on productivity metrics. Note, for brevity, we do not always specify that we consider cases in which RPF is provided on productivity metrics (as opposed to other performance metrics, such as quality), but this should be assumed. RPF compares an individual worker's performance to his or her peers within the organization (Blanes i Vidal and Nossol 2011). Nearly one-third of U.S. corporations provide RPF to their workers (McGregor 2006, Nordstrom et al. 1991). We distinguish RPF, which is feedback provided specifically to the members within an organization, from public reporting, which is performance data made available to those external to an organization. For a review of public reporting in health care settings, see Fung et al. (2008).

When RPF on a productivity metric is disclosed *publicly* among workers within an organization (i.e., with each worker's name displayed alongside his or her productivity metric), workers may be able to determine how to improve their productivity by identifying high-productivity coworkers who have developed best practices around workflow. In addition, workers may also be more motivated to improve their productivity because public RPF makes performance differences across workers more salient. This is in contrast to when RPF is disclosed *privately*, in which the same information about relative performance is provided but the identities of coworkers are kept anonymous. The potential benefits of public RPF on improving worker productivity may be especially relevant for complex service systems in which standardized work is not in place, due to the resulting variation in how workers approach their work. Despite the frequent use of public RPF, scant research has investigated its impact on worker productivity. Therefore, we ask the following research questions: *What is the effect of public RPF on worker productivity? Does this effect vary by whether the work is standardized, and what are the implications for service quality? What are the mechanisms through which public RPF affects worker productivity?*

In this paper, we explore these questions in the context of a hospital ED, which is a complex service organization with significant variability in patients' needs and inherent variability in processing times across physicians due to their ability to exercise discretion in carrying out their work (McCarthy et al. 2012). Here, certain work tasks are standardized, but many are not. For example, for patients presenting with symptoms indicative of a stroke, there are clear protocols for diagnosis and treatment in the ED. However, for patients presenting with abdominal pain, which could be indicative of a high severity condition, standardized protocols are typically not in place. Improvements in the management of workflow and associated reductions in variability are expected to have a significant positive effect on operational performance in this setting (Soremekun et al. 2011). Such improvements are particularly needed in EDs given the high demand for ED care (Pitts et al. 2010) that coexists with declining reimbursement rates (Morganti et al. 2013) and high rates of uncompensated care (Centers for Medicare & Medicaid Services 2002).

For our identification strategy, we leverage an exogenous change to the way RPF is provided to physicians at one of two EDs within the same health care system. During the initial time period of our study, physicians at both EDs were provided private RPF in the form of a ranked histogram about the median ED length of stay (LOS) of their patients—an important productivity metric in an ED setting. With this private RPF, each physician's identity remained anonymous and was represented with a code number that only he or she knew. Beginning in August 2010, physicians at one ED (but not the other) received the same RPF information in a public manner where each physician's name was listed next to his or her median LOS on the ranked histogram. Thus, with public RPF, physicians were able to identify high-productivity peers who could share best practices regarding workflow. Using a difference-in-differences approach, we first estimate the impact on worker productivity of changing from private to public RPF. We then assess how this effect varies by whether the patient's condition has a standardized work protocol. In addition, we examine the effects of this change on service quality. Finally, assuming that the hypothesized positive impact of public RPF on productivity exists, we consider three mechanisms that may explain this effect: motivation to be at the top of the performance distribution, motivation to avoid the shame of being at the bottom of the performance distribution, and the identification and diffusion of best practices around workflow.

Our results show that, on average, public RPF is associated with an 8.6% increase in physician productivity. We find that the increase in productivity is not statistically significant for patient conditions with standardized processes but *is* statistically significant for patient conditions without standardized processes. In addition, we find a slight decrease in the amount of care provided to patients and no significant reduction in clinical quality or patient satisfaction. Based on our analyses of heterogeneous treatment effects and changes in the variation in processing times, the explanation that is most consistently supported for why public RPF leads to improved productivity is that public RPF enables the identification and diffusion of best practices around workflow. Data from physician interviews and observations reinforce this conclusion.

This paper makes several contributions to both theory and practice. First, we contribute to the operations management literature on feedback (e.g., Bendoly 2013, Schultz et al. 1999) by examining the effects of public RPF in a field setting. To our knowledge, there have been few studies of public RPF, and most have been conducted in laboratory settings. This line of research is important both theoretically and practically because feedback provides an additional lever for influencing worker behavior that compliments operations management's typical focus on standardized work. Second, we illustrate that focusing on improving the management of *workflow*, as opposed to the standardization of *work tasks*, may be particularly useful when workers have discretion in how to carry out their work (Hopp et al. 2009). In doing so, we build on the operations literature that examines workflow management in health care settings (Dobson et al. 2013, KC 2014). Third, we illuminate conditions under which public RPF is most likely to lead to improved productivity. We identify the importance of providing such feedback on metrics that can be improved by the spread of best practices rather than simply reflective of differences in individual ability. We also find that providing public RPF on productivity metrics is more likely to result in productivity gains when workers are carrying out unstandardized work tasks as opposed to ones that are already standardized. Finally, we provide insight into the mechanisms through which public RPF has a positive impact on productivity.

2.2 Related Literature and Hypotheses

2.2.1 Productivity Improvement in Service Organizations

Operations management scholars have examined various approaches to improve productivity at the level of the individual worker. One commonly employed approach is standardized work (Spear and Bowen 1999), which has been discussed in the operations management literature since Taylor's (1911) seminal research in this area. Standardized work reduces process variability, which in turn helps improve productivity by reducing customer waiting times (Hopp and Spearman 2000, Kingman 1961) and facilitating learning through process control (Bohn 1995, Jaikumar 2005). The positive effects of standardized work on quality and productivity have been well documented in studies spanning several industries, including health care (Andritsos and Tang 2014, Chandrasekaran et al. 2012, Shah et al. 2008), automotive (Ohno 1988), retail (DeHoratius and Raman 2008), banking (Frei et al. 1999, Staats and Gino 2012), and software (Narayanan et al. 2009, Staats et al. 2011).

Of course, standardized work only helps improve productivity and quality if employees are compliant with the specified procedure. Research suggests that compliance with established procedures is often low, regardless of industry (Anand et al. 2012, Staats et al. 2015). For example, in pharmaceutical manufacturing plants, noncompliance with operational routines is identified in more than half of all inspections conducted by the Food and Drug Administration (Anand et al. 2012). Similarly, despite it being widely established that handwashing leads to improved patient outcomes, average compliance rates with handwashing guidelines are documented to be consistently below 50% (Centers for Disease Control and Prevention 2002). Electronic monitoring has been examined as a potential way to ensure greater compliance (Pierce et al. 2015, Staats et al. 2015), but this is not without challenges. In a paper examining the effectiveness of electronic monitoring to encourage greater hand hygiene compliance, Staats et al. (2015) finds that compliance rates fall *below* pre-monitoring levels after electronic monitoring is discontinued. This suggests that monitoring efforts should be implemented with care and sustained managerial commitment.

The many benefits of standardized work suggest that managers should strive to develop, implement, and ensure compliance with standardized work. However, in complex service systems with significant heterogeneity in customer needs, standardization may not always be feasible to develop and implement at the level of each specific customer type. Furthermore, service workers often have to prioritize among different customers who need service simultaneously, making it difficult to develop absolute rules about which tasks to perform and in which order (KC 2014, Tucker and Spear 2006, Wang et al. 2015). In such settings, it may be helpful to consider instead how workers can best manage their *workflow* across all their tasks. One approach may be to select the appropriate level of multitasking that allows for faster service rates without sacrificing quality (KC 2014). Prior research on service workers has found that workers who perform a variety of tasks enjoy productivity benefits over employees with repetitive tasks (Narayanan et al. 2009, Staats and Gino 2012). Another approach may be to manage interruptions by adhering to optimal prioritization policies between new customers and those already in the system (Dobson et al. 2013).

Given the context-dependent nature of these decisions around the management of workflow, an effective approach to identifying best practices in a given setting may be to have the frontline workers define these approaches themselves. Yet, with each worker employing a different set of practices, which are not always fully transparent to others, these practices may be difficult to identify and isolate. Furthermore, employees might be inclined to adhere to their own approaches, even if their methods are non-optimal (Berwick 2003). How might managers help motivate and enable workers to better manage their workflow across all tasks?

2.2.2 Relative Performance Feedback in Service Organizations

One way to help identify best practices around workflow—especially those that are developed by the frontline workers themselves—may begin with identifying the most productive workers in an organization. Managers can enable workers to identify their most productive peers by publicly disclosing individual-level RPF on productivity metrics among workers within an organization. By identifying these workers and their approaches for managing workflow, other workers within the organization may be able to identify a set of best practices to adopt and be motivated to change their practices around workflow.

Even when privately disclosed, RPF leverages the effects of social comparison by providing workers with visibility into the full distribution of all workers' performance. Social comparison has been shown to have powerful effects on worker behaviors. Some workers find social comparisons to be motivating, especially when monetary incentives are linked to relative performance. When provided social comparisons, workers tend to increase their productivity, as has been shown by prior studies in both laboratory (Charness et al. 2014, Kuhnen and Tymula 2012) and field settings (Blanes i Vidal and Nossol 2011, Cowgill 2015). However, social comparisons can lead to negative effects as well, in which workers exert less effort as opposed to more (Ashraf et al. 2014, Barankay 2012). For example, in a field study of furniture salespeople, Barankay (2012) finds that low-ranked workers exert less effort when presented with feedback on rank due to the effect of a low ranking on an individual's self-image. In another field study with individuals training to become Zambian health workers, Ashraf et al. (2014) finds that social comparisons negatively affect the performance of low-performing trainees, as measured by their test scores on a training course exam. Perhaps even worse, some research finds that workers receiving unfavorable social comparisons are likely to engage in deceptive behaviors to artificially inflate their reported performance (Edelman and Larkin 2015, Moran and Schweitzer 2008). Although this may improve a worker's reported relative performance measure, it has negative implications for the organization's performance and trust among coworkers (Dunn et al. 2012).

Because of these varying potential implications, care is needed when deciding how to implement RPF. First, RPF can be disclosed either privately or publicly amongst the workers. With private RPF, a worker cannot identify which individual coworker exhibited a specific performance level, and the worker knows that likewise her coworkers cannot link her own performance result to her. In contrast, with public RPF, a worker can link each performance level to an individual coworker and knows that her coworkers can also see how she herself has performed. Second, managers can link monetary incentives and compensation to relative performance. Perhaps unsurprisingly, many studies find that RPF motivates workers to work harder when monetary incentives are linked to relative performance (Tafkov 2013). Some studies also find that, even without monetary incentives, RPF can lead to increases in productivity by stoking intrinsic competition (Blanes i Vidal and Nossol 2011, Roels and Su 2014). In this paper, we focus on privately versus publicly disclosing RPF without any monetary incentives for productivity. If this approach is successful at increasing productivity, it may be easier and more cost-effective to implement than having to financially incentivize workers for their productivity.

2.2.3 Private Versus Public Relative Performance Feedback

Much of the prior research on RPF has focused on examining the effects of *private* RPF, either in the presence or absence of monetary incentives. To our knowledge, Tafkov (2013) and Hannan et al. (2013) are the only studies that examine the relationship between public RPF and worker performance. In a laboratory setting with participants solving multiplication problems for pay, Tafkov (2013) finds that participants solve more problems correctly when they are provided public as opposed to private RPF. Although this study finds improvements in performance with public RPF, it is unclear whether such effects would hold in a field setting with complex tasks, various approaches for carrying out these tasks, and workers who know one another. Hannan et al. (2013) builds on these results in a multi-task environment and shows that when workers can allocate their time across two different types of tasks, RPF leads workers to distort their time allocation away from the optimal allocation (50/50 in the experiment) to favor the task on which they performed better in an earlier round. The authors find that the effort distortion effect is stronger with public

RPF than private RPF, and persists even if it decreases their remuneration for the experiment. These studies highlight that public RPF provides strong motivation for workers to change their behaviors.

In this paper, we examine how public RPF affects workers' productivity, on average, compared to private RPF when workers are operating in a complex service organization. From the perspective of the worker, there are two key differences between public versus private RPF: (a) the knowledge that coworkers can see his or her own level of performance, and (b) the knowledge of each coworker's level of performance. The first of these two elements may motivate workers to increase their productivity more than when provided private RPF, due to a heightened level of social pressure (Mas and Moretti 2009, Schultz et al. 1999). Receiving public RPF is likely to have a greater effect on productivity than private RPF due to the increased salience of being monitored (Staats et al. 2015). This effect manifests not only when the worker's productivity level is visually observable to others (Schultz et al. 1999) or to one's high-performing peers in particular (Mas and Moretti 2009), but at all times. The second element may equip workers with the necessary knowledge to identify top performers and their best practices around workflow in a credible way, which may further improve workers' productivity on average. This leads us to our first hypothesis:

Hypothesis 1 (H1): Public RPF leads to a decrease in processing time, on average.

Assuming that public RPF does lead to an improvement in worker productivity, we are interested in assessing whether this productivity gain is particularly salient when standardized work is not in place and there is variation in how workers approach their work. This helps to determine whether providing public RPF may be a useful tool for improving productivity in complex service systems in which standardized work is difficult to implement—the type of setting in which we are most interested in this study.

Take for example a hospital ED. In this setting, there is no standardized work protocol for the diagnosis and treatment of patients presenting with symptoms of abdominal pain. Physicians develop their own workflow that fits their own needs and the needs of the other patients under their care. This results in significant variation in workflow across physicians, leading to lower average worker productivity than would occur if all workers employed the best practices of their high-productivity peers.

At the same time, some medical conditions have standardized work protocols. For example, there is a clear protocol for diagnosis and treatment in the ED for patients presenting with symptoms indicative of a heart attack or stroke. When processes for standardized work are already in place, there may not be much room for improvement from adopting the best practices of high-productivity peers. Thus, we expect public RPF to be particularly helpful in improving productivity when standardized work is not in place:

Hypothesis 2 (H2): The decrease in processing time resulting from a shift from private to public RPF *is greater when standardized processes are not in place compared to when they are.*

The hypothesized benefit of public RPF on worker productivity would be called into question if it were accompanied by a decrease in quality. After all, managers seek to improve productivity *without* eroding service quality. Consequently, we also examine, but do not hypothesize about, the impact of public RPF on service quality.

There are several potential mechanisms that may result in increased productivity when switching from private to public RPF. First, workers may be motivated to be at the top of the relative performance distribution. Past research suggests that individuals care about the prestige of being a top-ranked performer (Charness et al. 2014). Furthermore, this concern about high rank is innate (Zizzo 2002). Given the greater utility experienced when attaining a higher rank on a relative performance distribution, workers who have access to public RPF may exhibit ahead-seeking behaviors, which is when they seek to outperform others by increasing their own performance (Roels and Su 2014). Workers who are already toward the top of the distribution are most likely to exhibit such behaviors, since the additional effort required to remain in the top is modest compared to what a mid- or bottom-performing worker would need to exert (Boudreau et al. 2016). That said, it is possible that motivation to be at the top is no stronger when RPF is publicly as opposed to privately disclosed. This may be because workers are intrinsically motivated to be at the top even if their peers are ignorant of their high level of performance. Furthermore, workers may care about their ranking even if it remains private because their managers know how their performance compares with their peers' performance (Nagin et al. 2002).

A second potential mechanism is workers' motivation to avoid being at the bottom of the relative performance distribution. Bottom-ranked workers may exert greater effort and seek to improve their productivity in order to avoid the shame or embarrassment of being in last place. When RPF is disclosed publicly as opposed to privately, last-place aversion becomes more salient for these workers, as their bottom-ranked status becomes disclosed to others (Kuziemko et al. 2014). Nevertheless, it is also possible that these workers become discouraged by upward social comparisons, as has been documented in the context of retirement savings (Beshears et al. 2015), health worker trainings (Ashraf et al. 2014), fruit picking (Bandiera et al. 2013), and furniture sales (Barankay 2012). Thus, it is unclear whether bottom-ranked workers will exert more or less effort to move up in the rankings, and whether the motivation will be stronger with public or private RPF.

A third potential mechanism is the identification and diffusion of best practices that is enabled by public RPF. Public RPF makes the top performers easily identifiable, which in turn enables workers to learn best practices either by observing how the top performers work or by directly asking them to share their productivity tips. Public RPF also helps overcome an important barrier to best practice diffusion, which is the perceived credibility of the information. When best practices are shared while RPF is privately disclosed, the "anonymous" information may lack the credibility necessary to change behaviors. Public RPF thus reduces barriers to both the identification and diffusion of best practices, enabling and motivating workers to seek out reliable ways to improve productivity (Szulanski 1996). This identification of top performers and their best practices enables peer-based learning, which has been found to be more important for productivity improvement than learning by doing (Chan et al. 2014). In a study of department store salespeople, Chan et al. (2014) finds that workers learn from each other through direct observation and active teaching when they are working on the same shift alongside one another. Similarly, Kuziemko et al. (2014) finds that top-ranked workers are willing to engage in behaviors to help low-ranked people improve their performance.

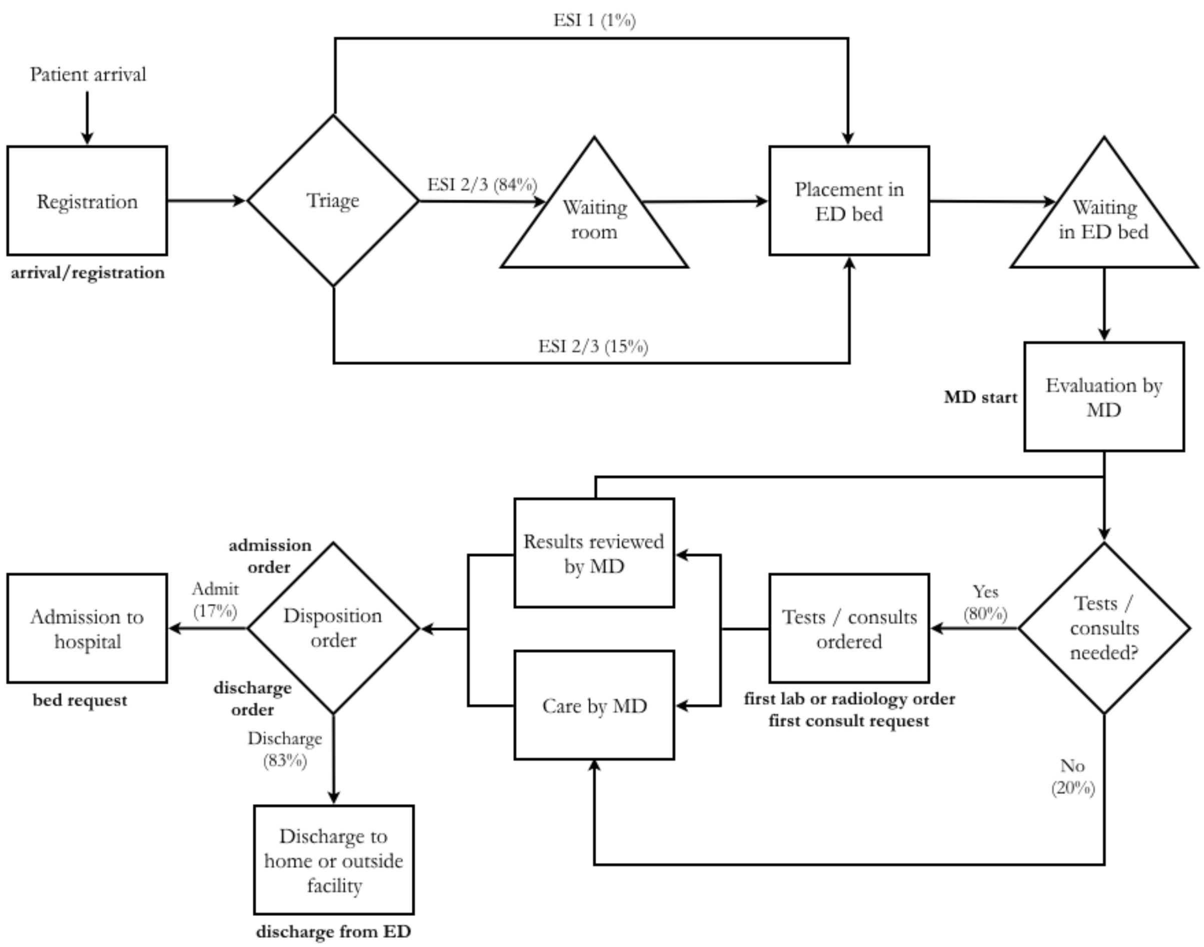
These three potential mechanisms need not be mutually exclusive of one another in explaining how public RPF leads to improvements in productivity. In fact, it is likely that each of these mechanisms contributes to some extent to the overall effect of public RPF on worker productivity. While our primary objective in this study is to examine the effect of public RPF on worker productivity and the degree to which this effect varies by whether standardized work is already in place, we also explore which of these mechanisms may serve as the leading explanation for what may be driving the effects of public RPF on worker productivity.

2.3 Setting and Data

2.3.1 Research Setting

We examine our research questions in the context of a hospital ED. In this setting, there is significant variability in patient needs and inherent variability in processing times across physicians (McCarthy et al. 2012). In addition, while certain work tasks are standardized, many are not.

The ED is a setting in which a key objective is to attain shorter waiting times and LOS while simultaneously minimizing adverse outcomes (McClelland et al. 2011). When LOSs are longer, EDs become crowded, which prevents new arrivals from being seen in a timely manner (i.e., longer waiting times), and ultimately increases the risk of adverse patient outcomes (Guttmann et al. 2011, Schull et al. 2015). In addition, longer waiting times and LOS are associated with lower levels of patient satisfaction (Spaite et al. 2002).



Notes. For process measures with a corresponding time stamp in the electronic health record, we note the time stamp in bold.

Figure 2.1. Schematic representation of the patient flow process in an ED

Figure 2.1 is a schematic representation of a typical patient journey through the ED. A patient arrives either via ambulance or as a walk-in, at which point an ED clerk registers the patient. A triage nurse then obtains vital signs, collects the chief complaint, and assigns a 5-level Emergency Severity

Index (ESI) triage category with 1 being the most urgent and 5 being the least urgent (Gilboy et al. 2011). Unless the patient is of ESI level 1 (in which case the patient is immediately taken to the resuscitation room) or there is an available ED bed, the patient returns to the waiting room until an ED bed becomes available. Once the patient is placed in the bed, he or she is evaluated by the physician. After this initial evaluation, the physician places orders for lab tests, radiology tests, specialty consults, or therapies to be conducted. Note, it is possible to deviate from this flow pattern as tests can be ordered by a nurse or before the physician sees the patient (Batt and Terwiesch 2015). However, in our setting, this occurs less than 5% of the time, and tests are nearly always ordered by physicians. Once these are carried out and results of tests and consults are available, the physician assesses the patient's need for further care and makes a disposition decision. This can be to admit the patient to the hospital or to discharge the patient home or to an outside facility.

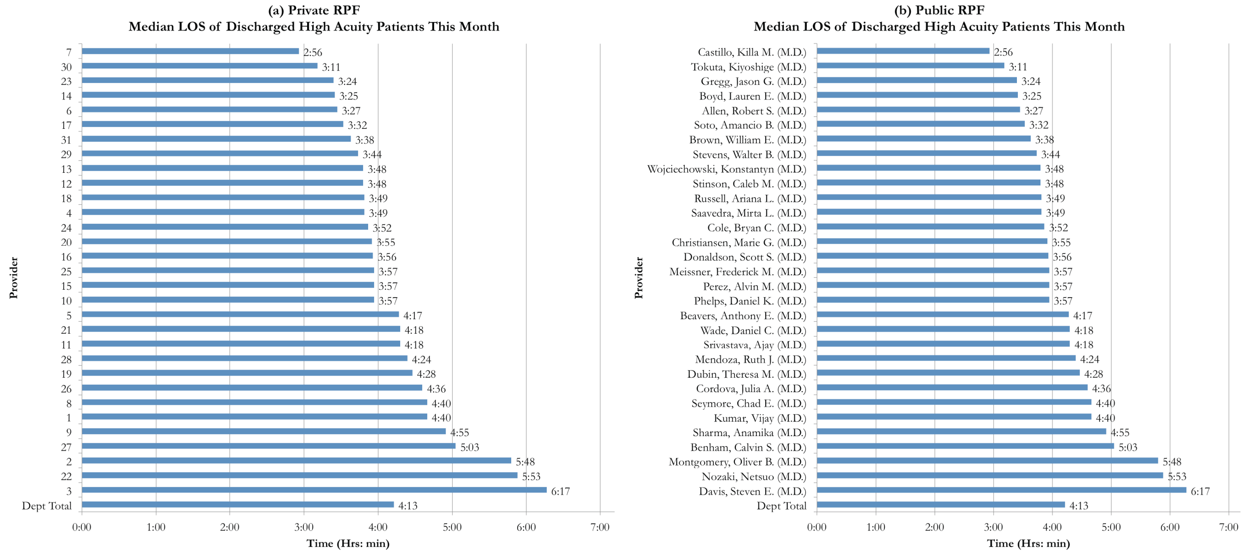
For this study, we leverage an exogenous change to the way RPF was disclosed to physicians at one of two EDs. We describe this change in greater detail in chapter 2.3.2. The two EDs, which we refer to as Treatment ED and Control ED, both belong to the same not-for-profit, integrated health care system in California. They are located less than 15 miles apart and serve a similar catchment area. Each has a large patient load, with more than 70,000 patient encounters each in 2010.

Though the medical centers within the health care system are interdependent and cooperative, they are each independently managed and have the ability to design their own practice structures and initiatives. The physician practice group that staffs each ED is specific to the ED; in other words, different physicians staff the two EDs. Nevertheless, Treatment ED and Control ED have four key commonalities. First, both EDs employ physicians on a salaried basis, with opportunities for additional shifts but no additional compensation for ordering more tests or working more hours on a shift than scheduled. This enables us to disentangle the effect of public RPF on productivity from the effect of a monetary incentive to outperform one's coworkers. Second, to ensure fairness among physicians in terms of their average workload, both EDs use a round robin routing policy that fairly allocates patients to physicians independent of physician work speed or idle time (Patel and Vinson 2005). Third, prior to leaving at the end of the shift, physicians at both EDs are required to discharge, or at least complete a care plan for, each of the patients assigned to them. To make this possible, physicians have two to four hours (depending on the scheduled shift length) of protected time at the end of the scheduled shift during which they are not assigned any new patients. Fourth, both EDs have a fast track to which less urgent and non-urgent patients (ESI levels 4 and 5) are routed for their care. On a given shift, physicians work in either the main ED or the fast track.

2.3.2 Relative Performance Feedback at Treatment ED and Control ED

Prior to August 2010, both Treatment ED and Control ED provided private RPF to their respective group of physicians and encouraged physicians to reduce their patients' LOS. At each ED, private RPF was presented in the form of a ranked histogram of the median LOS of the patients of a given acuity level (e.g., ESI levels 1, 2, and 3) treated by each physician in that facility during a given period of time. Each physician had his or her own bar on the histogram and was identified by a code number (see Figure 2.2). Physicians knew their own numbers, but not the numbers for other physicians. Thus, the histogram provided each physician with information about his or her own ranking among all physicians in the ED, but did not provide any insight into which specific other physician was a top- or bottom-ranked physician.

At Treatment ED, there were two LOS metrics presented to physicians in monthly reports: (a) the median LOS of all high acuity patients (ESI levels 1 – 3) who were subsequently discharged and (b) the median LOS of all low acuity patients (ESI levels 4 and 5). At Treatment ED, these metrics were presented to all physicians at the monthly staff meeting. During the "State of the ED" update portion of the meeting, Treatment ED's Chief of Emergency Medicine presented and discussed



Notes. Each ranked histogram reports, by physician, the median LOS of all ESI level 1 through 3 patients over a one-month period who were subsequently discharged. The examples are constructed using actual data from Treatment ED. For the public RPF report, names have been disguised using a fake name generator.

Figure 2.2. Example of monthly report with private versus public RPF

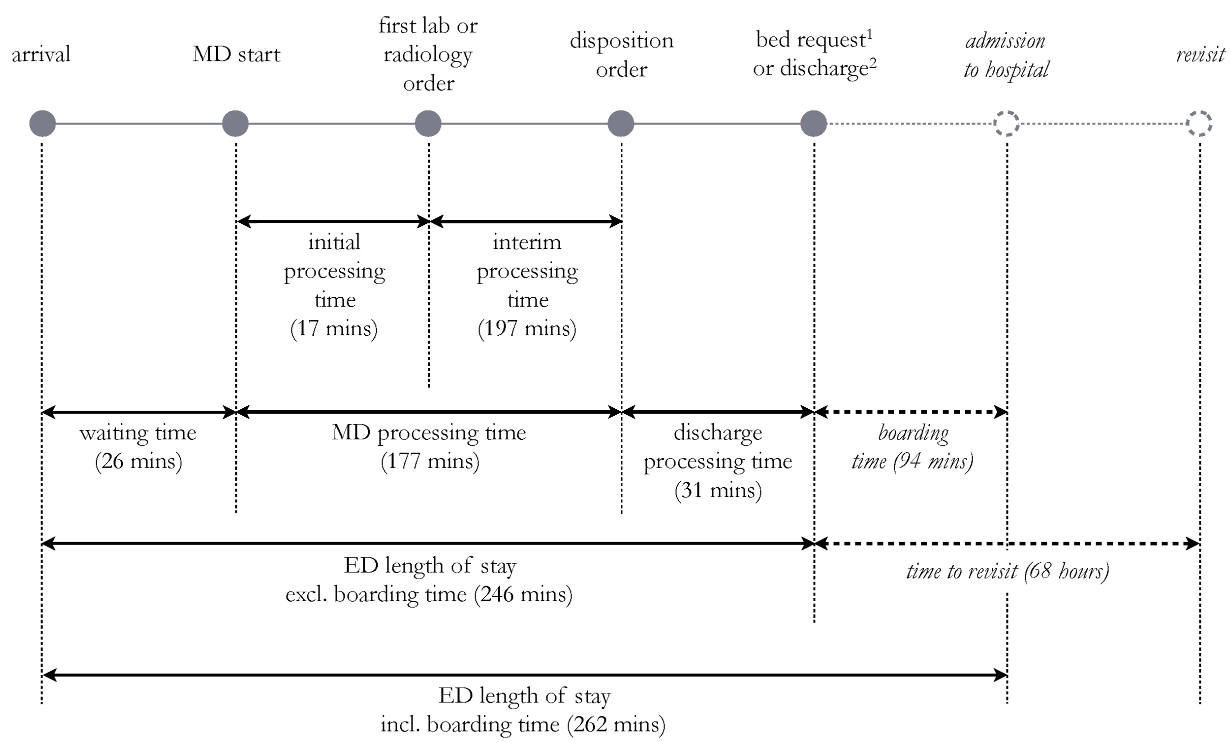
these productivity metrics. In particular, the Chief highlighted the metrics of top-ranked physicians (i.e., physicians with the shortest median LOS), although she did not identify these physicians by name. At the conclusion of the staff meeting, the ranked histogram report was posted in the staff lounge until a new report replaced it the following month. The same two LOS metrics were also presented to physicians at Control ED. Rather than at monthly meetings, these metrics were presented to each physician at Control ED twice a year at a one-on-one performance evaluation meeting with the Chief or Assistant Chief of Emergency Medicine. At both Treatment ED and Control ED, the Chiefs did not use the information to punish or shame low performers.

In August 2010, only Treatment ED transitioned to providing public RPF (see Figure 2.2). This change was implemented after Treatment ED's Chief saw an example of non-blinded productivity metrics of sales clerks being provided at a large department store. The motivation to change from private to public RPF was therefore an exogenous shock to the physicians. With public RPF, the monthly reports at Treatment ED included the same metrics as before, but now each bar in the ranked histogram was accompanied by a physician's name in lieu of a code number. Thus, with public RPF, each physician at Treatment ED could see his or her own productivity metric and how he or she performed relative to all other Treatment ED physicians, as well as identify high performers from whom he or she could learn best practices regarding workflow. During the State of the ED updates at monthly staff meetings, Treatment ED's Chief continued to highlight top-ranked physicians, now sometimes by name. In addition, top-ranked physicians—whose identities were no longer anonymous—were encouraged to share efficiency tips at the meeting, and all physicians were encouraged to shadow top-ranked physicians to learn best practices. Two main efficiency tips emerged from these meetings: (a) ordering lab and radiology tests as early in the care delivery process as possible (efficiency tip 1) and (b) beginning the discharge instructions and encounter note as soon as possible after the initial patient evaluation so that it takes less time to complete this documentation once test results are available (efficiency tip 2).

At Control ED, no changes were made to the reporting of private RPF throughout the study period. Note, throughout the study period, Control ED maintained a page on its internal website with efficiency tips for reducing patient LOS. These tips were solicited by Control ED's Chief and posted anonymously. The two efficiency tips that emerged from Treatment ED's staff meetings were among the efficiency tips listed on Control ED's internal website. Throughout the study period, no other changes were made with respect to the processes and resources in the high acuity area at only one ED and not the other.

2.3.3 Data

The data for our main analyses come from the electronic health records at Treatment ED and Control ED. Our data are de-identified and consist of all patient encounters of ESI levels 1, 2, and 3 from January 2009 to December 2011. The data include, but are not limited to, patient encounter-level information regarding the patient's time of arrival and departure, ESI level, attending physician, disposition, and time stamps for several process measures of the patient flow through the ED (see Figure 2.3). In our data, the attending physician of record is the physician who is initially assigned to the patient, even if the patient is handed off to an oncoming physician at the conclusion of the shift. Note, at both EDs, the initial physician is responsible for completing a care plan for each patient who is handed off to an oncoming physician. We exclude patients with no attending physician or ESI level listed on their record, and patients who had a LOS of less than one minute. Altogether, we



[1] For patients admitted to the hospital.
[2] For patients discharged to home or to an outside facility.
Notes. Time durations (in parentheses) are mean times of high acuity patient encounters at Treatment ED and Control ED from January 2009 to December 2011. Subcomponents of time durations do not necessarily add up to the larger component because not all patients experience each process (e.g., 20% of patients do not receive a lab or radiology order, 83% of patients are not admitted to the hospital and thus do not experience boarding time). N = 279,025 patient encounters.

Figure 2.3. Patient flow in the ED

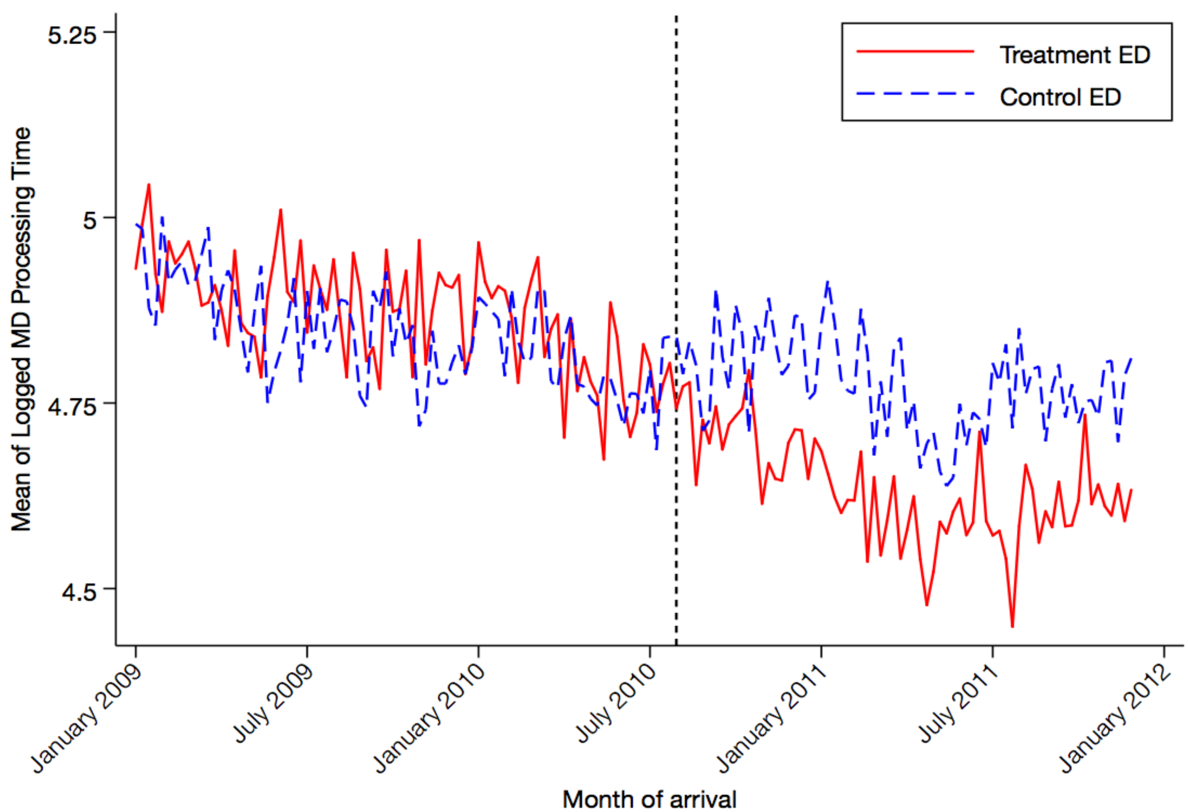
exclude 1,832 of 303,014 observations or 0.6% of the overall sample. For our analyses, we also exclude data from August 2010 to account for a washout period because the exact date of the month when public RPF was made available to physicians is unknown. In addition, we limit our sample to the patients seen by physicians whose home facility was the facility to which the patient presented. Physicians who work in a facility other than their home facility tend to be those based at another ED within the health care system who are brought in to cover a portion of a shift when the facility's own physicians are not able to staff the ED (e.g., during monthly staff meetings). The resulting final sample consists of 279,025 patient encounter-level observations.

We link the electronic health record data with data from the health care system's patient satisfaction survey. This survey is designed, developed, and administered by the health care system to all patients who have been treated in one of its EDs. Surveys are distributed within 72 hours of their visit by email or regular postal mail, and the system-wide response rate in 2010 was 25%. The survey assesses patient satisfaction with regard to interactions with the physician, other health care providers (e.g., nurse, other staff), the ED overall, and ancillary services (e.g., pharmacy, laboratory, radiology). For our analyses, we focus on three survey questions that measure patients' satisfaction with their physician: (1) the physician's skills and ability, (2) the patient's confidence that the physician provided the care and services the medical condition required, and (3) how well the physician listened and explained what was being done and why. We also obtain data on two general experience measures: (1) the total time spent in the ED and (2) how well the patient's needs were met. Each of the measures are rated using a 5-point Likert scale, ranging from 1 = Poor to 5 = Excellent.

In addition, we collected data from semi-structured interviews with 41 ED physicians (21 at Treatment ED and 20 at Control ED) of varying levels of productivity. These interviews were conducted by the first and second authors in July 2015, at which time Treatment ED physicians were still receiving public RPF reports and Control ED physicians were still receiving private RPF reports. Interviews addressed physicians' management of workflow to reduce LOS and their experiences with the RPF system in place. In addition to several other questions, we asked each physician to list the different ways in which they manage their workflow to reduce patients' LOS without hurting quality; what information they take away from the RPF reports; how the RPF reports have affected their workflow, if at all; and whether they are motivated to be towards the top of the relative performance distribution. Each interview was recorded and transcribed. The first author reviewed all interview transcripts to identify illustrative quotes for each of the interview questions relevant to our analyses. Using a survey scale, we also asked each physician to indicate the extent to which he or she felt ashamed, deserving of criticism, stupid, self-conscious, or embarrassed when seeing the RPF reports. These terms collectively comprise an index for shame (Brown et al. 2009), and were interspersed among other terms capturing positive affects. Terms originally come from the Positive and Negative Affect Schedule (Watson et al. 1988) and the Personal Feelings Questionnaire (Harder and Lewis 1987). Note, we are unable to link these interview data with the electronic health record data per the Institutional Review Board's requirement to de-identify all data. The first and second author also observed the workflow of two physicians each at Treatment ED and Control ED, respectively: one physician who is typically top-ranked and one physician who is typically bottom-ranked. The relative ranking of each physician was not revealed until after the observations had been completed.

2.4 Empirical Strategy

We rely on a difference-in-differences approach to estimate the causal effect of public RPF on physicians' productivity. By measuring the difference in differences between Treatment ED and Control ED over time, this approach allows us to control for characteristics unobservable to the researcher that may also impact performance (Imbens and Wooldridge 2009). The implementation of public RPF constitutes our treatment, which was introduced in August 2010 as an exogenous shock to physicians at Treatment ED. Because this approach accounts for differences in baseline levels of physician productivity across the two facilities, it adjusts for any potential differences that may arise due to the difference in the frequency with which RPF reports were distributed at Treatment ED and Control ED throughout the study period.



Notes. This figure depicts trends in raw data that are unadjusted for covariates, physician fixed effects, or time trends. Data are collapsed to the week level for ease of presentation. Vertical dashed line indicates the time when public RPF was implemented at Treatment ED (August 2010).

Figure 2.4. Mean of logged MD processing time of high acuity patients

To justify the use of this methodological approach, we first examine a key assumption: whether MD processing time exhibits parallel trends before the intervention at Treatment ED and Control ED (Abadie 2005). We address this assumption by testing the difference in trends at Treatment ED and Control ED in the pre-intervention period using monthly time trends. The results support the assumption of parallel trends, as the difference in trends is not statistically significant at conventional levels ($p \approx 0.51$). This assumption is also visually supported by Figure 2.4, which shows roughly parallel trends between unadjusted logged MD processing times at Treatment ED and Control ED in the 19 months prior to the introduction of public RPF at Treatment ED.

2.4.1 Effect of Public RPF on Physician Productivity

We begin by examining the change in physician productivity when RPF is disclosed publicly as opposed to privately. As a proxy for productivity, we measure MD processing time, which begins with the time the physician commences care and ends with the disposition order for admission or discharge (see Figure 2.3). We estimate the following fixed-effects log-linear model at the patient encounter level:

$$\ln\left(MDProc_{ijt}\right) = \beta_0 + \beta_1 Treat_{ij} \times Post_t + \delta \mathbf{X}_{ijt} + \theta_t + \alpha_j + \varepsilon_{ijt} \tag{2.1}$$

In Equation (2.1), i indexes each patient encounter, j indexes each physician, and t indexes time in month-year periods. $Treat_{ij}$ equals one for patients seen by a physician at Treatment ED and zero for patients seen by a physician at Control ED. $Post_t$ equals one after public RPF was implemented (beginning September 2010) and zero before (up to July 2010). We set $Post_t$ to missing for August 2010 to account for a washout period. $\mathbf{X}_{ijt}$ is a vector of control variables that includes time-varying patient- and ED-level covariates known to impact the set of outcome measures we employ. Specifically, to adjust for heterogeneity across patient types, we control for patient age, gender, and ESI level. To account for the inverted U-shaped relationship between ED workload and worker performance (KC and Terwiesch 2009, Tan and Netessine 2014), we include both a linear and quadratic term for the total number of patients in the ED at the beginning of each patient encounter. To account for time-varying levels of staffing in the ED, we include the total number of physicians working on shift, with shifts operationalized as morning (7am-2pm), afternoon (3pm-10pm), and overnight (11pm-6am) shifts. In addition, we control for the time of the day and the day of the week of the patient encounter, which may have implications for the availability of resources external to the ED. θ_t accounts for time trends by controlling for the month-year of the observation. The physician fixed effect α_j allows us to control for time-invariant aspects of physicians and controls for all between-physician variance such that our model explores within-physician variance. ε_{ijt} are heteroskedasticity-robust standard errors clustered by physician, which addresses the serial correlation problem and relaxes the assumption that standard errors are identically distributed and independent of each other (Bertrand et al. 2004, Wooldridge 2010). Each of the main effects for $Treat_{ij}$ and $Post_t$ are omitted due to the former being perfectly collinear with the physician fixed effects (since each physician only works at either Treatment ED or Control ED) and the latter being perfectly collinear with the month-year fixed effects (since $Post_t$ always equals zero in the month-years before August 2010 and one in the month-years after August 2010).

To test H1, our coefficient of interest is β_1. The term β_1 represents the difference in logged MD processing time between patients presenting to Treatment ED and Control ED before and after the introduction of public RPF at Treatment ED. Therefore, β_1 captures the impact of moving from a system in which RPF is disclosed privately to one in which it is disclosed publicly. We log $MDProc_{ijt}$ after adding one because the distribution of this variable is right-skewed.

2.4.2 Standardized Work as a Moderator of the Effect of Public RPF on Physician Productivity

To examine if the effect of public RPF on physician productivity varies by whether work is standardized, we repeat the estimation of Equation (2.1) on two subsamples of the data: (a) patients who received a diagnosis of a heart attack (i.e., acute myocardial infarction (AMI)) or stroke, and (b) all others. We isolate AMI and stroke as the two conditions for which standardized processes are in place based on discussions with ED physicians and administrators who indicate that these were the only two conditions for which standardized protocols existed and were adhered to during the study period.

In testing H2, our coefficient of interest is β_1. We predict β_1 will not be statistically significant with the sample of patients diagnosed with AMI or stroke, and will be negative and significant with the sample of all other patients.

2.4.3 Effects of Public RPF on Care Intensity, Clinical Quality, and Patient Satisfaction

Given the importance of finding the optimal balance between operational efficiency and quality in service settings like the ED (Anand et al. 2010, Batt and Terwiesch 2015, Hopp et al. 2007), we also explore whether there are any negative implications for care intensity, clinical quality, and patient satisfaction that result from publicly disclosing RPF. Using the available disposition data, we construct the following four binary indicators as proxies for care intensity: whether a lab test was ordered, whether a radiology test was ordered, whether a specialty consult was requested, and whether the patient was admitted to the hospital. Due to data limitations, we are unable to capture the number and type of lab tests, radiology tests, or specialty consults that were placed. We also use the following two binary indicators as proxies for clinical quality: whether the patient died in the ED and whether the patient returned to an ED within the health care system's network within 72 hours.

Instead of limiting our definition of revisit to a patient's subsequent visit to the same facility as the initial encounter, we use a broader definition that captures a subsequent visit to any ED within the health care system's network to account for the fact that patients are not limited in their choice of which ED to visit. To assess the change in care intensity and clinical quality associated with the implementation of public RPF, we estimate the following fixed-effects logit model:

$$\ln\left[\frac{\Pr\left(Q_{ijt}=1\mid X_{ijt}\right)}{1-\Pr\left(Q_{ijt}=1\mid X_{ijt}\right)}\right]=\rho_0+\rho_1 Treat_{ij}\times Post_t+\delta\mathbf{X}_{ijt}+\theta_t+\alpha_j \tag{2.2}$$

In Equation (2.2), Q_{ijt} equals one if patient encounter i with physician j at time t involved a lab test (or a radiology test, or a specialty consult, or a hospital admission, or an in-ED death, or an ED revisit, respectively), and zero otherwise. All other variables remain the same as in Equation (2.1). Monte Carlo studies, like Greene (2004), have shown bias due to the incidental parameters problem is large when the number of observations per group is small, and this bias declines quickly as the number of observations per group increases. Greene's (2004) conclusions were based on simulations with the length of the panel T ranging from 2 to 20. In this study, the number of observations per physician, T_j, has an average value of 1,000. Since each physician has many observations, the bias due to the incidental parameters problem is likely to be quite small in this setting. Nevertheless, we repeat our estimation using a linear probability model and note the lack of substantive difference in our findings when using either of these specifications. To account for the nonlinearity introduced by the logit model, we obtain the average marginal treatment effect by calculating the marginal effect for each observation in the data and then averaging these results.

To assess potential changes in the level of patient satisfaction, we use data from the patient satisfaction survey described in chapter 2.3.3. Using each of the three measures of patients' satisfaction with their physician and the two general experience measures as the dependent variable, respectively, we estimate a fixed-effects ordered logit model to account for the ordinal nature of the dependent variables. This ordered logit model takes a similar form as Equation (2.2), but allows the dependent variable to take values ranging from 1 to 5.

2.4.4 Consideration of Potential Mechanisms

As noted in chapter 2.2.3, there are three potential mechanisms that may lead to an increase in worker productivity when switching from private to public RPF: motivation to be at the top of the relative performance distribution, motivation to avoid being at the bottom of the relative performance distribution, and the identification and diffusion of best practices that is enabled by public RPF. Empirically separating these explanations is a challenging task given they need not be mutually exclusive of one another. Acknowledging this, we explore which of these mechanisms may serve as the leading explanation, rather than trying to estimate the extent to which each independently explains the hypothesized improvement in worker productivity.

For this, we examine (a) heterogeneous treatment effects by a physician's ranking on the relative performance distribution and (b) between- and within-physician variation in initial and interim processing times, respectively. For the latter, we focus on the variation in initial and interim processing times (as opposed to total MD processing time) because each directly corresponds to the specific portion of processing time that would be affected by the adoption of efficiency tips 1 and 2, respectively (see Figure 2.3). Here, initial processing time begins with the time the physician commences care and ends with the first lab or radiology order. Interim processing time begins with the first lab or radiology order and ends with the disposition order for admission or discharge.

Heterogeneous Treatment Effects. Examining heterogeneous treatment effects of public RPF on physician productivity allows us to assess whether the implementation of public RPF has a differential effect on the productivity of physicians who were initially top-ranked or bottom-ranked, respectively. If the leading explanation is the motivation to be at the top of the RPF distribution, we would see a differential increase in productivity among top-ranked physicians. In contrast, if the leading explanation is either the motivation to avoid the shame of being bottom-ranked or identifying and diffusing the best practices of top-ranked physicians, we would see a differential increase in productivity among bottom-ranked physicians as a result of public RPF implementation.

To examine the heterogeneous effects of public RPF on the productivity of top-ranked versus bottom-ranked physicians, we estimate the following fixed-effects log-linear model at the patient encounter level:

$$\begin{aligned}\ln\left(MDProc_{ijt}\right) = {} & \lambda_0 + \lambda_1 T3_j \times Post_t + \lambda_2 B3_j \times Post_t + \lambda_3 Treat_{ij} \times Post_t \\ & + \lambda_4 T3_j \times Treat_{ij} \times Post_t + \lambda_5 B3_j \times Treat_{ij} \times Post_t \\ & + \delta \mathbf{X}_{ijt} + \theta_t + \alpha_j + \varepsilon_{ijt}\end{aligned} \qquad (2.3)$$

In Equation (2.3), $T3_j$ ($B3_j$) is a binary indicator for physician j that equals one for initially top-ranked (bottom-ranked) physicians and zero otherwise. We define a physician's initial level of productivity using his or her median monthly ranking in the pre-intervention period. Drawing from the ranked histograms on median LOS of all discharged high acuity patients, we use the median ranking from the 19 pre-intervention months to determine the top-ranked and bottom-ranked physicians at each facility.

For our analyses, we focus on the top 3 and bottom 3 median rankings, respectively, because the rankings of physicians whose median pre-intervention rankings fall within these categories are the most stable. Figure 2.5 shows that physicians with a median ranking above 5 are among the top 3 physicians more than 50% of the time during the 19 pre-intervention period months and are never among the bottom 3. In addition, physicians with a median ranking of 24 or below are never among the top 3 physicians during the 19 pre-intervention months, but are among the bottom 3 in more than 50% of the 19 pre-intervention months. Thus, membership in the top 3 and bottom 3 is fairly stable.

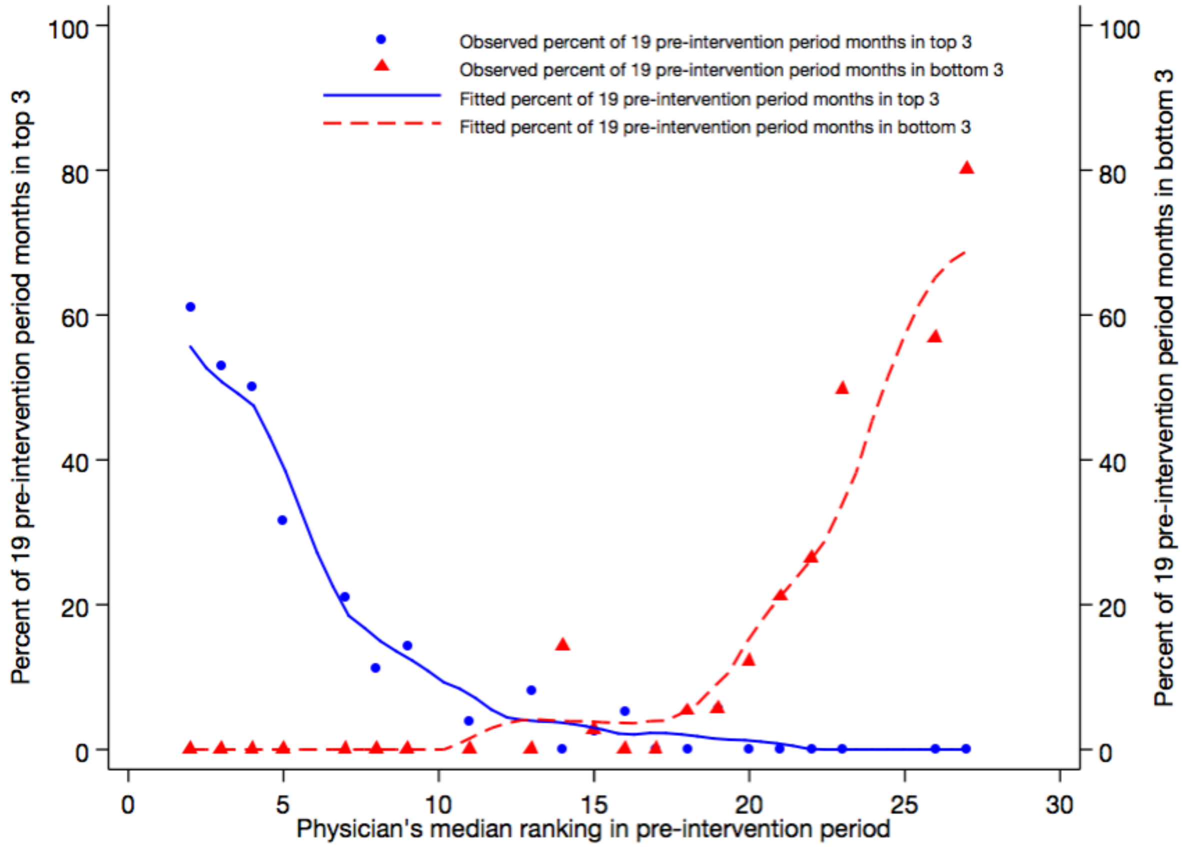


Figure 2.5. Ranking stability by physician's median ranking in pre-intervention period

All other variables remain the same as in Equation (2.1). The main effects for $T3_j$, $B3_j$, and $Treat_{ij}$, and the interaction effects $T3_j \times Treat_{ij}$ and $B3_j \times Treat_{ij}$ are omitted due to perfect collinearity with the physician fixed effects, and the main effect for $Post_t$ is omitted due to perfect collinearity with the month-year fixed effects. Our coefficients of interest are λ_4 and λ_5, which each represents the additional effect of being top-ranked or bottom-ranked, relative to being in neither category, on the impact on MD processing time of moving from private to public RPF.

Change in Between- and Within-Physician Variation in Processing Times. Examining the change in the variation in initial and interim processing times *between* physicians helps us understand how the implementation of public RPF affects the distribution of ways in which different physicians manage their workflow. If the leading explanation for improved productivity is either physicians' motivation to avoid the shame of being bottom-ranked or identifying and diffusing the best practices of top-ranked physicians, we would see a decrease in the between-physician variation in processing times. Specifically, if physicians are mimicking their high-performing peers by adopting their best practices, the set of practices used by physicians should become more homogeneous and we would expect a reduction in the between-physician variation in processing times. In contrast, if the leading explanation is the motivation to be top-ranked, we would see an increase in the level of between-physician variation.

Examining *within*-physician variation allows us to assess whether public RPF affects the distribution of ways in which each physician manages his or her workflow. If the leading explanation is that physicians are consistently adopting the best practices of their top-ranked peers, we would see a decrease in the within-physician variation of initial and interim processing times. However, if the leading explanation is either the motivation to be at the top of the distribution or to avoid being at the bottom of it and physicians are merely speeding up rather than differently managing their workflow, we would not expect to see a significant change in the within-physician variation in processing times.

To measure variation in initial and interim processing times between and within physicians, we use the coefficient of variation (CV), which is a unitless measure equal to the standard deviation divided by the mean of a random variable. For variation between physicians, we calculate the CV for initial processing time and interim processing time, respectively, for each facility-day-level observation m by dividing the standard deviation of each measure at each facility on a given day by the mean of the measure at the same unit of observation. For variation within physicians, we calculate the CV for initial processing time and interim processing time, respectively, for each physician-shift-level observation n by dividing the standard deviation of each measure for each physician on a given shift by the mean of the measure at the same unit of observation. Using these measures of variation, we assess the change in variation associated with the implementation of public RPF by estimating the following log-linear model:

$$\ln\left(CV_{mt}\right) = \varphi_0 + \varphi_1 Treat_m + \varphi_2 Treat_m \times Post_t + \delta \mathbf{X}_{mt} + \theta_t + \varepsilon_{mt} \tag{2.4}$$

$$\ln\left(CV_{nt}\right) = \tau_0 + \tau_1 Treat_n + \tau_2 Treat_n \times Post_t + \delta \mathbf{X}_{nt} + \theta_t + \varepsilon_{nt} \tag{2.5}$$

In Equations (2.4) and (2.5), *m* indexes each facility-day level observation and *n* indexes each physician-shift level observation. Covariates are measured as the means of each covariate described in chapter 2.4.1 at the corresponding level of observation. All other variables remain the same as in Equation (2.1). The main effect for $Post_t$ is omitted due to perfect collinearity with θ_t. Our coefficients of interest are φ_2 and τ_2, which each represent the change in logged CV between and within physicians, respectively, associated with the implementation of public RPF. We log CV to interpret these estimates as percent changes.

2.5 Results

Table 2.1 presents means for all patient- and ED-level covariates included in the empirical models stratified by facility (Treatment ED or Control ED) and time period (pre-intervention or post-intervention). At both Treatment ED and Control ED, approximately 80% of high acuity patients presenting from January 2009 to December 2011 were of ESI level 3 and only 1% of high acuity patients were of ESI level 1. The average age of patients was 45 years and 59% of patients were female. On average, there were 31 to 36 patients in the ED at a given time with 4 to 5 physicians on a given shift. Approximately 40% of shifts were morning shifts and 45% were afternoon shifts. Patients were slightly more likely to present to the ED on Mondays, Tuesdays, and Saturdays compared to other days of the week. On a given shift, each physician saw 14 patients on average and was multitasking across 4 to 5 patients at a given time (not shown in table).

Table 2.1. Summary statistics

	Treatment ED			Control ED		
Variables	**Pre**	**Post**	**Diff**	**Pre**	**Post**	**Diff**
Patient Level						
% ESI Level 1	1%	1%	0.01%	1%	1%	0.001%
% ESI Level 2	18%	23%	5.7%	19%	22%	3.1%
% ESI Level 3	82%	76%	-6.3%	81%	78%	-3.0%
Age (Years)	45	43	-1.1	46	47	0.5
% Female	59%	59%	0.1%	59%	59%	0.1%
Total patients in ED	36	32	-3.3	31	32	1.2
ED Level						
Total physicians on shift	4	5	0.4	4	5	0.5
% Morning shift	38%	37%	-0.7%	38%	40%	1.3%
% Afternoon shift	45%	46%	0.7%	44%	44%	-0.8%
% Overnight shift	18%	18%	0.003%	17%	17%	-0.5%
% Sunday	14%	15%	0.5%	14%	14%	-0.04%
% Monday	16%	15%	-0.3%	15%	15%	-0.02%
% Tuesday	15%	15%	0.2%	14%	14%	-0.05%
% Wednesday	13%	14%	0.3%	14%	15%	0.3%
% Thursday	14%	14%	-0.03%	14%	14%	-0.09%
% Friday	14%	14%	-0.4%	14%	14%	0.02%
% Saturday	15%	14%	-0.3%	14%	14%	-0.1%
Daily Patient Volume	131	147	16	140	145	5

Notes. N = 279,025 patient encounters.

Between the pre- and post-intervention periods, changes to the distribution of patient acuity and gender were very similar at both facilities, though there were small differences in the changes to the average age of patients and the total number of patients in the ED at a given time. At both sites, there was a small increase in the proportion of ESI 2 patients and a corresponding decrease in the proportion of ESI 3 patients due to an updated training on how to appropriately assign ESI levels.

In addition, though the average daily patient volume increased at both facilities, this increase was more pronounced at Treatment ED. All of these differences are adjusted for in our empirical specification, and thus we account for any differential changes between Treatment ED and Control ED.

We examine the correlations between each of the control variables, none of which has correlations close to 0.80, thus minimizing concerns about collinearity. We check for multicollinearity by calculating variance inflation factors (VIF) as well. The largest VIF in our empirical model is 6.8 and the mean VIF is 2.2, both of which fall below the conventional threshold of 10 (Hair et al. 1998). This suggests that multicollinearity is not a concern (Wooldridge 2012).

2.5.1 Effect of Public RPF on Physician Productivity

We estimate Equation (2.1) to examine how public RPF affects physician productivity. In the first two columns of Table 2.2, we estimate Equation (2.1), first without and then with month-year fixed effects to account for time trends. We find that the estimate for the effect of public RPF on logged MD processing time remains stable at -8.6% ($p < 0.01$). This 8.6% decrease corresponds to a 17-minute decrease in MD processing time for an average patient presenting to Treatment ED. This average effect of public RPF can be seen graphically as well. In Figure 2.4, the mean logged MD processing time of high acuity patients drops significantly immediately after the implementation of public RPF at Treatment ED. This supports H1, which predicts that public RPF leads to a decrease in processing time, on average.

In addition to productivity at the physician level, we also examine the effect of public RPF on system-level productivity. We find that the implementation of public RPF is associated with a significant increase in system-level productivity, as measured by ED LOS excluding boarding time (10.0% decrease, $p < 0.001$) and ED LOS including boarding time (11.3% decrease, $p < 0.001$),

Table 2.2. Average and heterogeneous effects of public RPF on physician productivity

	Average Effect					Heterogeneous Effect
VARIABLES	(1) Full Sample	(2) Full Sample	(3) AMI or Stroke	(4) Not AMI or Stroke	(5) Abdominal Pain	(6) Full Sample
Post X Treat	-0.086**	-0.086**	0.037	-0.086**	-0.124***	-0.080**
	(0.027)	(0.028)	(0.061)	(0.028)	(0.030)	(0.030)
Post X Treat X T3						-0.068
						(0.074)
Post X Treat X B3						-0.085**
						(0.030)
ESI level 2	0.269***	0.271***	0.076	0.276***	-0.069	0.270***
	(0.030)	(0.029)	(0.088)	(0.030)	(0.130)	(0.033)
ESI level 3	-0.120***	-0.122***	0.245**	-0.120***	-0.279*	-0.124***
	(0.033)	(0.033)	(0.088)	(0.033)	(0.131)	(0.037)
Age	0.011***	0.011***	-0.000	0.011***	0.013***	0.011***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Female	0.048***	0.048***	-0.002	0.049***	0.068***	0.047***
	(0.005)	(0.005)	(0.031)	(0.005)	(0.012)	(0.005)
Total patients	0.009***	0.006*	0.006	0.006*	0.003	0.005
	(0.003)	(0.002)	(0.011)	(0.002)	(0.005)	(0.002)
(Total patients)^2	-0.001***	-0.000**	-0.001	-0.000**	-0.000	-0.000**
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
MDs on shift	-0.010***	-0.005*	-0.014	-0.005*	-0.000	-0.005*
	(0.002)	(0.002)	(0.013)	(0.002)	(0.004)	(0.002)
Afternoon shift	-0.032***	-0.032***	0.039	-0.032***	-0.063***	-0.035***
	(0.008)	(0.007)	(0.034)	(0.007)	(0.012)	(0.008)
Overnight shift	-0.022*	-0.015	0.044	-0.014	-0.001	-0.018
	(0.010)	(0.010)	(0.053)	(0.010)	(0.016)	(0.011)
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	No	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	261,586	261,586	1,930	259,656	35,084	232,297
Adjusted R-squared	0.188	0.190	0.046	0.190	0.203	0.181

Notes. Regressions are fixed-effects log-linear difference-in-differences models estimated at the patient encounter level. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p<0.05, **p<0.01, ***p<0.001.

where boarding time is the amount of time that patients being admitted to the hospital spend waiting for an inpatient bed (Table 2.3 columns (1) and (2)). In addition, we find that the increase in physician productivity is not accompanied by a lower throughput rate. Following KC (2014), we measure throughput rate as the number of patients discharged by a physician on a given shift, adjusted for the duration of the shift. We find a modest increase in throughput rate (2.6%, $p < 0.05$), which suggests that physicians are not working less to compensate for being more productive (Table 2.3 column (3)).

Table 2.3. System-level effects of public RPF on ED LOS

VARIABLES	(1) Logged ED LOS Excluding Boarding Time	(2) Logged ED LOS Including Boarding Time	(3) Logged Throughput Rate
Post X Treat	-0.100***	-0.113***	0.026*
	(0.025)	(0.025)	(0.013)
ESI level 2	0.343***	0.296***	0.000
	(0.021)	(0.022)	(0.004)
ESI level 3	0.021	-0.054*	0.034***
	(0.024)	(0.023)	(0.005)
Age	0.007***	0.008***	-0.001***
	(0.000)	(0.000)	(0.000)
Female	0.032***	0.024***	0.001
	(0.003)	(0.003)	(0.001)
Total patients	0.023***	0.023***	0.013***
	(0.002)	(0.002)	(0.001)
(Total patients)^2	-0.001***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)
MDs on shift	-0.003	-0.001	-0.008***
	(0.004)	(0.004)	(0.001)
Afternoon shift	-0.029*	-0.030*	0.009*
	(0.012)	(0.012)	(0.004)
Overnight shift	0.025	0.034	0.004
	(0.020)	(0.020)	(0.005)
Time-varying controls	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes
Physician FE	No	No	Yes
Observations	274,339	278,676	196,326
Adjusted R-squared	0.122	0.147	0.211

Notes. Regressions are fixed-effects log-linear difference-in-differences models estimated at the patient encounter level. Throughput rate is operationalized as the number of patients discharged by a physician on a given shift, adjusted for the duration of the shift. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p<0.05, **p<0.01, ***p<0.001.

2.5.2 Standardized Work as a Moderator of the Effect of Public RPF on Physician Productivity

In columns (3) and (4) of Table 2.2, we repeat the estimation of Equation (2.1) on a subsample of patients with conditions with standardized processes (i.e., patients presenting with symptoms of an

AMI or stroke) and those with conditions without standardized processes (i.e., all other patients). We find that patients presenting with symptoms of an AMI or stroke experience no significant change in MD processing time after the implementation of public RPF ($p \approx 0.54$). In contrast, patients who do not fall into this category exhibit an 8.6% decrease in MD processing time with public RPF ($p < 0.01$). As a test of robustness, we also consider a third subsample that is limited to patients presenting with abdominal pain symptoms—an example cited by physicians as a frequent and potentially serious case for which standardized processes are not in place (column (5)). We find that these patients experience a 12.4% decrease in MD processing time with public RPF ($p < 0.001$). This strongly supports H2, which predicts that the decrease in processing time resulting from a shift from private to public RPF is greater when standardized processes are not in place compared to when they are.

2.5.3 Effect of Public RPF on Care Intensity, Clinical Quality, and Patient Satisfaction

Given the potential tradeoff between operational efficiency and service quality in service settings like the ED (Anand et al. 2010, Hopp et al. 2007, Netessine and Yakubovich 2012), we consider the impact of public RPF on care intensity, clinical quality, and patient satisfaction. We estimate Equation (2.2) to explore these effects and present results in Table 2.4. Because the binary-outcome logit regression model is nonlinear, we express results as marginal effects averaging across all values of the covariates in the data.

In columns (1) – (4), we examine the change in care intensity under public RPF by estimating the difference in the change in a patient's likelihood of having a lab test ordered, a radiology test ordered, a specialty consult requested, or being admitted to the hospital, respectively, at Treatment ED and Control ED. With regard to care intensity *in* the ED, we find that the implementation of public RPF is associated with a 2.9 percentage point differential decrease in the likelihood of having

Table 2.4. Marginal treatment effect of public RPF on care intensity and clinical quality

	Care Intensity				Clinical Quality	
VARIABLES	(1) Lab Ordered	(2) Radiology Ordered	(3) Consult Requested	(4) Admitted	(5) Expired	(6) Revisit 72 hours
Post X Treat	-0.029***	-0.038***	0.001	0.012**	-0.001*	-0.004***
	(0.009)	(0.007)	(0.006)	(0.005)	(0.000)	(0.000)
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277,970	278,473	277,945	278,039	264,876	271,956
Pseudo R-squared	0.106	0.103	0.140	0.154	0.553	0.033

Note: Regressions are fixed-effects logit difference-in-differences models estimated at the patient encounter level and coefficients are expressed as marginal effects. Controls not shown include ESI level, age, gender, current patient count, current patient count squared, total number of MDs on shift, arrival shift type, time of day, day of week, month-year fixed effects, and physician fixed effects. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p<0.05, **p<0.01, ***p<0.001.

a lab test ordered ($p < 0.001$), a 3.8 percentage point differential decrease in the likelihood of having a radiology test ordered ($p < 0.001$), and no significant change in the likelihood of having a specialty consult requested ($p \approx 0.90$). These results show that public RPF may be associated with a reduction in the intensity of care provided in the ED. Given top-ranked physicians were less likely to exhibit high levels of care intensity in the pre-intervention period, this decrease in in-ED care intensity may be suggestive of a convergence around best practices. When we examine changes in care intensity *beyond* the ED, we find that patients presenting to Treatment ED experience a 1.2 percentage point differential increase in the likelihood of being admitted to the hospital ($p < 0.01$). This suggests that there may be some degree of task shifting occurring, where ED physicians may be reducing MD processing time by, on the margin, deciding to admit a patient rather than keeping the patient in the ED for further observation.

Using a mediation model (Preacher and Hayes 2008), we examine the extent to which the decrease in in-ED care intensity and the increase in task shifting explain the observed decrease in

MD processing time. When we account for the four proxy measures as potential mediators, we find a statistically significant, but smaller decrease in processing time (-4.4%, $p < 0.05$), which is suggestive of partial mediation. That is, the change in care intensity contributes to the decrease in processing time, but does not fully explain the effect.

Without any additional information, it is difficult to determine whether this reduction in in-ED care intensity and some degree of task shifting is suggestive of lower quality care. If it were, we might find a corresponding increase in the likelihood of dying in the ED or returning to the ED. As shown in columns (5) and (6), we find that public RPF is associated with a 0.1 percentage point differential *decrease* in the likelihood of dying in the ED and a 0.4 percentage point differential *decrease* in the likelihood of returning to the ED. Though economically modest, both of these decreases are statistically significant at conventional levels ($p < 0.05$ and $p < 0.001$, respectively). These findings help reduce concern that public RPF may be associated with a decrease in clinical quality.

Table 2.5. Marginal treatment effect of public RPF on patient satisfaction

	Experience with Physician			General Experience	
VARIABLES	(1) MD Skills and Ability	(2) MD Provided Proper Care	(3) MD Listened and Explained	(4) Time Spent in ED	(5) Needs Met
Post X Treat	-0.002	-0.002	-0.004	-0.002	-0.001
	(0.002)	(0.003)	(0.003)	(0.005)	(0.004)
Time-varying controls	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes
Observations	70,274	73,107	72,963	75,578	77,005
Pseudo R-squared	0.015	0.016	0.016	0.013	0.010

Note: Regressions are fixed-effects logit difference-in-differences models estimated at the patient encounter level and coefficients are expressed as marginal effects. Controls not shown include ESI level, age, gender, current patient count, current patient count squared, total number of MDs on shift, arrival shift type, time of day, day of week, month-year fixed effects, and physician fixed effects. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. **p*<0.05, ***p*<0.01, ****p*<0.001.

It is also possible that reductions in MD processing time may affect patient satisfaction, concerning both their satisfaction with the physician and their general ED experience. On the one hand, patients may feel rushed by the physician and perceive that the physician is not taking enough time to listen to them or explain what is being done and why. On the other hand, being able to leave the ED sooner may lead patients to feel more satisfied. Table 2.5 reports the marginal effects resulting from estimating ordered logit models of patient satisfaction. In columns (1) – (3), we find no significant change in patient satisfaction regarding their experience with the physician. In columns (4) and (5), we find no significant change in patient satisfaction regarding their general ED experience ($p > 0.24$ for each measure).

2.5.4 Consideration of Potential Mechanisms

Heterogeneous Treatment Effects. In column (6) of Table 2.2, we report results from estimating Equation (2.3). We find that the effect of public RPF on MD processing time is significantly greater for physicians who ranked in the bottom 3 in the pre-intervention period than for mid-ranked physicians (i.e., physicians not in the top or bottom 3). In addition, we find that the differential effect for physicians who ranked in the top 3 is not statistically significant ($p \approx 0.63$). Specifically, bottom-ranked physicians attained an additional 8.5% decrease in MD processing time ($p < 0.01$) above and beyond the 8.0% decrease ($p < 0.01$) attained by mid-ranked physicians. In terms of effect size, this corresponds to an additional time savings of 20 minutes on average by bottom-ranked physicians at Treatment ED, above and beyond the 15-minute decrease in MD processing time attained by mid-ranked physicians at the same facility. Together, this points to a 35-minute time savings attained by bottom-ranked physicians at Treatment ED after the implementation of public RPF.

Table 2.6. Effect of public RPF on between- and within-physician variation in processing times

	Between Physicians		Within Physician	
VARIABLES	(1) Logged CV of Initial Processing Time	(2) Logged CV of Interim Processing Time	(3) Logged CV of Initial Processing Time	(4) Logged CV of Interim Processing Time
Post X Treat	-0.143***	-0.086***	-0.070***	0.010***
	(0.010)	(0.010)	(0.006)	(0.002)
ESI level 2	0.288	0.229	-0.026	-0.001
	(0.290)	(0.281)	(0.025)	(0.011)
ESI level 3	0.240	0.166	-0.040	0.005
	(0.286)	(0.278)	(0.027)	(0.011)
Age	0.001	0.002	-0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)
Female	-0.009	0.029	0.003	-0.001
	(0.054)	(0.053)	(0.006)	(0.002)
Total patients	0.000	0.051***	-0.003	0.003**
	(0.007)	(0.006)	(0.002)	(0.001)
(Total patients)^2	0.000	-0.002***	0.000***	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
MDs on shift	-0.009*	-0.003	-0.001	-0.003***
	(0.004)	(0.004)	(0.001)	(0.001)
Afternoon shift	0.064	0.026	-0.003	-0.002
	(0.048)	(0.047)	(0.005)	(0.002)
Overnight shift	-0.009	0.176**	-0.011	-0.006
	(0.067)	(0.065)	(0.007)	(0.003)
Time-varying controls	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes
Physician FE	No	No	Yes	Yes
Observations	2,128	2,128	27,049	27,049
Adjusted R-squared	0.637	0.591	0.579	0.951

Notes. Regressions are fixed-effects log-linear difference-in-differences models estimated at the facility-day level. Standard errors (in parentheses) are heteroskedasticity robust. **p*<0.05, ***p*<0.01, ****p*<0.001.

Change in Between- and Within-Physician Variation in Processing Times. In Table 2.6, we report results from estimating Equations (2.4) and (2.5). In columns (1) and (2), we see that the implementation of public RPF leads to a decrease in between-physician variation in both initial and interim processing times. The between-physician CV of initial processing times decreases by 14.3% (p < 0.001) and the between-physician CV of interim processing times decreases by 8.6% (p <

0.001). This suggests that different physicians may be converging on a similar set of practices with regards to how they structure their workflow. In columns (3) and (4), we see that the within-physician CV of initial processing time decreases by 7% (p < 0.001) whereas the within-physician CV of interim processing time increases by 1% (p < 0.001). This suggests that each physician may be taking a more consistent approach with regards to how he or she structures workflow during the initial processing time, but that there is a modest increase in the variation with which each physician structures his or her workflow during the interim processing time. In additional analyses that examine heterogeneous effects in the change in within-physician CV (not shown in table), we find that the decrease in within-physician CV in initial processing times is greater for bottom-ranked physicians than for top-ranked physicians (p < 0.01). For the change in within-physician CV in interim processing times, this difference is not statistically significant ($p \approx 0.72$). Nevertheless, on balance, we find that the within-physician CV of total MD processing time is reduced (-2.8%, p < 0.001).

Additional Considerations Using Qualitative Data. Using data from the physician interviews, we further consider each of the potential mechanisms for improved productivity. Though we are unable to test changes in interview responses over time as we do not have data from the pre-intervention period, we are able to compare cross-sectional differences in responses across the two facilities. Note, the physicians at Treatment ED and Control ED are employed by the same health care system, work in the same city, and are under the same senior ED leadership that exists at the level of the health care system. Consequently, physicians from the two EDs are subject to an overlapping set of organizational and contextual factors that serve to increase their comparability. We summarize the findings from these interviews in Table 2.7.

At Treatment ED, 52% of respondents reported feeling motivated to be top-ranked on the RPF reports in comparison to 53% of respondents at Control ED. Using a two-tailed t-test assuming unequal variance, we find that these proportions are not significantly different at conventional levels ($p \approx 0.99$). Respondents at both EDs also reported similar levels of shame as measured by the shame index. On a scale ranging from 1 (Low) to 5 (High), respondents at Treatment ED reported a mean level of shame of 1.52 and those at Control ED reported a mean level of 1.34. This difference is not statistically significant ($p \approx 0.18$). Since we would expect feelings of shame to be most salient among bottom-ranked physicians, we also compare the mean level of shame reported by physicians at Treatment ED and Control ED who are ranked in the bottom third; we find no statistically significant differences between their responses ($p \approx 0.47$). However, *within* each facility, we find that bottom-ranked physicians reported a higher level of shame than the top-ranked physicians, as we would expect. Collectively, these results do not support the notion that public RPF improves productivity by increasing the level of shame felt by the bottom-ranked physicians when compared to private RPF.

When asked to list the different ways in which he or she manages workflow to reduce patients' LOS, we find significant differences between respondents at Treatment ED and Control ED in their rates of employing workflow practices that correspond to efficiency tips 1 and 2. While 100% of respondents at Treatment ED reported placing lab and radiology orders as early as possible, only 60% of respondents at Control ED reported this to be the case ($p < 0.001$). Regarding early initiation of the discharge instructions and the encounter note, 60% of respondents at Treatment ED and 30% of respondents at Control ED reported employing this practice ($p < 0.001$). This is consistent with what we would expect to find if workers receiving public RPF were to identify and adopt the best practices around workflow that are employed by their top-ranked peers.

During the interviews, several physicians at Treatment ED shared examples of how they were able to learn from the best practices of their top-ranked peers once public RPF enabled them to identify these individuals (see illustrative quotes in Table 2.7). Specifically with regards to efficiency

Table 2.7. Summary of data from physician interviews

	Motivation to be top-ranked (1=Yes, 0=No)	Shame index (1=Low, 5=High)	Efficiency tip 1	Efficiency tip 2	Favor public RPF	Illustrative quotes about public and private RPF reports
Treatment (*N* = 21)	0.52	1.52	100%	60%	82%	"The open data [was] a factor in me identifying [Dr. X]… [Before when] a patient would come in with flank pain and I would think 'maybe it's a kidney stone, maybe not,' I would order a urinalysis. If the urinalysis shows blood in the urine, then I would do a CT scan. What would happen inevitably is that it would take 2 hours to get the urinalysis results. And then sure enough, there is blood in the urine, [and] then I would order the CT scan. So that would delay the patient. [Now], based on my exam and the history, [if] I think this patient has a kidney stone, I order everything for the kidney stone and I don't wait to determine if there's blood in the urine." "We ask people at the top who are successful and have them share their ideas." "I really do feel that it is so much more helpful to have it [public]… There is so much more data to be gained and understood as you look at the group and not just look at the number and have no idea what that means." "There is value for other EDs to follow suit. I think [public] reports are fun." "With [private RPF], it wasn't as helpful. Really all I could take away was where I was in comparison to the group. I definitely get more out of the data now than I did back then."
Control (*N* = 20)	0.53	1.34	60%	30%	33%	"With these [private] reports, I look at where I am versus everyone else. But I don't know what to do with it." "I don't know how I'd know how I could improve, like what are the limitations, what's causing that. I don't know what I [can] really take away from [private reports] other than 'Okay, that's my number. Let's try and improve it.'" "With [public RPF], I think it would be interesting to see which docs fell where in terms of assessing a pattern." "[With public RPF], I think I might look at who was doing 'better' or had a better metric and try to figure out what they were doing that I wasn't. It might introduce some healthy competition into things." "I think for the most part, our docs are working as hard as they can, and to have this additional pressure put on them, I don't know if it's the best thing to do in terms of morale."
***p*-value**	0.99	0.18	< 0.001	< 0.001	<0.01	

tip 1, several Treatment ED physicians stated that they had changed their practices such that they no longer "dribble ordered" tests (i.e., ordered tests serially in which the physician waited for the results of the first test prior to ordering a subsequent test), and instead ordered them simultaneously so that they could be processed in parallel. Though in some cases this may have resulted in inefficiencies from ordering unnecessary tests, several physicians suggested that this was unlikely to be the case because now they were more selective about whether to order a given test at all. Whereas before physicians would often order a test to rule out a diagnosis, now they tended to do this less often by not ordering tests that were only for that reason. In this setting, physicians had the confidence to alter their workflow in this way because the integrated health care system made it possible for patients to easily follow-up with an outpatient department if needed. These statements suggest that a patient's likelihood of receiving at least one test may have *decreased* after the intervention, which is consistent with our findings in columns (1) and (2) of Table 2.4.

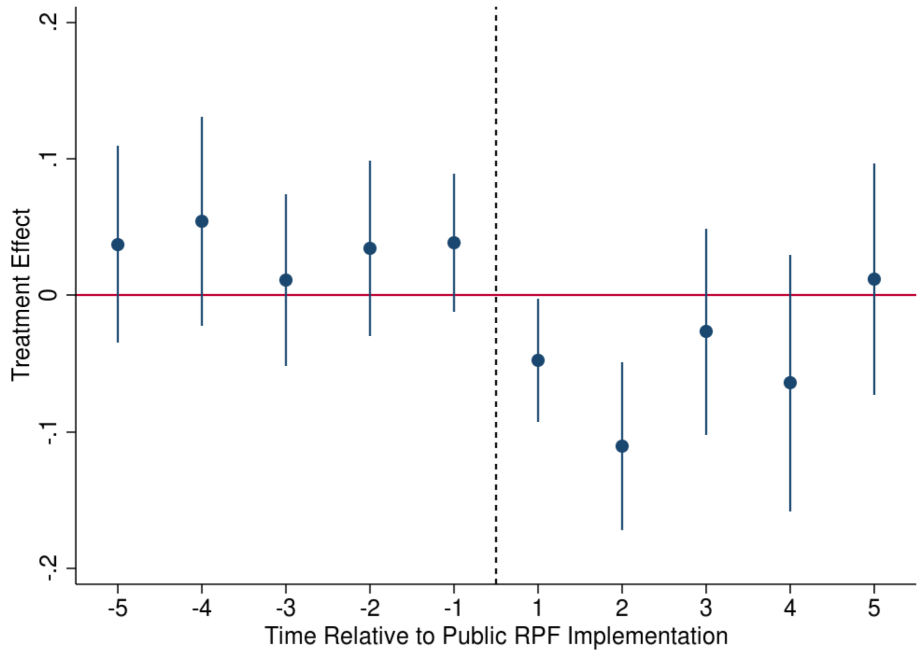
Also during these interviews, several Treatment ED physicians stated that knowing the identities of the physicians corresponding to various levels of performance improved the credibility of the RPF reports and the credibility of the efficiency tips, which ultimately motivated them to change their workflow. At Treatment ED, knowing the identities of top-ranked physicians via public RPF allowed physicians to assess whether these physicians provided high quality care (based on their own observations). When this was the case, other physicians were more willing to seek out and adopt their workflow practices. In contrast, at Control ED, several physicians reported skepticism around the value of the private RPF reports. Furthermore, despite there being an internal website with anonymized efficiency tips, none of the Control ED physicians we interviewed reported accessing these tips. This suggests that knowing the identities of the individuals sharing efficiency tips is an important factor in motivating peers to adopt best practices.

Finally, despite concerns raised by prior literature and critics of public RPF about its potential negative effects (e.g., demoralization, internal competition), the majority of respondents at Treatment ED (82%) favored keeping public RPF reports rather than reverting to private RPF reports. Interestingly, only 33% of respondents at Control ED reported that they would favor moving from private to public RPF reports, suggesting a discomfort with the *concept* of public RPF as opposed to the actual experience of it. In Table 2.7, we include illustrative quotes that capture respondents' thoughts in response to this question as well.

2.5.5 Robustness Checks and Specification Tests

We conduct a number of additional analyses to examine the robustness of our main findings. We can group our additional analyses into three main categories: tests of validity concerning our difference-in-differences approach, tests using alternative measures, and tests with sample exclusions. We conduct these analyses on all dependent variables, but focus our discussion here on results concerning physician productivity for brevity. Findings of additional analyses with the other dependent variables are not meaningfully different from those concerning physician productivity.

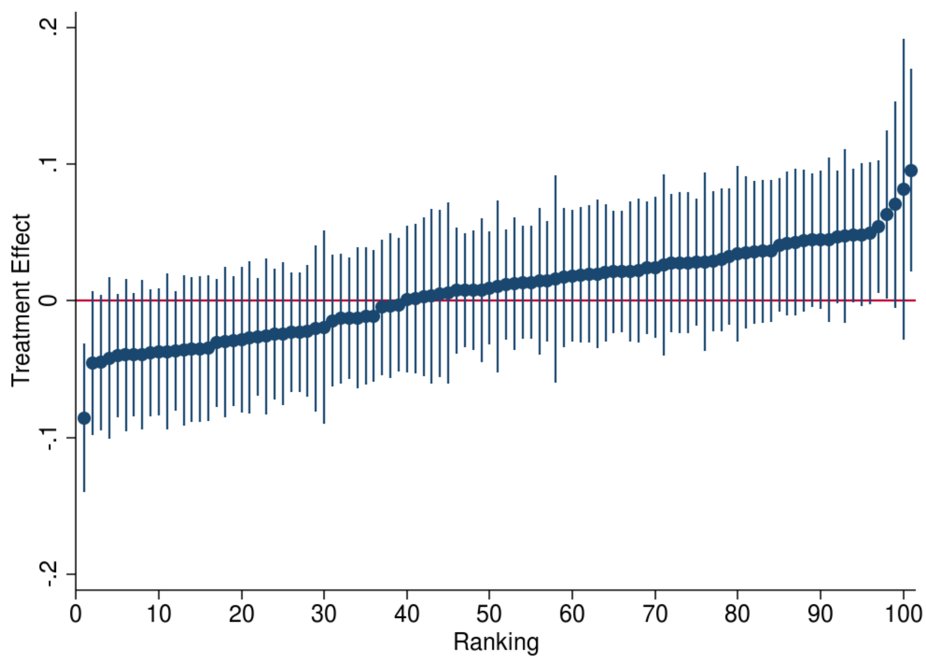
To test the internal validity concerns that arise from employing a difference-in-differences approach, we begin by examining whether MD processing time was already differentially decreasing at Treatment ED before public RPF was implemented. Following Autor (2003), we use a leads and lags model to explore the presence of pre-intervention time trends in our data. Specifically, we create five pre-period (i.e., lead) indicator variables corresponding to each of the 5 quarters before public RPF was implemented and 5 post-period (i.e., lag) indicator variables corresponding to each of the 5 quarters after implementation. We estimate a model similar to Equation (2.1) that replaces *Post* with the ten lead and lag indicator variables. In this model, the reference category is the quarter during which public RPF was implemented. As we see in Figure 2.6, none of the coefficients on the five lead indicator variables is statistically significant. This suggests that Treatment ED did not exhibit a differential decrease in MD processing times in advance of the implementation of public RPF.



Notes. Each point and its whiskers (95% confidence intervals) correspond to an estimation of Equation (2.1) as described in chapter 2.4.1 that replaces the intervention indicator *Post* with five lead variables and five lag variables. Lead and lag variables correspond to each of the 5 quarters before and after the intervention, respectively. The vertical dashed line indicates the time when public RPF was implemented at Treatment ED (August 2010).

Figure 2.6. Leads and lags model testing pre-intervention time trends

As another test of validity, we conduct placebo tests to assess whether our findings are merely artifacts of the structure of our data (Bertrand et al. 2004). Following the methodology employed by Pierce et al. (2015) and Staats et al. (2015), we randomly assign an implementation date and repeat the estimation of Equation (2.1) by replacing the actual implementation date with the placebo date. We repeat this 100 times and present the results in Figure 2.7. We find that only three of the 100 placebo models produces a coefficient that is statistically significant at the 5% level. Each of these coefficients is weaker in statistical significance compared to that of our true estimate.



Notes. Placebo intervention dates were randomly assigned 100 times. Each point and its whiskers (95% confidence intervals) correspond to an estimation of Equation (2.1) as described in chapter 2.4.1. The coefficient and confidence interval for the true data are represented on the far left (rank 1).

Figure 2.7. Placebo tests of the effect of public RPF on physician productivity

As a third test of validity, we repeat our estimation of Equation (2.1) while limiting our study period to nullify the possibility of other changes having differentially impacted the two EDs. Specifically, we limit the study period to span a period of 3 months, 6 months, and 12 months, respectively, before and after the intervention. Compared to our main findings, we find that the effect of public RPF implementation on physician productivity is weaker with a 3-month pre/post period, suggesting a ramp-up period, but then is stronger and fairly constant from 6 months pre/post onward, reducing concern that our results are driven by the time window used for the analysis (Table 2.8 columns (1) – (3)).

Table 2.8. Alternate model specifications

	Restricted Time Frame			Other		
VARIABLES	(1) 3 Months Pre/Post	(2) 6 Months Pre/Post	(3) 12 Months Pre/Post	(4) No Washout Period	(5) Linear DV	(5) Interrupted Time Series
Post X Treat	-0.067**	-0.105***	-0.095***	-0.083**	-16.657***	
	(0.026)	(0.026)	(0.028)	(0.027)	(2.807)	
ESI level 2	0.352***	0.324***	0.273***	0.275***	72.915***	0.440***
	(0.058)	(0.042)	(0.033)	(0.030)	(6.219)	(0.034)
ESI level 3	-0.021	-0.061	-0.129***	-0.117***	1.648	0.001
	(0.057)	(0.043)	(0.036)	(0.032)	(5.642)	(0.047)
Age	0.011***	0.011***	0.011***	0.011***	1.483***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.022)	(0.000)
Female	0.055***	0.049***	0.048***	0.048***	0.345	0.030***
	(0.008)	(0.006)	(0.005)	(0.005)	(1.001)	(0.007)
Total patients	0.010*	0.007*	0.005*	0.005*	1.758***	0.001
	(0.005)	(0.003)	(0.002)	(0.002)	(0.471)	(0.002)
(Total patients)^2	-0.001*	-0.000*	-0.000*	-0.000**	-0.088***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.025)	(0.000)
MDs on shift	-0.004	-0.006	-0.004	-0.004	-0.499	-0.010***
	(0.004)	(0.003)	(0.003)	(0.002)	(0.380)	(0.003)
Afternoon shift	-0.038**	-0.013	-0.022*	-0.031***	-1.112	-0.023*
	(0.013)	(0.010)	(0.009)	(0.007)	(1.169)	(0.012)
Overnight shift	-0.001	0.015	-0.002	-0.014	5.429**	-0.004
	(0.019)	(0.014)	(0.012)	(0.010)	(1.854)	(0.019)
Post						-0.088**
						(0.027)
Time						-0.002***
						(0.000)
Time after intervention						0.001
						(0.001)
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	45,222	91,123	181,935	269,270	261,586	126,196
Adjusted R-squared	0.174	0.179	0.189	0.189	0.097	0.198

Notes. Columns 1 – 4 are fixed-effects log-linear difference-in-differences models, Column 5 is a fixed-effects linear difference-in-differences model, and Column 6 is a fixed-effects log-linear interrupted time series model. All models are estimated at the patient encounter level. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. *p<0.05, **p<0.01, ***p<0.001.

Next, we consider several alternate model specifications. First, instead of accounting for a washout period by excluding data from August 2010, we do not impose this exclusion on the data (Table 2.8 column (4)). We obtain a similar set of findings, in which the implementation of public RPF is associated with an 8.3% decrease in MD processing time ($p < 0.01$).

Second, instead of using log-transformed dependent variables in estimating Equation (2.1), we use linear dependent variables measured in minutes. The results are robust to our main findings. In particular, we find that the implementation of public RPF leads to a 17-minute decrease in MD processing time (Table 2.8 column (5)), which closely corresponds to the effect size estimate discussed in chapter 2.5.1.

Third, we use an interrupted time series model rather than a difference-in-differences model to estimate the effect of public RPF on MD processing time. An interrupted time series model uses data that are collected at multiple time points before and after an intervention. Unlike a difference-in-differences model, which assumes parallel trends in the treatment and control groups before the intervention and only accounts for a change in the level (i.e., mean) of the treatment group as compared to the control group, an interrupted time series model can account for both a change in level and a change in trends (Shadish et al. 2002, Wagner et al. 2002). Using an interrupted time series model, we find a change in levels but not a change in trends as a result of public RPF at Treatment ED (Table 2.8 column (6)). Specifically, we find an 8.8% decrease in the level of MD processing time ($p < 0.001$) and a non-significant change in trends ($p \approx 0.4$). Thus, our results are robust to whether we employ a difference-in-differences model or an interrupted time series model.

Fourth, we employ a set of alternative thresholds for determining top-, mid-, and bottom-ranked physicians. Our results concerning the heterogeneous effects of public RPF on physician productivity are robust to a more restrictive threshold of top 2 and bottom 2. With this alternate threshold, each of the coefficients of interest is similar in direction and magnitude as in our main specification (Table 2.9). As we adopt a progressively less restrictive threshold of top 4 and bottom 4, and top 5 and bottom 5, respectively, the average effect embodied by the mid-performing physicians remains relatively robust but the additional effect of the low-performing physicians decreases in magnitude and statistical significance. This result is consistent with the fact that physician rankings are very stable in the extremes (top 3 and bottom 3) but less so beyond this

Table 2.9. Heterogeneous effect of public RPF on physician productivity with alternate thresholds

VARIABLES	(1) Top/Bottom 2	(2) Top/Bottom 4	(3) Top/Bottom 5
Post X Treat	-0.082**	-0.067*	-0.072*
	(0.029)	(0.027)	(0.030)
Post X Treat X Top	0.002	-0.237	-0.176
	(0.031)	(0.178)	(0.094)
Post X Treat X Bottom	-0.084**	-0.049	-0.022
	(0.029)	(0.035)	(0.034)
ESI level 2	0.270***	0.270***	0.270***
	(0.033)	(0.033)	(0.033)
ESI level 3	-0.124***	-0.123**	-0.123**
	(0.037)	(0.037)	(0.037)
Age	0.011***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)
Female	0.047***	0.047***	0.047***
	(0.005)	(0.005)	(0.005)
Total patients	0.005	0.005	0.005
	(0.002)	(0.002)	(0.002)
(Total patients)^2	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)
MDs on shift	-0.005	-0.005	-0.005
	(0.002)	(0.002)	(0.002)
Afternoon shift	-0.035***	-0.035***	-0.034***
	(0.008)	(0.008)	(0.008)
Overnight shift	-0.018	-0.019	-0.018
	(0.011)	(0.010)	(0.010)
Time-varying controls	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes
Observations	232,297	232,297	232,297
Adjusted R-squared	0.181	0.181	0.181

Notes. Regressions are fixed-effects log-linear difference-in-differences models estimated at the patient encounter level. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. **p*<0.05, ***p*<0.01, ****p*<0.001.

Table 2.10. Average effect of public RPF on physician productivity with various exclusions

VARIABLES	(1) Excluding Patients With LOS > 48 hours	(2) Excluding Patients Receiving Psychiatry Consult	(3) Excluding ESI Level 1 Patients	(4) Excluding Expired Patients	(5) Excluding LWBS, AMA, Eloped Patients
Post X Treat	-0.085**	-0.084**	-0.086**	-0.086**	-0.086**
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
ESI level 2	0.268***	0.199***	0.393***	0.268***	0.270***
	(0.029)	(0.028)	(0.014)	(0.029)	(0.029)
ESI level 3	-0.120***	-0.114***		-0.125***	-0.124***
	(0.033)	(0.032)		(0.033)	(0.032)
Age	0.011***	0.011***	0.011***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.048***	0.052***	0.047***	0.048***	0.047***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Total patients	0.006*	0.005*	0.006*	0.006*	0.006*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(Total patients)^2	-0.000**	-0.000**	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MDs on shift	-0.005*	-0.004	-0.005*	-0.005*	-0.005*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Afternoon shift	-0.032***	-0.037***	-0.032***	-0.032***	-0.032***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Overnight shift	-0.015	-0.027**	-0.015	-0.015	-0.015
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Time-varying controls	Yes	Yes	Yes	Yes	Yes
Day of week FE	Yes	Yes	Yes	Yes	Yes
Month-year FE	Yes	Yes	Yes	Yes	Yes
Physician FE	Yes	Yes	Yes	Yes	Yes
Observations	261,481	255,942	260,238	261,569	260,944
Adjusted R-squared	0.191	0.195	0.191	0.190	0.191

Notes. Regressions are fixed-effects log-linear difference-in-differences models estimated at the patient encounter level. Standard errors (in parentheses) are heteroskedasticity robust and clustered by physician. **p*<0.05, ***p*<0.01, ****p*<0.001.

threshold, as seen in Figure 2.5.

Lastly, we apply various exclusions to our study sample to create a more homogeneous patient population and to better control for factors outside a physician's control. Specifically, we exclude patients with a LOS excluding boarding time greater than 48 hours (n = 206) and patients who received a psychiatry consult (n = 11,312) because the extended MD processing times of most of these patients are typically driven by placement logistics rather than by a physician's level of productivity. We exclude patients of ESI level 1 (i.e., patients needing resuscitation) (n = 2,213) and patients who died in the ED (*n* = 669) because their MD processing times are likely to be driven by factors other than physician productivity. Lastly, we exclude patients who left without being seen (*n* = 3,185), patients who left against medical advice, and patients who eloped because their LOSs are likely to be truncated as a result of their departure. Our results are robust to each of these exclusions (Table 2.10).

2.6 Discussion and Conclusions

In this paper, we identify public RPF as an effective tool for improving productivity by fostering standardized workflow in complex service systems. In our field setting, public RPF was disclosed to physicians in a way that it (a) provided them with specific information on their levels of productivity and (b) helped to improve their understanding of how to improve their productivity while (c) accounting for natural productivity fluctuations over time by reporting data at the month level—three key considerations identified by Netessine and Yakubovich (2012) for a successful application of RPF systems. Using patient encounter data from hospital EDs, we find that public RPF is associated with a significant improvement in productivity and no significant reduction in overall quality. We find that public RPF is particularly helpful in improving productivity when standardized work is not in place. Additional analyses suggest that these improvements are most likely driven by the identification and diffusion of best practices that is enabled by public RPF.

Specifically, we find that public RPF is associated with an 8.6% decrease in MD processing time on average. When stratified by a physician's median ranking in the pre-intervention period, we find that top- and mid-ranked physicians exhibit an 8.0% decrease in MD processing time and bottom-ranked physicians exhibit an additional 8.5% decrease in MD processing time. Though we do not find evidence of lower levels of clinical quality or patient satisfaction associated with the implementation of public RPF, we do find a decrease in in-ED care intensity as measured by the likelihood of ordering at least one lab test or radiology test. Unfortunately, we are unable to examine changes in the *number* of tests ordered as a more precise measure of care intensity, as we do not have access to this information. However, interviews with physicians suggest that the decrease in the likelihood of ordering at least one test stems from physicians acting with greater confidence on their diagnoses and being more selective about whether to order tests when they may only serve to rule out diagnoses.

2.6.1 Theoretical Contributions

This paper makes several contributions to the literatures on feedback, best practice transfer, and workflow management. First, our paper answers the call for field research on the impact of public—as opposed to private—feedback (Blanes i Vidal and Nossol 2011). We do this by examining the effect of publicly disclosing RPF, a type of feedback, to ED physicians internal to an organization over a period of 3 years. We find that the small and financially costless change of replacing code numbers with worker names can lead to a substantial improvement in productivity. This study illustrates an example of how behavioral levers can be used to improve productivity by changing the way in which discretionary workers manage their work.

Second, in our field setting, we find that public RPF that is provided at a monthly interval is sufficient to improve worker productivity. This is in contrast to the experiments in Bendoly (2013) and Schultz et al. (1999), which provide continual performance feedback during task execution. Establishing the optimal frequency of feedback is beyond the scope of this paper and left for future research. On the one hand, shorter time frames between feedback reports may increase adherence to best practices. However, this may also increase workers' levels of stress (Bendoly 2013, Schultz et al. 1999) and increase the level of spurious variation in the rankings, thus reducing their usefulness (Netessine and Yakubovich 2012).

Third, we find that public RPF is most helpful when the productivity metrics on which the relative rankings are based are ones that can be improved by the spread of best practices rather than reflective of differences in individual ability. In our study, we find that the productivity gains from feedback are much greater for the slowest workers than for the fastest workers. This is similar to Schultz et al. (1999) but in contrast to Ashraf et al. (2014) and Barankay (2012), which find that bottom-ranked workers become discouraged and expend less effort as a result of receiving feedback about their poor performance. The key difference between our study and these latter two studies is that in the latter, bottom-ranked workers are not equipped with efficiency tips on how to improve as they are in our study. Thus, if public RPF is to be leveraged to foster improvements in productivity, the metrics on which the relative rankings are based must be carefully selected to be ones that reflect the extent to which best practices are employed.

Finally, with regard to workflow management, our study suggests that when workers have discretion in how to carry out their work and prioritize among different tasks (Hopp et al. 2009), focusing on improving the management of workflow may be particularly useful. We build on Dobson et al.'s (2013) study of the importance of workflow rules around which patient to serve (e.g., to bring in a new patient or to discharge an existing patient) to show that workflow decisions within the course of caring for a given patient (e.g., ordering all tests at once rather than serially) can also have a significant impact on worker productivity.

2.6.2 Implications for Practice

Our study suggests ways in which organizations may improve performance by leveraging existing efforts around performance feedback. We propose that public RPF can equip workers to learn from their peers how to better manage their workflow, particularly when work tasks are not standardized. Because this type of best practice sharing is most likely to occur internally within an organization or work group as opposed to across organizational boundaries, future research could examine whether public RPF—in which relative performance data are made available to workers within an organization or work group—is more effective than public reporting—in which performance data are made available to the external public—at spreading best practices and improving performance.

Managers currently employing public RPF may be able to further leverage its productivity benefits by segmenting the relative performance distributions to reflect worker performance when carrying out standardized versus unstandardized work tasks. This may facilitate the identification of best practices for carrying out unstandardized work tasks by highlighting workers who do particularly well with those types of tasks.

In the ED setting in particular, we find significant productivity benefits associated with implementing public RPF. We find that each high acuity patient seen by an average physician at Treatment ED experiences a time savings of 17 minutes on average in terms of their MD processing time. For those seen by an initially bottom-ranked physician at Treatment ED, the corresponding time savings is 35 minutes on average. With approximately 140 high acuity patients presenting to the ED each day, this is roughly equivalent to an average time savings of 40 patient-hours per day. With 3 hours being the average MD processing time for each patient, this suggests that Treatment ED could see an additional 13 patients per day without investing in any additional resources. Given the significant need for efficiency gains in EDs across the country, these time savings and their associated cost implications have significant practical implications.

2.6.3 Limitations and Future Research

This study has limitations and its results should be interpreted accordingly. First, in this study, we cannot separately identify the effect of implementing public RPF from the effect of sharing top performers' efficiency tips. This is because efficiency tips were systematically shared at Treatment

ED only once public RPF enabled the identification of top performers, and not when RPF was disclosed privately. To assess whether these effects are complementary or independent, future work could identify the effects of the following four conditions in a laboratory setting: no efficiency tips with private RPF, no efficiency tips with public RPF, efficiency tips with private RPF, and efficiency tips with public RPF. Furthermore, future work could examine whether the credibility instilled in the efficiency tips by attributing them to high performers mediates the effect of public RPF on improved productivity, as is suggested by the interviews.

Second, because this study was conducted in the field, it is difficult to ascertain that the identities of coworkers remained completely anonymous under the private RPF condition. Because the code number associated with each physician typically did not change from month to month, it is possible for physicians to have had a reasonable guess regarding the identities of the top-performing physicians and to have asked these physicians for efficiency tips. Though these discussions did not happen in a structured manner during staff meetings when RPF was disclosed privately, they may have happened informally outside of these meetings given the improvement culture that existed at these EDs. Nevertheless, if this were the case, then our findings would be an underestimate of the actual effects associated with public RPF.

Third, due to restrictions on data access, we are limited in our ability to conclusively explain how the intervention affected the total number and type of tests ordered by physicians at Treatment ED. Because a system that is separate from the main electronic health record system captures the data on the number and type of tests ordered, and because we would need patient identifiers to link these two datasets, we were unable to access these data for this study. We leave further exploration of changes in test ordering behaviors as an area for future research.

Fourth, the generalizability of our findings with regard to the changes in care intensity may be limited due to the specifics of our field setting. Because the health system in which these EDs belong is a managed care entity, which is one that seeks to reduce unnecessary health care costs, a reduction in care intensity has *positive* revenue implications as opposed to negative revenue implications that would be typical in a fee-for-service system. Future work should examine how public RPF and other levers to improve productivity may affect the intensity of care in fee-for-service systems, in which services are paid for separately and therefore providers are incentivized to increase their care intensity.

2.6.4 Conclusions

In recent years, more organizations have been employing public RPF (Netessine and Yakubovich 2012). There are examples of such implementations not only in health care (Gorman 2015) but also in other service industries, such as the restaurant industry (Vanek Smith 2015). A deeper understanding of the conditions under which public RPF leads to improved operational performance is essential to determining when and how to adopt these practices in the hopes of improving productivity. Our paper provides initial evidence suggesting that public RPF may lead to improved productivity by enabling low-performing workers to better manage their workflow by adopting the best practices of their high-performing peers. This is essential to improving the productivity of many organizations in which it is difficult to standardize each specific work task. Furthermore, being able to improve the performance of the bottom tier of employees is an important objective in many industries, such as health care and education, where poor performance of any employee can have serious repercussions for their customers. We look to future work to identify other ways in which organizations can attain these goals without sacrificing quality.

CHAPTER 3

Cohort Turnover and Operational Performance:

The July Phenomenon in Teaching Hospitals

3.1 Introduction

Nearly all managers must deal with the consequences of employee turnover within their organizations. Despite the importance of this issue, academic attention has been disproportionately focused on the *causes* rather than *consequences* of turnover (Glebbeek and Bax 2004, Mobley 1982). One possible explanation for the limited number of studies on the effects of turnover is the difficulty in answering this question empirically. Turnover is typically an endogenous phenomenon that may occur for various reasons not observed by the researcher. For example, more productive workers may be more likely to remain with a company longer than less productive ones (Jovanovic 1979). Under such circumstances, it is difficult to make causal inferences concerning turnover's effect on performance using organization-level data. Those studies that do examine the consequences deliver mixed findings. Though some find that employee turnover exhibits a negative effect on performance through lower levels of productivity (Batt 2002), higher costs (Drexler and Schoar 2014), smaller profit margins (Ton and Huckman 2008), worse customer service (Kacmar et al. 2006, Ton and Huckman 2008), and lower job satisfaction (Whitt 2006), others find that the effect of turnover on performance depends on contextual factors such as how knowledge is embedded in an organization's structure (Hausknecht and Holwerda 2013, Rao and Argote 2006) or the degree of process conformance (Ton and Huckman 2008).

A second issue concerning the effect of turnover on performance is that turnover itself appears in multiple forms. Many organizations face a continuous stream of *individual turnover* in which employees leave and are replaced by new workers at various points throughout the year. In such settings, there is no one particular time during the year when managers are required to train and orient a large portion of their workforce. In contrast, other organizations bring on new employees in large numbers at discrete points in the year. For example, law and consulting firms tend to start most of their new employees in late summer or early fall. These new employees must all be trained and integrated into the firm at one time. In the law and consulting examples, the potential negative effects of the large inflow of new workers may be buffered by the fact that firms do not face the simultaneous exit of large portions of their experienced workers. Rather, these departures occur in a roughly continuous manner throughout the year.

An extreme, though not uncommon, form of this discrete scenario is what we term *cohort turnover*—the planned simultaneous exit of a large number of experienced employees and a similarly sized entry of new workers—and serves as the focus of this study. Cohort turnover is related to, but distinct from, *collective turnover*, which refers to the aggregate departure of employees within an entity regardless of whether the departure was planned in advance or was accompanied by a similarly sized entry of newcomers (Hausknecht and Trevor 2011). Examples of cohort turnover can be found in changeovers that occur between military units in combat, political administrations, and residents and fellows in teaching hospitals. Given the number of individuals transitioning into or out of employment at a specific point in time, cohort turnover raises concerns about adverse effects on operational performance due to factors such as lower levels of team familiarity (Huckman and Staats

2011, Huckman et al. 2009), operational disruption (Krueger and Mas 2004) or the loss of the tacit knowledge (Polanyi 1966) held by departing workers. However, because cohort turnover is, by definition, anticipated and supervisors typically remain in place, one might expect this particular type of turnover not to have a negative impact on operational performance. In short, the impact of cohort turnover on operational performance is not obvious on an *a priori* basis.

In this paper, we consider cohort turnover among resident physicians in teaching hospitals. Residency represents a new physician's first assignment following medical school and typically lasts from three to five years depending on the physician's area of specialization. At (or slightly before) the beginning of every July, the most senior residents move on to permanent medical positions or fellowships (for further training in a sub-specialty) at other hospitals, and recent medical school graduates arrive as first-year residents. Every summer, this cohort turnover leads to a discrete reduction in the average experience of the labor force at teaching hospitals. In addition, it may disrupt established teams of physicians and other caregivers within hospitals. Although most attending physicians and nurses who supervise and serve as an operational safeguard for the work of residents typically do not change roles at this time of year, either of the above effects may have negative consequences for the two major determinants of a hospital's operational performance: resource utilization (a proxy for cost) and clinical outcomes (a measure of quality). This cohort turnover, colloquially referred to as the "July phenomenon," is often mentioned in the lore of medical professionals. To date, the clinical literature presents a mixed and inconclusive picture of whether the July phenomenon exists, and if so, for which outcomes.

We examine the impact of the July turnover on hospital operational performance using data on all patient admissions from a large, multi-state sample of U.S. hospitals over a 12-year period. When comparing trends in teaching hospitals to those in non-teaching (i.e., control) hospitals over the course of the year, we find a significant increase in resource utilization (as measured by risk-adjusted average length of stay (LOS)) associated with the cohort turnover. These effects are increasing in a hospital's level of teaching intensity (as measured by the number of residents per hospital bed), which reflects the degree to which a hospital relies on residents. In terms of clinical quality, we find limited evidence of a decrease in performance (as measured by risk-adjusted mortality rates) in hospitals with a higher level of teaching intensity (i.e., major teaching hospitals), but not in those with a lower level of teaching intensity (i.e., minor teaching hospitals). In major teaching hospitals, we also find evidence of a substantial anticipation effect that manifests as a gradual decrease in operational performance that begins as early as the March prior to a given cohort turnover in July. This anticipation effect may result from a transition of responsibilities in the last several months of the academic year that is not coupled with the same degree of precaution that accompanies the cohort turnover in July.

Using additional data on hospital nursing quality and staffing levels, we also identify managerial levers that hospitals may use to mitigate their decrease in operational performance both at the time of and in the months leading up to the July turnover. Specifically, we find that hospitals may be able to dampen these negative effects on operational performance by improving their overall quality of nursing and increasing their intensity of quality assurance related to the work of the residents. Though neither of these levers is costless to implement, each has significant and lasting implications for improved hospital productivity and quality beyond just the month of July.

This paper contributes to the management literature on turnover and performance by empirically examining the effects of cohort turnover. We define cohort turnover as a distinct phenomenon and find that it leads to greater resource utilization but not necessarily a lower level of clinical quality in teaching hospitals relative to non-teaching hospitals. The limited effect on clinical quality may be attributable to the fact that a cohort turnover event is anticipated, thereby allowing hospitals to prepare for it. Consistent with this possibility, we identify the presence of a substantial anticipation effect, which is important for an organization to acknowledge and prepare for in assessing the impact of cohort turnover on operational performance. We identify structures and processes that can facilitate knowledge transfer from departing to entering workers and, ultimately, mitigate the negative effects of cohort turnover. An improved understanding of the implications of cohort turnover may help inform analytical models of staffing and scheduling, which tend not to address the operational costs of turnover and tend to model turnover as a stochastic event (Boudreau et al. 2003, Gans and Zhou 2002).

3.2 Cohort Turnover and Performance

Cohort turnover is distinct from individual and collective turnover in that it is planned in advance, happens on a large scale, and involves the simultaneous exit of a large number of experienced employees and a similarly sized entry of new workers. These changeovers occur as a matter of policy, regardless of the underlying productivity of the workers involved. Given the large-scale nature of cohort turnover, it has the potential to affect an organization's operational performance significantly. Yet there is little known about the effects of cohort turnover on operational performance, as much of the prior work on turnover in the management literature considers only individual turnover (Campbell et al. 2012).

One reason for the lack of attention to cohort turnover may be the fact that it occurs less frequently than individual turnover. Nevertheless, it takes place in several important settings beyond annual resident turnover in teaching hospitals, including military deployments, changes of political administrations, and labor strikes. A second reason may be the assumption that the answers to questions concerning its effects are obvious; given the sheer magnitude of the change brought by cohort turnover, one might assume that it *must* have a detrimental impact on operational performance. Despite its magnitude, however, cohort turnover often occurs in a predictable fashion and the affected organizations should, theoretically, have time to anticipate and prepare for it. For example, attending physicians and nurses in teaching hospitals—being aware of the turnover that occurs each July—may focus more intently on supervising and checking the quality of the work of new residents at that time of the year. As a result, the impact of cohort turnover in settings where the formal and informal supervisory staff does not change is not obvious on an *a priori* basis.

In the management literature, there are arguably two analogs to cohort turnover that have been previously studied. One is the turnover of CEOs and top management teams (Cao et al. 2006, Messersmith et al. 2014, Tushman and Rosenkopf 1996). Though this involves the turnover of a limited number of individuals, the significance of their roles within the organization suggests that such turnover may have a significant impact on an organization's operational performance. Nevertheless, perhaps due to the endogenous nature of this type of turnover event, this relationship has been difficult to identify and empirical findings to date deliver mixed results. Another is collective turnover, which refers to the aggregate departure of employees regardless of whether the departure was planned or accompanied by a similarly sized entry of newcomers (Hausknecht and Trevor 2011). This construct has yet to be studied empirically, though one conceptual paper has suggested five potential moderators of the relationship between collective turnover and performance: leaver proficiencies, time dispersion, positional distribution, remaining member proficiencies, and newcomer proficiencies (Hausknecht and Holwerda 2013).

Looking beyond the management literature, there is a body of research in the medical literature that examines the effects of the annual resident turnover in teaching hospitals on various clinical outcomes. To date, this literature presents mixed and inconclusive findings (Young et al. 2011). While several studies suggest that patients admitted in July have similar mortality outcomes (van Walraven et al. 2011) and morbidity outcomes (Ford et al. 2007) as patients presenting in other months, other studies show the opposite, suggesting that patients exhibit worse outcomes in July in terms of mortality rates (Englesbe et al. 2007), morbidity (Haller et al. 2009), medication error rates (Phillips and Barker 2010), and hospital-acquired complications (Wen et al. 2015). In addition to these mixed findings, several of the individual papers in this literature face either methodological or contextual limitations, which we address below.

As a whole, the existing management and medical literature thus suggests that cohort turnover could have either a positive or negative overall impact on operational performance. On one hand, cohort turnover may have a negative impact on operational performance, given that several prior studies have found a negative relationship between individual turnover and productivity (Huselid 1995, Staw 1980). When individuals leave an organization and others must be recruited in their place, the organization may experience a decline in operational performance due to the resulting operational disruption and lower level of team familiarity (Huckman and Staats 2011, Huckman et al. 2009, Krueger and Mas 2004, Staw 1980). This may interfere with organizational learning to the extent that locally acquired knowledge is difficult to disseminate and lower levels of experience lead to lower levels of productivity (Adler and Clark 1991, Argote et al. 2003, Lapré et al. 2000). In addition, organizations may experience further declines in operational performance due to the demoralization of remaining workers and the costs incurred in selecting, recruiting, and training new workers (Staw 1980). If individual turnover results in these negative effects on operational performance, the turnover of a large cohort of individuals may amplify the magnitude of this negative effect.

On the other hand, because cohort turnover is a predictable phenomenon that is planned far in advance, organizations may be able to preempt and mitigate these negative effects on operational performance. For example, organizations may systematically transfer information from departing workers to new workers in an attempt to smooth the transition (Drexler and Schoar 2014). Organizations may also temporarily increase their staffing or heighten the intensity of quality assurance related to the work of new employees around the time of the cohort turnover to buffer against the anticipated loss of productivity. Furthermore, cohort turnover may lead to an overall *increase* in performance to the extent that outgoing workers may be experiencing burnout and new workers may be motivated to exert high levels of effort (Staw 1980). Given that operational performance is a function of not only skill but also effort, turnover may improve average operational performance (Dalton and Todor 1979, Staw 1980). In fact, some prior work suggests that operational performance can degrade when turnover is *too low* due to employee stagnation (Abelson and Baysinger 1984, Glebbeek and Bax 2004, Whitt 2006).

We expect that both the negative and positive effects of cohort turnover may exist concurrently. Nevertheless, on average we expect the negative effects to dominate because several of the potential positive effects are conditional on an organization taking proactive measures to mitigate the anticipated negative effects. Therefore, assuming that organizations are likely to take the path of least resistance and behave passively, we hypothesize the following:

Hypothesis 1: Cohort turnover has a negative impact on operational performance.

Next, we consider the magnitude of this effect as a function of the size of the cohort turnover. By definition, cohort turnover occurs on a large scale with a group of workers turning over at a pre-specified time. If there is, indeed, an overall negative effect of cohort turnover on operational performance, we would expect the magnitude of this effect to be increasing in the size of the cohort relative to the size of the organization. This is because a larger relative cohort would indicate that a greater proportion of the organization's workforce is turning over. In fact, it may be the case that any negative effect of cohort turnover on operational performance only appears when the relative size of the cohort exceeds a certain threshold. We hypothesize:

Hypothesis 2: Cohort turnover has a greater negative impact on operational performance in organizations that experience a greater relative magnitude of cohort turnover.

If cohort turnover has an overall negative effect on an organization's operational performance, how might organizations be able to mitigate these effects? One way to buffer against this negative impact may be to adopt organizational structures and processes that facilitate the transfer of knowledge from departing to entering workers. These structures and processes include, but are not limited to, the greater use of standard operating procedures (Ton and Huckman 2008), an increase in the quality of workers who are not turning over and can transfer knowledge to new workers, and an increase in the intensity of quality assurance by workers who are not turning over and can serve as an operational safeguard for the work of new workers. We thus hypothesize the following mitigation effect:

Hypothesis 3: The negative impact of cohort turnover on operational performance is less pronounced in organizations with structures and processes that facilitate knowledge transfer from departing to entering workers.

3.3 Setting and Data

3.3.1 Research Setting

We study cohort turnover in the context of teaching hospitals, which have two primary objectives: the provision of high quality medical care and the training of new physicians. These related but distinct objectives overlap within residency programs. During their three to five years of residency at roughly 1,000 teaching hospitals in the country, residents represent an important piece of a hospital's system for delivering care (Association of American Medical Colleges 2009).

Patient care in teaching hospitals is provided by teams of medical professionals that include attending physicians, nurses, fellows, residents, and medical students. Much of the care is delivered by a resident, who supervises medical students and is supervised by an attending physician with or without additional supervision from a more senior resident or a fellow. Nurses also play a key role in delivering patient care and they help new residents learn how to deliver care effectively, efficiently, and safely. The daily activities of residents include admitting, diagnosing, treating, and discharging patients.

Residency programs in the United States are structured like schools. Each class of residents enters together at the beginning of the academic year, and senior members of each program graduate together at the end of the academic year. For most residency programs, the year officially begins on July 1 and ends the following June 30, though the annual transition does not occur all on one day. Typically, hospitals complete the transition over a two-to-three week period, lasting from the middle of June through the first week of July. This turnover creates potential transitional challenges in teaching hospitals—even for residents in the middle years of their programs—as each cohort of physicians becomes comfortable with new roles and responsibilities.

3.3.2 Data

The primary source of data for this analysis is the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS) for each year from 1997 to 2008. NIS contains discharge-level data for all inpatient cases at a sample of roughly 20% of the community hospitals[1] in the United States. Depending on the year, NIS includes information for hospitals from between 22 and 42 states (Agency for Healthcare Research and Quality 2013). For each patient, NIS provides information on patient age and gender, admission source, expected primary payer (i.e., Medicare, Medicaid, private including HMO, self pay, no charge, and other), LOS, in-hospital mortality, diagnosis-related group (DRG), and comorbidities.

We link the NIS data with information from each corresponding year's AHA Annual Survey of

---

[1] Community hospitals are defined by the NIS and the American Hospital Association (AHA) as "…'all nonfederal, short-term, general, and other specialty hospitals, excluding hospital units of institutions.' Included… are specialty hospitals such as obstetrics-gynecology, ear-nose-throat, short-term rehabilitation, orthopedic, and pediatric. Excluded are long-term hospitals, psychiatric hospitals, and alcoholism/chemical dependency treatment facilities" (Healthcare Cost and Utilization Project 1999).

Hospitals, which includes data on the operating and financial characteristics for nearly all of the more than 5,000 acute care hospitals in the United States. In addition to several other items, the AHA data provide information on the number of hospital beds and full-time equivalents (FTEs) for residents and nurses at each facility in a given year. Using this information, we are able to construct a measure of teaching intensity, which captures the relative magnitude of the cohort turnover at a given hospital. We calculate teaching intensity as the number of FTE residents per hospital bed. We note that because the AHA data on the number of FTE residents are only available at the hospital level and not at the level of the specialty within a hospital, we construct our measure of teaching intensity at the hospital level so that there is a "match" in the level at which the key independent and dependent variables are observed.

Though we do not have readily available measures of the use of standard operating procedures, we do have access to two proxies for assessing the structures and processes that facilitate knowledge transfer from departing to entering workers. First, we use the quality of nursing care at a hospital as a proxy for the quality of workers who are not turning over and can transfer knowledge to new workers. The quality of nursing care at a hospital is an important and relevant measure because it reflects the average capability of the nurses who help new residents learn how to deliver care effectively, efficiently, and safely while the new residents are still building familiarity with their new roles. For this measure, we use data from the American Nurses Credentialing Center's (ANCC) magnet recognition program. According to the ANCC, magnet recognition is a nationally recognized credential for quality in patient care and excellence in nursing that arises from "transformational leadership, structural empowerment, [and] exemplary professional practice" (American Nurses Credentialing Center 2015). As of July 2015, approximately 400 of the 5,000 hospitals in the United States had received magnet recognition. Once recognized, the credential lasts for four years, and most hospitals are successful in getting the credential renewed from that point forward.

Our second proxy is the relative staffing level of nurses to residents at a given hospital. This measure captures the potential intensity of quality assurance by workers who are not turning over and can serve as an operational safeguard for the work of new residents. We use a nursing-related measure because nurses play an important role in assuring the quality of resident work. This is due to the fact that nurses have both clinical and hospital-specific process knowledge that is likely helpful to new residents. For example, nurses in teaching hospitals work closely with residents and have many opportunities to guide new residents in preparing patients for procedures, double-checking the dosages of medications ordered, engaging in challenging conversations with patients and families, and other duties. To operationalize this measure, we use the AHA data to construct a measure of the intensity of quality assurance, which we calculate as the ratio of FTE nurses to FTE residents. Although this measure reflects only the *potential* (not actual) level of quality assurance by nurses, it represents a reasonable approximation of quality assurance, as the majority of nurses in teaching hospitals work closely with residents (Vallis et al. 2004).

3.4 Empirical Methodology

To examine the impact of the July turnover on hospital performance, as measured by resource utilization and clinical quality, we employ a difference-in-differences framework that compares changes in our operational performance measures over the course of the year in teaching hospitals relative to those in the baseline of non-teaching hospitals. Unlike individual and collective turnover, the exogenous nature of cohort turnover allows for the use of this approach to identify relative changes in LOS (proxy for resource utilization) and mortality rate (proxy for clinical quality) over the course of the year. This method of using non-teaching hospitals as a control group enables us to separate changes that are driven by the cohort turnover in July from those that are driven by unobserved factors—such as illness severity and seasonal variation—that are also present in non-teaching hospitals.

3.4.1 Hospital Categories

The source of identification in our empirical analysis is the varying degree to which certain types of hospitals rely on residents. To test Hypothesis 1, we divide hospitals into two categories: non-teaching hospitals and teaching hospitals. To test Hypotheses 2 and 3, we further divide teaching hospitals into two sub-categories: minor teaching hospitals and major teaching hospitals.

Non-teaching hospitals are those not listed as teaching hospitals in the NIS. These facilities have very few, if any, residents. As such, we would not expect them to be affected by the cohort turnover in July. Minor teaching hospitals are those that are listed as teaching hospitals in the NIS and have teaching intensities (i.e., FTE residents per hospital bed) that are less than 0.25. Major teaching hospitals are those facilities listed in the NIS as teaching hospitals and with teaching intensities equal to or greater than 0.25. This threshold for teaching intensity is used by the Medicare Payment Advisory Commission (MedPAC) to distinguish minor and major teaching facilities (MedPAC 2002). We repeat our analyses using Jena et al.'s (2013) higher threshold of 0.60 residents per bed as the boundary between minor and major teaching hospitals. Due to the small percentage of hospitals (2.3% of total hospitals) that are considered major teaching hospitals under this latter definition, we maintain MedPAC's 0.25 threshold. We note, however, the lack of substantive difference in our main findings when using either of these threshold values.

We present descriptive statistics for each of the three hospital categories as well as for the entire sample in Table 3.1. The first row illustrates the differences in average teaching intensity across the three groups. The average teaching intensities of non-teaching and minor teaching hospitals are similar (0.02 and 0.07, respectively) while that of major teaching hospitals (0.55) is substantially larger than that for either of the other two categories. In terms of both measures of facility size—

hospital beds and admissions per year—hospitals get progressively larger as teaching intensity increases.

Table 3.1. Descriptive statistics by hospital type

	Non-Teaching		Minor Teaching		Major Teaching		Full Sample	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Residents Per Inpatient Bed	0.02	0.04	0.07	0.09	0.55	0.33	0.13	0.25
Inpatient Hospital Beds	224	142	402	236	580	244	348	239
Inpatient Admissions/Year	10,824	7,423	19,719	11,001	28,718	12,281	17,057	11,824
Patient Age	49.3	8.4	46.5	9.2	43.6	8.5	47.3	9.0
Medicaid Admissions /Total Admissions	17%	12%	16%	13%	24%	15%	18%	13%
Medicare Admissions /Total Admissions	40%	13%	34%	12%	29%	11%	36%	13%
Risk-adjusted Average Length of Stay (Days)	4.4	0.9	4.7	0.7	5.2	0.8	4.7	0.9
Risk-adjusted Mortality Rate	2.1%	0.6%	2.1%	0.5%	2.2%	0.5%	2.1%	0.6%
Observations (hospital-years)	5,537		1,369		431		7,337	
Percentage of Total Sample	75.5%		18.7%		5.9%		100.0%	

Notes. Observations are at the hospital-year level and cover the 12-year period from 1997 to 2008.
Source: NIS, 1997-2008.

3.4.2 Risk Adjustment of Dependent Variables

To account for systematic differences in the level of patient severity at non-teaching, minor teaching, and major teaching hospitals, we risk adjust the dependent variables: LOS and mortality rate. Table

3.1 illustrates that such differences in patient characteristics exist across hospital types in our sample. Rows (4) – (6) of Table 3.1 illustrate that teaching intensity is correlated with the demographics of a hospital's patient base. In particular, non-teaching hospitals attract older patients (49.3 years) than either type of teaching hospital (46.5 and 43.6 years for minor and major teaching hospitals, respectively), possibly due to the differences in location of the different types of hospitals, on average. In addition to having younger patients, major teaching hospitals also have a higher percentage of Medicaid patients than the other groups. These relationships are consistent with the fact that many teaching hospitals are located in densely populated cities.

We note that, to the extent that these differences in risk across types of hospitals remain constant over the course of the calendar year, the risk adjustment we perform using the clinical and demographic characteristics of individual patients would not be necessary to identify the effect of cohort turnover. It is possible, however, that risk differences across types of hospitals are *not* constant over the calendar year. For example, to the degree that, within older populations, relatively healthy individuals tend to move from cold climates in northeastern states—which tend to have a high concentration of teaching hospitals—to warmer southern and western states during the winter months, the age-adjusted mortality risk for the hospitalized population in the northeast will increase *ceteris paribus* during this period of the year. Our approach to risk adjustment using individual clinical and demographic characteristics addresses this concern.

To risk adjust our dependent variables, we adopt an approach that has been widely used in the operations management literature on productivity and quality of care (Huckman and Pisano 2006, Huckman 2003, KC and Staats 2012). As covariates in our equations to calculate expected LOS and mortality, respectively, we include patient age, age squared, gender, and an indicator for Medicaid as the primary payment source. The Medicaid variable is included as a proxy for the patient's socioeconomic status. We also include each patient's primary diagnosis or procedure as captured by one of more than 400 diagnosis-related groups (DRGs) assigned by the Centers for Medicare & Medicaid Services; our regressions include fixed effects for each of these DRGs. Finally, we include the Charlson index, a measure of comorbidities that increase a patient's risk of mortality (Charlson et al. 1987). Given our DRG fixed effects, the Charlson index captures the within-diagnosis severity of a patient's illness.

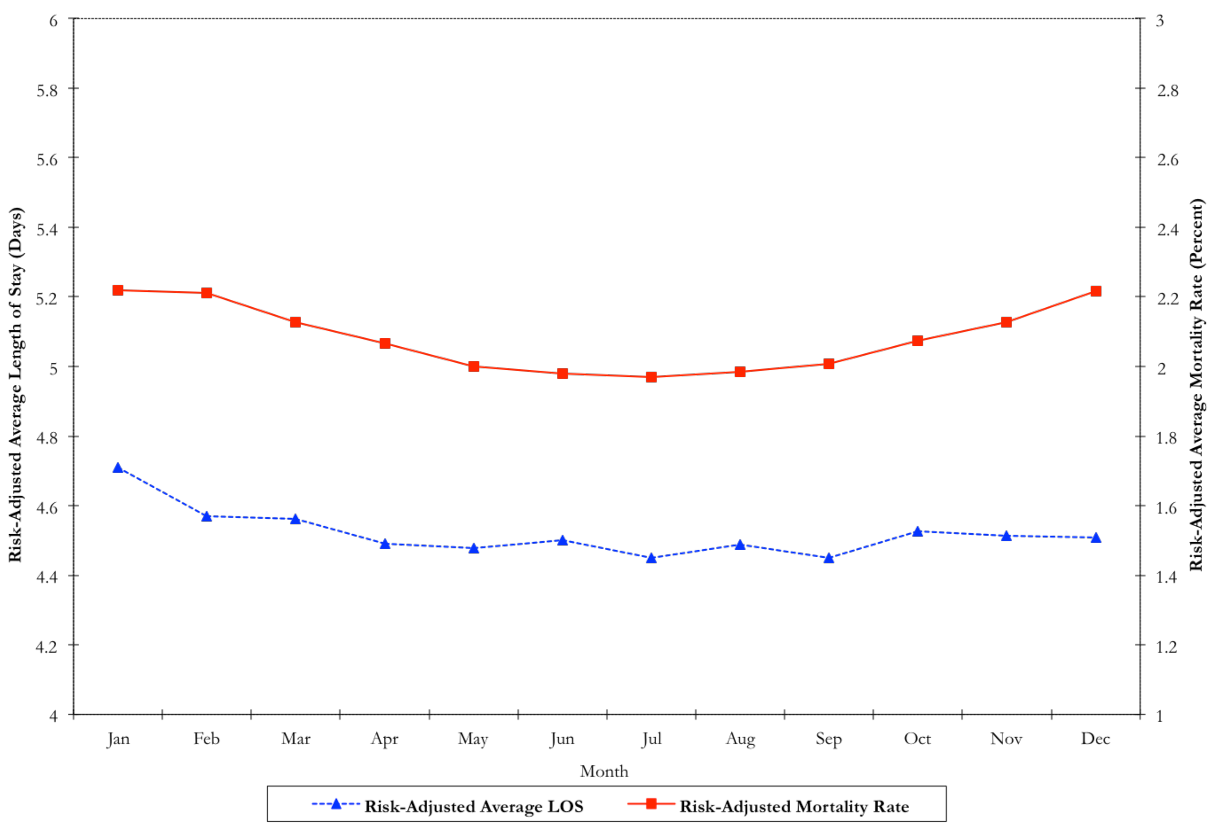
For LOS, we use a simple linear regression to calculate expected values. Given that the in-hospital mortality variable is binary, we use logistic regression to obtain the estimated probability of death for each patient discharge. These equations are run separately for each calendar year for computational ease, as each year has approximately 6 million observations. The observed and expected values for LOS and mortality are then averaged by hospital and month. The risk-adjusted value of each dependent variable is calculated as the ratio of the observed-to-expected rate for a given hospital-year. For example, the risk-adjusted LOS ($RALOS_{h,m,t}$) is:

$$RALOS_{h,m,t} = \frac{OLOS_{h,m,t}}{ELOS_{h,m,t}} \times \overline{OLOS_t} \tag{3.1}$$

where $OLOS_{h,m,t}$ and $ELOS_{h,m,t}$ are the observed and expected LOS, respectively, for hospital h in month m of year t. $\overline{OLOS_t}$ is the average observed LOS for the entire sample in year t and is used to normalize the value of $RALOS_{h,m,t}$.

In rows (7) and (8) of Table 3.1, we present the risk-adjusted average LOS and mortality rate, respectively, for each type of hospital. We find that risk-adjusted average LOS increases with teaching intensity. This trend is consistent with the claim that major teaching hospitals tend to attract the most complex cases among the three groups. In addition, we find that the risk-adjusted mortality rate is higher in major (2.2%) teaching facilities compared to non- and minor teaching facilities (2.1%).

In Figure 3.1, we see that risk-adjusted LOS and risk-adjusted mortality vary over the course of the calendar year. Risk-adjusted mortality is relatively high in the winter months, declines until the summer months, and then begins increasing during the fall. This pattern has been noted by epidemiologists (e.g., Gemmell et al. 2000) and attributed to a range of factors including the impact of seasonal disease (e.g., influenza and respiratory illness). Similarly, risk-adjusted LOS also shows seasonal patterns. Key to the empirical strategy in our paper is the use of non-teaching hospitals as a control for these seasonal changes in outcomes, which should affect all hospitals regardless of teaching status. With this approach, we can calculate "de-seasoned" trends in risk-adjusted LOS and risk-adjusted mortality for teaching hospitals to determine the potential effect of the July turnover.



Source. NIS, 1997-2008.

Figure 3.1. Risk-adjusted average length of stay and mortality rate by month, 1997-2008

3.4.3 Empirical Specification

Our analyses rely on a difference-in-differences framework that follows the relative changes in risk-adjusted average LOS and risk-adjusted mortality for the different groups of hospitals over the course of the year. To test Hypothesis 1, we estimate the following model:

$$Y_{h,m,t} = \alpha_h + \delta_t + \mu_m + \sum_{m=1}^{12} \beta_m \cdot \left(\mu_m \times TCH_{h,m,t} \right) + \varepsilon_{h,m,t} \tag{3.2}$$

Y represents the dependent variable of interest (i.e., risk-adjusted average LOS or risk-adjusted mortality), α_h is a vector of hospital fixed effects, δ_t is a vector of year fixed effects, and μ_m is a vector of fixed effects for each month of the year. *TCH* is an indicator for teaching hospitals, which is interacted with each of the month fixed effects. The main effect on *TCH* is absorbed by the hospital fixed effects because specific hospitals are classified as either non-teaching or teaching hospitals for the entirety of our sample period.[2] The coefficients β_m on the month-*TCH* interactions capture the extent to which any seasonal pattern that is found for teaching hospitals differs from that for the non-teaching controls. Given that the residency changeover begins in late June for many hospitals, we compare the change in the dependent variable from May (reference month) to July for teaching hospitals to the similar change for non-teaching hospitals to measure the impact of the July turnover. This difference is captured by β_7, the coefficient on $\mu_7 \times TCH$. Each of the hospital-month level observations is weighted by the total number of cases for the hospital-month pair to account for the fact that all values of the dependent variable are averages. We analyze the data at the hospital-month level for computational tractability, as we would otherwise have approximately 77 million observations. Finally, the standard errors are clustered by hospital to address potential lack of independence in the error term, $\varepsilon_{h,m,t}$.

[2] Some hospitals, though very few, move across the threshold between minor and major teaching hospitals during our sample period. We assigned these hospitals to the category of teaching intensity that matched the majority of the months for which they were present in our sample.

To test Hypothesis 2, we estimate the following model:

$$Y_{h,m,t} = \alpha_h + \delta_t + \mu_m + \sum_{m=1}^{12} \gamma_{1,m} \cdot \left(\mu_m \times MIN\_TCH_{h,m,t}\right) + \sum_{m=1}^{12} \gamma_{2,m} \cdot \left(\mu_m \times MAJ\_TCH_{h,m,t}\right) + \varepsilon_{h,m,t} \quad (3.3)$$

Y, α_h, δ_t, μ_m, and $\varepsilon_{h,m,t}$ remain the same as in Equation (3.2). The two summation terms are vectors of interactions between the indicators for minor and major teaching hospitals, respectively, and the month effects. Analogous to the case in Equation (3.2), the main effects on minor and major teaching hospitals (*MIN_TCH* and *MAJ_TCH*, respectively) are absorbed by the hospital fixed effects because specific teaching hospitals are classified as either minor or major teaching hospitals for the entirety of our sample period. Here, the coefficients $\gamma_{1,m}$ ($\gamma_{2,m}$) on the $\mu_m \times MIN\_TCH$ ($\mu_m \times MAJ\_TCH$) interactions capture the extent to which any seasonal pattern that is found for minor (major) teaching hospitals differs from that for non-teaching controls. We examine the coefficients $\gamma_{1,7}$ and $\gamma_{2,7}$ on $\mu_7 \times MIN\_TCH$ and $\mu_7 \times MAJ\_TCH$, respectively, to determine whether the July turnover more negatively impacts the productivity of teaching hospitals with higher teaching intensity than those with lower teaching intensity.

To test Hypothesis 3 using our first proxy, we estimate the following model:

$$\begin{aligned} Y_{h,m,t} = {} & \alpha_h + \delta_t + \mu_m + \sum_{m=1}^{12} \lambda_{1,m} \cdot \left(\mu_m \times MIN\_TCH_{h,m,t}\right) \\ & + \sum_{m=1}^{12} \lambda_{2,m} \cdot \left(\mu_m \times MAJ\_TCH\_NONMAGNET_{h,m,t}\right) \\ & + \sum_{m=1}^{12} \lambda_{3,m} \cdot \left(\mu_m \times MAJ\_TCH\_MAGNET_{h,m,t}\right) + \varepsilon_{h,m,t} \end{aligned} \quad (3.4)$$

Y, α_h, δ_t, μ_m, *MIN_TCH*, and $\varepsilon_{h,m,t}$ remain the same as in Equation (3.3). We stratify major teaching hospitals into those that have never been (*MAJ_TCH_NONMAGNET*) and those that have been (*MAJ_TCH_MAGNET*) magnet certified since the beginning of the recognition program in 1991. ANCC magnet recognition serves as a proxy for organizational structures and processes that facilitate knowledge transfer from departing to entering workers that is independent of a hospital's staffing levels. Here, the coefficients $\lambda_{1,m}$ on the $\mu_m \times MIN\_TCH$ interactions are identical to the coefficients $\gamma_{1,m}$ on the $\mu_m \times MIN\_TCH$ interactions in Equation (3.3) because the stratification of the major teaching hospitals does not affect the minor teaching hospitals. The coefficients $\lambda_{2,m}$ ($\lambda_{3,m}$) on the $\mu_m \times MAJ\_TCH\_NONMAGNET$ ($\mu_m \times MAJ\_TCH\_MAGNET$) interactions capture the extent to which any seasonal pattern that is found for major teaching hospitals without (with) magnet recognition differs from that for non-teaching controls. We examine the coefficients $\lambda_{2,7}$ and $\lambda_{3,7}$ on $\mu_7 \times MAJ\_TCH\_NONMAGNET$ and $\mu_7 \times MAJ\_TCH\_MAGNET$, respectively, to determine whether the July turnover differentially impacts the performance of major teaching hospitals without and with ANCC magnet recognition.

To test Hypothesis 3 using our second proxy, we estimate the following model:

$$\begin{aligned}
Y_{h,m,t} &= \alpha_h + \delta_t + \mu_m + \rho_1 \cdot MAJ_TCH_LO_QA_{h,m,t} \\
&\quad + \rho_2 \cdot MAJ_TCH_HI_QA_{h,m,t} + \sum_{m=1}^{12} \rho_{3,m} \cdot \left(\mu_m \times MIN_TCH_{h,m,t}\right) \\
&\quad + \sum_{m=1}^{12} \rho_{4,m} \cdot \left(\mu_m \times MAJ_TCH_LO_QA_{h,m,t}\right) \\
&\quad + \sum_{m=1}^{12} \rho_{5,m} \cdot \left(\mu_m \times MAJ_TCH_HI_QA_{h,m,t}\right) + \varepsilon_{h,m,t}
\end{aligned} \tag{3.5}$$

Y, α_h, δ_t, μ_m, MIN_TCH, and $\varepsilon_{h,m,t}$ remain the same as in Equations (3.3) and (3.4). We stratify major teaching hospitals into those with lower and higher intensities of potential quality assurance ($MAJ\_TCH\_LO\_QA$ and $MAJ\_TCH\_HI\_QA$, respectively). The final two summation terms are vectors of interactions between the two sub-categories of major teaching hospitals and the month effects. The coefficients $\rho_{4,m}$ ($\rho_{5,m}$) on the $\mu_m \times MAJ\_TCH\_LO\_QA$ ($\mu_m \times MAJ\_TCH\_HI\_QA$) interactions capture the extent to which any seasonal pattern that is found for major teaching hospitals with low (high) intensities of quality assurance differs from that for non-teaching controls. We examine the coefficients $\rho_{4,7}$ and $\rho_{5,7}$ on $\mu_7 \times MAJ\_TCH\_LO\_QA$ and $\mu_7 \times$

MAJ_TCH_HI_QA, respectively, to determine whether the July turnover differentially impacts the performance of major teaching hospitals with lower versus higher intensities of quality assurance.

We use the median of FTE nurses per FTE resident as the boundary between major teaching facilities with lower versus higher intensities of quality assurance. We also run a version of this analysis using thresholds of thirds rather than halves and note the lack of any substantive differences in our findings using either of these threshold values.

3.5 Results and Discussion

3.5.1 The Impact of Cohort Turnover on the Operational Performance of Teaching Hospitals

Table 3.2 presents the coefficients $\beta_{1,m}$ on the month-*TCH* interactions from Equation (3.2), which represent the change in the dependent variable in each month relative to May for all teaching hospitals relative to the same change for non-teaching hospitals. A positive coefficient $\beta_{1,m}$ indicates that, on average, teaching hospitals experience a larger increase in the outcome measure in a particular month (relative to May) than do non-teaching hospitals in the same month (also relative to May). For example, in column (1), the value of 0.036 for the coefficient on teaching hospitals in September, $\beta_{1,9}$, suggests that the change in LOS from May to September is 0.036 days greater for teaching hospitals than for non-teaching hospitals. To simplify the discussion, we do not always reiterate that the effects and coefficients being discussed are relative to non-teaching hospitals, but this should be assumed. We also use "LOS" to refer to risk-adjusted LOS and "mortality" to refer to risk-adjusted mortality.

Risk-adjusted LOS. In terms of LOS (Table 3.2 column (1)), the coefficient $\beta_{1,7}$ on teaching hospitals in July is 0.034 and is significant at the 1% level. This result offers support for Hypothesis

1: the cohort turnover of residents in July is associated with a negative performance effect with respect to resource utilization as measured by LOS. Given the average LOS for teaching hospitals is 4.84 days, these results suggest that costs increase by roughly 0.7% in teaching hospitals following the July turnover if we assume that LOS is proportional to hospital costs (Fine et al. 2000).

For each month in July through September, the estimated coefficients $\beta_{1,m}$ on teaching hospitals remain significantly greater than the May baseline. By October, the change in LOS for teaching hospitals relative to non-teaching hospitals returns to being statistically indistinguishable from the May baseline. The coefficients $\beta_{1,m}$ on teaching hospitals in August and September are each not

Table 3.2. Effects of cohort turnover on hospital performance, using single teaching category

	Change in Dependent Variable Relative to Non-Teaching Baseline (Reference=May)			
	Risk-Adjusted LOS		Risk-Adjusted Mortality	
Teaching				
January	-0.014	(0.012)	-0.035	(0.022)
February	-0.019	(0.011) *	-0.064	(0.021) ***
March	-0.012	(0.010)	-0.029	(0.021)
April	-0.009	(0.009)	-0.027	(0.021)
June	0.001	(0.014)	0.021	(0.019)
July	0.034	(0.010) ***	0.035	(0.020) *
August	0.033	(0.010) ***	0.014	(0.019)
September	0.036	(0.010) ***	0.022	(0.020)
October	0.011	(0.012)	0.016	(0.019)
November	0.015	(0.011)	-0.010	(0.019)
December	0.003	(0.012)	-0.010	(0.021)
Mean of Dependent Variable				
Teaching	4.84		2.09	
Observations	87,707		87,707	
Adjusted R^2	0.719		0.335	

Notes. *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The level of observation is the hospital-month. All regressions include fixed effects for hospital, year, and month, though these coefficients are not shown in the table for ease of presentation. Standard errors (in parentheses) are heteroskedasticity robust and clustered by hospital. Observations are weighted by the total number of cases for the relevant hospital-month.

significantly different from that for teaching hospitals in July, which suggests that relative LOS for teaching hospitals increases in July and remains at that higher level for a few months. In addition, each of the coefficients on teaching hospitals in December through April is significantly smaller than each of the July through September coefficients on teaching hospitals at the 1% level. This general reduction in the estimated coefficients, $\beta_{1,m}$, over the course of the academic year suggests that residents at teaching hospitals may benefit from experience-based improvement in performance (i.e., learning) as measured by resource utilization (Reagans et al. 2005).

Risk-adjusted mortality. In addition to resource utilization, we also consider the impact of the July turnover on clinical quality. Column (2) of Table 3.2 presents the coefficients $\beta_{1,m}$ on the month-*TCH* interactions from our estimation of Equation (3.2) using mortality as the dependent variable. We find that teaching hospitals experience a change in their relative mortality rates in July ($\beta_{1,7}$ = 0.035) that is significant only at the 10% level. These findings offer limited support for Hypothesis 1 with respect to mortality. Nevertheless, we do find evidence of learning over the course of the academic year, as each of the coefficients on teaching hospitals in January through April is significantly lower than that for teaching hospitals in July at the 3% level.

Anticipation effect. We also find new evidence of an anticipation effect that begins earlier in the calendar year than the actual cohort turnover in July. For both LOS and mortality, after a period of decline in operational performance at teaching hospitals from September through February, we find that relative LOS and mortality begins to increase gradually starting in March until July. The magnitude of each month-to-month change from February to March, March to April, April to May, and May to June is not statistically distinguishable from that between June and July, suggesting that the increase in mortality is beginning before the actual cohort turnover in July. Given the relatively smooth and continuous nature of this pattern over the course of first half the year—where an initial increase in LOS and mortality, respectively, is not followed by a subsequent decrease until after the

July turnover—it is consistent with an anticipation effect related to the July turnover.

Though our data do not allow for conclusive explanations of what drives this anticipation effect, anecdotal evidence suggests that the increase in LOS and mortality at teaching hospitals relative to non-teaching hospitals during this anticipatory period may be explained by either or both of the following: (a) a gradual transition to greater responsibility in the last several months of the academic year for those residents who will remain at a given hospital during the next year (i.e., preparing for the upcoming cohort turnover in July) or (b) a decline in performance at teaching hospitals as senior residents or fellows "wind down" their appointments and become involved with the process of preparing to transition to new positions at other hospitals. With respect to the first explanation, some residency programs begin giving first-year residents more clinical responsibilities toward the end of their first academic year to prepare them to assume their roles as second-year residents in July. This may involve the first-year resident being the first to triage a new patient, assuming the care of patients coming from the ED, and initiating the diagnostic workups on patients. Compared to earlier in the academic year, first-year residents complete the above tasks either independently or with less oversight from second-year residents. However, also unlike earlier in the academic year, in these several months prior to the July turnover, there may not be an expectation of an increase in LOS and mortality and, thus, teaching hospitals may not have extra precautions in place. Whether due to a transition of responsibility or a decline in performance, it seems that the decline in operational performance that is commonly attributed to the period immediately following the July turnover may actually impact hospital performance beginning several months *prior* to the actual turnover event.

3.5.2 The Impact of Cohort Turnover on the Operational Performance of Minor vs. Major Teaching Hospitals

In Table 3.3, we present coefficients $\gamma_{1,m}$ and $\gamma_{2,m}$ on the month-*MIN_TCH* and month-*MAJ_TCH* interactions from Equation (3.3), which represent the change in the dependent variable for minor and major teaching hospitals, respectively, relative to the same change for non-teaching hospitals, using May as the baseline month. These results are shown graphically in Figures 3.2 and 3.3 for the change in the risk-adjusted LOS and risk-adjusted mortality, respectively, for minor and major teaching hospitals relative to the change for non-teaching hospitals.

Risk-adjusted LOS. In terms of LOS (Table 3.3 column (1)), the coefficient on minor teaching hospitals in July, $\gamma_{1,7}$, is 0.028 and is significant at the 5% level. This result suggests an increase in LOS for minor teaching hospitals of roughly 0.6% (relative to the average LOS of 4.73 days for minor teaching hospitals) following the July turnover. The estimated coefficients, $\gamma_{1,m}$, on LOS for minor teaching hospitals exhibit an overall, yet non-monotonic, decline in magnitude during the months from July to February, though the coefficients in August, September, November, and December are not each significantly different from that for July. By January, LOS in minor teaching hospitals is back to its value in the baseline May period, and the coefficients for minor teaching hospitals in January through April are each significantly lower than each of the July, August, and September coefficients at the 5% level. This general reduction in the estimated coefficients, $\gamma_{1,m}$, over the course of the academic year suggests experience-based improvement in performance as measured by LOS, our measure of resource utilization and cost.

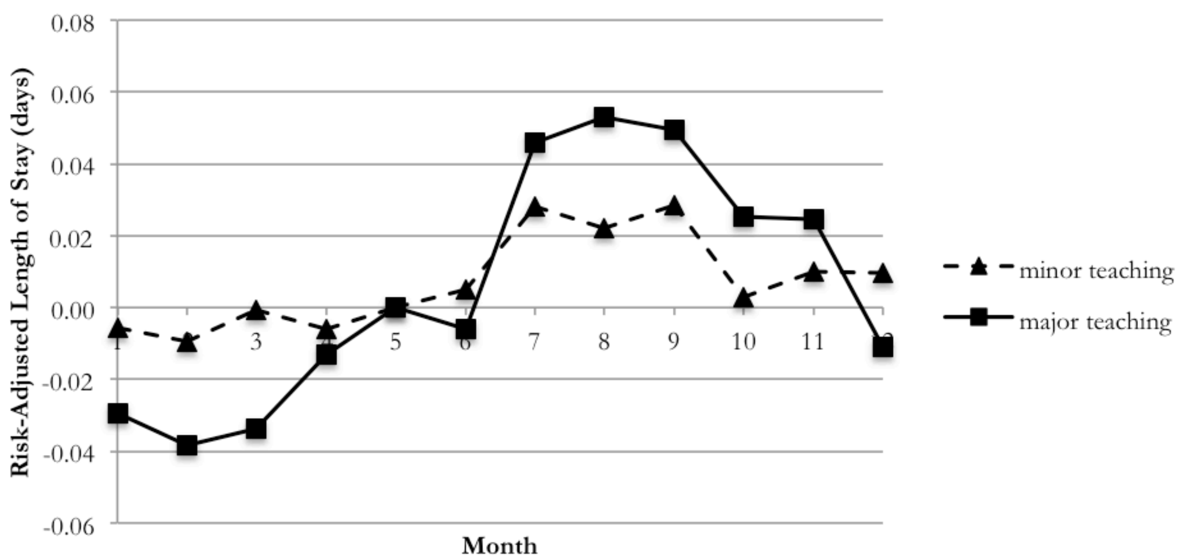
Consistent with Hypothesis 2, we find that major teaching facilities experience a greater increase in LOS relative to minor teaching hospitals. Specifically, major teaching hospitals experience a positive and significant (at the 1% level) increase in LOS relative to that for non-teaching hospitals following the July turnover, and the effect remains for approximately three months. The magnitude

Table 3.3. Effects of cohort turnover on hospital performance, using minor and major teaching categories

	Change in Dependent Variable Relative to Non-Teaching Baseline (Reference=May)			
	Risk-Adjusted LOS		Risk-Adjusted Mortality	
Minor Teaching				
January	-0.006	(0.013)	-0.012	(0.026)
February	-0.009	(0.011)	-0.033	(0.025)
March	-0.001	(0.010)	-0.002	(0.025)
April	-0.006	(0.010)	-0.012	(0.025)
June	0.005	(0.015)	0.023	(0.021)
July	0.028	(0.011) **	0.030	(0.023)
August	0.022	(0.011) **	0.003	(0.021)
September	0.028	(0.011) **	0.032	(0.023)
October	0.003	(0.012)	0.013	(0.021)
November	0.010	(0.011)	0.001	(0.021)
December	0.009	(0.012)	-0.001	(0.024)
Major Teaching				
January	-0.029	(0.017) *	-0.081	(0.031) ***
February	-0.038	(0.018) **	-0.123	(0.027) ***
March	-0.034	(0.020) *	-0.083	(0.026) ***
April	-0.013	(0.015)	-0.056	(0.025) **
June	-0.006	(0.017)	0.018	(0.026)
July	0.046	(0.016) ***	0.044	(0.026) *
August	0.053	(0.016) ***	0.034	(0.027)
September	0.049	(0.015) ***	0.003	(0.028)
October	0.025	(0.018)	0.020	(0.027)
November	0.025	(0.017)	-0.032	(0.027)
December	-0.011	(0.019)	0.029	(0.027)
Mean of Dependent Variable				
Minor Teaching	4.73		2.06	
Major Teaching	5.20		2.21	
Observations	87,707		87,707	
Adjusted R^2	0.719		0.336	

Notes. *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The level of observation is the hospital-month. All regressions include fixed effects for hospital, year, and month, though these coefficients are not shown in the table for ease of presentation. Standard errors (in parentheses) are heteroskedasticity robust and clustered by hospital. Observations are weighted by the total number of cases for the relevant hospital-month.

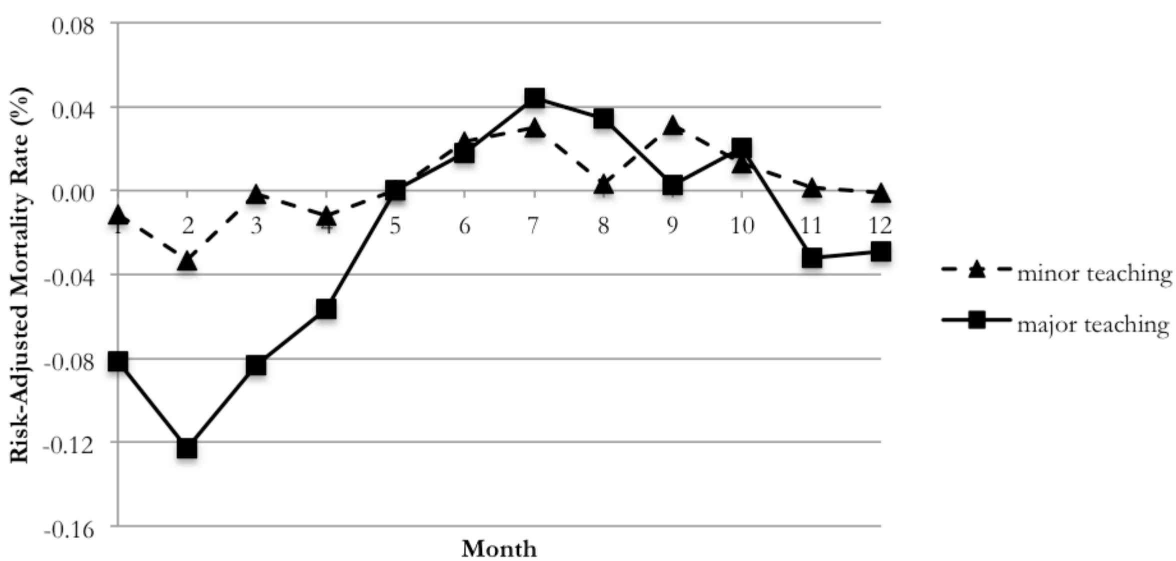
of this July effect ($\gamma_{2,7} = 0.046$) is nearly twice that for minor teaching hospitals ($\gamma_{1,7} = 0.028$), though the two coefficients are not significantly different at the 5% level. This effect for major teaching hospitals represents a 0.9% increase relative to the average LOS for such facilities (5.20 days). As with minor teaching hospitals, the effects for major teaching hospitals in each of the months from August through February decline in magnitude over time, though the coefficients ($\gamma_{2,m}$) in August through November are not statistically distinguishable from the July estimate for major teaching hospitals. By January, LOS falls to the point where it is significantly lower than that for May, and this coefficient is significantly lower than each of the July through November coefficients on major teaching hospitals at the 1% level. These results provide additional support for the contention that residents learn over the course of the academic year.



Notes. Values indicate the change in the risk-adjusted length of stay for minor and major teaching hospitals relative to the baseline change for non-teaching hospitals.

Figure 3.2. Change in risk-adjusted length of stay relative to non-teaching baseline

Risk-adjusted mortality. In column (2) of Table 3.3, we consider the impact of the July turnover on mortality in minor and major teaching hospitals compared to non-teaching hospitals. We find that minor teaching hospitals do not experience significant changes in mortality compared to non-teaching hospitals during the course of the academic year, and that major teaching facilities experience an increase in their relative mortality in July ($\gamma_{2,7}$ = 0.044) that is significant only at the 10% level. Evidence of learning is again present in the coefficients for the remainder of the academic year. Although the levels in August through December are not significantly different from the May baseline, the levels for November and December are each significantly lower than that for July at the 1% level. In addition, each of the coefficients ($\gamma_{2,m}$) on major teaching hospitals in January through April is significantly lower than each of those for August through October at the 1% level, which is suggestive of learning over the course of the academic year.



Notes. Values indicate the change in the risk-adjusted mortality for minor and major teaching hospitals relative to the baseline change for non-teaching hospitals.

Figure 3.3. Change in risk-adjusted mortality relative to non-teaching baseline

Anticipation effect. For major teaching hospitals, we again find evidence of a substantial turnover anticipation effect. After a period of decline in relative LOS from August through

February, we find that relative LOS in major teaching hospitals begins to increase gradually starting in March before reaching its peak in August. Each change from February to March, March to April, April to May, and May to June is not statistically distinguishable from the change between June and July, suggesting a "smooth" increase in relative LOS in major teaching hospitals that appears to be beginning before the actual cohort turnover in July.

With regard to mortality, this effect is even more pronounced. Following a period of decline in relative mortality from July through February, we find that relative mortality in major teaching hospitals begins to increase gradually starting in March before reaching its peak in July. Each change from February to March, March to April, April to May, and May to June is also not statistically distinguishable from the change between June and July.

3.5.3 Mitigating the Impact of Cohort Turnover on Operational Performance

We estimate Equations (3.4) and (3.5) to examine the extent to which hospitals may be able to mitigate the negative impact of the July turnover on operational performance. In columns (1) and (2) of Table 3.4, we present coefficients $\lambda_{2,m}$ and $\lambda_{3,m}$ on the month-*MAJ_TCH_NONMAGNET* and month-*MAJ_TCH_MAGNET* interactions from Equation (3.4), which represent the change in the dependent variable in each month relative to May for major teaching hospitals without and with ANCC magnet recognition, respectively, relative to the same change for non-teaching hospitals. In columns (3) and (4) of Table 3.4, we present coefficients $\rho_{4,m}$ and $\rho_{5,m}$ on the month-*MAJ_TCH_LO_QA* and month-*MAJ_TCH_HI_QA* interactions from Equation (3.5), which represent the change in the dependent variable in each month relative to May for major teaching hospitals with low and high intensities of quality assurance, respectively, relative to the same change for non-teaching hospitals.

Table 3.4. Effects of cohort turnover on hospital performance, stratifying major teaching hospitals by ANCC magnet recognition and intensity of potential quality assurance

	Change in Dependent Variable Relative to Non-Teaching Baseline (Reference=May)			
	ANCC Magnet Recognition (Yes v No)		Intensity of Quality Assurance (High v Low)	
	Risk-Adjusted LOS	Risk-Adjusted Mortality	Risk-Adjusted LOS	Risk-Adjusted Mortality
Major Teaching; No/Low				
January	-0.054 (0.023) **	-0.043 (0.037)	-0.044 (0.023) *	-0.079 (0.039) **
February	-0.065 (0.020) ***	-0.104 (0.032) ***	-0.061 (0.021) ***	-0.141 (0.033) ***
March	-0.061 (0.023) ***	-0.083 (0.032) **	-0.085 (0.021) ***	-0.087 (0.028) ***
April	-0.024 (0.019)	-0.031 (0.033)	-0.044 (0.020) **	-0.086 (0.028) ***
June	-0.018 (0.020)	0.021 (0.034)	-0.025 (0.020)	0.022 (0.034)
July	0.050 (0.020) **	0.074 (0.033) **	0.066 (0.019) ***	0.060 (0.033) *
August	0.058 (0.020) ***	0.050 (0.037)	0.061 (0.021) ***	0.022 (0.034)
September	0.050 (0.019) ***	0.028 (0.040)	0.067 (0.021) ***	0.036 (0.036)
October	0.028 (0.023)	0.041 (0.036)	0.033 (0.026)	0.061 (0.033) *
November	0.012 (0.022)	-0.022 (0.036)	0.029 (0.022)	0.011 (0.035)
December	-0.007 (0.023)	0.004 (0.034)	-0.012 (0.021)	-0.010 (0.032)
Major Teaching; Yes/High				
January	0.005 (0.022)	-0.136 (0.049) ***	-0.020 (0.024)	-0.072 (0.042) *
February	0.000 (0.026)	-0.150 (0.040) ***	-0.018 (0.027)	-0.100 (0.039) **
March	0.004 (0.032)	-0.084 (0.037) **	0.019 (0.032)	-0.083 (0.038) **
April	0.001 (0.022)	-0.092 (0.031) ***	0.021 (0.023)	-0.029 (0.035)
June	0.011 (0.019)	0.013 (0.034)	0.013 (0.020)	0.011 (0.033)
July	0.040 (0.021) *	0.001 (0.035)	0.027 (0.023)	0.025 (0.036)
August	0.046 (0.021) **	0.012 (0.035)	0.040 (0.022) *	0.042 (0.037)
September	0.048 (0.021) **	-0.033 (0.033)	0.028 (0.020)	-0.030 (0.039)
October	0.021 (0.024)	-0.010 (0.032)	0.011 (0.022)	-0.028 (0.040)
November	0.043 (0.025) *	-0.046 (0.032)	0.016 (0.026)	-0.077 (0.039) **
December	-0.016 (0.031)	-0.075 (0.036) **	-0.011 (0.028)	-0.057 (0.042)
Mean of Dependent Variable				
Major Teaching; No/Low	5.34	2.26	5.30	2.24
Major Teaching; Yes/High	4.96	2.13	5.09	2.17
Observations	87,707	87,707	87,707	87,707
Adjusted R^2	0.719	0.335	0.719	0.336

Notes. *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The level of observation is the hospital-month. All regressions include fixed effects for hospital, year, and month, though these coefficients are not shown in the table for ease of presentation. All regressions also include interactions between the minor teaching hospital category and month effects, which are also not shown in the table for ease of presentation but are identical to the coefficients for minor teaching hospitals shown in Table 3. Standard errors (in parentheses) are heteroskedasticity robust and clustered by hospital. Observations are weighted by the total number of cases for the relevant hospital-month.

ANCC magnet recognition. In column (1) of Table 3.4, we consider the differential impact of the July turnover on the LOS of major teaching hospitals without and with ANCC magnet recognition. We find that the coefficient on non-magnet certified major teaching hospitals in July, $\lambda_{2,7}$, is 0.050 ($p < 0.05$) whereas the same coefficient on magnet certified major teaching hospitals in July, $\lambda_{3,7}$, is 0.040 ($p < 0.10$). We note that the estimated magnitude of this coefficient for magnet certified facilities is less than that of the coefficient for non-magnet certified facilities, though the two coefficients are not significantly different from each other ($p < 0.75$). In addition, while there is evidence of an anticipation effect at non-magnet certified major teaching hospitals, we find no evidence of an anticipation effect at magnet certified major teaching hospitals. This suggests that magnet certified major teaching hospitals are able to mitigate some of the negative effects of the July turnover on LOS.

In column (2) of Table 3.4, we see the differential impact of the July turnover on mortality in major teaching hospitals without and with ANCC magnet recognition. Here, the coefficient on non-magnet certified major teaching hospitals in July, $\lambda_{2,7}$, is 0.074 ($p < 0.05$) whereas the same coefficient on magnet certified major teaching hospitals in July, $\lambda_{3,7}$, is 0.001 and not statistically significant at conventional levels ($p < 0.97$). These two coefficients are statistically significantly different from each other at the 10% level. Thus, while the performance of magnet certified major teaching hospitals as measured by clinical quality does not decline at the time of the July turnover, non-magnet certified facilities experience a significant increase in their relative mortality in July. This increase of 0.074 percentage points represents a 3.3% increase relative to the average mortality rate of 2.26% for non-magnet certified major teaching hospitals. Nevertheless, for both major teaching hospitals without and with ANCC magnet recognition, we continue to find evidence of an anticipation effect that begins in March and persists until the July turnover.

Potential quality assurance. In column (3) of Table 3.4, we consider the differential impact of the July turnover on the LOS of major teaching hospitals with low versus high intensities of potential quality assurance by nurses. We find that the July coefficient on major teaching hospitals with low intensities of quality assurance, $\rho_{4,7}$, is 0.066 ($p < 0.01$) whereas the same coefficient on facilities with high intensities of quality assurance, $\rho_{5,7}$, is 0.001 and not statistically significant at conventional levels ($p < 0.25$). We note the magnitude of this coefficient for those facilities with low intensities of quality assurance is significantly greater than that of the coefficient for facilities with high intensities of quality assurance, though the two coefficients are not significantly different from each other at conventional levels ($p < 0.19$). In addition, while there is evidence of a strong anticipation effect at major teaching hospitals with low intensities of quality assurance, we find no evidence of an anticipation effect at those facilities with high intensities of quality assurance. This suggests that facilities with high intensities of quality assurance are able to mitigate some of the negative effects of the July turnover on LOS.

Column (4) presents the differential impact of the July turnover on mortality in major teaching hospitals with low versus high intensities of quality assurance. Here, the coefficient on major teaching hospitals in July with low intensities of quality assurance, $\rho_{4,7}$, is 0.060 and statistically significant at the 10% level whereas the same coefficient on facilities with high intensities of quality assurance, $\rho_{5,7}$, is 0.025 and not statistically significant at conventional levels ($p < 0.50$). We note the magnitude of this coefficient for those facilities with low intensities of quality assurance is more than twice that of the coefficient for facilities with high intensity of quality assurance, although these two coefficients are not significantly different from each other ($p < 0.44$). For both facilities with low and high intensities of quality assurance, we continue for find evidence of an anticipation effect that begins in March.

Altogether, these analyses suggest that major teaching hospitals that invest in higher levels nursing quality and higher intensities of quality assurance may be able to mitigate the negative effects of the July turnover on operational performance, particularly with respect to the effects around the time of the July turnover. These facilities do not exhibit a decrease in performance with regards to LOS or mortality relative to non-teaching controls that is significant at the 5% level. In addition, we note that major teaching hospitals that are either magnet certified or have high intensities of quality assurance are also able to buffer against the anticipation effect with respect to LOS, which continues to manifest in the months preceding the July turnover at non-magnet certified facilities and those with low intensities of quality assurance. Nonetheless, even those facilities that are either magnet certified or have high intensities of quality assurance experience an anticipation effect with respect to mortality, which may indicate that even these facilities are not sufficiently buffering against the negative impact on quality that accompanies the transition of responsibilities that occurs in the months preceding the July turnover. This offers some degree of support for Hypothesis 3, which suggests that organizations with structures and processes that facilitate knowledge transfer from departing to entering workers may be able to mitigate the negative effects of cohort turnover on operational performance.

3.5.4 Extensions and Robustness

Though suggestive of declines in operational performance following cohort turnover, our findings are potentially consistent with alternate explanations. Therefore, we examine the robustness of our main findings and extend our results through several additional analyses (tables available in the online supplement).[3]

---

[3] In addition to the analyses presented in chapter 3.5.4, we also considered evaluating the robustness of our findings by assessing the relative changes in LOS and mortality in various subsets of the data (e.g., medical

Testing for patient self-selection and increased transfers. One alternate explanation of our findings is that elective patients may recognize July to be a time of transition for teaching hospitals and may decide to avoid those facilities at that time of the year. Under the reasonable assumption that these elective patients tend to be healthier than those who lack choice regarding their admission to the hospital, this self-selection by patients (on dimensions that are potentially unobservable to researchers) could leave teaching hospitals with relatively sicker patient populations at precisely the time we estimate their resource use to be increasing and their outcomes to be declining. If such selection were occurring, we would be mistaken to assume that the effects we observe were simply due to cohort turnover in July.

We offer three tests of the selection hypothesis in columns (1) – (3) of Table 3.5. First, if elective patients are, in fact, selecting away from teaching hospitals in July, those hospitals should experience a decline in their number of admissions relative to non-teaching facilities in July. Second, if the selection away from teaching hospitals leaves them with relatively sicker patient populations in July, teaching hospitals should experience an increase in the *expected* (not risk-adjusted) mortality and LOS relative to non-teaching facilities in July. We thus estimate three separate regressions of the same form as Equation (3.3) but with the number of hospital admissions, expected mortality, and expected LOS respectively, as the dependent variable.

With the number of hospital admissions as the dependent variable (column (1)), the results are mixed. In particular, the July coefficient on minor teaching hospitals is negative and significant ($\gamma_{1,7} = -6.9$, $p < 0.05$), though its magnitude is quite small and represents a decrease of only 0.55%

versus surgical patients, patients with high-mortality diagnoses such as acute myocardial infarction or stroke). However, because the AHA data only provides residency data at the hospital level and not at the level of specific diagnoses or clinical specialties (e.g., cardiology residents or orthopedic surgery residents), we are not able to accurately "match" a hospital's teaching intensity in a specific area with its risk-adjusted performance in that same area. Thus, we limit our presentation of additional analyses in this chapter to those for which we are able to match the level at which the key independent and dependent variables are observed.

Table 3.5. Effects of cohort turnover on hospital performance, using minor and major teaching categories for tests of patient self-selection and increased transfers

	Change in Dependent Variable Relative to Non-Teaching Baseline (Reference=May)							
	Total Admissions		Expected Mortality		Expected LOS		Transfer Rate (Transfers/Total Admissions*100)	
Minor Teaching								
January	7.5	(3.4) **	-0.008	(0.018)	-0.010	(0.013)	-0.04	(0.05)
February	-56.1	(5.2) ***	0.024	(0.018)	-0.002	(0.014)	0.02	(0.05)
March	14.9	(3.1) ***	0.012	(0.015)	-0.010	(0.011)	-0.02	(0.04)
April	-21.4	(2.8) ***	-0.017	(0.017)	-0.015	(0.011)	-0.01	(0.03)
June	-17.1	(2.7) ***	0.001	(0.016)	0.009	(0.011)	-0.04	(0.03)
July	-6.9	(3.2) **	0.016	(0.020)	0.013	(0.013)	-0.03	(0.04)
August	-4.0	(3.4)	0.043	(0.017) **	0.018	(0.018)	-0.03	(0.04)
September	-26.2	(3.4) ***	0.023	(0.017)	0.003	(0.012)	-0.09	(0.04) **
October	6.1	(3.4) *	0.029	(0.017) *	-0.006	(0.013)	-0.12	(0.06) **
November	-40.2	(4.0) ***	-0.001	(0.017)	-0.025	(0.016)	-0.15	(0.06) ***
December	-24.0	(4.0) ***	0.002	(0.018)	-0.023	(0.013) *	-0.09	(0.05) *
Major Teaching								
January	11.5	(13.8)	-0.060	(0.020) ***	-0.060	(0.026) **	-0.04	(0.10)
February	-118.8	(16.7) ***	-0.014	(0.020)	-0.045	(0.026) *	0.08	(0.08)
March	22.2	(11.1) **	-0.024	(0.018)	-0.025	(0.026)	0.14	(0.07) **
April	-44.8	(6.8) ***	-0.037	(0.019) **	0.021	(0.029)	0.05	(0.05)
June	-28.5	(6.2) ***	0.000	(0.016)	0.015	(0.020)	-0.01	(0.05)
July	-0.3	(11.9)	0.020	(0.021)	0.005	(0.025)	-0.03	(0.06)
August	2.9	(11.8)	0.047	(0.018) ***	0.002	(0.025)	-0.01	(0.07)
September	-47.8	(13.4) ***	0.044	(0.020) **	-0.021	(0.026)	-0.11	(0.09)
October	12.5	(13.6)	0.035	(0.027)	0.011	(0.054)	-0.31	(0.14) **
November	-67.4	(13.8) ***	0.006	(0.035)	-0.018	(0.046)	-0.28	(0.12) **
December	-53.2	(15.6) ***	-0.067	(0.021) ***	-0.073	(0.026) ***	-0.11	(0.11)
Mean of Dependent Variable								
Minor Teaching	1,244		2.09		4.75		4.14	
Major Teaching	2,031		1.98		4.92		5.24	
Observations	87,710		87,707		87,707		87,710	
Adjusted R^2	0.977		0.744		0.773		0.785	

Notes. *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The level of observation is the hospital-month. All regressions include fixed effects for hospital, year, and month, though these coefficients are not shown in the table for ease of presentation. Standard errors (in parentheses) are heteroskedasticity robust and clustered by hospital. Observations are weighted by the total number of cases for the relevant hospital-month.

relative to the average of 1,244 admissions per month for minor teaching hospitals. Meanwhile, the

July coefficient on major teaching hospitals—where we see the most substantial effects of cohort turnover in our base regressions—is not statistically significantly different from zero at conventional levels ($p < 0.99$). For the second regression with expected mortality as the dependent variable (column (2)), we do not find significant evidence of patient selection in July as measured by a change in expected mortality. For the third regression with expected LOS as the dependent variable (column (3)), we again do not find evidence of patient selection in July as measured by a change in expected LOS. Altogether, there is no clear evidence of patient selection away from the hospitals that are most affected by the July turnover.

Our analysis of admissions is related to a second potential explanation for our main results: that major teaching hospitals are receiving a higher percentage of patients transferred from minor teaching or non-teaching hospitals during the summer months due to potential excess capacity at teaching hospitals during warmer months. If one were to assume that these transfer patients were relatively sick compared to those typically seen at teaching hospitals in the summer months, one might expect both LOS and mortality to increase at the teaching hospitals receiving them. In column (4) of Table 3.5, we repeat our analysis using the percentage of cases transferred from another hospital as the dependent variable. We do not find a systematic change in relative transfer rates for either minor or major teaching hospitals around the July turnover.

Results for patients admitted from the Emergency Department. Although we do not find strong support for the selection hypothesis, we conduct an additional robustness check by limiting our sample to only the inpatient cases that arrive through a hospital's ED. These cases, which constitute 36% to 48% of inpatient cases depending on the year, are arguably less susceptible to endogeneity concerns than elective cases scheduled in advance because they are less likely to involve a patient choosing among hospitals. Further, ED cases are more serious on average than those that do not enter through the ED, resulting in higher average values for LOS and mortality (bottom two

Table 3.6. Effects of cohort turnover on hospital performance, using minor and major teaching categories for patients admitted from the Emergency Department

	Change in Dependent Variable Relative to Non-Teaching Baseline (Reference=May)			
	Risk-Adjusted LOS		Risk-Adjusted Mortality	
Minor Teaching				
January	0.020	(0.016)	-0.010	(0.056)
February	-0.004	(0.015)	-0.022	(0.050)
March	-0.002	(0.014)	0.014	(0.055)
April	-0.008	(0.014)	0.004	(0.054)
June	0.023	(0.013) *	0.028	(0.044)
July	0.052	(0.014) ***	0.064	(0.045)
August	0.031	(0.015) **	-0.001	(0.044)
September	0.019	(0.017)	0.047	(0.046)
October	0.007	(0.015)	0.037	(0.043)
November	0.027	(0.016) *	0.027	(0.042)
December	0.020	(0.016)	-0.016	(0.049)
Major Teaching				
January	-0.029	(0.030)	-0.102	(0.056) *
February	-0.062	(0.030) **	-0.176	(0.053) ***
March	-0.020	(0.041)	-0.070	(0.051)
April	-0.017	(0.022)	-0.060	(0.049)
June	0.026	(0.019)	0.073	(0.052)
July	0.081	(0.022) ***	0.142	(0.051) ***
August	0.065	(0.026) **	0.083	(0.060)
September	-0.001	(0.029)	-0.029	(0.059)
October	0.050	(0.029) *	0.018	(0.062)
November	0.029	(0.029)	-0.013	(0.061)
December	0.022	(0.031)	0.008	(0.060)
Mean of Dependent Variable				
Minor Teaching	5.13		3.22	
Major Teaching	5.78		3.33	
Observations	76,647		76,643	
Adjusted R^2	0.811		0.235	

Notes. *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The level of observation is the hospital-month. All regressions include fixed effects for hospital, year, and month, though these coefficients are not shown in the table for ease of presentation. Standard errors (in parentheses) are heteroskedasticity robust and clustered by hospital. Observations are weighted by the total number of cases for the relevant hospital-month.

rows of Table 3.6) than for the overall population that includes elective patients (bottom two rows of Table 3.3). As a result, we would expect the magnitude of the July effect to be even greater than those from our main results.

We estimate a regression of the same form as Equation (3.3) on this subset of cases. The results are similar to our main results (Table 3.6). In terms of risk-adjusted LOS, the July coefficients on minor and major teaching hospitals are 0.052 and 0.081, respectively, both of which are significant at the 1% level. In addition, in line with our expectations, each of these coefficients is greater in magnitude than those from our main results. The estimated coefficients on relative LOS in minor and major teaching hospitals exhibit a general decline in magnitude during the months from July to December, which is suggestive of learning over the course of the academic year. We also continue to find an anticipation effect at major teaching hospitals that begins in March prior to the July turnover.

The results for mortality are also similar to our main results, although here we find evidence of a significant July effect in major teaching hospitals. The July coefficient on mortality in major teaching hospitals is 0.142, which represents a 4.3% increase relative to the average mortality rate of 3.33 for major teaching hospitals. The estimated coefficients on relative mortality in major teaching hospitals exhibit a general decline in magnitude during the months from July to February, after which there is a gradual increase until peaking in July. This is suggestive of learning over the course of the academic year and a turnover anticipation effect that begins in the several months preceding July.

Results using alternate model specifications. A potential concern with our main empirical specification is that is relies on observations at the hospital-month level. Given mortality is a relatively rare event in most hospitals, the use of monthly mortality rates at the hospital level may result in a noisy measure of clinical quality. To address this possibility, we modify our main specification in Equation (3.3) to include a vector of fixed effects for seven multi-month periods during the year as opposed to each month of the year. We specify these multi-month periods as January through February, March through April, May, June, July through August, September through October, and November through December. As in our main specification, we isolate June as a transitional period because the residency changeover begins in late June for many hospitals.

Our findings highlight the robustness of our main results (Table 3.7). We find evidence of a substantial increase in relative LOS at teaching hospitals in the period just following the cohort turnover. The July-August coefficients on minor and major teaching hospitals are 0.025 and 0.049, respectively, which are significant at the 5% and 1% levels, respectively. The magnitude of this effect increases with the hospital's teaching intensity, with the July-August coefficient on major teaching hospitals being significantly different from the minor teaching coefficient for the same period at the

Table 3.7. Effects of cohort turnover on hospital performance, using minor and major teaching categories with multi-month periods

	Change in Dependent Variable Relative to Non-Teaching Baseline (Reference=May)			
	Risk-Adjusted LOS		Risk-Adjusted Mortality	
Minor Teaching				
Jan-Feb	-0.007	(0.011)	-0.022	(0.023)
Mar-Apr	-0.004	(0.009)	-0.007	(0.023)
June	0.005	(0.015)	0.023	(0.021)
Jul-Aug	0.025	(0.010) **	0.017	(0.020)
Sep-Oct	0.016	(0.010)	0.022	(0.019)
Nov-Dec	0.010	(0.010)	0.000	(0.020)
Major Teaching				
Jan-Feb	-0.034	(0.015) *	-0.101	(0.025) ***
Mar-Apr	-0.024	(0.015)	-0.070	(0.022) ***
June	-0.006	(0.017)	0.018	(0.026)
Jul-Aug	0.049	(0.014) ***	0.039	(0.023) *
Sep-Oct	0.037	(0.015) **	0.012	(0.025)
Nov-Dec	0.007	(0.017)	-0.031	(0.023)
Mean of Dependent Variable				
Minor Teaching	4.73		2.06	
Major Teaching	5.20		2.21	
Observations	87,707		87,707	
Adjusted R^2	0.719		0.334	

Notes. *,**, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The level of observation is the hospital-month. All regressions include fixed effects for hospital, year, and month, though these coefficients are not shown in the table for ease of presentation. Standard errors (in parentheses) are heteroskedasticity robust and clustered by hospital. Observations are weighted by the total number of cases for the relevant hospital-month.

10% level. This increase in LOS is sustained before it returns to the May level in the November-December period. Again, we observe a slight anticipation effect in major teaching hospitals, where there is an increase in relative LOS between the March-April period and the May baseline. For mortality, we find no evidence of a significant change in the July-August period at minor teaching hospitals. At major teaching hospitals, we observe an increase in relative mortality in the July-August period ($\gamma_{2,7\text{-}8} = 0.039$) that is significant at the 10% level. In addition, we find evidence of an anticipation effect that begins in the March-April period.

Another potential concern is that the ratio of FTE nurses per FTE resident may be highly correlated with the teaching intensity of a hospital. To address this issue, we conduct additional analyses in which we match hospitals on teaching intensity (FTE residents per bed) and exploit the remaining variation in the intensity of quality assurance by separating each matched pair into a higher- and lower-quality assurance hospital based on their relative values of FTE nurses per FTE resident. With this alternate specification, the correlation between teaching intensity and quality assurance reduces from 0.47 to 0.32. This approach yields results that are similar to our main findings.

Finally, because LOS and mortality could also be affected by the extent to which a hospital is more crowded (i.e., operating at a higher level of capacity utilization), we also estimate a model that accounts for hospital crowding. We assess whether a hospital is more or less crowded compared to its average level by using the number of total admissions in a month. In addition, we include the squared term of the total number of admissions in a month as an additional covariate to account for potential non-linearity in the effect of hospital crowding on LOS and mortality. We find our main results to be robust to this alternate specification.

3.6 Discussion and Conclusions

In this paper, we utilize data on hospital inpatient discharges to examine the impact of cohort turnover on operational performance. Specifically, we examine the impact of the July turnover of resident physicians in teaching hospitals on hospital operational performance as measured by resource utilization and clinical quality. We find that that the effects of the cohort turnover appear not only in the months during and after the turnover event, but also in the months *preceding* the turnover event.

First, at the time of the cohort turnover event in July and in the following months, we find a reduction in operational performance at teaching hospitals as evidenced by an increase in resource utilization, but only limited evidence of a decrease in clinical quality. At minor teaching hospitals, we find a relative increase in risk-adjusted LOS that lasts for a few months but no accompanying increase in the risk-adjusted mortality rate. At major teaching hospitals, we find a significant increase in risk-adjusted LOS and limited evidence of an increase in risk-adjusted mortality.

Second, in the months preceding the cohort turnover, we find that teaching hospitals exhibit a gradual decrease in operational performance relative to non-teaching hospitals, which presents as an increase in the relative risk-adjusted LOS and mortality rate beginning in March. This pre-July decrease in operational performance at teaching hospitals may result from institutional efforts to preempt the turnover by increasing the responsibilities of remaining workers (e.g., making first-year residents responsible for holding the admission pager) and from departing workers winding down their current duties and becoming involved in the process of transitioning to new positions at other hospitals. We refer to this phenomenon as an anticipation effect.

Particularly at major teaching hospitals, the effect of the July turnover on resource utilization is substantial. Average risk-adjusted LOS—our proxy for resource utilization and, therefore, cost—for the average major teaching hospital increases by 0.9% following the July turnover and remains higher for several months after July. Given the slim profit margins of most teaching hospitals in the U.S. (American Hospital Association 2015), this is an economically significant increase for multiple months each year. The increase in risk-adjusted mortality is small but also not trivial, being roughly 2.0% (percent, not percentage points) in July. Determining the social cost of this increase in mortality requires assumptions—beyond the scope of this paper—about the expected longevity and quality of life of these individuals in the absence of the July turnover.

3.6.1 Contributions

Our findings contribute to the literature on turnover, performance, and productivity in several ways. First, we define and examine cohort turnover as a phenomenon distinct from individual or collective turnover and empirically estimate its impact on operational performance. Most studies on turnover tend to focus on the turnover of individuals rather than large cohorts. In addition, they typically consider the causes of turnover rather than examining their consequences, perhaps because the endogenous nature of individual turnover makes the phenomenon difficult to study empirically. Nevertheless, it is important to examine the impact of the turnover of cohorts on performance because of the sizable nature of such events and their occurrence in several organizational settings including hospitals, political administrations, and military deployments.

Second, we consider the broader implications of cohort turnover by looking beyond the immediate time window of the turnover event. Because cohort turnover is, by definition, planned in advance and known to occur at a specific point in time, it may impact operational performance in advance of the actual turnover event. If there were a decline in operational performance that begins earlier than the actual turnover itself, failing to note this earlier decline in performance would result in an underestimate of the true magnitude of the effect associated with the cohort turnover. Accordingly, we identify a decline in hospital operational performance in the several months leading up to the July turnover. We define this as an anticipation effect, and are aware of no prior work that examines whether the effect of cohort turnover might precede its formal occurrence.

Third, we improve upon the methodological approaches employed by papers in the medical literature that have previously examined the effect of the July turnover on various clinical outcomes. Specifically, many prior studies on this topic do not adjust for risk, adjust for seasonal variation, or use suitable concurrent controls. All of these are necessary adjustments given the observational design of these studies. To adjust for these factors, we risk adjust our dependent variables and use a control group of non-teaching hospitals as a baseline category. In addition, prior work in the medical literature typically examines the presence of a July effect for a specific category of high-risk patients (e.g., very low weight infants, intensive care unit patients, spinal surgery patients, trauma patients). While this focus on a narrow study population helps reduce patient heterogeneity, it may limit the generalizability of the research findings to the extent that residents' level of responsibility for treatment may vary across clinical areas, particularly early in their post-graduate training. Further, because the clinical outcomes that these studies aim to explain are typically measured at the department level—whereas teaching intensities are based on hospital-level figures that may vary across specific clinical areas within a hospital (e.g., a hospital may have a more substantial residency program in surgery than in radiology)—it is not clear that there is always a "match" in the level at which the dependent and key independent variables are observed. We avoid this mismatch by assessing our outcome measures at the hospital level.

3.6.2 Managerial Implications and Conclusions

Our results suggest that cohort turnover should be managed not only at the time of, but also during the period leading up to, the actual turnover event. We also find that operational performance declines may be mitigated if appropriately managed. These mitigation efforts may require managers to consider the size of the cohort relative to the size of the organization.

To mitigate the negative effects of cohort turnover on operational performance, managers should consider implementing structures and processes that facilitate knowledge transfer from departing to entering workers. This can be done by increasing the overall quality of the workforce that is not turning over and can, thus, transfer knowledge to new workers. Although this may require a significant investment in terms of financial and organizational resources, it is likely to have positive spillover effects beyond an organization's increased ability to withstand the negative effects of cohort turnover. Another approach, which may have a more immediate impact, would be to increase the intensity of quality assurance. One way to do this is by increasing the relative staffing level of those workers who are not subject to the cohort turnover who are in a position to serve as an operational safeguard for the output of new workers. In the case of teaching hospitals, these may be the nurses who work closely with residents but do not tend to turn over as a cohort in July. Given our finding that the effects of cohort turnover are increasing in its relative intensity, hospital managers might consider increasing the nurse-to-resident ratios especially in those departments that are more resident intensive within the facility.

One question not resolved by this study is the degree to which managers should be *concerned* about turnover-related declines in operational performance. On one hand, these declines likely reflect the costs associated with valuable on-the-job training. On the other, they may be larger than necessary to obtain the desired training benefit for new physicians. In the case of teaching hospitals, we thus are not arguing that an optimal residency system would result in *no* systematic change in operational performance throughout the year. It is likely that no system can guarantee residents will be as productive and high performing at the beginning of their tenure as they will be at its end. Ultimately, a key question is whether declines in operational performance are higher than necessary to train new workers efficiently and whether those effects can be mitigated. Our findings suggest that this may be possible, assuming that the costs of improving the knowledge transfer do not outweigh the resulting benefits in terms of cost and quality performance.

The question of whether there are optimal levels of pre-turnover preparation and post-turnover employee oversight in the face of significant on-the-job training is an important issue worthy of further study in contexts both within and outside the hospital industry. In this paper, we provide some examples of potential approaches for mitigating the impact of cohort turnover. We welcome future work to extend this line of inquiry by considering other ways to mitigate these effects, such as by better utilizing information technology and increasing teamwork and coordination within work groups. Ultimately, even if organizations are not able to reduce the cohort turnover they face, they may be able to take steps to better manage its effects.

REFERENCES

Abadie, A. 2005. Semiparametric difference-in-differences estimators. *Rev. Econ. Stud.* **72**(1) 1–19.

Abelson, M. A., B. D. Baysinger. 1984. Optimal and dysfunctional turnover: Toward an organizational level model. *Acad. Manag. Rev.* **9**(2) 331–341.

Adler, P. S., K. B. Clark. 1991. Behind the learning curve: A sketch of the learning process. *Manage. Sci.* **37**(3) 267–281.

Agency for Healthcare Research and Quality. 2013. Overview of the Nationwide Inpatient Sample (NIS). *Healthc. Cost Util. Proj.* Available at: http://www.hcup-us.ahrq.gov/nisoverview.jsp.

American Hospital Association. 2015. Trends in hospital financing. *Trends Affecting Hospitals and Health Systems*, chap. 4.

American Nurses Credentialing Center. 2015. Magnet recognition program overview. Available at: http://www.nursecredentialing.org/Magnet/ProgramOverview [Accessed July 22, 2015].

Anand, G., J. Gray, E. Siemsen. 2012. Decay, shock, and renewal: Operational routines and process entropy in the pharmaceutical industry. *Organ. Sci.* **23**(6) 1700–1716.

Anand, K. S., M. F. Pac, S. K. Veeraraghavan. 2010. Quality-speed conundrum: Trade-offs in customer-intensive services. *Manage. Sci.* **57**(1) 40–56.

Andritsos, D. A., C. S. Tang. 2014. Linking process quality and resource usage: An empirical analysis. *Prod. Oper. Manag.* **23**(12) 2163–2177.

Anupindi, R., S. Chopra, S. D. Deshmukh, J. A. Van Mieghem, E. Zemel. 2005. *Managing Business Process Flows: Principles of Operations Management* 2nd ed. Upper Saddle River, NJ, Prentice-Hall.

Argote, L., B. Mcevily, R. Reagans. 2003. Managing knowledge in organizations: An integrative framework and review of emerging themes. *Manage. Sci.* **49**(4) 571–582.

Armony, M., A. R. Ward. 2010. Fair dynamic routing in large-scale heterogeneous-server systems. *Oper. Res.* **58**(3) 624–637.

Ashraf, N., O. Bandiera, S. S. Lee. 2014. Awards unbundled: Evidence from a natural field experiment. *J. Econ. Behav. Organ.* **100** 44–63.

Association of American Medical Colleges. 2009. Key facts about teaching hospitals. Tech. Rep. February, Association of American Medical Colleges, Washington, DC.

Ata, B., J. A. Van Mieghem. 2008. The value of partial resource pooling: Should a service network be integrated or product-focused? *Manage. Sci.* **55**(1) 115–131.

Autor, D. H. 2003. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *J. Labor Econ.* **21**(1) 1–42.

Bandiera, O., I. Barankay, I. Rasul. 2013. Team incentives: Evidence from a firm level experiment. *J. Eur. Econ. Assoc.* **11** 1079–1114.

Barankay, I. 2012. Rank incentives: Evidence from a randomized workplace experiment. Working paper, The Wharton School, University of Pennsylvania, Philadelphia.

Bassamboo, A., R. S. Randhawa, J. A. Van Mieghem. 2010. Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. *Manage. Sci.* **56**(8) 1285–1303.

Batt, R. 2002. Managing customer services: Human resource practices, quit rates, and sales growth. *Acad. Manag. J.* **45**(3) 587–597.

Batt, R. J., C. Terwiesch. 2015. How a service process adapts to load: An econometric analysis of patient treatment in the Emergency Department. Working paper, The Wharton School, University of Pennsylvania, Philadelphia.

Bendoly, E. 2013. Real-time feedback and booking behavior in the hospitality industry: Moderating the balance between imperfect judgment and imperfect prescription. *J. Oper. Manag.* **31**(1-2) 62–71.

Benjaafar, S. 1995. Performance bounds for the effectiveness of pooling in multi-processing systems. *Eur. J. Oper. Res.* **87**(2) 375–388.

Bertrand, M., E. Duflo, S. Mullainathan. 2004. How much should we trust differences-in-differences estimates? *Q. J. Econ.* **119**(1) 249–275.

Berwick, D. M. 2003. Errors today and errors tomorrow. *N. Engl. J. Med.* **348**(25) 2570–2572.

Beshears, J., J. J. Choi, D. Laibson, B. C. Madrian, K. L. Milkman. 2015. The effect of providing peer information on retirement savings decisions. *J. Finance* **70**(3) 1161–1201.

Best, T. J., B. Sandıkçı, D. D. Eisenstein, D. O. Meltzer. 2015. Managing hospital inpatient bed capacity through partitioning care into focused wings. *Manuf. Serv. Oper. Manag.* **17**(2) 157–176.

Blanes i Vidal, J., M. Nossol. 2011. Tournaments without prizes: Evidence from personnel records. *Manage. Sci.* **57**(10) 1721–1736.

Bohn, R. E. 1995. Noise and learning in semiconductor manufacturing. *Manage. Sci.* **41**(1) 31–42.

Boudreau, J., W. J. Hopp, J. O. McClain, L. J. Thomas. 2003. On the interface between operations and human resources management. *Manuf. Serv. Oper. Manag.* **5**(3) 179–202.

Boudreau, K. J., K. R. Lakhani, M. Menietti. 2015. Performance responses to competition across skill-levels in rank order tournaments: Field evidence and implications for tournament design. *RAND J. Econ.* **47**(1) 140–165.

Brown, M. Z., M. M. Linehan, K. A. Comtois, A. Murray, A. L. Chapman. 2009. Shame as a prospective predictor of self-inflicted injury in borderline personality disorder: A multi-modal analysis. *Behav. Res. Ther.* **47**(10) 815–822.

Cachon, G. P. 2002. Supply chain coordination. S. C. Graves, T. de Kok, eds. *Handb. OR MS, Supply Chain Manag.* North-Holland, The Netherlands, Elsevier Science Publishers, 229–340.

Cachon, G. P., F. Zhang. 2007. Obtaining fast service in a queueing system via performance-based allocation of demand. *Manage. Sci.* **53**(3) 408–420.

Campbell, B. A., M. Ganco, A. M. Franco, R. Agarwal. 2012. Who leaves, where to, and why worry? Employee mobility, entrepreneurship and effects on source firm performance. *Strateg. Manag. J.* **33**(1) 65–87.

Cao, Q., L. M. Maruping, R. Takeuchi. 2006. Disentangling the effects of CEO turnover and succession on organizational capabilities: A social network perspective. *Organ. Sci.* **17**(5) 563–576.

CDC - National Center for Health Statistics. 2016. FastStats. Available at: http://www.cdc.gov/nchs/fastats/default.htm [Accessed March 9, 2016].

Centers for Disease Control and Prevention. 2002. Guideline for hand hygiene in health-care settings. *Morb. Mortal. Wkly. Rep.* **51**.

Centers for Medicare & Medicaid Services. 2002. Medicare program; Revisions to payment policies under the physician fee schedule for calendar year 2003 and inclusion of Registered Nurses in the personnel provision of the critical access hospital emergency services requirement for frontier areas and Remote Locations. *Fed. Regist.* **67**(251) 79972.

Centers for Medicare & Medicaid Services. 2015. NHE fact sheet. Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet.html [Accessed March 9, 2016].

Chan, D. C. Forthcoming. Teamwork and moral hazard: Evidence from the emergency department. *J. Polit. Econ.*

Chan, T. Y., J. Li, L. Pierce. 2014. Learning from peers: Knowledge transfer and sales force productivity growth. *Mark. Sci.* **33**(4) 463–484.

Chandrasekaran, A., C. Senot, K. K. Boyer. 2012. Process management impact on clinical and experiential quality: Managing tensions between safe and patient-centered healthcare. *Manuf. Serv. Oper. Manag.* **14**(4) 548–566.

Charlson, M. E., P. Pompei, K. L. Ales, C. R. MacKenzie. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J. Chronic Dis.* **40**(5) 373–383.

Charness, G., D. Masclet, M. C. Villeval. 2014. The dark side of competition for status. *Manage. Sci.* **60**(1) 38–55.

Corbett, C. J., K. Rajaram. 2006. A generalization of the inventory pooling effect to nonnormal dependent demand. *Manuf. Serv. Oper. Manag.* **8**(4) 351–358.

Cowgill, B. 2015. Competition and productivity in employee promotion contests. Working paper, Haas School of Business, University of California Berkeley, Berkeley.

Dalton, D. R., W. D. Todor. 1979. Turnover turned over: An expanded and positive perspective. *Acad. Manag. Rev.* **4**(2) 225–235.

Debo, L. G., L. B. Toktay, L. N. Van Wassenhove. 2008. Queuing for expert services. *Manage. Sci.* **54**(8) 1497–1512.

DeHoratius, N., A. Raman. 2008. Inventory record inaccuracy: An empirical analysis. *Manage. Sci.* **54**(4) 627–641.

Deo, S., A. Jain, P. K. Pendem. 2014. Pacing work in the presence of goals and deadlines: Econometric analysis of an outpatient department. Working paper, Indian School of Business, Hyderabad, India.

van Dijk, N. M., E. van der Sluis. 2008. To pool or not to pool in call centers. *Prod. Oper. Manag.* **17**(3) 296–305.

van Dijk, N. M., E. van der Sluis. 2009. Pooling is not the answer. *Eur. J. Oper. Res.* **197**(1) 415–421.

Dobson, G., T. Tezcan, V. Tilson. 2013. Optimal workflow decisions for investigators in systems with interruptions. *Manage. Sci.* **59**(5) 1125–1141.

Doroudi, S., R. Gopalakrishnan, A. Wierman. 2011. Dispatching to incentivize fast service in multi-server queues. *ACM SIGMETRICS Perform. Eval. Rev.* **39**(3) 43–45.

Drexler, A., A. Schoar. 2014. Do relationships matter? Evidence from loan officer turnover. *Manage. Sci.* **60**(11) 2722–2736.

Duflo, E. 2001. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *Am. Econ. Rev.* **91**(4) 795–813.

Dunn, J., N. E. Ruedy, M. E. Schweitzer. 2012. It hurts both ways: How social comparisons harm affective and cognitive trust. *Organ. Behav. Hum. Decis. Process.* **117**(1) 2–14.

Edelman, B., I. Larkin. 2015. Social comparisons and deception across workplace hierarchies: Field and experimental evidence. *Organ. Sci.* **26**(1) 78–98.

Englesbe, M. J., S. J. Pelletier, J. C. Magee, et al. 2007. Seasonal variation in surgical outcomes as measured by the American College of Surgeons-National Surgical Quality Improvement Program (ACS-NSQIP). *Ann. Surg.* **246**(3) 456–462.

Eppen, G. D. 1979. Note - Effects of centralization on expected costs in a multi-location newsboy problem. *Manage. Sci.* **25**(5) 498–501.

Fine, M. J., H. M. Pratt, D. S. Obrosky, et al. 2000. Relation between length of hospital stay and costs of care for patients with community-acquired pneumonia. *Am. J. Med.* **109**(5) 378–385.

Ford, A. A., B. T. Bateman, L. L. Simpson, R. B. Ratan. 2007. Nationwide data confirms absence of "July phenomenon" in obstetrics: It's safe to deliver in July. *J. Perinatol.* **27**(2) 73–76.

Frei, F. X., R. Kalakota, A. J. Leone, L. M. Marx. 1999. Process variation as a determinant of bank performance: Evidence from the retail banking study. *Manage. Sci.* **45**(9) 1210–1220.

Fung, C. H., Y.-W. Lim, S. Mattke, C. Damberg, P. G. Shekelle. 2008. Systematic review: The evidence that publishing patient care performance data improves quality of care. *Ann. Intern. Med.* **148**(2) 111–123.

Gans, N., G. Koole, A. Mandelbaum. 2003. Telephone call centers: Tutorial, review, and research prospects. *Manuf. Serv. Oper. Manag.* **5**(2) 79–141.

Gans, N., Y.-P. Zhou. 2002. Managing learning and turnover in employee staffing. *Oper. Res.* **50**(6) 991–1006.

Gemmell, I., P. McLoone, F. A. Boddy, G. J. Dickinson, G. C. Watt. 2000. Seasonal variation in mortality in Scotland. *Int. J. Epidemiol.* **29**(2) 274–279.

Georgakopoulos, D., M. Hornick, A. Sheth. 1995. An overview of workflow management: From process modeling to workflow automation infrastructure. *Distrib. Parallel Databases* **3**(2) 119–153.

Gilbert, S. M., Z. K. Weng. 1998. Incentive effects favor nonconsolidating queues in a service system: The principal-agent perspective. *Manage. Sci.* **44**(12-Part-1) 1662–1669.

Gilboy, N., T. Tanabe, D. Travers, A. Rosenau. 2011. *Emergency Severity Index (ESI): A Triage Tool for Emergency Department Care, Version 4, Implementation Handbook*. Rockville, MD, Agency for Healthcare Research and Quality.

Glebbeek, A. C., E. H. Bax. 2004. Is high employee turnover really harmful? An empirical test using company records. *Acad. Manag. J.* **47**(2) 277–286.

Gorman, A. 2015. How one hospital reduced unnecessary C-sections. *The Atlantic*.

Green, L. V, V. Nguyen. 2001. Strategies for cutting hospital beds: The impact on patient service. *Health Serv. Res.* **36**(2) 421–442.

Greene, W. 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econom. J.* **7** 98–119.

Guttmann, A., H. J. Schultz, M. J. Vermeulen, T. A. Stukel. 2011. Association between waiting times and short term mortality and hospital admission after departure from emergency department: Population based cohort study from Ontario, Canada. *BMJ* **342** d2983.

Hackman, J. R., G. R. Oldham. 1976. Motivation through the design of work: Test of a theory. *Organ. Behav. Hum. Perform.* **16**(2) 250–279.

Hair, J. F., R. E. Anderson, R. L. Tatham, W. C. Black. 1998. *Multivariate Data Analysis* 5th ed. Upper Saddle River, NJ, Prentice Hall.

Haller, G., P. S. Myles, P. Taffe, T. V Perneger, C. L. Wu. 2009. Rate of undesirable events at beginning of academic year: Retrospective cohort study. *BMJ* **339**(October 13) b3974–b3974.

Hannan, R. L., G. P. McPhee, A. H. Newman, I. D. Tafkov. 2013. The effect of relative performance information on performance and effort allocation in a multi-task environment. *Account. Rev.* **88**(2) 553–575.

Harder, D. W., S. J. Lewis. 1987. The assessment of shame and guilt. J. N. Butcher, C. D. Spielberger, eds. *Adv. Personal. assessment, Vol. 6*. Hillsdale, NJ, Erlbaum, 89–114.

Hasija, S., E. Pinker, R. A. Shumsky. 2010. Work expands to fill the time available: Capacity estimation and staffing under Parkinson's law. *Manuf. Serv. Oper. Manag.* **12**(1) 1–18.

Hausknecht, J. P., J. A. Holwerda. 2013. When does employee turnover matter? Dynamic member configurations, productive capacity, and collective performance. *Organ. Sci.* **24**(1) 210–225.

Hausknecht, J. P., C. O. Trevor. 2011. Collective turnover at the group, unit, and organizational levels: Evidence, issues, and implications. *J. Manage.* **37**(1) 352–388.

Healthcare Cost and Utilization Project. 1999. *The HCUP Nationwide Inpatient Sample (NIS), Release 6, 1997*. Rockville, MD.

Holden, R. J. 2011. Lean Thinking in emergency departments: A critical review. *Ann. Emerg. Med.* **57**(3) 265–278.

Hopp, W. J., S. M. R. Iravani, F. Liu. 2009. Managing white-collar work: An operations-oriented survey. *Prod. Oper. Manag.* **18**(1) 1–32.

Hopp, W. J., S. M. R. Iravani, G. Y. Yuen. 2007. Operations systems with discretionary task completion. *Manage. Sci.* **53**(1) 61–77.

Hopp, W. J., M. L. Spearman. 2000. *Factory Physics: Foundations of Manufacturing Management* 2nd ed. Burr Ridge, IL, McGraw-Hill.

Huckman, R. S. 2003. The utilization of competing technologies within the firm: Evidence from cardiac procedures. *Manage. Sci.* **49**(5) 599–617.

Huckman, R. S., G. P. Pisano. 2006. The firm specificity of individual performance: Evidence from cardiac surgery. *Manage. Sci.* **52**(4) 473–488.

Huckman, R. S., B. R. Staats. 2011. Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manuf. Serv. Oper. Manag.* **13**(3) 310–328.

Huckman, R. S., B. R. Staats, D. M. Upton. 2009. Team familiarity, role experience, and performance: Evidence from Indian software services. *Manage. Sci.* **55**(1) 85–100.

Huselid, M. A. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Acad. Manag. J.* **38**(3) 635–672.

Hyytiä, E., S. Aalto. 2013. Round-robin routing policy: Value-functions and mean performance with job- and server-specific costs. *Proc. 7th Int. Conf. Perform. Eval. Methodol. Tools. (ValueTools '13).* Turin, Italy, ACM Press.

Imbens, G. W., J. M. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *J. Econ. Lit.* **47**(1) 5–86.

Jaikumar, R. 2005. From filing and fitting to flexible manufacturing: A study in the evolution of process control. *Found. Trends Technol. Inf. Oper. Manag.* **1**(1).

Jena, A. B., E. C. Sun, J. A. Romley. 2013. Mortality among high risk patients with acute myocardial infarction admitted to U.S. teaching-intensive hospitals in July: A retrospective observational study. *Circulation* **128**(25) 2754–2763.

Jouini, O., Y. Dallery, R. Nait-Abdallah. 2008. Analysis of the impact of team-based organizations in call center management. *Manage. Sci.* **54**(2) 400–414.

Jovanovic, B. 1979. Job matching and the theory of turnover. *J. Polit. Econ.* **87**(5) 972–990.

Kacmar, K. M., M. C. Andrews, D. L. Van Rooy, R. C. Steilberg, S. Cerrone. 2006. Sure everyone can be replaced… but at what cost? Turnover as a predictor of unit-level performance. *Acad. Manag. J.* **49**(1) 133–144.

KC, D. S. 2014. Does multitasking improve performance? Evidence from the emergency department. *Manuf. Serv. Oper. Manag.* **16**(2) 168–183.

KC, D. S., B. R. Staats. 2012. Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manuf. Serv. Oper. Manag.* **14**(4) 618–633.

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

Keith, K. D., J. J. Bocka, M. S. Kobernick, R. L. Krome, M. A. Ross. 1989. Emergency department revisits. *Ann. Emerg. Med.* **18**(9) 964–968.

Kingman, J. F. C. 1961. The single server queue in heavy traffic. *Math. Proc. Cambridge Philos. Soc.* **57**(04) 902–904.

Klassen, R. D., L. J. Menor. 2007. The process management triangle: An empirical investigation of process trade-offs. *J. Oper. Manag.* **25**(5) 1015–1034.

Kleinrock, L. 1976. *Queueing Systems, Volume 2: Computer Applications*. New York, NY, John Wiley & Sons.

Krueger, A. B., A. Mas. 2004. Strikes, scabs, and tread separations: Labor strife and the production of defective Bridgestone/Firestone tires. *J. Polit. Econ.* **112**(2) 253–289.

Kuhnen, C. M., A. Tymula. 2012. Feedback, self-esteem, and performance in organizations. *Manage. Sci.* **58**(1) 94–113.

Kuziemko, I., R. W. Buell, T. Reich, M. I. Norton. 2014. Last-place aversion: Evidence and redistributive implications. *Q. J. Econ.* **129** 105–149.

Lapré, M. A., A. S. Mukherjee, L. N. Van Wassenhove. 2000. Behind the learning curve: Linking learning activities to waste reduction. *Manage. Sci.* **46**(5) 597–611.

Link, S., E. Naveh. 2006. Standardization and discretion: Does the environmental standard ISO 14001 lead to performance benefits? *IEEE Trans. Eng. Manag.* **53**(4) 508–519.

Loch, C. 1998. Operations management and reengineering. *Eur. Manag. J.* **16**(3) 306–317.

Mandelbaum, A., M. I. Reiman. 1998. On pooling in queueing networks. *Manage. Sci.* **44**(7) 971–981.

Mas, A., E. Moretti. 2009. Peers at work. *Am. Econ. Rev.* **99**(1) 112–145.

Mazzocato, P., C. Savage, M. Brommels, H. Aronsson, J. Thor. 2010. Lean thinking in healthcare: A realist review of the literature. *Qual. Saf. Health Care* **19**(5) 376–382.

McCarthy, M. L., R. Ding, J. M. Pines, et al. 2012. Provider variation in fast track treatment time. *Med. Care* **50**(1) 43–49.

McClelland, M. S., D. Lazar, V. Sears, M. Wilson, B. Siegel, J. M. Pines. 2011. The past, present, and future of urgent emergency department flow improvement. *Acad. Emerg. Med.* **18** 1392–1399.

McGregor, J. 2006. The struggle to measure performance. *Bloom. Businessweek*. Available at: http://www.bloomberg.com/bw/stories/2006-01-08/the-struggle-to-measure-performance [Accessed May 21, 2015].

Medicare Payment Advisory Commission. 2002. Report to the Congress: Medicare payment policy. Washington, DC.

Messersmith, J. G., J.-Y. Lee, J. P. Guthrie, Y.-Y. Ji. 2014. Turnover at the top: Executive team departures and firm performance. *Organ. Sci.* **25**(3) 776–793.

Mobley, W. H. 1982. *Employee Turnover: Causes, Consequences, and Control*. Reading, MA, Addison-Wesley.

Moran, S., M. E. Schweitzer. 2008. When better is worse: Envy and the use of deception. *Negot. Confl. Manag. Res.* **1**(1) 3–29.

Morganti, K. G., S. Bauhoff, J. C. Blanchard, et al. 2013. *The Evolving Role of Emergency Departments in the United States*. Santa Monica, CA.

Nagin, D. S., J. B. Rebitzer, S. Sanders, L. J. Taylor. 2002. Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *Am. Econ. Rev.* **92**(4) 850–873.

Narayanan, S., S. Balasubramanian, J. M. Swaminathan. 2009. A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Manage. Sci.* **55**(11) 1861–1876.

Netessine, S., V. Yakubovich. 2012. The Darwinian workplace. *Harv. Bus. Rev.* **90**(5) 25–28.

Nordstrom, R., P. Lorenzi, R. V. Hall. 1991. A review of public posting of performance feedback in work settings. *J. Organ. Behav. Manage.* **11**(2) 101–124.

Ohno, T. 1988. *Toyota Production System: Beyond Large-scale Production*. Portland, OR, Productivity Press.

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

Patel, P. B., D. R. Vinson. 2005. Team assignment system: Expediting emergency department care. *Ann. Emerg. Med.* **46** 499–506.

Phillips, D. P., G. E. C. Barker. 2010. A July spike in fatal medication errors: A possible effect of new medical residents. *J. Gen. Intern. Med.* **25**(8) 774–779.

Pierce, L., D. Snow, A. McAfee. 2015. Cleaning House : The impact of information technology monitoring on employee theft and productivity. *Manage. Sci.* **61**(10) 2299–2319.

Pitts, S. R., E. R. Carrier, E. C. Rich, A. L. Kellermann. 2010. Where Americans get acute care: Increasingly, it's not at their doctor's office. *Health Aff.* **29**(9) 1620–1629.

Polanyi, M. 1966. *The Tacit Dimension*. New York, NY, Anchor Day Books.

Porteus, E. L. 1990. Stochastic inventory theory. D. P. Heyman, M. J. Sobel, eds. *Handb. OR MS, Vol. 2*. North-Holland, The Netherlands, Elsevier Science Publishers, 605–652.

Preacher, K. J., A. F. Hayes. 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* **40**(3) 879–891.

Pronovost, P. J., D. Needham, S. M. Berenholtz, et al. 2006. An intervention to decrease catheter-related bloodstream infections in the ICU. *N. Engl. J. Med.* **355**(26) 2725–2732.

Rao, R. D., L. Argote. 2006. Organizational learning and forgetting: The effects of turnover and structure. *Eur. Manag. Rev.* **3**(2) 77–85.

Raz, D., B. Avi-Itzhak, H. Levy. 2006. Fairness considerations of scheduling in multi-server and multi-queue systems. *Proc. 1st Int. Conf. Perform. Eval. Methodolgies Tools (ValueTools '06)*. New York, New York, USA, ACM Press, 39.

Reagans, R., L. Argote, D. Brooks. 2005. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Manage. Sci.* **51**(6) 869–881.

Roels, G., X. Su. 2014. Optimal design of social comparison effects: Setting reference groups and reference points. *Manage. Sci.* **60**(3) 606–627.

Rothkopf, M. H., P. Rech. 1987. Perspectives on queues: Combining queues is not always beneficial. *Oper. Res.* **35**(6) 906–909.

Schull, M., M. Vermeulen, A. Guttmann, T. Stukel. 2015. Better performance on length-of-stay benchmarks associated with reduced risk following emergency department discharge: an observational cohort study. *Can. J. Emerg. Med.* **17**(3) 253–262.

Schultz, K. L., D. C. Juran, J. Boudreau. 1999. The effects of low inventory on the development of productivity norms. *Manage. Sci.* **45**(12) 1664–1678.

Shadish, W. R., T. D. Cook, D. T. Campbell. 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

Shah, R., S. M. Goldstein, B. T. Unger, T. D. Henry. 2008. Explaining anomalous high performance in a health care supply chain. *Decis. Sci.* **39**(4) 759–789.

Soremekun, O. A., C. Terwiesch, J. M. Pines. 2011. Emergency medicine: An operations management view. *Acad. Emerg. Med.* **18**(12) 1262–1268.

Spaite, D. W., F. Bartholomeaux, J. Guisto, et al. 2002. Rapid process redesign in a university-based emergency department: Decreasing waiting time intervals and improving patient satisfaction. *Ann. Emerg. Med.* **39**(2) 168–177.

Spear, S. J., H. K. Bowen. 1999. Decoding the DNA of the Toyota Production System. *Harv. Bus. Rev.* **77**(5) 96–106.

Staats, B. R., D. J. Brunner, D. M. Upton. 2011. Lean principles, learning, and knowledge work: Evidence from a software services provider. *J. Oper. Manag.* **29**(5) 376–390.

Staats, B. R., H. Dai, D. Hofmann, K. L. Milkman. 2015. Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. Working paper.

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

Staw, B. M. 1980. The consequences of turnover. *J. Occup. Behav.* **1**(4) 253–273.

Szulanski, G. 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strateg. Manag. J.* **17**(Winter Special Issue) 27–43.

Tafkov, I. D. 2013. Private and public relative performance information under different compensation contracts. *Account. Rev.* **88**(1) 327–350.

Tan, T. F., S. Netessine. 2014. When does the devil make work? An empirical study of the impact of workload on worker productivity. *Manage. Sci.* **60**(6) 1574–1593.

Taylor, F. W. 1911. *The Principles of Scientific Management*. New York, NY, Harper & Brothers.

Ton, Z., R. S. Huckman. 2008. Managing the impact of employee turnover on performance: The role of process conformance. *Organ. Sci.* **19**(1) 56–68.

Tucker, A. L., I. M. Nembhard, A. C. Edmondson. 2007. Implementing new practices: An empirical study of organizational learning in hospital intensive care units. *Manage. Sci.* **53**(6) 894–907.

Tucker, A. L., S. J. Spear. 2006. Operational failures and interruptions in hospital nursing. *Health Serv. Res.* **41**(3 I) 643–662.

Tushman, M. L., L. Rosenkopf. 1996. Executive succession, strategic reorientation and performance growth: A longitudinal study in the U.S. cement industry. *Manage. Sci.* **42**(7) 939–953.

U.S. Bureau of Labor Statistics. 2016a. Healthcare practitioners and technical occupations. Available at: http://www.bls.gov/oes/current/oes290000.htm [Accessed March 9, 2016].

U.S. Bureau of Labor Statistics. 2016b. Healthcare support occupations. Available at: http://www.bls.gov/oes/current/oes310000.htm [Accessed March 9, 2016].

U.S. Senate. 2010. *Patient Protection and Affordable Care Act*. H.R. 3590 - 111th Congress.

Vallis, J., A. Hesketh, S. Macpherson. 2004. Pre-registration house officer training: A role for nurses in the new Foundation Programme? *Med. Educ.* **38**(7) 708–716.

Vanek Smith, S. 2015. I, waiter. *Planet Money*. Available at: http://www.npr.org/templates/transcript/transcript.php?storyId=407086723.

Wagner, A. K., S. B. Soumerai, F. Zhang, D. Ross-Degnan. 2002. Segmented regression analysis of interrupted time series studies in medication use research. *J. Clin. Pharm. Ther.* **27**(4) 299–309.

van Walraven, C., A. Jennings, J. Wong, A. J. Forster. 2011. Influence of house-staff experience on teaching-hospital mortality: The "July Phenomenon" revisited. *J. Hosp. Med.* **6**(7) 389–394.

Wang, L., I. Gurvich, J. A. Van Mieghem, K. J. O'Leary. 2015. Collaboration and professional labor productivity: An empirical study of physician workflows in a hospital. Working paper, Kellogg School of Management, Northwestern University, Evanston.

Wang, X., L. G. Debo, A. Scheller-Wolf, S. F. Smith. 2010. Design and analysis of diagnostic service centers. *Manage. Sci.* **56**(11) 1873–1890.

Watson, D., L. A. Clark, A. Tellegen. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* **54**(6) 1063–1070.

Wen, T., F. J. Attenello, B. Wu, A. Ng, S. Y. Cen, W. J. Mack. 2015. The July Effect: An analysis of never events in the Nationwide Inpatient Sample. *J. Hosp. Med.* **10**(7) 432–438.

Whitt, W. 2006. The impact of increased employee retention on performance in a customer contact center. *Manuf. Serv. Oper. Manag.* **8**(3) 235–252.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data: Second Edition* 2nd ed. Cambridge, Massachusetts, MIT Press.

Wooldridge, J. M. 2012. *Introductory Econometrics: A Modern Approach* 5th ed. Mason, OH, South-Western Cengage Learning.

Young, J. Q., S. R. Ranji, R. M. Wachter, C. M. Lee, B. Niehaus, A. D. Auerbach. 2011. "July effect": Impact of the academic year-end changeover on patient outcomes: a systematic review. *Ann. Intern. Med.* **155**(5) 309–315.

Zizzo, D. J. 2002. Between utility and cognition: The neurobiology of relative position. *J. Econ. Behav. Organ.* **48**(1) 71–91.