Essays on Emergency Department Physician Performance

HARVARD | BUSINESS | SCHOOL

DISSERTATION ACCEPTANCE CERTIFICATE

(To be placed in First Library Copy)

The undersigned, approved by the Chair of the Doctoral Programs, have examined a dissertation entitled

Essays on Emergency Department Physician Performance

presented by **Raha Imanirad**

candidate for the degree of Doctor of Business Administration and hereby certify that it is worthy of acceptance.

Signature Ananth Raman

Ananth Raman, Co- Chair

Signature Soroush Sa

Soroush Saghafian, Co-Chair

Signature

Ryan Buell

Date 6/30/2020

Page left intentionally blank.

Essays on Emergency Department Physician Performance

A dissertation presented

by

Raha Imanirad

to

The Technology and Operations Management Unit

in partial fulfillment of the requirements

for the degree of

Doctor of Business Administration

in the subject of

Operations Management

Harvard University

Cambridge, Massachusetts

June 2020

Dissertation Advisor: Professor Soroush Saghafian Raha Imanirad

Essays on Emergency Department Physician Performance

Abstract

In this dissertation, I examine the problem of physician performance evaluation and investigate ways to improve the performance of physicians in the context of an Emergency Department (ED) setting. In the first chapter — co-authored with Soroush Saghafian and Stephen Traub — we use Data Envelopment Analysis (DEA) to develop models for evaluating physician effectiveness and efficiency. We apply our DEA models to a large dataset of care delivered by ED physicians and derive effectiveness and efficiency scores for the physicians in our dataset. Using the generated DEA scores, we then conduct a second-stage analysis in which we use a Tobit framework to identify factors that are associated with higher levels of physician effectiveness and efficiency.

In the second chapter — co-authored with Soroush Saghafian and Stephen Traub — we conduct a large-scale empirical investigation into whether and how physicians who work during the same shift affect each other's performance. We find strong empirical evidence that physicians affect each other's speed and quality in our setting. We identify spillover from peers' utilization of shared resources as the main driver of the observed effects and show that during high-volume shifts, the magnitude of the effects increases. We draw conclusions from our results and discuss how they can be utilized by hospital administrators to improve the overall performance of physicians.

In the third chapter — co-authored with Soroush Saghafian and Stephen Traub — we address the question: To which shift should the ED's high-performing

physicians be assigned? Specifically, we empirically examine how assigning a high-performing group of physicians to different shifts of the day affects the daily performance of the ED. Our results demonstrate that assigning a group of high-performing physicians to the first shift of the day has the highest impact on the daily performance of the ED. We further show that physicians' performance in the earlier shifts of the day has a "domino effect" throughout the rest of the day.

Together these studies provide insights into ED physician performance and shed light on potential ways to improve performance through assigning the right mix of physicians to the right shifts.

Contents

List of Figures

List of Tables

Acknowledgments

Writing this dissertation would not have been possible without the help and support of many individuals. First and foremost, I am indebted to my committee members, Ananth Raman, Soroush Saghafian, and Ryan Buell, for their encouragement, patience, and guidance throughout this adventure. I am forever grateful to have had the opportunity to learn from them, not only about research but also about life. I am also profoundly thankful to my first research mentors, Wade Cook and Julian Scott Yeomans, for introducing me to the world of research.

I wish to extend my sincere appreciation to Dr. Stephen Traub at the Mayo Clinic Arizona for his generosity and invaluable feedback. I benefited enormously from our collaborations and his infectious passion for improving operational efficiency in emergency departments.

Thank you to the Research Computing Services and in particular to Xiang Ao who has been a tremendous source of support throughout my doctoral studies. I am also grateful for the generous financial support from the HBS Doctoral Programs. My heartfelt appreciation to Jennifer Mucciarone, Angela Valvis, Marais Young, Kathy Randel, Keith Foster, John Korn, LuAnn Langan, and Trevor Bierig for helping me navigate the doctoral journey. To all my fellow TOMers, past and present: Thank you for the camaraderie, encouragement, and memories that I will always cherish.

Lastly, but most importantly, to my family: Thank you for giving me the courage to pursue my dreams. For your love and support, I am forever grateful.

Dedicated to all healthcare providers on the frontlines of the COVID-19 battle.

Introduction

The burgeoning expense and complexity of the current care delivery system have prompted healthcare organizations to improve the effectiveness and efficiency of care delivery. Given that most decisions regarding resource utilization are made by physicians, improving the healthcare delivery system requires an understanding of the effectiveness and efficiency of care delivered by physicians. Although this topic is of great interest to policymakers, researchers, and hospital managers, rigorous methods of evaluating physician effectiveness and efficiency have proven elusive.

In this dissertation, I examine the problem of physician performance evaluation in the context of an Emergency Department (ED) setting. Furthermore, I explore ways to improve physician performance through operational changes that require no additional investments. Specifically, the work in this dissertation derives insights into ways to construct an optimal mix of physicians to be assigned to the same shift and examines how assigning a high-performing group of physicians to different shifts of the day affects the daily performance of the ED. This work provides practice-related implications for hospital administrators as it sheds light on potential ways to improve the performance of physicians in hospital EDs.

In Chapter 1, "Who is an Effective and Efficient Physician? Evidence from Emergence Medicine" — co-authored with Soroush Saghafian and Stephen Traub — we use Data Envelopment Analysis (DEA) to develop models that gauge physician performance in terms of effectiveness and efficiency. We apply our DEA models to a large dataset of care delivered by ED physicians and generate effectiveness and efficiency scores for the physicians in our dataset. In order to validate our generated DEA scores, we use Machine Learning (ML) algorithms to predict the effectiveness and efficiency of the physicians in our dataset. We observe a 76% overlap between the results derived from the ML approach and those obtained from our DEA models. We then use the derived DEA scores along with Tobit analysis to identify the distinguishing behaviors of physicians who perform highly on the effectiveness and efficiency metrics. In addition, we use this framework to examine the influence of peers on a focal physician's effectiveness and efficiency. We find that highly effective physicians order less tests compared to their peers and maintain their effectiveness when working under high workloads. We also observe that highly efficient physicians order less tests on average and become even more efficient during high-volume shifts. Importantly, our results indicate a statistically significant positive relationship between a physician's effectiveness and efficiency scores, suggesting that effectiveness and efficiency in care delivery should be viewed as complements. Finally, we find evidence of peer influence on a focal physician's effectiveness and efficiency, which suggests an opportunity to improve system performance by taking physicians' relative characteristics into account when determining the group of physicians that should be scheduled during the same shift.

In Chapter 2, "Do Physicians Influence Each Other's Performance? Evidence from the Emergency Department" — co-authored with Soroush Saghafian and Stephen Traub — we examine whether and how physicians who work alongside each other during the same shift affect each other's performance. We find strong empirical evidence that physicians affect each other's speed and quality in our setting. Specifically, our results show that a faster peer has a negative effect on a focal physician's average speed while a slower peer has a positive effect on the average speed of a focal physician. Similarly, we find that a higher-quality peer negatively affects a focal physician's average quality while a lower-quality peer positively influences the average quality of a focal physician. We identify spillover from peers' utilization of shared resources as the main driver of the observed effects and show that during high-volume shifts (i.e., when the shared resources are most constrained), the magnitude of the effects increases. We provide further evidence for the resource spillover mechanism by showing that physicians influence their peers' speed and quality through affecting their test order count and admission rate, respectively.

In Chapter 3, "Which Shift Matters the Most? Evidence from the Emergency Department" — co-authored with Soroush Saghafian and Stephen Traub — we utilize a day-level dataset collected from the ED of a leading U.S. hospital, and address a simple but important question: To which shift (e.g., first, second, etc.) should the ED's high-performing physicians be assigned? Answering this question requires identifying whether and how assigning high-performing physicians to different shifts of the day affects the daily performance of the ED. We evaluate the performance of the ED in terms of speed, quality, and admission rate, where speed and quality are measured using the average patient LOS and 72-hour rate of return, respectively. Our results show that assigning a group of physicians with a higher-than-average aggregate speed to the first shift results in a 8.2-minute improvement in the average daily speed of the ED. While we find weak statistically significant evidence of the first-shift effect with respect to physician quality, we find statistically significant evidence that assigning a group of physicians with a higher-than-average aggregate admission rate to the first shift has the highest impact on the average daily admission rate of the ED. We also examine whether our results are sensitive to heterogeneity in ED volume and day of the week. We find that the first-shift effect is stronger on high-volume days and on weekdays. Our findings can be useful in the area of physician scheduling as they highlight the importance of assigning high-performing physicians to the first shift of the day.

Chapter 1

Who is an Effective and Efficient Physician? Evidence from Emergence Medicine

1.1 Introduction

Motivation. Healthcare spending is projected to rise to 19.9% of the GDP by 2025 (Keehan et al. 2017), spurring interest in finding new ways to improve both the effectiveness and efficiency of care delivery. As most decisions regarding utilization of healthcare services are ultimately made by frontline clinicians (Tsugawa et al. 2017), understanding and evaluating provider performance could help to identify sources of waste in the healthcare sector. Although care delivery performance measurement initiatives have proliferated in recent years, there are few rigorous methods for evaluating the effectiveness and efficiency of physicians. A method for evaluating the effectiveness and efficiency of physicians is especially needed for understanding what the effective and efficient physicians do differently than their peers. This understanding of best practices can, in turn, result in training more effective and efficient physicians, and thereby, improve the performance of the healthcare sector.

In this study, we focus on care delivery in hospital Emergency Departments (EDs). Specifically, we collect a large dataset of care delivered by ED physicians that includes more than 115,000 patient visits. We employ Data Envelopment Analysis (DEA) — a linear programming optimization technique that provides a multi-dimensional evaluation tool — to develop scores related to physician effectiveness and efficiency. We validate our generated DEA scores by making use of various Machine Learning (ML) algorithms, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Random Forest (RF), a Generalized Linear Model (GLM), and Least Absolute Shrinkage and Selection Operator (LASSO). Our results show that there is a 76% overlap between the results derived from the best ML approach and those obtained from our DEA models, giving us confidence about the validity of our DEA models. Unlike the ML algorithms, however, DEA provides an important advantage in terms of *interpretability*, since it offers a clear input-output view of a physician's performance and avoids any "black-box" operations. Thus, it can be easily communicated to (a) hospital administrators who are interested in improving the effectiveness and efficiency of care delivery in their hospitals, and (b) physicians who are interested in improving their own individual performance.

In order to learn about what the high-performing physicians do differently than other physicians, and thereby generate insights into best practices, we conduct a second-stage analysis in which we use our DEA scores along with a Tobit framework to identify factors (e.g., test order count, experience, etc.) associated with higher levels of performance. Furthermore, we use our framework to study how physicians influence each other's effectiveness and efficiency. In particular, we make use of our DEA models and consider various peer characteristics, including relative effectiveness, efficiency, gender, and type of medical degree (MD vs. DO) to examine whether and how these peer characteristics affect a focal physician's effectiveness and efficiency.

Data and Setting. Our data consist of detailed care delivery information associated with 115,350 patient visits in a leading U.S. hospital. Our partner ED is equipped with an emergency medicine team comprising 32 board-certified physicians and more than 70 registered nurses. All patients in our partner ED are algorithmically assigned to physicians upon arrival through an automated rotational patient assignment process (Traub et al. 2016). This workflow essentially removes all patient selection biases or preferences of physicians in "cherry-picking" their patients.

All patients who were identified in the electronic health record system as having been seen by an ED physician between July 12, 2012, and July 31, 2016 were included in our dataset. Patient-specific data include demographic (age, gender, race) and insurance information. Encounter-level data include laboratory tests, chief complaint, Emergency Severity Index (ESI) level (a five-level triage scale that categorizes patients according to their acuity levels), day of the ED visit, and time of the day, among others. To avoid distortion of the results by outliers, we excluded 4 physicians with relatively low patient volumes (fewer than 200 visits over the 4-year period) from our analyses. Our final dataset comprises 110,325 patient-visit-level observations.

Research Questions. We address four research questions as follows. Research Question 1: Are effectiveness and efficiency of a physician substitutes (negatively correlated) or complements (positively correlated)? Research Question 2: What is the relationship between effectiveness/efficiency of a physician and various characteristics, including those of the physician (e.g., test order count, experience, job tenure), patients (e.g., race, gender, age, ESI), and the environment (e.g., ED volume/workload)? Research Question 3: What do highly effective and efficient physicians do differently than their peers? Research Question 4: How do physician peers influence each other's effectiveness and efficiency? Addressing these questions enables us to (a) shed light on factors that affect physicians' effectiveness and efficiency, and (b) provide actionable insights into ways physicians' effectiveness and efficiency can be improved.

Main Findings. Regarding Research Question 1, our results indicate that a conventional wisdom about the efficiency-effectiveness tradeoff in the healthcare sector might not be true. This conventional wisdom suggests that improving the efficiency of care delivery comes at the price of lowering effectiveness. Contrary to this conventional wisdom, we find that overall, there is a statistically significant positive association between physicians' efficiency and effectiveness scores. This implies that physicians who are efficient in care delivery are also more likely to provide effective care (and vice-versa). Our results, hence, suggest that physician effectiveness and efficiency serve as *complements* and not *substitutes*.

With respect to Research Question 2, we find that a physician's efficiency score is negatively associated with his/her average number of test orders per patient visit and positively correlated with his/her experience (measured in number of years after graduation from medical school). This implies that efficient physicians are those who (a) order less tests, and (b) are more experienced. In addition, we observe a statistically significant negative relationship between a physician's effectiveness and his/her job tenure (measured in number of years the physician has worked in our partner ED). This finding might be related to a selection bias: the ED might have imposed higher hiring standards in recent years or simply has been able to attract physicians with higher effectiveness levels. However, our finding might also be due to a difference in motivation levels of new hires versus those of existing physicians.

Newly hired employees typically are more motivated than existing employees to perform well (Hackman and Oldham 1980, Kass et al. 2001, Bruursema et al. 2011). As such, they might inherently score higher on the effectiveness metric. Our dataset is insufficient for distinguishing between these potential hiring and motivation differences (which are both difficult to measure and hidden to us). Nevertheless, our finding that job tenure negatively impacts physicians' effectiveness provides an important avenue for future research to shed light on mechanisms that might improve effectiveness of care delivery (e.g., motivational training programs, providing performance-based incentives for physicians with long job tenure, or making use of specific hiring procedures). Furthermore, our results show that patient characteristics have little, if any, effect on a physician's effectiveness and efficiency. We also find that high workloads (during high-volume shifts) have a negative effect on a physician's effectiveness and a positive effect on a physician's efficiency.

Addressing Research Question 3, our findings suggest that highly effective physicians order less tests on average compared to their peers. Our results also indicate that during high-volume shifts, highly effective physicians are able to maintain their effectiveness compared to their peers. Similarly, our results show that highly efficient physicians have a lower average test order count and become even more efficient under high workloads.

Finally, addressing our last research question (Research Question 4), our findings suggest that working alongside more effective and efficient peers is negatively associated with improving a focal physician's effectiveness and efficiency, respectively. This is consistent with the findings in Saghafian et al. (2019), which studies the influence of physicians on each other's performance using a different methodology. The authors show that a "resource spillover" effect caused by the existence of shared resources with limited capacities in the ED is the mechanism driving the observed peer influence.

Implications. Our results have various implications for both hospital administrators and physicians. In particular, our DEA models provide hospital administrators with a transparent and easy-to-understand scoring system to evaluate the effectiveness and efficiency of care delivered in their hospitals. Similarly, our models allow individual physicians to identify their weaknesses and realize the advantages of following what the highly effective and efficient physicians do in their practice. We expect well-designed training programs to be able to facilitate this learning process. In addition, our results have implications for physician scheduling programs, where hospital administrators need to decide upon the set of physicians who should work during the same shift. In particular, our analyses of our DEA scores show that effectiveness and efficiency scores of a physician decline while working alongside more effective and efficient peers, respectively. This observation can be incorporated in future scheduling programs and utilized as a mechanism for improving the overall performance of physicians. Finally, as noted earlier, our results provide an important avenue for future research to explore and implement mechanisms for performance improvement including designing motivational training programs, providing performance-based incentives for physicians with long job tenure, or making use of specific hiring procedures.

1.2 Related Studies

Evaluating the performance of physicians has gained attention in research as health policymakers look for ways to drive quality improvement and increase physicians' accountability for achieving quality goals. Most lines of research on the topic of physician performance evaluation have focused on specific patient conditions.

For example, Glickman et al. (2008) use clinical measures such as performing a diagnostic electrocardiogram (ECG) for syncope in patients older than 60 years as a criterion for physician performance measurement. Hess et al. (2011) utilize physician performance measures such as completion of retinal and foot exams and blood pressure test to assess the quality of care provided to diabetic patients. However, the findings generated from such studies may not be generalizable to settings such as EDs where there exists heterogeneity in patient population. Other studies have evaluated behavioral aspects of physician performance using questionnaires (Smith et al. 2004) and patient chart audits (Goulet et al. 2002). Qualitative metrics, however, are difficult to measure and may cause bias in performance evaluation.

Various performance-specific measures have been used to assess the performance of ED physicians. A review of the literature highlights ED time intervals such as the time between patient arrival to initial clinical assessment, Length of Stay (LOS), as well as the percentage of patients who leave without being seen, re-admissions within 72 hours and mortality/morbidity rate as most frequently used performance measures in EDs (see Fernandes et al. 1997, Spaite et al. 2002). Using pure performance measures for evaluating ED physicians, however, does not account for the amount of resources utilized by physicians. In a setting such as an ED, where resources are shared and constrained, physicians' utilization of shared resources could influence their effectiveness and efficiency. Hence, using a methodology such as DEA, which incorporates resource utilization into performance evaluation, lends itself well to evaluating physician performance in EDs.

DEA has been applied in a variety of healthcare settings including hospitals (Sherman 1984, Grosskopf and Valdmanis 1987), veterans administration medical centers (Harrison and Ogniewski 2005), and organ procurement organizations (Ozcan et al. 1999) to evaluate the relative performance of healthcare institutions.

While performance evaluation of hospitals has been explored in prior literature (Hollingsworth 2008, Varabyova and Schreyögg 2013, Castelli et al. 2015, Zheng et al. 2018), the performance of physicians has proven to be more difficult to assess because of diversity in patient mix and treatments, and differences among specialties, among others (Storfa and Wilson 2015). Hence, macro parameters and proxies such as billing and reimbursement are often used to capture physician performance (Johannessen et al. 2017). For example, Wagner et al. (2003) propose DEA models focused on cost containment by using admission and patient visit payments as input variables. Collier et al. (2006) use the total billable charges attributed to physicians as one of the outputs of their proposed model. The authors, however, assume uniform resource utilization among physicians. Other studies use costs of treating specific patient conditions such as sinusitis (Pai et al. 2000) and asthma (Ozcan 1998) in their suggested DEA models.

Our study contributes to this literature by proposing two DEA models for evaluating physician effectiveness and efficiency. Our choice of the models' input and output variables reduces the risk of overfitting to our study setting and increases generalizability of the models to any ED setting. Specifically, we do not use parameters specific to patient health conditions or physician practice style in our models. Rather, we investigate the effects of patient- and physician-specific factors on physician performance in a second-stage analysis, where we identify characteristics of effective and efficient physicians.

Our work is also related to studies on speed-quality tradeoffs. Anand et al. (2011) use a queueing framework to examine the speed-quality tradeoff in a customer-intensive service setting and study how service providers make the optimal speed-quality tradeoff. Saghafian et al. (2018) study the speed-quality tradeoffs in a telemedical physician triage system in the context of an ED setting. Several other studies have examined the interactions between speed and quality of service in different settings including call centers (Hasija et al. 2008) and medical diagnostic services (Wang et al. 2010). Our work contributes to this stream of literature by examining the relationship between physician effectiveness and efficiency using metrics derived from the DEA methodology.

1.3 DEA Models

DEA, first introduced by Charnes et al. (1978), is a methodology useful in evaluating the relative performance of a set of decision making units (DMUs) in a multiple input, multiple output setting. A DMU can be viewed as an entity responsible for converting a number of inputs into a set of outputs (Cooper et al. 2007). Contrary to a central tendency approach, which evaluates units relative to an average performer, DEA computes a DMU's relative performance by using the best-performing units as the basis for comparison. One of the key advantages of DEA over other regression-based statistical methods is that it does not require specification of any functional relationship (e.g., a specific linear or non-linear model) between inputs and outputs. As a result, DEA can uncover information that remains hidden from other parametric methodologies, and hence, might capture a more complete picture of a DMU's performance relative to the resources it uses. As a data-driven approach, however, DEA is vulnerable to data errors and outliers.

The conventional input-oriented DEA methodology evaluates each DMU j in the population based upon a set of inputs $\{x_{ij}\}_{i=1}^{I}$ and outputs $\{y_{rj}\}_{r=1}^{R}$ by assuming a proportional reduction in all inputs while maintaining a fixed level of outputs. According to an output-oriented approach, this methodology provides for a proportional expansion in outputs rather than a reduction in inputs while keeping inputs constant. For the goals of this study, we make use of both the input- and output-oriented mechanisms.

The original DEA model is based on a Constant Returns to Scale (CRS) methodology. The input-oriented CRS model takes the following form:

$$\begin{aligned} max \quad & \theta = \frac{\sum_r u_r y_{rj_o}}{\sum_i \nu_i x_{ij_o}} \\ s.t. \quad & \frac{\sum_r u_r y_{rj}}{\sum_i \nu_i x_{ij}} \leq 1 \qquad j = 1, ..., n, \\ & u_r, \nu_i \geq 0, \qquad r = 1, ..., R; \quad i = 1, ..., I, \end{aligned} \tag{1.1}$$

where y_{rj_o} and x_{ij_o} represent the output(s) and input(s) of DMU j_o, respectively, and $\{u_r\}_{r=1}^{R}$ and $\{\nu_i\}_{i=1}^{I}$ are decision variables representing the most favorable set of weights for the DMU under evaluation. The constraints ensure that, when this set of weights is applied to each DMU in the population, no unit's efficiency exceeds 1. The maximum value obtained for DMU j_o is that unit's DEA score, and a value of 1 signifies a frontier-efficient unit. Contrary to composite scoring methods which apply a single set of weights to each unit in the population, DEA assigns a different set of weights to each DMU under evaluation. Hence, it avoids the subjective nature of weight assignment in multi-objective problems.

In order to evaluate the performance of physicians, we develop two DEA models: (1) an *effectiveness DEA model* (see Section 1.3.1), and (2) an *efficiency DEA model* (see Section 1.3.2). To improve the power of our statistical analyses and ensure enough variation across the models' input/output parameters, we conduct our analyses at the physician-year level. Specifically, we design our DMUs so that they each capture a physician's performance in a particular year. To this end, we construct a dataset that includes 106 physician-year observations. Our DEA models, therefore, evaluate the effectiveness and efficiency of individual physician i who uses hospital resources to deliver care in a given year t relative to his/her peers. Furthermore, since there is no reason to believe that an increase in inputs results in a proportional change in outputs in our effectiveness and efficiency DEA models, we have used the Banker-Charnes-Cooper (BCC) model (Banker et al. 1984) which extends the CRS model to allow for variable returns to scale. We tested this assumption by making use of Simar and Wilson's (Simar and Wilson 2002, Simar and Wilson 2011) returns-to-scale tests for both the effectiveness and efficiency DEA models.

The choice of the input and output variables in each model is based on the view of the physician as a "production entity" utilizing hospital resources (inputs) to generate effective and efficient care (outputs). It is important to note that there is no objective definition of the 'right' variables to use as inputs and outputs. We have chosen to define the models' inputs and outputs in terms of parameters (a) that best reflect a physician's performance, (b) for which there is at least face validity and some level of agreement among physicians, and (c) that are discussed in the literature of Emergency Medicine and ED operations as common performance measures. For example, to define our output variables, we note that efficiency in the ED can be measured in multiple ways. We primarily focus on a physician's average contact-to-disposition time (the time from when the physician initiates the first contact with the patient until the time a disposition order is issued for the patient), because all else equal a lower contact-to-disposition time means that a higher number of patients can be moved through the ED per unit of time (i.e., a higher ED throughput). Given that ED crowding has reached epidemic proportions in the last several years, improving physicians' contact-to-disposition time has become even more important (Salway et al. 2017).

Similarly, we consider the percentage of discharged patients who do not return to the ED within 72 hours as our primary output variable for our effectiveness DEA model. Returns to the ED within 72 hours of discharge may result from a sub-optimal (i.e., ineffective) first visit, in which not all medical issues were sufficiently identified or addressed. The 72-hour rate of return has been proposed as a measure of quality in the Emergency Medicine literature (see, e.g., Abualenain et al. 2013, Pham et al. 2011, Klasco et al. 2015) although using it for measuring quality of care is controversial. Nevertheless, to check the robustness of our results, we repeat our analyses using different combinations of input/output variables for both our effectiveness and efficiency DEA models, and observe that our main results hold (see Section 1.7 for our robustness checks).

Furthermore, as noted earlier, we validate our DEA scores through comparison with the results obtained using various ML algorithms (see Section 1.4) that do not necessarily rely on the same set of variables used in our DEA models. In particular, unlike our DEA models, these ML algorithms are given the entire dataset and are able to either use it as a whole or select the important variables using some predetermined regularization techniques. The fact that we obtain similar results from our DEA models and the ML algorithms gives us further confidence about the validity of our DEA models.

Finally, we note that due to the nature of the automated rotational patient assignment algorithm implemented in our partner hospital, which randomly assigns arriving patients to physicians, risk-adjustments of outcome measures are likely not essential. Nevertheless, in our statistical analyses we control for various patient characteristics that might affect physician performance (see Section 1.5).

1.3.1 Effectiveness DEA Model

Our main effectiveness DEA model uses the following set of variables as inputs and outputs. As noted earlier, in our robustness checks, we test the validity of our main DEA models by repeating our analyses using different combinations of input/output variables.

Output:

- *Rate of discharged patients who do not return within 72 hours*: Since a high 72-hour return rate is an undesirable indicator of care delivery effectiveness in the ED (see, e.g., Abualenain et al. 2013, Pham et al. 2011, Klasco et al. 2015), we use the proportion of a physician's discharged patients who do not return to the ED within 72 hours of their initial discharge as the model's output variable.

Input:

- *Average patient Length of Stay (LOS):* This variable captures the total time patients spend in the ED from registration to discharge.

For the effectiveness model, we choose the output-oriented DEA approach based on which the conceptual goal is to maximize outputs for a given level of inputs. Specifically, we compare physicians' percentage of patients who are discharged home after their ED visit and do not return within 72 hours (output) for a given level of LOS (input), where LOS can be viewed as a surrogate measure for using hospital resources (e.g., using diagnostic test services, ED beds, etc.). Intuitively, physicians who score higher on the effectiveness metric are those with a lower 72-hour rate of return for a fixed level of overall use of ED resources measured by the surrogate variable, LOS. From a patient perspective, this roughly means that the service is considered to be more effective if the chance of returning to the ED (e.g., due to an unresolved issue) is minimized per hour spent in the ED.[1] Both the LOS and 72-hour rate of return metrics have been used in the literature as valid performance measures (see, e.g., Chilingerian 1995, Fiallos et al. 2017). We refer to the θ scores generated by the DEA model with the above input-output parameters as physicians' *effectiveness scores*. Similarly, we refer to the θ scores generated by the DEA model with the input-output parameters described in the next section as physicians' *efficiency scores*.

1.3.2 Efficiency DEA Model

Our main efficiency DEA model uses the following set of variables as inputs and outputs.

Outputs:

- *Low ESI-level patients*: Percentage of patients served by the physician who have ESI levels 1 and 2 (i.e., high-acuity patients);
- *Patients older than 65*: Percentage of patients served by the physician who are older than 65.

Input:

- *Average contact-to-disposition time:* This variable denotes the time from the physician's initial contact with the patient to the time that a disposition order is issued.

[1]In EDs, the service is provided by a specific physician who is in charge of the patient, and the ED service is very rarely composed of teamwork among physicians (see, e.g., Saghafian et al. 2012, Saghafian et al. 2019, and the references therein). Thus, a patient's outcomes are directly related to the physician who serves him/her.

For the efficiency model, we use an input-oriented approach based on which physicians with higher efficiency scores in this setting are those who have a lower average contact-to-disposition time for a given mix of patients they serve. Low-ESI patients and those older than 65 are known to be patients that have a relatively higher contact-to-disposition time compared to other patients (Latham and Ackroyd-Stolarz 2014). Thus, assuming that two physicians serve the same mix of patients (ratio of low-ESI and older patients to other patients), the one who can maintain a lower contact-to-disposition time, will have a higher throughput (a widely-used measure of operational efficiency).

Our selection of the efficiency model's input/output variables described above is mainly based on our discussions with ED physicians.[2] In particular, our discussions indicate that while a physician's ability to serve patients efficiently might be attributable to his/her cognitive skills, his/her average contact-to-disposition time given a fixed mix of low-ESI and older patients s/he sees can serve as a valid proxy for measuring such skills. We also note that while we have chosen patient LOS as the effectiveness model's input variable, our choice for the efficiency model's input variable is the average contact-to-disposition time. The reason is that LOS captures the total time a patient spends in the ED, which is not fully controllable by the physician. In contrast, contact-to-disposition time is at the discretion of physicians. Finally, we note while LOS and contact-to-disposition time are positively correlated, the fact that our effectiveness and efficiency models use different DEA orientations ensures that any potential relationship between physicians' effectiveness and efficiency scores is not merely due to the inherent dependency between these variables.

[2]One of the authors of this paper is the chairman of the ED of our partner hospital, which is a leading hospital in the U.S.

1.3.3 Physician-Pair DEA Models

Our DEA models described in the previous sections allow us to capture individual physicians' effectiveness and efficiency, and answer our first three research questions (Research Questions 1, 2, and 3). In order to also examine the effects of peers' presence on a focal physician's effectiveness and efficiency scores (Research Question 4), we use a variation of the proposed DEA models in which each DMU comprises physician i who has worked alongside his/her peer physician j in year t. Our physician-pair DEA models, hence, capture a focal physician i's average effectiveness and efficiency while working alongside his/her peer physician j in year t. We identify a focal physician's peers as those physicians who have worked alongside the focal physician during the same shifts. We then construct a dataset comprising every combination of focal-peer physician pairs corresponding to each year of our study period. This leaves us with 2,268 physician-pair observations (DMUs) that we use in our physician-pair analysis. Making use of all of our four DEA models (individual and physician-pair effectiveness and efficiency models), in turn, enables us to provide answers to our four research questions (see Section 1.6).

1.4 Machine Learning (ML) Algorithms

To test the validity of our generated DEA scores, in addition to re-running our DEA models with different sets of input/output variables (see Section 1.7), we make use of various ML algorithms including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Random Forest (RF), a Generalized Linear Model (GLM), and Least Absolute Shrinkage and Selection Operator (LASSO). We first compare these algorithms in terms of their performance in predicting the effectiveness and efficiency of physicians. We do so via 5-fold cross-validation, which allows us to measure the average out-of-sample performance of these algorithms across different training and test sets. We label the highly effective and efficient physicians in the training sets as those with lower-than-average 72-hour rate of return and contact-to-disposition time, respectively. The input variables (potential predictors) that are used by the ML algorithms include various patient characteristics (age, gender, race, ESI), physician characteristics (average test order count, job tenure, admission rate, etc.), and ED characteristics (e.g., ED volume). A summary statistics of these variables is presented in Table 1.1. We omit the 72-hour rate of return and average contact-to-disposition time variables from the set of potential predictors in the effectiveness and efficiency ML models, respectively, since these represent the outcome variables (i.e., what the algorithms are asked to predict).

We compare the predictive power of the ML algorithms using the Area Under the Curve (AUC) measure as well as classification accuracy and the Kappa coefficient (which adjusts for the effect of random chance on accuracy). These measures (calculated using 5-fold cross-validation) are presented in Figures 1.1-1.4. As demonstrated in these figures, the RF algorithm results in the highest AUC, accuracy, and Kappa measures compared to the other algorithms. We, therefore, use the RF model to predict the highly effective and efficient physicians in the test sets. We then compare the predictions made by the RF algorithm to those derived from our DEA models. To this end, we use the average effectiveness and efficiency DEA scores to categorize physicians into the following four groups:

Group 1: Highly effective / Highly efficient;

Group 2: Highly efficient / Lowly effective;

Group 3: Highly effective / Lowly efficient;

Group 4: Lowly efficient/ Lowly effective.

Table 1.1: Summary Statistics - ML Variables

Variable	Mean	SD	Min	Max
Patient Characteristics				
Older than 65 (%)	45	2.88	39	58
Female (%)	53	1.83	48	58
White (%)	91	1.54	87	95
ESI Levels 1 and 2 (%)	15	2.13	7.5	21
Physician Characteristics				
Test Order Count	144.13	24.37	87.55	215.39
Experience (Years)	22.16	7.49	6	39
Job Tenure (Years)	8.38	6.01	0	18
Admission Rate (%)	0.11	0.03	0.05	0.20
Over-Calling Rate (%)	0.18	0.05	0.08	0.33
Under-Calling Rate (%)	0.04	0.02	0	0.11
LOS (Minutes)	235.02	26.84	180.64	297.81
72-hr Rate of Return (%)	0.03	0.01	0.01	0.06
Contact-to-Disposition Time (Minutes)	144.13	24.37	87.55	215.39
ED Characteristics				
ED Volume (Patients per Physician Shift)	23.77	4.90	12.20	41.85

Note: $N = 106$. Observations are at the physician-year level.

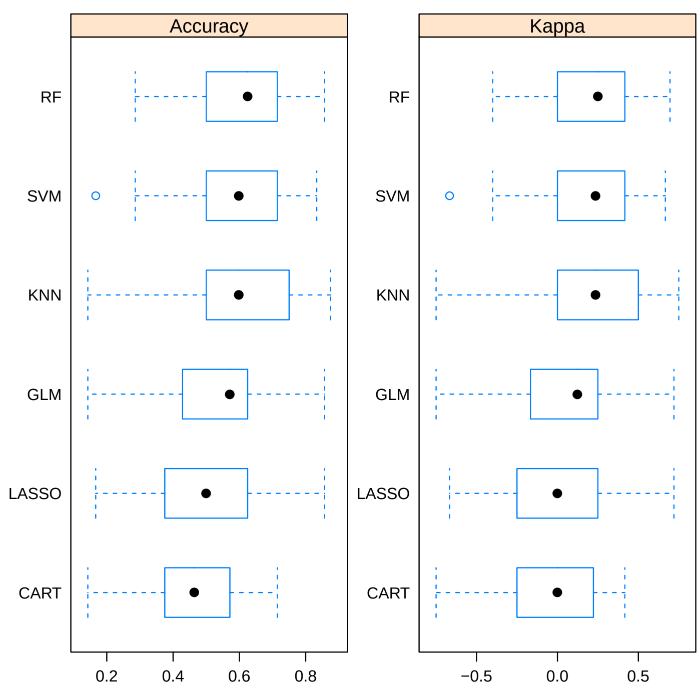


Figure 1.1: Accuracy and Kappa Measures of Effectiveness ML Models

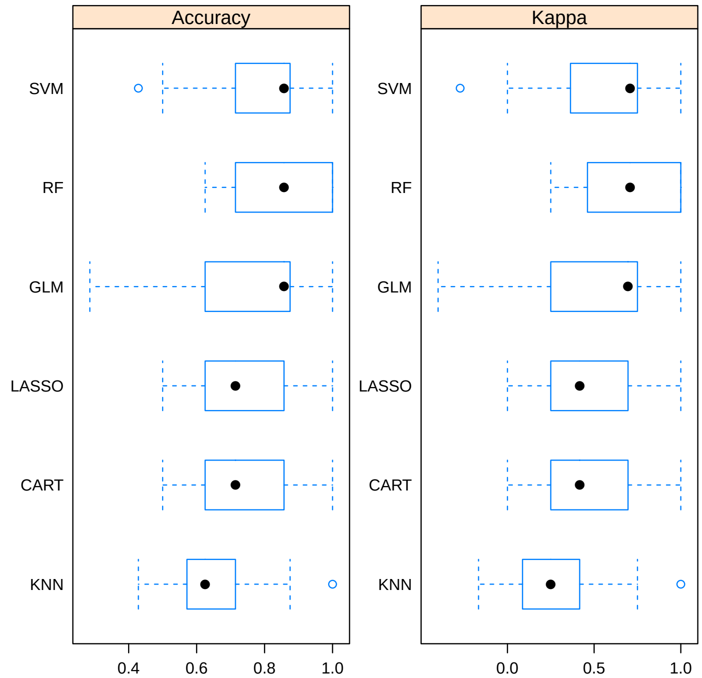


Figure 1.2: Accuracy and Kappa Measures of Efficiency ML Models

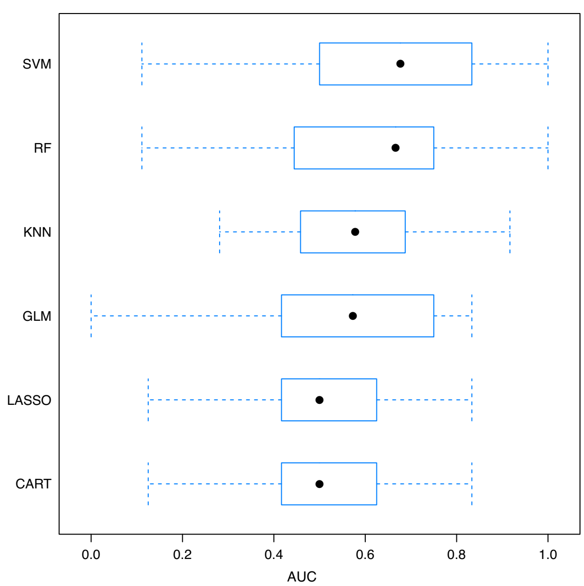


Figure 1.3: Effectiveness ML Models

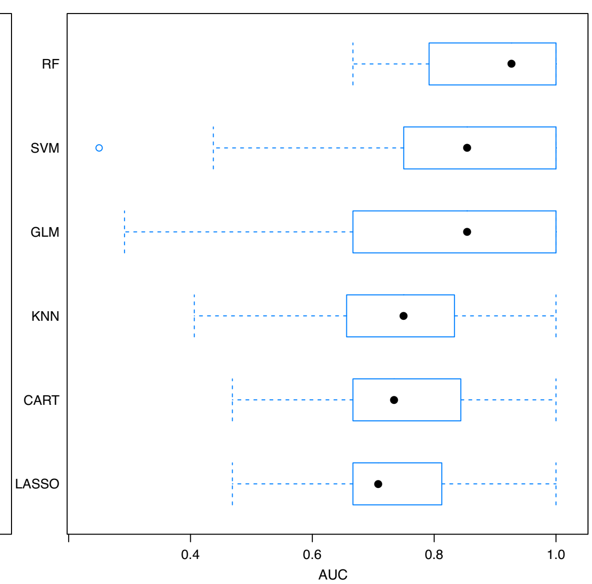


Figure 1.4: Efficiency ML Models

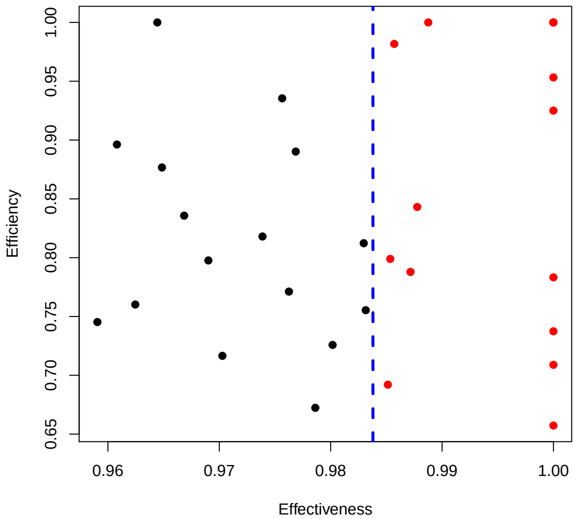
Independently, we use the predictions obtained from the best ML model — the RF algorithm — to classify physicians into the aforementioned four groups. We then compare the classifications derived from the DEA and ML approaches as illustrated in Figures 1.5-1.8. In these figures, red data points indicate highly effective and efficient physicians. We find an average 76% overlap between the classifications obtained via the DEA and ML approaches. This finding validates the accuracy of our proposed DEA models to a great extent. This is especially the case since the RF algorithm uses a different set of input variables compared to those used in our DEA models. For example, Figures A1 and A2 in Appendix A present the variable importance graphs corresponding to the RF effectiveness and efficiency models, respectively.[3] As illustrated in these figures, the RF algorithm's selection of important variables is completely different than that of our DEA models. Yet, the results obtained from the RF model significantly overlap with those of our DEA models. This gives us confidence about the validity of our DEA models.

1.5 Statistical Methodology

To gain insights into our Research Questions 1-3, we regress the generated DEA scores of physician i in year t (θ_{it}) (defined in Section 1.3), on a set of explanatory variables related to physician, patient, and ED characteristics. The regression model takes the following general form:

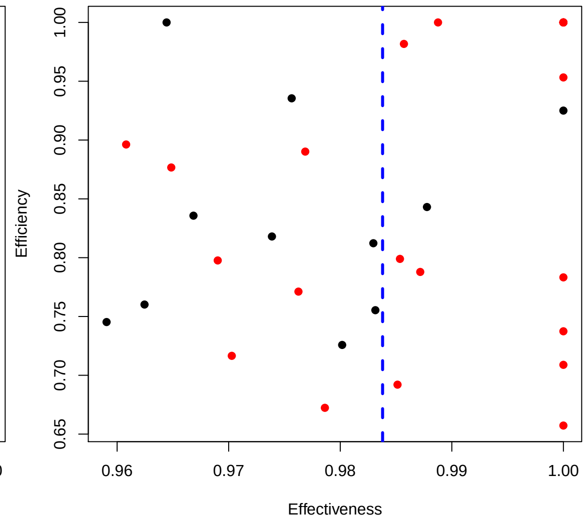
$$\theta_{it} = \beta_1 U_{it} + \beta_2 W_{it} + \beta_3 E_{it} + \gamma_t + \epsilon_{it}, \tag{1.2}$$

[3]These figures demonstrate the mean decrease in node impurity (the Gini coefficient) such that a higher Gini coefficient denotes higher variable importance.



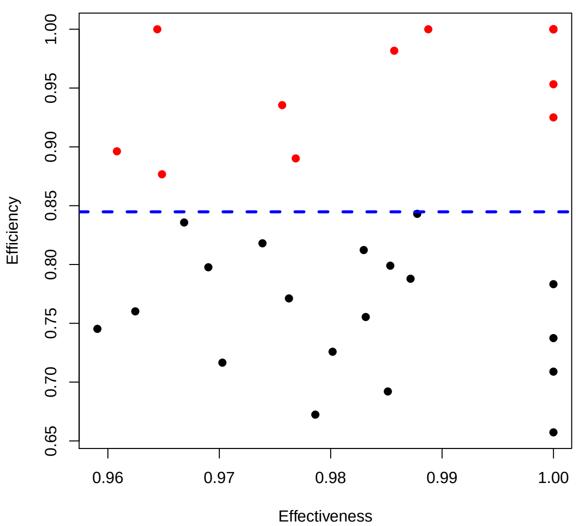
Note: The dotted blue line denotes the average effectiveness score.

Figure 1.5: DEA Effectiveness Classification



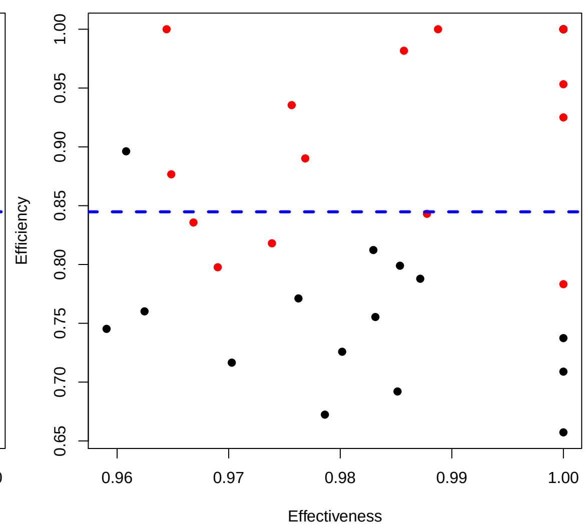
Note: The dotted blue line denotes the average effectiveness score.

Figure 1.6: ML Effectiveness Classification



Note: The dotted blue line denotes the average efficiency score.

Figure 1.7: DEA Efficiency Classification



Note: The dotted blue line denotes the average efficiency score.

Figure 1.8: ML Efficiency Classification

where U_{it} and W_{it} denote vectors of physician and patient characteristics. E_{it} indicates the average ED volume of those shifts that physician i is assigned to in year t and γ_t denotes year fixed effects. ϵ_{it} is a statistical noise.

In order to examine the potential influence of peers' characteristics on a focal physician's average effectiveness and efficiency (Research Question 4), we make use of the following regression model:

$$\theta_{ijt} = \beta_1 Z_{ijt} + \beta_2 U_{ijt} + \beta_3 W_{ijt} + \beta_4 E_{ijt} + \sigma_{it} + \gamma_t + \epsilon_{ijt}, \tag{1.3}$$

where θ_{ijt} (defined in Section 1.3.3) denotes physician-pair DEA scores corresponding to focal physician i while working alongside peer physician j in year t. Z_{ijt} represents indicator variables coded as 1 if peer physician j has a higher effectiveness score, a higher efficiency score, different medical degree, or is of the opposite gender compared to focal physician i. σ_{it} denotes physician-year fixed effects.

In order to estimate the coefficients in (1.2) and (1.3), a regression technique other than the standard multivariate regression is needed. This is because the standard regression technique assumes a normal and homoscedastic distribution of the noise. However, since the DEA scores are between 0 and 1, our dependent variable is bounded and error terms may not satisfy these assumptions.

Tobit regression can be used whenever there is truncation, causing a mass of observations at a threshold value such as 0 or 1 (Chilingerian 1995). Unlike the case of truncation, however, DEA does not exclude observations greater than 1 (or below 0). Instead, it does not allow a DMU to be assigned a value outside the range (0, 1]. Hence, DEA easily fits the requirement of the Tobit model (Chilingerian 1995). Following the normalization approach of Greene (1993), which assumes a censoring point at zero, we transform the DEA scores to:

$$y_{it} = (1/\theta_{it}) - 1,$$

where θ_{it} is physician i's DEA score in year t. The transformed DEA scores then become the dependent variable that takes the form:

$$y_{it} = \begin{cases} B'x_{it} + u_{it}, & \text{if } y_{it} > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where B is a vector of coefficients and x_{it} is a vector of covariates, and u_{it} is the error term. To account for unobserved serial correlation in the DEA scores, which might arise as a result of calculating a DMU's DEA score by incorporating all other DMUs in the dataset, we use Simar and Wilson's bootstrap procedure (Simar and Wilson 1998) for bias-correction of the scores.

1.6 Results

To present our results, we first discuss our findings related to our Research Question 1: are effectiveness and efficiency of a physician substitutes (negatively correlated) or complements (positively correlated)? We then present our results related to our Research Question 2: what is the relationship between effectiveness/efficiency of a physician and various characteristics, including those of the physician (e.g., test order count, experience, tenure), patients (e.g., race, gender, age, ESI), and the environment (e.g., ED volume)? Next, we present our findings regarding our Research Question 3: what do highly effective and efficient physicians do differently than their peers? Finally, we discuss our results with respect to our Research Question 4: How do physicians influence each other's effectiveness and efficiency?

1.6.1 Effectiveness and Efficiency: Substitutes or Complements?

We begin our analysis by generating insights into our Research Question 1. We do so by examining the relationship between physicians' effectiveness and efficiency scores. Importantly, we find that higher scores on the efficiency metric do not lead to lower scores on the effectiveness metric, as conventional wisdom might suggest. Rather, there is a statistically significant positive relationship between the two scores (see Table 1.2) which suggests that effective physicians are more likely to be efficient as well. This is an important observation, especially in the view of traditional debates that argue healthcare providers cannot be effective and efficient at the same time. Indeed, our finding questions the validity of the conventional wisdom, and suggests that physician effectiveness and efficiency should be viewed as complements (not substitutes).

1.6.2 Physician Characteristics

To provide insights into Research Question 2, we next examine the relationship between physicians' DEA scores and their characteristics. As shown in Table 1.2, our results indicate a statistically significant negative relationship between a physician's effectiveness score and his/her job tenure. This observation implies that more tenured physicians have, on average, lower effectiveness scores. A reasonable initial assumption might be that as knowledge and skill increase with greater tenure, effectiveness will also improve (Ng and Feldman 2013). In contrast, our finding is more consistent with the literature on job design and motivation that suggests that, as job tenure increases, employees are likely to become less motivated at work (Hackman and Oldham 1980, Kass et al. 2001, Bruursema et al. 2011). However,

Table 1.2: Regression Results - Effectiveness Model - Individual Physician

	Dependent Variable: Effectiveness Score
Efficiency Score	0.0565^{***}
	(0.0110)
Job Tenure	-0.0003^{*}
	(0.0001)
Test Order Count	-0.0043^{**}
	(0.0015)
ED Volume	-0.0022^{**}
	(0.0008)
ED Volume × Test Order Count	0.0002^{**}
	(0.00006)
Contact-to-Disposition Time	-0.0001^{***}
	(0.00002)

Note: $N = 106$. Observations are at the physician-year level.
*p<0.05; **p<0.01; ***p<0.001

our results might also be related to the fact that the ED might have imposed higher hiring standards in recent years or simply has been able to attract more effective physicians. Due to lack of data, we are unable to differentiate between these or other potential reasons behind our finding. We hope future research can use other sources of data to shed light on the reason behind the negative relationship between job tenure and effectiveness.

Our results also indicate a negative relationship between a physician's effectiveness and his/her average number of test order count. This implies that effective physicians are those who order less tests, or more accurately, order tests more intelligently. The fact that physicians with lower number of ordered tests have higher scores on the effectiveness metric supports a theory that not only there exist inherent differences among physicians with respect to effectiveness, but that effectiveness of providers might be improved via training programs that enable providers to decrease their use of unnecessary tests.

Our results regarding physician efficiency are displayed in Table 1.3. The results indicate a statistically significant positive relationship between a physician's efficiency score and his/her experience level. This is consistent with the extant literature that suggests that efficiency improves with higher levels of experience.[4] In addition, our results show a negative correlation between a physician's efficiency and his/her average number of test order count, implying that a physician's test ordering behavior is a contributing factor to his/her efficiency (similar to his/her effectiveness).

1.6.3 Patient Characteristics

To provide further answers to Research Question 2, we also examine the relationship between a physician's DEA scores and characteristics of his/her patients. Our results presented in Table 1.4 show no statistically significant relationship between a physician's effectiveness score and his/her average patient characteristics. With regards to physician efficiency, the results presented in Table 1.5 show that physicians' average efficiency scores increase when they encounter older patients, although the size of the coefficient is small (0.04). Overall, our results are consistent with the relevant literature that suggests patient characteristics should ideally have little or no effect on DEA scores (Chilingerian 1995).

[4] For example, Venkataraman et al. (2018) show that more experienced surgeons are more efficient (evidenced by their patients' reduced LOS) in performing surgical procedures.

Table 1.3: Regression Results - Efficiency Model - Individual Physician

	Dependent Variable: Efficiency Score
Experience	0.0028^{*}
	(0.0012)
Test Order Count	-0.0249^{***}
	(0.0059)
ED Volume	0.0254^{***}
	(0.0043)
ED Volume × Test Order Count	-0.00012^{*}
	(0.0005)

Note: $N = 106$. Observations are at the physician-year level.
*p<0.05; **p<0.01; ***p<0.001

Table 1.4: Regression Results - Effectiveness Model - Patient Characteristics

	Dependent Variable: Effectiveness Score
Age	0.0011
	(0.0001)
ESI Level	-0.0514
	(0.0330)
Female	-0.0555
	(0.0719)
White	-0.0076
	(0.0633)

Note: $N = 106$. Observations are at the physician-year level.
*p<0.05; **p<0.01; ***p<0.001

Table 1.5: Regression Results - Efficiency Model - Patient Characteristics

	Dependent Variable: Efficiency Score
Age	0.0407***
	(0.0099)
ESI Level	-0.3212
	(0.3818)
Female	-0.5533
	(0.3693)
White	1.1730
	(1.0329)

Note: $N = 106$. Observations are at the physician-year level.
*p<0.05; **p<0.01; ***p<0.001

1.6.4 Environment Characteristics

In addition to physician and patient characteristics, we study the impact of environment characteristics on physicians' effectiveness and efficiency. Specifically, given the large body of literature examining the effects of high workloads on physician performance (KC and Terwiesch 2009, Powell et al. 2012, Berry Jaeker and Tucker 2017, Batt and Terwiesch 2017), we study whether and how physician effectiveness and efficiency are affected by high ED volume.

The results presented in Table 1.3 show that, on average, physician efficiency improves as ED volume increases. Furthermore, our results regarding physician effectiveness presented in Table 1.2 show that high workloads have a negative effect on physicians' average effectiveness scores. Consistent with the extant literature, our findings, thus, highlight the impact of high workloads on physician effectiveness and efficiency.

1.6.5 What Do Highly Effective and Efficient Physicians Do Differently?

Our results presented in the previous sections provide insights into our Research Questions 1 and 2. We now turn our attention to our Research Question 3, and generate insights into the characteristics of highly effective and efficient physicians, defined as those physicians with a higher-than-average effectiveness and efficiency DEA scores, respectively. To this end, we run model (1.2) on sub-samples of highly effective and efficient physicians. We present our results corresponding to highly effective and efficient physicians in Tables 1.6 and 1.7, respectively.

As presented in Table 1.6, we find a negative relationship (though weakly statistically significant) between highly effective physicians' effectiveness scores and their average test order count. This implies that, compared to other physicians, highly effective physicians are able to order tests more intelligently and eliminate the unnecessary tests. While we establish a negative relationship between average physician effectiveness and ED volume in Section 1.6.4, we find no statistically significant evidence that high ED volume impacts the effectiveness of highly effective physicians. This finding, thus, suggests that highly effective physicians maintain their effectiveness under high workloads.

The results presented in Table 1.7 provide statistically significant evidence of a positive association between physician efficiency and ED volume among highly efficient physicians, suggesting that a highly efficient physician's efficiency improves during high-volume shifts. In addition, our results show that highly efficient physicians order less tests on average compared to their peers.

Table 1.6: Regression Results - Effectiveness Model - Highly Effective Physicians

	Dependent Variable: Effectiveness Score
LOS	-0.0003***
	(0.00007)
Test Order Count	-0.0009
	(0.0006)
ED Volume	0.0007
	(0.0006)

Note: $N = 46$. Observations are at the physician-year level.
*p<0.05; **p<0.01; ***p<0.001

Table 1.7: Regression Results - Efficiency Model - Highly Efficient Physicians

	Dependent Variable: Efficiency Score
Contact-to-Disposition Time	-0.00514***
	(0.0006)
Test Order Count	-0.0120**
	(0.0045)
ED Volume	0.0117**
	(0.0042)

Note: $N = 40$. Observations are at the physician-year level.
*p<0.05; **p<0.01; ***p<0.001

1.6.6 Peer Influence

We now provide insights into our Research Question 4. Our results presented in Table 1.8 show a statistically significant negative relationship between a focal physician's effectiveness and the presence of a more effective peer. This finding suggests that, all else equal, scheduling a physician with a more effective peer during the same shift results in a decrease in the physician's effectiveness. Similarly, the

Table 1.8: Regression Results - Effectiveness Model - Physician-Pair

	Dependent Variable: Effectiveness Score
More Efficient Peer	0.0007
	(0.0022)
More Effective Peer	-0.009^{***}
	(0.0021)
Different-Degree Peer	-0.004
	(0.003)
Opposite-Gender Peer	0.0016
	(0.003)

Note: N = 2,268. Observations are at the physician pair-year level.
*p<0.05; **p<0.01; ***p<0.001

regression results regarding peers' relative efficiency scores displayed in Table 1.9 show that the presence of a more efficient peer is associated with a decrease in a focal physician's efficiency. These findings suggest that more effective and efficient providers can have a negative influence on their peers' effectiveness and efficiency, respectively. Our results are in line with the findings in (Saghafian et al. 2019) in which, using a different statistical methodology, the authors provide evidence of opposite-directional peer influence, and highlight the importance of incorporating peer influence in physician scheduling. Finally, our results do not provide statistically significant evidence of peer influence with respect to physicians' relative gender and medical degree on a focal physician's average effectiveness and efficiency.

1.7 Robustness Checks

In this section, we provide alternative models for evaluating physicians' effectiveness and efficiency by making use of different sets of input/output variables.

Table 1.9: Regression Results - Efficiency Model - Physician-Pair

	Dependent Variable: Efficiency Score
More Efficient Peer	-0.006^{*}
	(0.0023)
More Effective Peer	0.004
	(0.0021)
Different-Degree Peer	0.0002
	(0.0031)
Opposite-Gender Peer	0.0024
	(0.0031)

Note: N = 2,268. Observations are at the physician pair-year level.
*p<0.05; **p<0.01; ***p<0.001

1.7.1 Alternative Effectiveness Model

In order to ensure that our results with respect to physician effectiveness are not sensitive to the choice of the output variable (72-hour rate of non-return), we repeat our analysis using an alternative set of variables. Specifically, we make use of a physician's over- and under-calling rates in addition to the 72-hour rate of non-return patient visits. We define a physician's over-calling rate as the percentage of patients admitted by him/her from the ED to the hospital who were subsequently discharged from the hospital within 12 hours of admission. Similarly, we choose the percentage of patients admitted by a physician to the hospital (from the ED) who were upgraded from a floor bed to an intermediate care unit or ICU bed within 12 hours of admission as a proxy for how often the physician under-calls his/her patients' illness severity. Since the over- and under-calling rates would be considered undesirable outputs, we use the 12-hour non-discharge and 12-hour non-upgrade rates as output variables. We choose the physician's average number of test order counts as the model's input variables. This effectiveness model's variables, thus, include:

Outputs:

- *Rate of discharged patients who do not return within 72 hours;*
- *Rate of admitted patients who are not discharged within 12 hours;*
- *Rate of patients admitted to a floor/ward bed who are not upgraded within 12 hours.*

It should be noted that the first output variable above is suitable for measuring performance with regards to discharged patients, while the other two output variables capture performance with respect to admitted patients. The choice of threshold numbers (72 and 12) is made based on observations made in the literature (see, e.g., Keith et al. 1989, Gordon et al. 1998, and the references therein) as well as conversations with ED physicians. In addition, we perform sensitivity analyses on these thresholds by changing each of them within a range, and observe that our main results still hold.

Inputs:

- *Radiology order count*: Average number of the physician's radiology orders per patient visit;
- *Ultrasound order count*: Average number of the physician's ultrasound orders per patient visit;
- *MRI order count*: Average number of the physician's MRI orders per patient visit.

A positive correlation between the input and the output variables confirms our choice of the model's input variables. Similar to our main effectiveness model, we utilize an output-oriented approach. We re-run our second-stage Tobit regression analysis using the scores derived from this model and observe that our main results hold.

1.7.2 Alternative Efficiency Model

Similar to our robustness test for the effectiveness model, we re-run our analysis using an alternative efficiency model defined as follows:

Output:

- *Throughput:* Average number of patients seen by the physician per shift.

Inputs:

- *High ESI-level patients*: Percentage of patients served by the physician who have ESI levels 4 and 5 (i.e., low-acuity patients);
- *Patients younger than 65*: Percentage of patients served by the physician who are younger than 65.

For our alternative efficiency model, we choose an output-oriented DEA model based on which efficient physicians are identified as those who have a higher throughput rate for a given mix of patients. Based on our discussions with ED physicians, throughput — the average number of patients served by a provider per unit of time — possesses significant face validity for exploratory analysis. Our input variables in this model comprise a low-acuity and younger patient mix which, on average, requires less time to treat. We re-run our statistical analysis using this alternative efficiency model and observe that our findings are consistent with our main results discussed in Section 1.6.

1.8 Conclusions

Using evidence from emergency medicine, we develop and analyze metrics for measuring physicians' effectiveness and efficiency. We then use our metrics to generate insights into the relationship between physician performance and factors related to patient, physician, environment, and peer physicians. Unlike what the conventional wisdom suggests, our findings show that a physician's effectiveness and his/her efficiency are positively associated. In addition, we find that more effective physicians have lower-than-average test order count and job tenure. We also find that efficient physicians have, on average, a lower test order count per patient visit and more years of experience compared to their peers. In addition, we find that during high-volume shifts, a physician's efficiency improves while his/her effectiveness declines.

We identify some of the characteristics of highly effective and efficient physicians. Our findings indicate that highly effective physicians order less tests compared to their peers. We show that highly effective physicians are able to maintain their effectiveness under high workloads more so than their peers. In addition, we find that highly efficient physicians have a lower test order count per patient visit and are more efficient during high-volume shifts compared to their peers. Furthermore, our results provide evidence for the existence of peer influence, and suggest that the presence of more effective and efficient peers has negative effects on a focal physician's effectiveness and efficiency, respectively.

We believe that our analysis serves as an early step to explore issues related to physician effectiveness and efficiency. Importantly, we do not believe that the metrics we develop are the only ways to measure physician effectiveness or efficiency.[5] That is, our work does not provide a definitive calculus for determining who is (or is not) an effective or efficient physician, but rather uses analytical techniques to explore these issues in an early attempt to better understand them. Nevertheless, our findings shed light on potential new ways to improve the effectiveness and efficiency of healthcare delivery. For example, our results can help individual physicians identify their weaknesses and learn about what the highly effective and efficient physicians do differently. Similarly, well-designed training programs can use our findings to facilitate this learning process. Furthermore, our findings can prove useful in the area of physician scheduling as they demonstrate how peer influence can play an important role in effective and efficient care delivery. Thus, our insights on peer influence can be used to understand which physicians should be scheduled during the same shift so as to improve performance without increasing resources.

Finally, we note that our analyses in this paper are purely based on quantitative data. Future research can improve the strength and applicability of our effectiveness and efficiency models by considering qualitative factors as well. Future work can also provide a more complete picture of the channels through which a physician's effectiveness and efficiency can be improved. Given the importance of understanding factors that can improve the effectiveness and efficiency of physicians, we hope to see more future studies in these veins.

[5] For example, one may improve our scores by also including aspects of patient satisfaction that correlate with higher provider performance levels.

Chapter 2

Do Physicians Influence Each Other's Performance? Evidence from the Emergency Department

2.1 Introduction

Decisions regarding how to schedule physicians during the same shift are made everyday in hospitals. Making such decisions, however, without considering whether and how physicians who work alongside each other affect each other's performance could have significant implications. In this study, we examine how physicians who work during the same shift influence each other's speed and quality in the context of an Emergency Department (ED) setting. An ED provides an interesting study setting where physicians aim to optimize speed (to sustain a reasonable flow in the interest of those patients waiting) while maintaining quality for the patient being seen in a shared resource environment (Emergency Department Cases 2015). Thus, understanding whether and how physicians influence each other's speed and quality in EDs can generate insights into physician pairing and scheduling methods that can ultimately lead to more effective and efficient care delivery mechanisms.

In order to identify and quantify the potential influence of peers, we address the question of whether peer physicians' relative performance (measured in terms of speed and quality) affects a focal physician's performance. While prior research has identified peer networks using physical proximity (Manchanda et al. 2008) and social networks (Trusov et al. 2010), we define a focal physician's peers in our setting as those physicians who are scheduled to work alongside the focal physician during the same shift.

We measure physician performance in terms of speed and quality using the Length of Stay (LOS) and 72-hour return metrics, respectively. A patient's LOS captures the time from when the patient checks into the ED to the time when s/he leaves. A shorter LOS implies that more patients can be moved through the ED per unit time. Therefore, LOS serves as a valid proxy for measuring an ED physician's speed. The 72-hour return metric indicates patients' return to the ED within 72 hours of their initial discharge. When patients return to the ED, it is possible that during their first visit not all their medical issues were sufficiently addressed. Although controversial, this metric has been proposed and used as a measure of quality in the Emergency Medicine literature (Abualenain et al. 2013, Pham et al. 2011, Klasco et al. 2015). Nevertheless, we also re-run our analyses using two other quality metrics that measure how often a physician over- and under-calls his/her patients' illness severity, and observe similar results to those obtained by using the 72-hour rate of return metric.

Our results establish statistically significant evidence of the existence of peer physician influence in our setting. Specifically, we find that, on average, a faster peer has a negative effect on a focal physician's speed. Our results also document a slower peer's positive effect on a focal physician's average speed. In addition, a higher-quality peer is shown to negatively impact a focal physician's quality, and a lower-quality peer is found to positively affect a focal physician's quality, on average.

We explore two potential mechanisms that might be driving our results: social influence and resource spillover. Our findings indicate that spillover from physicians' utilization of shared ED resources is the main driver of the observed effects. In particular, we find that the magnitude of the documented effects increases during high-volume shifts (i.e., when resources are more constrained), suggesting that the existence of shared limited resources in the ED plays an important role in how physicians affect each other's speed and quality. This insight has potential implications in a variety of services in which workers utilize shared scarce resources, and sheds light on the connection between constrained capacity and influence of workers on each other's performance.

We further explore the resource spillover mechanism by examining the effects of peers on a focal physician's average test order count and admission rate. Our results show that a faster peer increases a focal physician's average test orders per patient visit by allowing the focal physician to utilize the test services as needed. Ordering more tests, in turn, results in a decline in the focal physician's speed. A slower peer, however, blocks the focal physician from using the test services in a timely manner. This reduces the average number of tests ordered by the focal physician, making him/her faster. With regards to peer influence on quality, we find that higher-quality peers have a negative effect and lower-quality peers have a positive effect on a focal physician's admission rate. Given the positive relationship between a physician's admission rate and his/her quality in our setting, we infer that in the presence of a higher-quality peer (who has a higher admission rate), a focal physician may not have access to the resources needed in order to admit his/her patients. As such, his/her admission rate and, in turn, quality decreases on average. Similarly, working alongside a lower-quality peer (who has a lower admission rate) results in an increase in the focal physician's admission rate and, hence, quality, on average.

To correctly estimate the impact of physicians on each other's speed and quality, we consider potential sources of endogeneity and confounding in our setting. For example, although there exists no systematic scheduling scheme in our setting, physicians' preferences in shift assignments might cause endogeneity concerns. We conduct robustness tests to address physicians' selection into peer groups and mitigate the concern of spurious correlations with omitted variables. Moreover, we use the nearest-neighbor propensity score matching without replacement to construct matched samples of physicians that achieve balance across a set of observable covariates related to patient and ED characteristics including patient age, gender, race, Emergency Severity Index (ESI) level (a five-level triage scale with 1 indicating the most urgent and 5 denoting the least urgent case), and ED volume. We re-run our analyses on these matched samples of physicians that achieve balance on all observable covariates. Our inferences remain the same.

Our findings have important practice-related implications for improving the operations of EDs. Given the large body of literature documenting the adverse effects of workload on physicians' performance (KC and Terwiesch 2009, Powell et al. 2012, Berry Jaeker and Tucker 2017, Batt and Terwiesch 2017), our study offers a potential way to alleviate the negative impact of high workloads by highlighting the importance of incorporating peer influence into physician scheduling and staffing models. Specifically, our findings suggest that scheduling physicians alongside peers with whom they utilize shared resources more efficiently would have a positive effect on the performance of physicians. Furthermore, our results could have significant financial implications for hospitals. Given the mounting pressure on hospitals to reduce costs (e.g., payment reforms), healthcare providers aim to reduce LOS and increase the number of patients they serve per bed per unit of time. In particular, considering that in an ED, a 15-minute decrease in LOS could result in $1.4 million additional revenue for a hospital (The Becker's Hospital Review 2016), our findings could lead to substantial savings for hospital EDs while maintaining a high level of care quality.

2.2 Related Studies

Our study is mainly related to two streams of literature: studies on how workers influence each other's performance, and the operations management literature surrounding physicians' speed and quality. Within the first stream, a large body of research has examined the impact of peers on an individual's performance in a diverse set of occupations including supermarket cashiers (Mas and Moretti 2009), physicians (Chan 2016), sales teams (Chan et al. 2014), and scientists (Waldinger 2012). Mas and Moretti (2009) study peer effects among cashiers in a supermarket chain and attribute the positive effects of productive peers on a worker's productivity to increased social pressure. Jackson and Bruegmann (2009) and Azoulay et al. (2010) find evidence of peer effects that are induced by knowledge spillover. Negative effects of peers have also been documented in the literature. For example, Steinbach and Tatsi (2016) compare the performance of workers in a group production process working alone and in the presence of peers. The authors identify free-riding as the main channel through which negative peer effects emerge (see also Cornelissen et al. 2017).

Peer effects among physicians have been studied in prior research using exogenous sources of variation in peer characteristics. For example, Iyengar et al. (2015) examine peer effects in the context of prescription choices of physicians and find that peer influence can affect both the trial and repeat prescription orders of a risky new drug. In a different setting, Huesch (2011) examines intra- and inter-group practice spillovers among a group of cardiologists by observing their use of a new medical device, and presents strong evidence for intra-group peer influence. While our study is related to how peers influence each other's performance, the focus of our paper is not to estimate peer effects. Rather, we examine whether and how physicians who work alongside each other during the same shifts influence each other's performance.

Our study is also related to the behavioral operations management literature surrounding worker speed and quality. A large body of literature has documented how employees adjust their service behavior in response to certain situations by either speeding up or slowing down (Powell and Schultz 2004, Do et al. 2018). For example, Schultz et al. (1998) find that workers speed up when they are the cause of disruptions (blocking and starving) in the flow of work. Several studies have examined the behavioral effects of workload on physician performance. KC and Terwiesch (2009) show that hospital employees speed up as load level increases. KC and Terwiesch (2012) also provide evidence of a negative association between the occupancy level of a cardiac intensive care unit and patient LOS due to early discharge of patients from the hospital. Armony et al. (2015) find evidence of slow-down and speed-up in EDs and propose plausible explanations for the slow-down effect including fatigue, shared resources being overstrained, and medical staff overload.

The effects of workload on physician quality have also been documented in the literature. Kuntz et al. (2015) show a non-linear relationship between hospital workload and mortality rates. Powell et al. (2012) find that high workloads result in a reduction in physician diligence over paperwork and, in turn, yield less revenue per patient. Our work builds upon these studies by demonstrating how physicians affect each other's speed and quality, and highlights the need to consider peer influence in staffing and planning models. In addition, our findings suggest that during high-volume shifts when resources are more constrained, the influence of peers increases in magnitude. Given that high congestion levels are linked to both longer patient LOS (Kuntz et al. 2011) and higher re-admission rates (Anderson et al. 2012), our insights offer hospital administrators a potential strategy to alleviate these negative consequences by making use of peer influence.

2.3 Empirical Setting and Data

We utilize a large dataset collected from the ED of our partner hospital, which is one of the leading hospitals in the U.S. Our data include 115,350 patient visits and are associated with 32 ED physicians who have served patients in our partner hospital. An automated rotational patient assignment algorithm (Traub et al. 2016) randomly assigns all arriving patients to physicians in our partner hospital's ED. This randomization process mitigates the concern of physicians' selection of patients and related potential cherry-picking behaviors that can influence physician performance. All visits from July 12, 2012, to July 31, 2016 that were associated with patients who were identified in the Electronic Medical Record as having been seen by an ED physician were included in our analysis. Our dataset comprises patient-specific information including demographic (age, gender, race), encounter-level information such as the number of ordered diagnostic tests, chief complaint, and ESI as well as detailed timestamps capturing patients' movements through the ED from registration to discharge. A summary statistics of the variables used in our analyses is presented in Table 2.1. We excluded 2,914 patient visits with missing values from our analyses. In addition, we removed all observations

Table 2.1: Summary Statistics

Variable	Mean	SD	Min	Max
Patient Age	58.64	20.89	1	105
Female Patient (%)	53	2	50	58
White Patient (%)	91	1	88	94
Patient ESI	2.98	0.57	1	5
IV Order Count	3.13	2.12	0	32
Ultrasound Order Count	1.28	0.50	0	5
Radiology Order Count	1.20	0.59	0	11
MRI Order Count	1.69	0.91	0	6
CT Order Count	0.32	0.57	0	8
Lab Order Count	11.74	6.53	0	136
Contact-to-Disposition Time (Minutes)	141.76	127.28	10	12953
LOS (Minutes)	232.24	187.44	30	24694

Note: N = 110,325. Observations are at the patient-visit level.

associated with 4 physicians who had fewer than 200 patient visits over the 4-year study period. This leaves us with a final dataset comprising 110,325 patient visits.

2.4 Methodology

To examine whether and how physicians influence each other in our setting, we model how a focal physician's performance, measured in terms of speed and quality, is affected by the presence of his/her peers. Specifically, our unit of analysis is focal physician i who works alongside his/her peer physician j while treating patient k at time t. The outcomes of interest which capture physician i's speed and quality at time t are the LOS and the 72-hour return of patient k, respectively. We define a focal physician's peer group at time t as all physicians who are scheduled to work in the ED at the same time.

Our dataset provides us with the identities of the main physicians associated with each patient visit. Using this information, we are able to infer the identities of peers corresponding to each patient k's visit by identifying all physicians for whom there exists at least one assigned patient in our dataset whose contact-to-disposition time (the time from initial physician contact to the time when a disposition decision is issued) overlaps with that of patient k. We then construct a dataset comprising all possible combinations of focal-peer physician pairs. This leaves us with 304,877 observations. Since our goal is to derive insights that could be useful in the area of physician scheduling (improving performance by making use of suitable physician pairs during the same shifts), we have chosen a pairwise comparison approach for easier interpretation of our results.

We examine the effect of peer physician j's characteristics on focal physician i's performance by introducing treatment variables coded as 1 if peer physician j is faster or higher-quality compared to focal physician i. Hence, our treatment group consists of all physician pairs i-j such that peer physician j is faster or higher-quality relative to focal physician i and the control group comprises all other physician pairs. We use binary variables to indicate physicians' relative performance characteristics (faster, higher-quality) for easier interpretation of our results and generating clear insights that can be utilized in practice in the area of physician scheduling. We use the quartiles of physician speed and quality measures to compare physicians along these dimensions. We evaluate physicians' relative speed and quality using their average patient LOS and 72-hour return rate, respectively.[1] Figures 2.1 and 2.2 illustrate the distributions of the average speed and quality measures of the physicians in our dataset, respectively. In order to account for possible variations in physicians' performance measures across our study period, we measure a physician's performance at time t using his/her patient visits prior to time t.

[1] In Section 2.9, we re-run our analyses using different measures of quality.

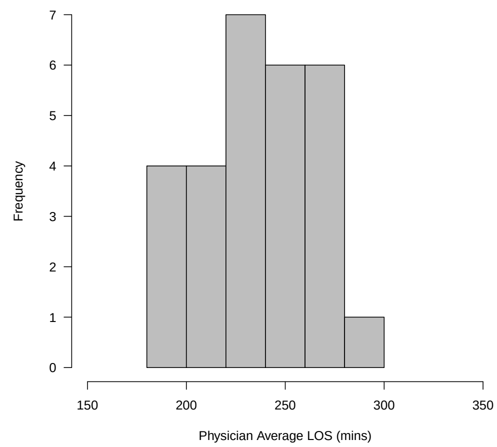


Figure 2.1: Distribution of Physicians' Average LOS (in Minutes)

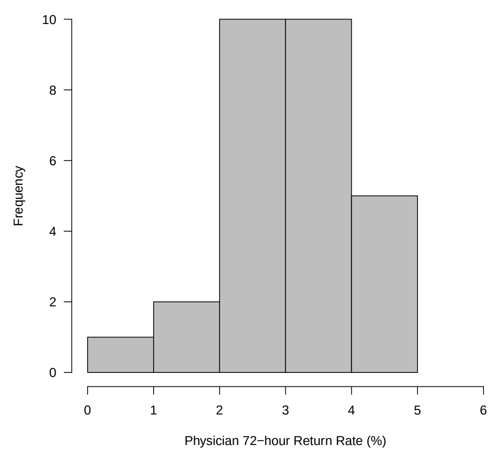


Figure 2.2: Distribution of Physicians' 72-Hour Rate of Return (in Percentage)

We control for patient k's characteristics including age, gender, race, and ESI level as well as focal physician i's characteristics with respect to patient k's visit at time t such as hospital admission (binary variable indicating whether the patient was admitted to the hospital after the ED visit) and the number of tests ordered. It is especially important to control for these factors because, as it is indicated in previous literature (see, e.g., Saghafian et al. 2014 and the references therein), there is a high level of variation in terms of patient complexity in EDs (i.e., the amount of work each patient brings).

In addition, we control for the familiarity between the focal-peer physician pairs. Similar to Huckman et al. (2009), we define physician i's familiarity with peer physician j at time t as the total number of minutes focal physician i has spent working alongside his/her peer j prior to time t. For the initial calculation of the physician familiarity metric, we use all observations associated with the first year of our sample study and exclude these observations from our final sample. This leaves us with 253,922 observations.

In addition to controlling for patient and physician characteristics, we control for ED volume at time t. We include hour, day, month, and year fixed effects to control for any unobserved time-varying effects as well as physician fixed effects that absorb all time-invariant physician characteristics. We cluster the error terms at the focal physician level to account for autocorrelation in the data.

We estimate the influence of peers in our setting using the following regression model:

$$Y_{ijkt} = \beta_1 Treat_{ijt} + \beta_2 P_{ikt} + \beta_3 R_{ikt} + \beta_4 E_{it} + \beta_5 Q_{ijt} + \gamma_t + \sigma_i + \epsilon_{ijkt}, \tag{2.1}$$

where Y_{ijkt} represents focal physician i's outcome of interest with respect to patient k's visit at time t while working alongside peer physician j. $Treat_{ijt}$ denotes treatment variables corresponding to the relative characteristics of physicians i and j at time t. P_{ikt} and R_{ikt} refer to vectors of physician i and patient k's characteristics at time t, respectively. E_{it} represents ED volume and Q_{ijt} refers to the familiarity between physicians i and j at time t. γ_t represents time fixed effects and σ_i denotes physician fixed effects. ϵ_{ijkt} is a statistical noise. We use OLS and logistic regression models to estimate the influence of peers on a focal physician's speed and quality, respectively.

2.5 Results and Discussion

Table 2.2 presents the effect estimates of faster and slower peers on a focal physician's speed and quality.[2] Our results demonstrate that in the presence of a faster peer, a focal physician's average patient LOS increases by 5.2 minutes. Similarly, we observe that a focal physician's average LOS decreases by 5.1 minutes while working with a slower peer. As shown in Table 2.2, we do not find statistically significant evidence of the effects of faster and slower peers on a focal physician's average quality.

Table 2.3 presents the effect estimates of higher- and lower-quality peers on a focal physician's average speed and quality. We document a statistically significant negative effect of a higher-quality peer and a statistically significant positive effect of a lower-quality peer on a focal physician's average quality. We do not, however, find statistically significant evidence of the influence of higher- and lower-quality peers on a focal physician's average speed. To ensure that our insights are not due to the measure of quality we use (the 72-hour rate of return), in Section 2.9 we derive the effect estimates of higher- and lower-quality peers using two alternative quality measures. Our findings reveal that the insights regarding the influence of higher- and lower-quality peers on a focal physician's quality are not sensitive to how a physician's quality is measured.

2.6 Endogeneity in Physician Shift Assignment

Estimation of peer influence in our setting is complicated by the non-random

[2]Complete regression results are presented in Appendix B.

Table 2.2: Speed Effect Estimates

	LOS	Rate of Return
Faster Peer	5.2402***	0.0004
	(0.6709)	(0.0267)
Slower Peer	-5.1070***	0.0170
	(0.6293)	(0.0263)
Observations	253,922	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 2.3: Quality Effect Estimates

	LOS	Rate of Return
Higher-Quality Peer	-0.7553	0.0850*
	(0.6697)	(0.0459)
Lower-Quality Peer	0.6836	-0.2036***
	(0.8513)	(0.0533)
Observations	253,922	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

assignment of physicians to shifts, which allows for the possibility of unobserved characteristics to confound the relationship between the treatment and the outcome. Although the unsystematic nature of physician assignment to shifts in our setting mitigates the potential endogeneity issue, we conduct two tests to address this concern.

In the first test, we address physicians' self-selection in peer groups by constructing a sub-sample of observations in which shift assignments are as close to random. Specifically, we construct a sub-sample of physicians' atypical patient visits. We define a physician's atypical patient visits as those which break out of a physician's scheduling pattern and hence could be viewed as a result of an exogenous shock (e.g., late change of schedule, physician calling in sick, etc.) to the physician assignment system. We identify these atypical observations using the least number of interactions (less than 8% of a physician's total patient visits) between each physician and his/her peers across our study period. Specifically, for each physician in our dataset, we identify those peers with whom the physician has had the least number of interactions across our 4-year study period. We then include all observations associated with the physician and the identified peers in the subsample. We re-run our analysis on this sub-sample and find the results (presented in Tables 2.4 and 2.5) to be consistent with our main findings. This suggests that our results are likely not derived by physicians' self-selection into peer groups.

In the second test, we examine whether high-performing physicians are assigned to high-volume shifts. Specifically, for each patient k's visit at time t, we examine whether the assigned physician's performance relative to his/her peers is correlated with ED volume at time t. A positive correlation would indicate that high-performing physicians are assigned to high-volume shifts. To test this, we make use of the following model:

$$E_{ikt} = \beta_1 HighPerformer_{it} + \beta_2 P_{ikt} + \beta_3 R_{ikt} + \gamma_t + \epsilon_{ikt}, \tag{2.2}$$

where E_{ikt} denotes ED volume at time t of patient k's visit (as described earlier, ED volume at time t indicates the number of patients being seen by all physicians other than physician i who is assigned to patient k). $HighPerformer_{it}$ is an indicator variable coded as 1 if physician i is a higher-than-average performer in terms of speed and/or quality relative to his/her peers. P_{ikt} and R_{ikt}, as indicated before, represent vectors of patient k and physician i's characteristics at time t. Lastly, γ_t denotes time fixed effects and ϵ_{ikt} is a statistical noise.

Table 2.4: Speed Effect Estimates - Atypical Subsample

	LOS	Rate of Return
Faster Peer	6.3821***	0.0853
	(1.6332)	(0.1080)
Slower Peer	-3.9734*	0.0574
	(2.2060)	(0.0826)
Observations	27,248	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 2.5: Quality Effect Estimates - Atypical Subsample

	LOS	Rate of Return
Higher-Quality Peer	-0.0703	0.2769**
	(0.5571)	(0.1096)
Lower-Quality Peer	0.6872	-0.2187*
	(0.9795)	(0.1228)
Observations	27,248	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

It should be noted that we run model (2.2) two times; once where $HighPerformer_{it}$ indicates that physician i is a higher-than-average physician with respect to speed and once where $HighPerformer_{it}$ indicates that physician i is a high-performing physician with respect to quality. Our results (presented in Tables 2.6 and 2.7) provide no statistically significant evidence that high-performing physicians are assigned to high-volume shifts.

Taken together, the results of both tests address the concern associated with physicians' selection into peer groups and confirm that endogeneity concerns are plausibly mitigated in our setting.

Table 2.6: High-Performing Physicians (Speed) and ED Volume

	ED Volume
High-Speed Physician	0.2485
	(0.3076)
Observations	110,325
Time Fixed Effects	Yes
Controls	Yes
Note:	*p<0.05; **p<0.01; ***p<0.001

Table 2.7: High-Performing Physicians (Quality) and ED Volume

	ED Volume
High-Quality Physician	-0.0968
	(0.2735)
Observations	110,325
Time Fixed Effects	Yes
Controls	Yes
Note:	*p<0.05; **p<0.01; ***p<0.001

Finally, as noted earlier, arriving patients in our setting are randomly assigned to physicians through an automated rotational patient assignment algorithm (Traub et al. 2016). Thus, concerns related to assignment of patients to physicians are also largely mitigated in our setting.

2.7 Robustness Checks

In this section, we present robustness checks to test the validity of our findings and the approaches that establish them.

2.7.1 Propensity Score Matching

In order to ensure that all focal-peer physician pairs in our sample have similar distributions across all observable covariates related to patient and ED characteristics, we use matching to construct well-matched samples of physician pairs. Specifically, we use the nearest-neighbor propensity score matching without replacement within a specified caliper width.[3] We have chosen to match on all patient- and ED-related observable covariates in our data including patient age, gender, race, ESI level, and ED volume. Tables C1-C4 in Appendix C present the mean baseline values of all covariates across the treatment and control groups as well as the standardized mean difference between the treatment and control groups. We find that the distribution of ED volume and some of the covariates related to patient characteristics including age and ESI level are relatively unbalanced across the treatment and control groups.

Tables C5-C8 in Appendix C illustrate how matching improves the balance in the means of the matching variables across the treatment and control samples. In order to ensure that our analysis is not sensitive to the choice of our matching technique, we use alternative matching approaches including one-to-one matching with and without replacement and coarsened exact matching. In each case, our inferences remain unchanged.

We re-run model (2.1) on matched samples of physician pairs. The results presented in Tables 2.8 and 2.9 show that our main findings remain unchanged.

2.7.2 Alternative Model Specification

To ensure robustness of our results to different model specifications, we re-run our analysis using an alternative specification of model (2.1). Specifically, given the evidence provided in the related literature for the impact of ED congestion on performance (e.g., KC and Terwiesch 2009, Kuntz et al. 2015), we include both linear and quadratic forms of ED volume in our model. Specifically, we make use of

[3]We use a caliper width of 0.1 times the pooled standard deviation of the logit of the propensity score (Rosenbaum and Rubin 1985).

Table 2.8: Speed Effect Estimates - Matched Samples

	LOS	Rate of Return
Faster Peer	1.3717***	0.0073
	(0.4804)	(0.0229)
Slower Peer	-5.1593***	0.0387
	(0.5864)	(0.0324)
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 2.9: Quality Effect Estimates - Matched Samples

	LOS	Rate of Return
Higher-Quality Peer	-0.7491	0.0755*
	(0.6085)	(0.0390)
Lower-Quality Peer	0.4425	-0.1598***
	(0.6008)	(0.0476)
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

the following model:

$$Y_{ijkt} = \beta_1 Treat_{ijt} + \beta_2 P_{ikt} + \beta_3 R_{ikt} + \beta_4 E_{it} + \beta_5 E_{it}^2 + \beta_6 Q_{ijt} + \gamma_t + \sigma_i + \epsilon_{ijkt}, \quad (2.3)$$

The regression results presented in Tables 2.10 and 2.11 confirm our main findings. Specifically, we observe statistically significant evidence of opposite-directional peer influence with respect to physicians' relative speed and quality.

2.8 Mechanisms

The results presented earlier show that both slower and lower-quality peers have positive effects on a focal physician's average performance while faster and higher-quality peers negatively impact the performance of a focal physician. In this

Table 2.10: Speed Effect Estimates - Alternative Model Specification

	LOS	Rate of Return
Faster Peer	3.8208***	-0.0022
	(0.8223)	(0.0279)
Slower Peer	-3.4183***	0.0215
	(0.7998)	(0.0302)
Observations	253,922	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 2.11: Quality Effect Estimates - Alternative Model Specification

	LOS	Rate of Return
Higher-Quality Peer	-0.5782	0.1204***
	(0.8749)	(0.0437)
Lower-Quality Peer	1.0472	-0.1761***
	(1.3315)	(0.0512)
Observations	253,922	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

section, we explore two mechanisms which may drive these observed effects: social influence and resource spillover.

2.8.1 Social Influence

Peers can influence individuals through a number of social channels including peer pressure, higher aspirations, and social norms. The relevant literature suggests that peers exert their influence through these channels when they serve as a commitment device imposing some social cost on an individual whom they observe (Buechel et al. 2018). They can have a "pulling-up" effect on individuals performing poorly or can have a "chocking" effect leading to under-performance.

To examine whether social influence is the main driver of our findings, we test whether the magnitude of the documented effects depends on the frequency of interactions among focal-peer physician pairs. If two physicians are rarely scheduled during the same shifts, it is less likely they would work alongside each other in the future. Hence, it is unlikely that they would be responsive to some social cost they might impose on each other (Mas and Moretti 2009). To test this hypothesis, we divide our data into two sub-samples according to the physician familiarity metric. That is, we construct two sub-samples of patient visits: one associated with focal-peer physician pairs who score higher than average on the familiarity metric, and one pertaining to the pairs who score lower than average on this metric. We conduct our matching and regression analyses on both sub-samples. Comparing the magnitude of the effects across the two sub-samples (presented in Tables 2.12 and 2.13) provides no evidence that social influence is the driving force behind our results.

2.8.2 Resource Spillover

Our findings might be attributed to physicians' utilization of shared resources such as laboratory services, nurses, and hallways. A setting such as an ED where shared (and limited) resources are often utilized, resembles a queuing system in which a server can be impacted by spillover from other servers (Gerla and Kleinrock 1980, Batt and Terwiesch 2017). For example, if a server is faster to use resources (e.g., issue tests), s/he can hinder the ability of his/her peers from using the same resources in a timely manner (for multi-stage ED queueing models with limited resources, see, e.g., Saghafian et al. 2012, Huang et al. 2015, and the references therein). Thus, a high-performing peer's negative effect on a focal physician's performance could be derived through a resource spillover effect.

Table 2.12: Effect Estimates - Below-Average Familiarity

	LOS	Rate of Return
Faster Peer	0.8542*	-0.0098
	(0.4698)	(0.0329)
Slower Peer	-2.9700***	0.0641
	(1.1083)	(0.0292)
Higher-Quality Peer	-1.3397	0.1603***
	(0.9491)	(0.0530)
Lower-Quality Peer	0.4147	-0.2763***
	(1.1249)	(0.0743)
Observations	126,960	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 2.13: Effect Estimates - Above-Average Familiarity

	LOS	Rate of Return
Faster Peer	1.5088*	0.0124
	(0.7987)	(0.0341)
Slower Peer	-2.0560***	0.0398
	(0.6804)	(0.0490)
Higher-Quality Peer	-0.3429	0.0503
	(1.1021)	(0.0526)
Lower-Quality Peer	0.5544	-0.0117
	(1.0734)	(0.0561)
Observations	126,962	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Furthermore, such resources are typically more binding during busy times when the ED volume is high. Hence, the spillover mechanism is expected to be more pronounced during busy periods. Therefore, to test whether resource spillover could be the mechanism driving our findings, we compare the magnitude of our documented effects across two sub-samples of observations: one pertaining to shifts with higher-than-average patient volume, and one corresponding to shifts with lower-than-average volume. From the results presented in Tables 2.14 and 2.15, we observe that the effects corresponding to high-volume shifts are indeed larger in magnitude compared to those associated with low-volume shifts. This suggests that resource spillover is more likely to be the driving force behind the documented opposite-directional effects.

Peer Influence on Test Order Count

Our analysis thus far suggests that physicians' utilization of shared ED resources is likely to be the driving force behind our results. While shared resources in EDs could comprise nurses, beds, and diagnostic test centers, our dataset provides us only with the number of tests ordered by physicians during each patient visit. Therefore, to further investigate the resource spillover mechanism, we examine whether physicians affect each other's performance through influencing each other's test ordering behavior. To this end, we make use of the following regression model:

$$O_{ijkt} = \beta_1 Treat_{ijt} + \beta_2 R_{ikt} + \beta_3 E_{it} + \beta_4 Q_{ijt} + \gamma_t + \sigma_i + \epsilon_{ijkt}, \tag{2.4}$$

where O_{ijkt} represents focal physician i's test order count associated with patient k's visit at time t.

The results presented in Table 2.16 provide evidence that physicians influence each other's speed through affecting each other's test ordering behavior. Specifically,

Table 2.14: Effect Estimates - Below-Average ED Volume

	LOS	Rate of Return
Faster Peer	2.0990***	0.0422
	(0.6046)	(0.0371)
Slower Peer	-2.3293***	-0.0107
	(0.4333)	(0.0397)
Higher-Quality Peer	-1.2995	0.0881*
	(0.3658)	(0.0509)
Lower-Quality Peer	0.5966	-0.1339*
	(0.5485)	(0.0726)
Observations	126,960	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 2.15: Effect Estimates - Above-Average ED Volume

	LOS	Rate of Return
Faster Peer	4.3900***	-0.0035
	(0.8955)	(0.0323)
Slower Peer	-4.2014***	0.0813
	(0.7952)	(0.0368)
Higher-Quality Peer	0.3032	0.1399***
	(0.9338)	(0.0478)
Lower-Quality Peer	0.9142	-0.2231***
	(1.0909)	(0.0678)
Observations	126,962	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 2.16: Peer Influence on Test Order Count

	Test Order Count
Faster Peer	0.1928***
	(0.0742)
Slower Peer	-0.1793***
	(0.0665)
Higher-Quality Peer	0.0346
	(0.0448)
Lower-Quality Peer	0.0129
	(0.0597)
Observations	253,922
Time Fixed Effects	Yes
Physician Fixed Effects	Yes
Controls	Yes
Note:	*p<0.05; **p<0.01; ***p<0.001

our results indicate that working alongside a faster peer increases a focal physician's average test orders while working with a slower peer results in a decrease in a focal physician's average test order count. However, as presented in Table 2.16, we find no statistically significant evidence that physicians influence each other's quality through affecting each other's test ordering behavior.

Table 2.17 presents the correlation of variables corresponding to physician and patient characteristics. We observe a positive correlation between the average LOS and test order count variables. Hence, it is likely that a faster peer utilizes less test orders on average compared to a focal physician. As a result, s/he allows the focal physician to make use of the test services in a timely manner. This, in turn, results in an increase in the focal physician's average test orders and a decrease in his/her average speed. Similarly, a slower peer, who orders more tests on average relative to a focal physician, hinders the ability of the focal physician to utilize the test services. Therefore, the focal physician's average test order count decreases, resulting in an increase in the average speed of the focal physician.

Table 2.17: Correlation Matrix of Patient- and Physician-Level Variables

	Avg LOS	Avg Rate of Return (%)	Admission Rate (%)	Female (%)	Avg Age	Avg ESI	White (%)	Avg Test Count
Avg LOS	1.00	-0.23	0.10	0.39	0.30	-0.11	-0.16	0.36
Avg Rate of Return (%)	-0.23	1.00	-0.19	0.06	-0.53	0.02	0.52	0.10
Admission Rate (%)	0.10	-0.19	1.00	-0.30	0.46	-0.06	0.12	0.23
Female (%)	0.39	0.06	-0.30	1.00	0.05	-0.35	0.33	0.04
Avg Age	0.30	-0.53	0.46	0.05	1.00	0.07	0.02	-0.09
Avg ESI	-0.11	0.02	-0.06	-0.35	0.07	1.00	-0.24	0.08
White (%)	-0.16	0.52	0.12	0.33	0.02	-0.24	1.00	-0.15
Avg Test Count	0.36	0.10	0.23	0.04	-0.09	0.08	-0.15	1.00

Peer Influence on Admission Rate

Our results presented in the previous section provide no statistically significant evidence that peers influence each other's quality through affecting each other's test order count. Given that the effect estimates of higher- and lower-quality peers increase in magnitude during high-volume shifts (as presented in Table 2.15), we consider other potential channels through which the resource spillover mechanism might operate. From Table 2.17, we observe that a physician's admission rate is negatively correlated with his/her 72-hour rate of return, and in turn, is positively correlated with his/her quality. Hence, a higher-quality peer in our setting has a higher patient admission rate, on average. As such, s/he would utilize more of the resources that are needed for admitting patients (e.g., inpatient beds) and hinder the focal physician from utilizing such resources as needed. This results in a decrease in the focal physician's admission rate and, accordingly, average quality. The same line of reasoning applies to a lower-quality peer with a lower admission rate, working alongside whom would allow the focal physician to admit more patients. This results in an increase in the focal physician's admission rate and an increase in his/her average quality. To test this hypothesis, we use the following logistic regression model:

$$A_{ijkt} = \beta_1 Treat_{ijt} + \beta_2 R_{ikt} + \beta_3 E_{it} + \beta_4 Q_{ijt} + \gamma_t + \sigma_i + \epsilon_{ijkt}, \tag{2.5}$$

where A_{ijkt} captures the log odds of focal physician i's admission decision regarding patient k's visit at time t. Table 2.18 presents the effect estimates of higher- and lower-quality peers on a focal physician's admission rate. The estimated coefficients, although not statistically significant, are negative and positive, respectively. Hence, the results indicate that higher-quality peers negatively affect a focal physician's admission rate while lower-quality peers have a positive effect on the admission rate of a focal physician. Our findings, thus, provide further evidence for the resource spillover mechanism and show that the influence of physicians on each other's quality is likely derived through their influence on each other's admission rate.

2.9 Alternative Quality Measures

In our analyses, we utilize the 72-hour return rate to measure physician quality. In this section, we estimate the effects of higher- and lower-quality peers using alternative quality metrics to ensure that our results are not merely due to the specific measure of quality we use. Specifically, we re-run our regression analysis using two alternative quality metrics: one capturing how often a physician over-calls his/her patients' illness severity and one related to how frequently a physician under-calls the severity of his/her patients' illness.

Contrary to the 72-hour return metric which evaluates a physician's quality with regards to his/her discharged patients, the over- and under-call metrics capture a physician's quality with respect to his/her admitted patients. We define the over-call metric as the percentage of a physician's patients who are admitted to the hospital by him/her but are discharged within 12 hours of their admission. Similarly, the under-call metric measures the percentage of a physician's patients who are admitted to the hospital by him/her but are upgraded from a floor bed to a more intensive area of care within 12 hours of their admission. Thus, these measures capture how

Table 2.18: Peer Influence on Admission Rate

	Admission Rate
Higher-Quality Peer	-0.0084
	(0.0165)
Lower-Quality Peer	0.0230
	(0.0160)
Observations	253,922
Time Fixed Effects	Yes
Physician Fixed Effects	Yes
Controls	Yes
Note:	*p<0.05; **p<0.01; ***p<0.001

well a physician makes the correct call regarding the needs and illness severity of his/her patients.[4]

Tables 2.19 and 2.20 present the effect estimates of higher- and lower-quality peers using the over- and under-call measures, respectively. In both cases, our inferences are similar to those made earlier using the 72-hour rate of return. Specifically, the effect estimates of higher- and lower-quality peers derived using physicians' over-calling rates (presented in Table 2.19), although not statistically significant, are positive and negative, respectively. This shows that consistent with our main findings, higher-quality peers negatively affect a focal physician's average quality while lower-quality peers have a positive effect on the average quality of a focal physician. Similarly, the results presented in Table 2.20 provide statistically significant evidence of the effects of higher- and lower-quality peers, captured using the under-calling rate, on a focal physician's average quality, respectively.

[4]Of note, the 12-hour threshold used for defining these metrics is based on inputs from ED physicians. However, we also perform sensitivity analyses on this threshold and observe that our main results hold.

Table 2.19: Quality Effect Estimates - Over-Calling Rate

	LOS	Over-Calling Rate
Higher-Quality Peer	-1.8123	0.0193
	(1.0624)	(0.1065)
Lower-Quality Peer	1.8464	-0.0273
	(1.1749)	(0.0890)
Observations	253,922	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table 2.20: Quality Effect Estimates - Under-Calling Rate

	LOS	Under-Calling Rate
Higher-Quality Peer	-1.3584	0.6956***
	(1.0719)	(0.1630)
Lower-Quality Peer	0.7552	-0.5078***
	(1.1258)	(0.1361)
Observations	253,922	
Time Fixed Effects	Yes	
Physician Fixed Effects	Yes	
Controls	Yes	
Note:	*p<0.05; **p<0.01; ***p<0.001	

2.10 Managerial Implications

We now summarize some of the main implications of our results. First, hospital administrators can benefit from our findings in constructing the optimal mix of physicians to schedule during the same shift. Overall, our results suggest that scheduling diverse peers during the same shift would positively affect the performance of physicians. This is consistent with the literature on teamwork that identifies team diversity as an important component of effective teams (Woehr et al. 2013, Zoogah et al. 2011). Second, the insights generated from our results could assist hospital administrators in the area of physician training. While scheduling high-performing physicians (in terms of quality) with lower-than-average performers could have a positive effect on the overall performance of physicians, it could also create learning opportunities for the low-performing physicians. More broadly, since most training programs require physicians to work alongside each other, our findings can be helpful in designing more effective training programs.

Finally, it is important to consider the financial implications of our findings for hospitals. Given the financial burden of prolonged ED LOS and unnecessary return visits on hospitals, our results may lead to significant cost savings for hospitals. This is because reducing LOS has both direct and indirect positive effects on the financial status of hospitals. It has a direct positive effect through decreasing the costs of patient care, facility and staffing expenses. It has an indirect positive effect by minimizing the risk of hospital-acquired infections and improving a variety of other patient safety metrics. Thus, reducing LOS by even a few minutes could have significant financial implications for hospitals (Krochmal and Riley 1994, The Becker's Hospital Review 2016). Similarly, reducing patient return rates or how often patients' illness severity is under- or over-called even by small amounts can have significant direct and indirect financial benefits for hospitals.

2.11 Limitations

It is important to note the limitations of our study. First, while we control for the observed factors that affect physician performance and conduct robustness tests to ensure that our results are not attributable to confounding effects, there might still be factors affecting physician performance that are unobservable in our dataset. Second, our analysis does not consider how learning among peers shapes the long-term influence of peers on physician performance. Prior research has shown that an individual's long-term performance improves over time as a result of learning from peers (Chan et al. 2014, Edmondson et al. 2001). Future research can explore potential learning effects that are induced by peer physicians. Finally, while we use simple measures to gauge physicians' speed and quality, we should note that there are various other metrics, both qualitative and quantitative, that can be used to measure physician performance. Future research can extend our analyses by using such measures and by removing some of the limitations of our study.

2.12 Conclusions

In this study, we examine the influence of peers on a focal physician's performance in an ED setting. We document statistically significant evidence of peer influence. In particular, our results demonstrate that a faster peer has a negative effect and a slower peer has a positive effect on a focal physician's speed, on average. Similarly, a higher-quality peer is found to negatively impact a focal physician's average quality while a lower-quality peer is shown to positively affect the average quality of a focal physician.

Our findings identify resource spillover from peers as the main driver of peer influence and indicate that diverse physicians utilize shared resources more efficiently. Furthermore, our findings show that physicians influence each other's speed and quality through affecting each other's test ordering behavior and admission rate, respectively.

Our findings have important practical implications for improving the performance of physicians by highlighting the need to consider peer influence as an important component of effective physician staffing strategies. In particular, our findings can be used by hospital administrators when designing (a) staffing and shift schedules, and (b) training programs. In both of these, understanding how physicians influence each other can have a significant impact on the overall performance of physicians.

Chapter 3

Which Shift Matters the Most? Evidence from the Emergency Department

3.1 Introduction

Negative effects of overcrowding in Emergency Departments (EDs) are well-documented in the literature (see, e.g., Anderson et al. 2012, Kuntz et al. 2015, Batt and Terwiesch 2017). Increased wait times, decreased physician productivity, increased likelihood of poor outcomes, and decreased patient and provider satisfaction are a few of the implications of overcrowding in EDs (Savage et al. 2015). Unsurprisingly, improving patient flow in hospital EDs has been shown to have a significant impact on quality of care as well as on patient satisfaction (Armony et al. 2015). As such, reducing congestion and improving patient flow in EDs are important objectives of ED physician scheduling and staffing models.

Although a large body of literature has examined staffing decisions in various settings (see, e.g., Gans et al. 2003, Aksin et al. 2007), there are a few empirical papers that have studied the effects of scheduling decisions on performance in healthcare settings. In this paper, we conduct an empirical investigation into understanding how assigning high-performing physicians to different shifts of the day affects the daily performance of the ED. Specifically, we address the question: to which shift should the ED's high-performing physicians be assigned?

We utilize a large dataset capturing daily operations of a single ED in a leading U.S. hospital over a 4-year period. We conduct a day-level analysis in which we examine how assigning high-performing physicians with respect to speed, quality, and admission rate to different shifts of the day affects the ED's average daily performance. We measure daily speed of the ED using the average daily patient Length of Stay (LOS), which we define as the time interval from patients' registration in the ED until discharge. We use the average 72-hour rate of return metric (the percentage of discharged patients who return to the ED within 72 hours of their discharge) as a proxy for measuring quality in the ED. Accordingly, we measure an ED physician's speed and quality using his/her average patient LOS and 72-hour rate of return, respectively.

Our results show that assigning a relatively faster group of physicians to the first shift of the day has the highest impact on the average daily speed of the ED. Specifically, our results suggest that assigning a group of physicians with a higher-than-average aggregate speed to the first shift results in a 8.2-minute improvement in the average daily speed of the ED. This is in comparison to the effect of assigning such physicians to the second shift that leads to a 4.8-minute improvement in the daily speed of the ED. Although we find no statistically significant evidence of the "first-shift effect" with respect to physician quality, we find that assigning physicians with a higher-than-average aggregate admission rate to the first shift results in the highest increase in the average daily admission rate of the ED. As such, our findings suggest that allocating a high-performing composition of physicians to the first shift of the day would have the greatest impact on the daily performance of the ED. We further demonstrate that physicians' performance in the earlier shifts of the day has a "domino effect" throughout the rest of the day. Specifically, we show that assigning high-performing physicians to the earlier shifts of the day affects the performance of physicians in the subsequent shifts.

We conduct robustness tests to validate our main findings. In the first test, we use the nearest-neighbor propensity score matching to construct matched samples of observations that achieve balance across a set of variables including patient age, gender, race, and Emergency Severity Index (ESI) (a five-level triage system denoting patient illness severity), as well as ED volume and physician shift preference. We define a physician's preference for a particular shift as how often s/he has worked in that shift over our 4-year study period. Our matching strategy, thus, ensures that we are comparing day-level observations that have similar distributions across observable characteristics related to patients, physicians, and the ED. We re-run our analysis of the first-shift effect on these matched samples and observe that our main findings remain unchanged. In the second test, we re-run our analysis using alternative measures of quality to ensure that our results are not sensitive to our choice of the quality metric (72-hour rate of return). We find the results to be consistent with our main findings.

Furthermore, we investigate whether our results are sensitive to heterogeneity in ED volume and day of the week. Specifically, we examine whether the magnitude of the documented effects changes on high-volume days and on weekends. Our results suggest that the magnitude of the first-shift effect is larger when ED volume is high. In addition, our results provide statistically significant evidence that the magnitude of the first-shift effect decreases on weekends.

Our findings can be useful to hospital administrators in making decisions regarding how to schedule physicians across different shifts of the day. Given the overwhelming evidence of the negative effects of ED congestion on provider performance and patient outcome (KC and Terwiesch 2009, Powell et al. 2012, Berry Jaeker and Tucker 2017, Batt and Terwiesch 2017), our insights can be useful in improving patient flow in EDs by highlighting the importance of improving the performance of the earlier shifts of the day.

3.2 Related Studies

Our research relates to two streams of literature: (a) optimal scheduling, (b) behavioral operations management. There is a large body of analytical literature on staffing/scheduling decisions. Most classical models developed in the operations management literature assume that workers are similar and independent of each other. Recent studies including Wallace and Whitt (2005), Ata and Van Mieghem (2009), and Ward and Armony (2013) have relaxed this assumption by incorporating worker heterogeneity in scheduling and planning models. Similarly, Arlotto et al. (2014) consider both worker heterogeneity and learning for staffing decisions.

A rich body of literature has focused on scheduling and appointment systems in healthcare settings (see, e.g., Cayirli and Veral 2009, Gupta and Denton 2008). A number of these studies are based on the assumption that patients are homogeneous (Begen and Queyranne 2011, Cayirli et al. 2012). Other studies have used variability in service duration to account for the heterogeneity in patient characteristics. For example, higher variance in treatment duration of patients has been shown to create higher variability in the system. As such, several studies have suggested sequencing patients based on their variance of service duration (Robinson and Chen 2003, and

Cayirli et al. 2008). For example, Wang (1999) argues that larger variability will lead to longer waiting times, so sequencing patients with smaller service duration variance first can reduce waiting times for subsequent patients.

A large body of literature has examined staffing and scheduling models in EDs (see, e.g., Saghafian et al. 2015 and the references therein). Sinreich et al. (2012) use two heuristic algorithms for staffing physicians, nurses and technicians. The authors show that the work schedules developed by these two algorithms result in a 20% to 64% decrease in patient waiting time and a 7% to 29% reduction in LOS. Yankovic and Green (2011) develop a variable finite-source queuing model representing the nursing system to approximate the actual interdependent dynamics of bed occupancy levels and demands for nursing. Patel and Vinson (2005) propose organizing ED staff members into teams consisting of one physician, two nurses, and one technician in a single suburban ED setting and report decreases in patient wait time and the Left Without Being Seen (LWBS) rate. Traub et al. (2015) show that a rotational assignment of patients to physicians results in a decrease in both LOS and LWBS.

Various studies have examined the staffing and scheduling problem empirically. For example, Fisher et al. (2006) and Chuang et al. (2016) find that staffing levels affect the conversion of traffic into sales in a retail setting. There are few empirical studies, however, that have examined the effects of staffing and scheduling decisions on performance in healthcare settings. Our study contributes to this literature by empirically examining whether and how incorporating physicians' relative performance in shift assignments affects the daily performance of the ED.

Within the second stream, our work is related to empirical studies on the effects of workload on performance. A large body of literature has examined the impact of workload on an individual's performance. Tan and Netessine (2014) investigate how workload affects workers' performance in a restaurant setting. Using a healthcare setting, KC and Terwiesch (2009) study the impact of workload on service time using operational data from patient transport services in cardiothoracic surgery. The authors show that workers speed up as workload increases, although this positive effect may disappear after prolonged periods of high workload. The authors further provide evidence of a negative association between the occupancy level of a cardiac intensive care unit and patients' LOS. They attribute their finding to the fact that high occupancy levels will force hospitals to discharge patients early. Our work is related to this stream of literature by demonstrating how physician performance is affected by the performance of physicians working in the previous shifts. Specifically, our findings shed light on a domino effect of physician performance throughout the day which increases in magnitude on high-volume days.

3.3 Setting and Data

Our research setting is a single ED in a large leading U.S. hospital. It is equipped with 24 rooms and has the capacity to use up to 9 hallway spaces, which are used during high-volume shifts. The ED is staffed 24 hours per day with board-certified emergency physicians. Our data comprises 110,325 patient visits corresponding to all patients who were seen by an ED physician from July 12, 2012 to July 31, 2016 in our partner ED. Our dataset is associated with 28 ED physicians, all with 3 years or more of post-residency emergency medicine experience. We observe a relatively constant physician labor over our 4-year study period. It should be noted that a rotational patient assignment system randomly assigns patients to physicians in our partner ED (Traub et al. 2016). The randomized assignment of patients to physicians mitigates the concern of patient selection bias in our setting.

Our dataset includes patient-visit-level information including patient characteristics (age, gender, race) and ESI level. Furthermore, our dataset provides us with timestamps detailing patients' time in the ED from registration to discharge. Table 3.1 presents a summary statistics of the variables included in our analyses.

In order to address our research question, we construct a day-level dataset that captures the daily performance of the ED over our 4-year study period. Specifically, we extract day-level information including patient characteristics, average daily LOS, and ED volume as well as identities of those physicians who worked during each shift of the day. Based on the current practice of our partner ED, we consider three shifts per day starting at 6:00 AM, 2:00 PM, and 10:00 PM. Descriptive statistics corresponding to shifts 1, 2, and 3 are presented in Tables 3.2-3.4, respectively. Our final dataset after removing missing values includes 1,277 day-level observations.

3.4 Empirical Analysis

In order to examine the effects of assigning high-performing physicians to different shifts of the day on the daily performance of the ED, we utilize our day-level dataset which captures the daily performance of the ED across all shifts. We evaluate the performance of the ED in terms of speed, quality, and admission rate. The dependent, independent, and control variables in our analyses are as follows.

Dependent Variables

Given that patient LOS in the ED captures the total time a patient spends in the ED, a shorter average daily LOS leads to an improvement in patient flow and reduction in patient queueing. Therefore, LOS constitutes a reasonable proxy for measuring the efficiency of an ED's operations. LOS has been used in the literature as a valid metric for evaluating efficiency in EDs (Chilingerian 1995, Fiallos et al.

Table 3.1: Summary Statistics

Variable	Mean	SD	Min	Max
Patient Age	58.64	20.89	1	105
Female Patient (%)	53	2	50	58
White Patient (%)	91	1	88	94
Patient ESI	2.98	0.57	1	5
IV Order Count	3.13	2.12	0	32
Ultrasound Order Count	1.28	0.50	0	5
Radiology Order Count	1.20	0.59	0	11
MRI Order Count	1.69	0.91	0	6
CT Order Count	0.32	0.57	0	8
Lab Order Count	11.74	6.53	0	136
LOS (Minutes)	232.24	187.44	30	24694

Note: N = 110,325. Observations are at the patient-visit level.

Table 3.2: Summary Statistics - Shift 1

Variable	Mean	SD
Patient Age	61.03	3.63
Female Patient (%)	0.54	0.08
White Patient (%)	0.92	0.05
Patient ESI	2.94	0.10
Test Order Count	11.92	2.00
LOS (Minutes)	253.30	35.83
Rate of Return (%)	0.02	0.03
Admission Rate (%)	0.11	0.06
ED Volume	26.64	7.19

Note: N = 1,259. Observations are at the shift level.

Table 3.3: Summary Statistics - Shift 2

Variable	Mean	SD
Patient Age	56.91	5.45
Female Patient (%)	0.55	0.13
White Patient (%)	0.91	0.07
Patient ESI	2.93	0.14
Test Order Count	11.53	2.74
LOS (Minutes)	232.53	45.51
Rate of Return (%)	0.02	0.04
Admission Rate (%)	0.11	0.08
ED Volume	21.33	6.23

Note: N = 1,172. Observations are at the shift level.

Table 3.4: Summary Statistics - Shift 3

Variable	Mean	SD
Patient Age	55.78	13.14
Female Patient (%)	0.51	0.31
White Patient (%)	0.91	0.18
Patient ESI	2.91	0.32
Test Order Count	12.80	6.24
LOS (Minutes)	246.73	174.61
Rate of Return (%)	0.02	0.09
Admission Rate (%)	0.14	0.21
ED Volume	11.24	9.01

Note: N = 1,492. Observations are at the shift level.

2017). We, thus, use the difference between the average LOS of day t and the departmental average daily LOS (measured as the average daily ED LOS across our 4-year study period) as the dependent variable. Similarly, we choose the difference between the average 72-hour rate of return of all patients seen by an ED physician on day t and the departmental average daily 72-hour rate of return to define our dependent variable with respect to quality, and the difference between the average admission rate of day t and the departmental average daily admission rate to define our dependent variable with respect to admission rate.

Independent Variables

Our independent variables capture the average permanent performance of all physicians working in a particular shift relative to the average physician permanent performance, where a physician's permanent performance is calculated using his/her average performance across our 4-year study period. Specifically, for each shift i on day t, we measure the average permanent speed of all physicians who worked in shift i on day t relative to the average permanent speed of all physicians across our 4-year study period. Similarly, our independent variables with respect to physician quality and admission rate are defined as the average permanent quality and admission rate of physicians assigned to shift i on day t relative to the average physician permanent quality and admission rate, respectively.

Control Variables

We control for characteristics of those patients who were seen by an ED physician on day t, including age, ESI level, gender, and race. In addition, we control for daily ED volume, which we define as the number of patients who registered between 5:30 AM on the day in question and 5:29 AM the next morning. We also control for physicians' shift preference. We define a physician's preference for a particular shift as the percentage of time s/he has spent working in that specific shift across our 4-year study period. For example, if 50% of a physician's patient visits were completed in the first shift, then his/her preference for shift 1 is determined to be 0.5.

We conduct a day-level analysis by making use of the following regression model:

$$Y_t = \beta_1 FirstShiftAbove_t + \beta_2 SecondShiftAbove_t + \beta_3 P_t + \beta_4 E_t + \beta_5 S_t + \gamma_t + \epsilon_t, \tag{3.1}$$

where Y_t represents the difference between the average performance of the ED on day t and the departmental average daily ED performance. $FirstShiftAbove_t$ is a binary variable coded as 1 if the average permanent performance of physicians assigned to shift 1 on day t is higher than average. Similarly, $SecondShiftAbove_t$ is an indicator variable denoting whether the average permanent performance of physicians assigned to shift 2 is higher than average. P_t refers to a vector of patient characteristics including average age, gender, race and ESI level. E_t and S_t represent average ED volume and physician shift preference on day t, respectively. Lastly, γ_t indicates time fixed effects.

3.5 Results

The regression results corresponding to assignment of physicians with a higher-than-average aggregate permanent speed, quality, and admission rate to shifts 1 and 2 are presented in Tables 3.5-3.7, respectively. As demonstrated in Table 3.5, our results show that the effect of assigning a faster group of physicians to the first shift results in a 8.2-minute improvement in the ED's average daily LOS. This is relatively a tangible effect, considering that reducing LOS in EDs by just a few minutes can have considerable financial implications (Krochmal and Riley 1994). In addition, our results show that the magnitude of the first shift's effect is larger compared to that of the second shift's effect. Furthermore, the coefficients of both the first and second shifts' effect estimates relative to the third shift are negative, implying that the magnitude of the effect of assigning a faster group of physicians to the last shift is smallest compared to that of the earlier shifts. Our finding, thus,

Table 3.5: Shift Effect Estimates - LOS

	Difference Between the Average Daily LOS and the Departmental Mean
Faster Shift 1	-8.1684***
	(2.2119)
Faster Shift 2	-4.7547**
	(1.7860)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3.6: Shift Effect Estimates - 72-Hour Rate of Return

	Difference Between the Average Daily Rate of Return and the Departmental Mean
Higher-Quality Shift 1	-0.0031
	(0.0017)
Higher-Quality Shift 2	0.0003
	(0.0014)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3.7: Shift Effect Estimates - Admission Rate

	Difference Between the Average Daily Admission Rate and the Departmental Mean
Higher-Admission Rate Shift 1	0.0153***
	(0.0026)
Higher-Admission Rate Shift 2	0.0089***
	(0.0026)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

suggests that assigning a mix of physicians with a higher-than-average aggregate permanent speed to the earlier shifts of the day has a higher impact on the daily performance of the ED compared to assigning this group of physicians to the later shifts.

Our results with respect to quality (presented in Table 3.6) provide weak statistically significant evidence of the effect of assigning a higher-quality mix of physicians to the first shift and no statistically significant evidence of the effect of such assignment to the second shift on the average daily quality of the ED. As indicated earlier, our metric for measuring quality is the 72-hour rate of return, which captures the percentage of discharged patients who return to the ED within 72 hours of their initial discharge. Because of the lagged nature of this metric, it is not surprising that we do not observe the effect of physician assignment with respect to quality, captured directly through the 72-hour rate of return, on the daily performance of the ED.[1]

Our results demonstrated in Table 3.7 show that the magnitude of the effect of assigning physicians with a higher-than-average aggregate permanent admission rate to the first shift is larger compared to that of the second shift's effect. In addition, we find that the effect estimates of both the first and second shifts relative to the third shift are positive. This finding suggests that the effect of assigning physicians with a higher-than-average aggregate permanent admission rate to the last shift is smallest in magnitude compared to the effects of such physician assignment to the earlier shifts.

Put together, our findings highlight the importance of assigning the right mix of physicians to the earlier shifts of the day. Particularly, our results suggest that assigning high-performing physicians with respect to speed and admission rate to the first shift of the day would have the greatest impact on the average daily

[1]As a robustness test, we re-run our analysis using alternative measures of quality and observe that our findings remain unchanged (see Section 3.6.2).

performance of the ED. It is important to note, however, that a high-performing group of physicians is not necessarily composed of fast or higher-quality physicians only. For instance, in Saghafian et al. (2019) the authors show that incorporating diversity with respect to ED physicians' quality into scheduling models can have a positive effect on the overall performance of physicians. In this paper, instead of studying the right mix of physicians that should be assigned to the same shift, we shed light on the impact of assigning high-performing physicians to different shifts of the day on the daily performance of the ED.

3.6 Robustness Tests

In this section, we conduct robustness tests to validate our approach and main findings. Specifically, we (a) re-run our analysis on matched samples of observations that achieve balance across observable covariates related to patient, physician, and ED characteristics, and (b) test the robustness of our results with regards to quality by re-running our analysis using alternative measures of quality.

3.6.1 Matching

In order to ensure that all day-level observations in our analysis are comparable across observable covariates corresponding to all three shifts of the day, we use the nearest-neighbor propensity score matching without replacement. We do this to construct matched samples of observations that achieve balance across a set of covariates related to patient characteristics (age, gender, race, ESI), ED volume, and physician shift preference associated with all three shifts of the day. We use a caliper width of 0.1 times the pooled standard deviation of the logit of the propensity score (Rosenbaum and Rubin 1985).

Tables D1-D3 in Appendix D present the pre-matching distributions of the matching variables across the treatment and control groups corresponding to the first shift. The treated group comprises all day-level observations such that the aggregate permanent performance of physicians assigned to the first shift is higher than average and the control group consists of all other day-level observations.

As expected, we observe that covariates related to ED volume and physician shift preference are relatively unbalanced across the treatment and control groups. Tables D4-D6 in Appendix D demonstrate the balanced (statistically indistinguishable) post-matching distributions of the matching variables across the treatment and control groups. Furthermore, Figures 3.1-3.3 demonstrate how matching has improved the balance of all matching covariates across all three shifts of the day. In each case, we observe that the post-matching distributions of the matching variables across the treatment and control groups are relatively balanced.

We run model (3.2) on these matched samples of observations to examine the first-shift effect on the daily performance of the ED :

$$Y_t = \beta_1 FirstShiftAbove_t + \beta_2 P_t + \beta_3 E_t + \beta_4 S_t + \gamma_t + \epsilon_t, \tag{3.2}$$

The results presented in Tables 3.8-3.10 are consistent with our main findings as they provide statistically significant evidence of the first-shift effect with respect to speed and admission rate.

3.6.2 Alternative Measures of Quality

Our analysis with respect to quality is conducted using the 72-hour rate of return metric (Section 3.5). Given the fact that using this metric as a proxy for measuring quality has been shown to be controversial, we test the robustness of our results by re-running our analysis using alternative measures of quality: physician

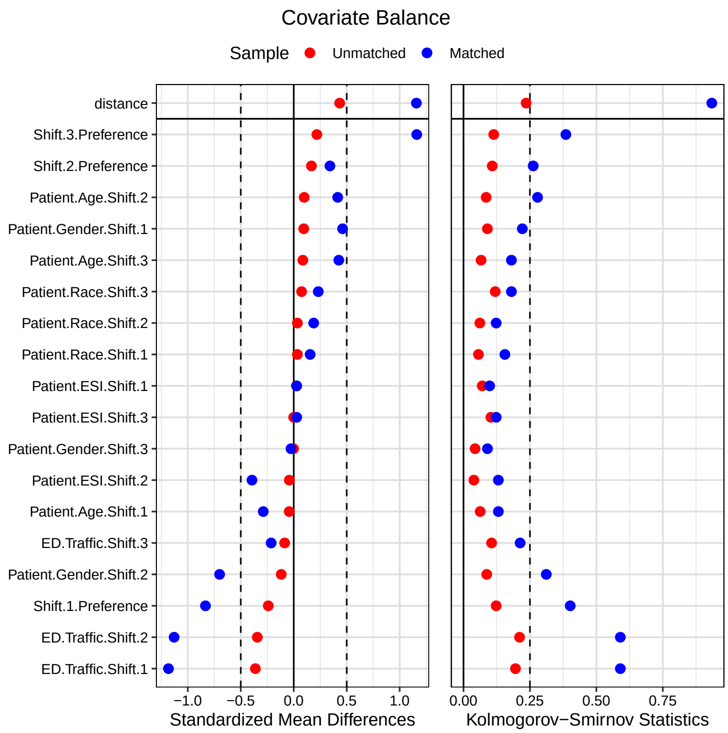


Figure 3.1: Summary Plot of Covariate Balance - Speed

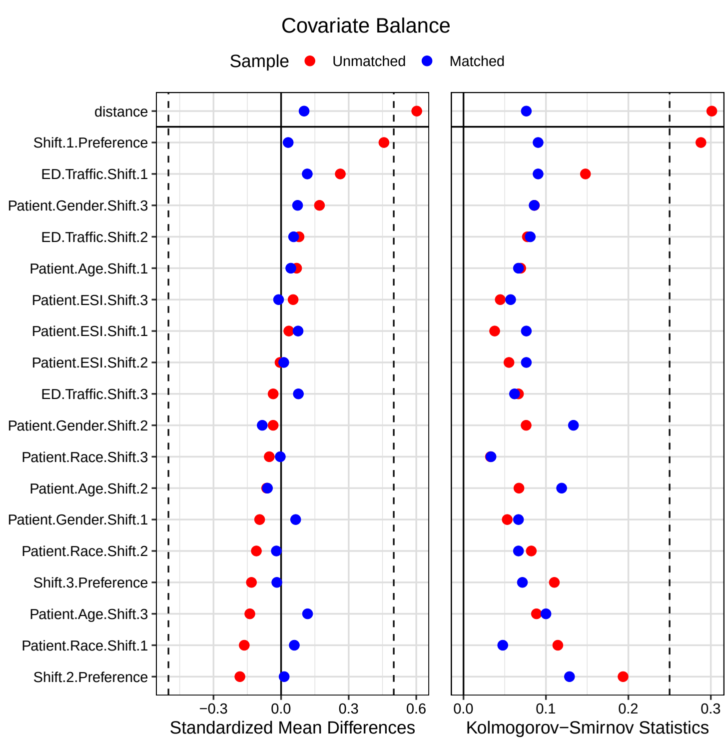


Figure 3.2: Summary Plot of Covariate Balance - Quality

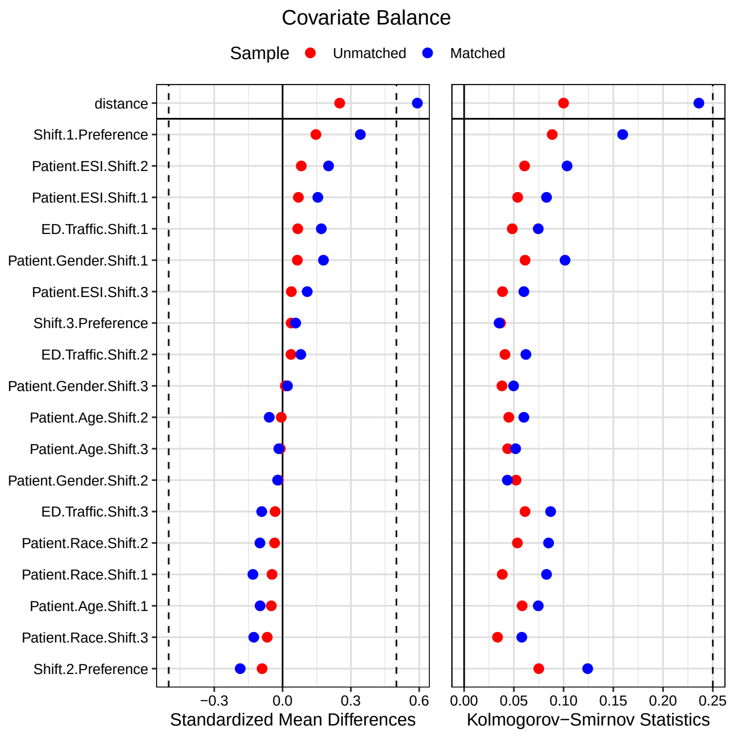


Figure 3.3: Summary Plot of Covariate Balance - Admission Rate

Table 3.8: Shift Effect Estimates - LOS - Matched Sample

	Difference Between the Average Daily LOS and the Departmental Mean
Faster Shift 1	-6.4148***
	(1.0552)
Faster Shift 2	-2.9136*
	(1.1966)
Observations	1,236
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3.9: Shift Effect Estimates - 72-Hour Rate of Return - Matched Sample

	Difference Between the Average Daily Rate of Return and the Departmental Mean
Higher-Quality Shift 1	-0.0032
	(0.0018)
Higher-Quality Shift 2	0.0002
	(0.0021)
Observations	420
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3.10: Shift Effect Estimates - Admission Rate - Matched Sample

	Difference Between the Average Daily Admission Rate and the Departmental Mean
Higher-Admission Rate Shift 1	0.0159***
	(0.0024)
Higher-Admission Rate Shift 2	0.0067**
	(0.0024)
Observations	1,188
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

under- and over-calling rates. We define a physician's under-calling rate as the percentage of the physician's patients who are admitted to the hospital (after their ED visit) by the physician and are upgraded to a more intensive level of care (e.g., ICU) within 12 hours of their initial admission. Similarly, we measure a physician's over-calling rate as the percentage of the physician's patients who are admitted to the hospital by the physician and are discharged within 12 hours of admission. The under- and over-calling rates, thus, measure how often a physician under- and over-calls his/her admitted patients' illness severity, respectively.

We run model (3.1) using the over- and under-calling rates to estimate the shift effects with respect to quality. Specifically, we examine whether assigning a high-performing group of physicians (measured in terms of their aggregate permanent under- and over-calling rates) to the earlier shifts of the day impacts the average daily under- and over-calling rates of the ED. The results presented in Tables 3.11 and 3.12 are consistent with our main findings as they show no statistically significant evidence that assigning high-performing physicians with respect to their under- and over-calling rates to the earlier shifts of the day affects the average daily under- and over-calling rates of the ED.

3.7 Heterogeneity in Shift Effects

In this section, we examine the heterogeneity in the shift effect estimates derived in Section 3.5 with regards to ED volume and day of the week.

3.7.1 ED Volume

Given the substantial evidence of the effects of workload on physician performance, we test whether and how our results change on high-volume days. To this end, we make use of the following model:

$$Y_t = \beta_1 FirstShiftAbove_t + \beta_2 SecondShiftAbove_t + \beta_3 FirstShiftAbove_t \times E_t + \beta_4 SecondShiftAbove_t \times E_t + \beta_5 P_t + \beta_6 E_t + \beta_7 S_t + \gamma_t + \epsilon_t, \tag{3.3}$$

where the interaction terms capture the extent to which the assignment of high-performing physicians to the first and second shifts differentially impacts the daily performance of the ED on high-volume days.

Tables 3.13 and 3.14 present the results with respect to physician speed and admission rate, respectively. Our results provide statistically significant evidence that the impact of assigning physicians with a higher-than-average aggregate speed and admission rate to the first shift of the day increases on high-volume days, suggesting that the effects of assigning these groups of physicians to the first shift

Table 3.11: Shift Effect Estimates - Over-Calling Rate

	Difference Between the Average Daily Over-Calling Rate and the Departmental Mean
Lower Over-Calling Rate Shift 1	-0.0029
	(0.0011)
Lower Over-Calling Rate Shift 2	-0.0002
	(0.0010)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3.12: Shift Effect Estimates - Under-Calling Rate

	Difference Between the Average Daily Under-Calling Rate and the Departmental Mean
Lower Under-Calling Rate Shift 1	-0.0021
	(0.0011)
Lower Under-Calling Rate Shift 2	-0.0004
	(0.0015)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3.13: Heterogeneity - ED Volume - Speed

	Difference Between the Average Daily LOS and the Departmental Mean
Faster Shift 1 × ED Volume	0.5965*
	(0.2640)
Faster Shift 2 × ED Volume	0.5459*
	(0.2296)
Faster Shift 1	-25.2781***
	(7.0224)
Faster Shift 2	-18.9005**
	(5.8712)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3.14: Heterogeneity - ED Volume - Admission Rate

	Difference Between the Average Daily Admission Rate and the Departmental Mean
Higher-Admission Rate Shift 1 × ED Volume	0.0090*
	(0.0048)
Higher-Admission Rate Shift 2 × ED Volume	0.0097
	(0.0051)
Higher-Admission Rate Shift 1	0.0301**
	(0.0102)
Higher-Admission Rate Shift 2	0.0177
	(0.0101)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

are more pronounced when ED volume is high.

3.7.2 Day of the Week

Understanding how the magnitude of the documented effects changes on weekends compared to weekdays could be useful in the area of physician scheduling and staffing. In order to examine the heterogeneity in the observed shift effect estimates with regards to the day of the week, we use the following model:

$$Y_t = \beta_1 FirstShiftAbove_t + \beta_2 SecondShiftAbove_t + \beta_3 FirstShiftAbove_t \times D_t + \beta_4 SecondShiftAbove_t \times D_t + \beta_5 P_t + \beta_6 E_t + \beta_7 S_t + \beta_8 D_t + \gamma_t + \epsilon_t, \quad (3.4)$$

where D_t is a binary variable coded as 1 if day t is a Saturday or Sunday.

Our results presented in Table 3.15 show a statistically significant negative coefficient on the interaction term ($FirstShiftAbove_t \times D_t$), suggesting that the effect of assigning a group of relatively faster physicians to the first shift of the day decreases on weekends. With regards to physicians' admission rate, our results (presented in Table 3.16) provide no statistically significant evidence of heterogeneity in the observed shift effects with respect to the day of the week.

Table 3.15: Heterogeneity - Day of the Week - Speed

	Difference Between the Average Daily LOS and the Departmental Mean
Faster Shift 1 × Weekend	-0.1824*
	(0.0853)
Faster Shift 2 × Weekend	-0.0750
	(0.0808)
Faster Shift 1	-9.2189***
	(2.2509)
Faster Shift 2	-5.1469**
	(1.8068)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3.16: Heterogeneity - Day of the Week - Admission Rate

	Difference Between the Average Daily Admission Rate and the Departmental Mean
Higher-Admission Rate Shift 1 × Weekend	0.0050
	(0.0060)
Higher-Admission Rate Shift 2 × Weekend	-0.0035
	(0.0060)
Higher-Admission Rate Shift 1	0.0138***
	(0.0031)
Higher-Admission Rate Shift 2	0.0095**
	(0.0031)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Overall, our findings provide evidence of heterogeneity in the documented shift effects with respect to ED volume and day of the week. Particularly, our results show that the magnitude of the first-shift effect increases on high-volume days (with respect to both speed and admission rate) and on weekdays (with respect to speed). These suggest that it is especially important to make use of high-performing physicians during the first shift of weekdays on which the ED historically has a higher patient demand. In our partner ED, for example, this translates to allocating high-performing physicians to the first shift of Wednesdays.

3.8 Propagation of the First-Shift Effect

Our results presented in Section 3.5 suggest that assigning physicians with a higher-than-average aggregate permanent speed and admission rate to the first shift of the day would have a greater impact on the daily performance of the ED compared to assigning these physicians to the second and last shifts. A setting such as an ED where there exist shared resources (e.g., test services, nurses, hallways) that are often constrained during busy times is similar to a queuing system such that a shift's performance is prone to be affected by spillover from earlier shifts (Gerla and Kleinrock 1980). Consequently, assigning high-performing physicians to the first shift could have a domino effect throughout the rest of the day: physicians' performance in the earlier shifts affects the performance of the ED in the later shifts of the day.

In order to test whether such a domino effect exists in our setting, we examine the effects of assigning physicians with a higher-than-average aggregate speed and admission rate to the first shift on the performance of the later shifts of the day. A positive correlation would imply that assigning a high-performing group of physicians to the earlier shifts of the day positively affects the performance of physicians assigned to the later shifts of the day. To this end, we make use of the following model:

$$Y_{t_1} = \beta_1 FirstShiftAbove_t + \beta_2 P_{t_1} + \beta_3 E_{t_1} + \beta_4 S_{t_1} + \gamma_{t_1} + \epsilon_{t_1}, \tag{3.5}$$

where Y_{t_1} indicates the average performance of the ED on day t across all shifts starting after shift 1. $FirstShiftAbove_t$ as indicated before is a binary variable coded as 1 if a high-performing group of physicians is assigned to the first shift. P_{t_1}, E_{t_1}, and S_{t_1} denote patient characteristics (average age, gender, race, ESI), average ED volume, and average physician shift preference associated with all shifts starting after the first shift, respectively.

The results presented in Tables 3.17 and 3.18 provide statistically significant evidence that the relative aggregate permanent speed and admission rate of physicians assigned to the first shift impact the performance of the later shifts with respect to speed and admission rate, respectively. This finding provides evidence for a domino effect of physician performance in our setting. In particular, our results shed light on the propagation effect of physician performance in a congestion-prone system such as an ED and highlight the impact of physician assignment to the first shift on the performance of the ED in the subsequent shifts. As demonstrated in Section 3.7.1, our results show that this effect is more pronounced on high-volume days.

3.9 Conclusions

In this paper, we study the effects of assigning high-performing physicians to different shifts of the day on the daily performance of the ED, where we evaluate the performance of the ED in terms of speed, quality, and admission rate.

Our results provide empirical evidence that allocating a group of high-performing physicians to the earlier shifts of the day has a greater impact on the daily performance of the ED compared to assigning these physicians to the later shifts. Our results further suggest that physicians' performance in the earlier shifts of the day has a domino effect throughout the rest of the day.

Table 3.17: Rest of the Day Shift Effect Estimates - LOS

	Rest of the Day Average LOS
Faster Shift 1	-5.2160***
	(0.9404)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table 3.18: Rest of the Day Shift Effect Estimates - Admission Rate

	Rest of the Day Average Admission Rate
Higher-Admission Rate Shift 1	0.0038***
	(0.0007)
Observations	1,277
Time Fixed Effects	Yes
Controls	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Furthermore, we examine whether the magnitude of the established shift effects changes on high-volume days and on weekends. Our results show that while the magnitude of the first-shift effect with respect to speed decreases on weekends compared to weekdays, the magnitude of the effects with respect to both speed and admission rate is greater on high-volume days. These findings suggest that it is especially important to allocate high-performing physicians to the first shift on weekdays and/or high-volume days.

It is important to consider the limitations of our study. First, while our results provide statistically significant evidence of the positive effects of assigning high-performers to the earlier shifts of the day, our insights are limited to a single ED. Future studies could improve the generalizability of our findings by examining a larger sample of EDs. Second, due to data limitations, we are unable to control for a number of factors including bed availability, identities of medical personnel assigned to each shift, among others. Capturing and including such factors in our analyses could improve the validity of our findings.

Our findings have important implications for improving the operations of hospital EDs. Specifically, our results highlight the importance of assigning the right mix of physicians to the earlier shifts of the day. In particular, our results can be useful to hospital administrators in the area of physician scheduling where decisions regarding assignment of physicians to different shifts of the day need to be made.

References

[1] Abualenain J, Frohna WJ, Smith M, Pipkin M, Webb C, Milzman D, Pines JM (2013) The prevalence of quality issues and adverse outcomes among 72-hour return admissions in the emergency department. *The Journal of Emergency Medicine* 45(2):281–288.

[2] Aksin OZ, Armony M, Mehrotra V (2007) The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management* 16(6):665–688.

[3] Anand KS, Paç MF, Veeraraghavan S (2011) Quality-speed conundrum: Trade-offs in customer-intensive services. *Management Science* 57(1):40–56.

[4] Anderson D, Golden B, Jank W, Wasil E (2012) The impact of hospital utilization on patient readmission rate. *Health Care Management Science* 15(1):29–36.

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

[6] Armony M, Israelit SH, Mandelbaum A, Marmor YN, Tseytlin Y, Yom-Tov GB (2015) Patient flow in hospitals: A data-based queuing-science perspective. *Stochastic Systems* 5(1):146–194.

[7] Ata B, Van Mieghem JA (2009) The value of partial resource pooling: Should a service network be integrated or product-focused? *Management Science* 55(1):115–131.

[8] Azoulay P, Graff Zivin JS, Wang J (2010) Superstar extinction. *Quarterly Journal of Economics* 125(2):549–589.

[9] Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30(9):1078–1092.

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

[11] Batt RJ, KC DS, Staats BR, Patterson BW (2019) The effects of discrete work shifts on a nonterminating service system. *Production and Operations Management* 28(6):1528–1544.

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

[13] Berry Jaeker JA, Tucker AL (2017) Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Science* 63(4):1042–1062.

[14] Bruursema K, Kessler SR, Spector PE (2011) Bored employees' misbehaviour: The relationship between boredom and counterproductive work behavior. *Work and Stress* 25(2):93–107.

[15] Buechel B, Mechtenberg L, Petersen J (2018) If I can do it, so can you! Peer effects on perseverance. *Journal of Economic Behavior and Organization* 155:301–314.

[16] Castelli A, Street A, Verzulli R, Ward P (2015) Examining variations in hospital productivity in the English NHS. *European Journal of Health Economics* 16(3):243–254.

[17] Cayirli T, Veral E, Rosen H (2008) Assessment of patient classification in appointment system design. *Production and Operations Management* 17(3):338–353.

[18] Cayirli T, Veral E (2009) Outpatient scheduling in health care: A review of the literature. *Production and Operations Management* 12(4):519–549.

[19] Cayirli T, Yang KK, Quek SA (2012) A Universal appointment rule in the presence of no-shows and walk-ins. *Production and Operations Management* 21(4):682–697.

[20] Chan DC (2016) Teamwork and moral hazard: Evidence from the emergency department. *Journal of Political Economy* 124(3):734–770.

[21] Chan DC (2018) The efficiency of slacking off: Evidence from the emergency department. *Econometrica* 86(3):997–1030.

[22] Chan TY, Li J, Pierce L (2014) Compensation and peer effects in competing sales teams. *Management Science* 60(8):1965–1984.

[23] Chan TY, Li J, Pierce L (2014) Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science* 33(4):463–484.

[24] Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research* 2(6):429–444.

[25] Chilingerian JA (1995) Evaluating physician efficiency in hospitals: A multivariate analysis of best practices. *European Journal of Operational Research* 80(3):548–574.

[26] Chuang HHC, Oliva R, Perdikaki O (2016) Traffic-based labor planning in retail stores. *Production and Operations Management* 25(1):96–113.

[27] Collier DA, Collier CE, Kelly TM (2006) Benchmarking physician performance, part 1. *Journal of Medical Practice Management* 21(4):185–189.

[28] Cooper WW, Seiford LM, Tone K (2007) *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* 2nd ed. (Springer, New York).

[29] Cornelissen T, Dustmann C, Schönberg U (2017) Peer effects in the workplace. *American Economic Review* 107(2):425–456.

[30] Do HT, Shunko M, Lucas MT, Novak DC (2018) Impact of behavioral factors on performance of multi-server queueing systems. *Production and Operations Management* 27(8):1553–1573.

[31] Edmondson AC, Bohmer RM, Pisano GP (2001) Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative Science Quarterly* 46(4):685–716.

[32] Emergency Department Cases (2015) Emergency Physician Speed-Solutions to Physician Productivity. Retrieved February 26, 2019, https://emergencymedicinecases.com/emergency-physician-speed-and-productivity-solutions/.

[33] Fernandes CMB, Price A, Christenson JM (1997) Does reduced length of stay decrease the number of emergency department patients who leave without seeing a physician? *Journal of Emergency Medicine* 15(3):397–399.

[34] Fiallos J, Patrick J, Michalowski W, Farion K (2017) Using data envelopment analysis for assessing the performance of pediatric emergency department physicians. *Health Care Management Science* 20(1):129–140.

[35] Fisher ML, Krishnan J, Netessine S (2006) Retail store execution: An empirical study. *Working Paper*, The Wharton School, University of Pennsylvania.

[36] Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management* 5(2):79–141

[37] Gerla M, Kleinrock L (1980) Flow control: A comparative survey. *IEEE Transactions on Communications* 28(4):553–574.

[38] Glickman SW, Schulman KA, Peterson ED, Hocker MB, Cairns CB (2008) Evidence-based perspectives on pay for performance and quality of patient care and outcomes in emergency medicine. *Annals of Emergency Medicine* 51(5):622–631.

[39] Gordon JA, An LC, Hayward RA, Williams BC (1998) Initial emergency department diagnosis and return visits: Risk versus perception. *Annals of Emergency Medicine* 32(5):569–573.

[40] Goulet F, Jacques A, Gagnon R, Bourbeau D, Laberge D, Melanson J, Ménard C, Racette P, Rivest R (2002) Performance assessment: Family physicians in Montreal meet the mark! *Canadian Family Physician* 48(8):1337–1344.

[41] Greene WH (1993) *Econometric Analysis.* 2nd ed. (Macmillan, New York).

[42] Grosskopf S, Valdmanis V (1987) Measuring hospital performance: A non-parametric approach. *Journal of Health Economics* 6(2):89–107.

[43] Gupta D, Denton BT (2008) Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions* 40(9):800–819.

[44] Hackman JR, Oldham GR (1980) *Work redesign.* Reading. (MA, Addison-Wesley).

[45] Harrison JP, Ogniewski RJ (2005) An efficiency analysis of veterans health administration hospitals. *Military Medicine* 170(7):607–611.

[46] Hasija S, Shumsky RA, Pinker EJ (2008) Call center outsourcing contracts under information asymmetry. *Management Science* 54(4):793–807.

[47] Hess BJ, Weng W, Lynn LA, Holmboe ES, Lipner RS (2011) Setting a fair performance standard for physicians' quality of patient care. *Journal of General Internal Medicine* 26(5):467–473.

[48] Hollingsworth B (2008) The measurement of efficiency and productivity of health care delivery. *Health Economics* 17(10):1107–1128.

[49] Huang J, Carmeli B, Mandelbaum A (2015) Control of patient flow in emergency departments, or multiclass queues with deadlines and feedback. *Operations Research* 63(4):892–908.

[50] Huckman RS, Staats BR, Upton DM (2009) Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Science* 55(1):85–100.

[51] Huesch MD (2011) Is blood thicker than water? Peer effects in stent utilization among Floridian cardiologists. *Social Science and Medicine* 73(12):1756–1765.

[52] Iyengar R, Van den Bulte C, Lee JY (2015) Social contagion in new product trial and repeat. *Marketing Science* 34(3):408–429.

[53] Jackson CK, Bruegmann E (2009) Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics* 1(4):85–108.

[54] Johannessen KA, Kittelsen SAC, Hagen TP (2017) Assessing physician productivity following Norwegian hospital reform: A panel and data envelopment analysis. *Social Science and Medicine* 175(C):117–126.

[55] Jouini O, Aksin OZ, Karaesmen AF, Aguir MS, Dallery Y (2015) Call center delay announcement using a newsvendor-like performance criterion. *Production and Operations Management* 24(4):587–604.

[56] Kass SJ, Vodanovich SJ, Callender A (2001) State-trait boredom: Relationship to absenteeism, tenure, and job satisfaction. *Journal of Business and Psychology* 16(2):317–327.

[57] KC DS, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* 55(9):1486–1498.

[58] KC DS, Terwiesch C (2012) An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing and Service Operations Management* 14(1):50–65.

[59] KC DS, Staats BR, Gino F (2013) Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery. *Management Science* 59(11):2435–2449.

[60] Keehan SP, Stone DA, Poisal JA, Cuckler GA, Sisko AM, Smith SD, Madison AJ, Wolfe CJ, Lizonitz JM (2017) National health expenditure projections, 2016-25: Price increases, aging push sector to 20 percent of economy. *Health Affairs* 36(3):553–563.

[61] Keith KD, Bocka JJ, Kobernick MS, Krome RL, Ross MA (1989) Emergency department revisits. *Annals of Emergency Medicine* 18(9):964–968.

[62] Klasco RS, Wolfe RE, Wong M, Edlow J, Chiu D, Anderson PD, Grossman SA (2015) Assessing the rates of error and adverse events in the ED. *The American Journal of Emergency Medicine* 33(12):1786–1789.

[63] Krochmal P, Riley TA (1994) Increased health care costs associated with ED overcrowding. *American Journal of Emergency Medicine* 12(3):265–266.

[64] Kuntz L, Mennicken R, Scholtes S (2011) Stress on the ward: An empirical study of the nonlinear relationship between organizational workload and service quality. *Working Paper*, Ruhr-Universitaet Bochum, Bochum, Germany.

[65] Kuntz L, Mennicken R, Scholtes S (2015) Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science* 61(4):754–771.

[66] Latham LP, Ackroyd-Stolarz S (2014) Emergency department utilization by older adults: A descriptive study. *Canadian Geriatrics Journal* 17(4):118–125.

[67] Manchanda P, Xie Y, Youn N (2008) The role of targeted communication and contagion in new product adoption. *Marketing Science* 27(6):961–976.

[68] Mas A, Moretti E (2009) Peers at work. *American Economic Review* 99(1):112–145.

[69] Ng TWH, Feldman DC (2013) Does longer job tenure help or hinder job performance? *Journal of Vocational Behavior* 83(3):305–314.

[70] Ozcan YA (1998) Physician benchmarking: Measuring variation in practice behavior in treatment of otitis media. *Health Care Management Science* 1(1):5–17.

[71] Ozcan YA, Begun JW, McKinney MM (1999) Benchmarking organ procurement organizations: A national study. *Health Services Research* 34(4):855–878.

[72] Pai CW, Ozcan YA, Jiang HJ (2000) Regional variation in physician practice pattern: An examination of technical and cost efficiency for treating sinusitis. *Journal of Medical Systems* 24(2):103–117.

[73] Patel PB, Vinson DR (2005) Team assignment system: Expediting emergency department care. *Annals of Emergency Medicine* 46(6):499–506.

[74] Pham JC, Kirsch TD, Hill PM, DeRuggerio K, Hoffmann B (2011) Seventy-two-hour returns may not be a good indicator of safety in the emergency department: A national study. *Academic Emergency Medicine* 18(4):390–397.

[75] Powell SG, Schultz KL (2004) Throughput in serial lines with state-dependent behavior. *Management Science* 50(8):1095–1105.

[76] Powell A, Savin S, Savva N (2012) Physician workload and hospital reimbursement. *Manufacturing and Service Operations Management* 14(4):512–528.

[77] Robinson LW, Chen RR (2003) Scheduling doctors' appointments: Optimal and empirically-based heuristic policies. *IIE Transactions* 35(3):295–307.

[78] Rosenbaum PR, Rubin DB (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39(1):33–38.

[79] Saghafian S, Hopp WJ, Van Oyen MP, Desmond JS, Kronick SL (2012) Patient streaming as a mechanism for improving responsiveness in emergency departments. *Operations Research* 60(5):1080–1097.

[80] Saghafian S, Hopp WJ, Van Oyen MP, Desmond JS, Kronick SL (2014) Complexity-augmented triage: A tool for improving patient safety and operational efficiency. *Manufacturing and Service Operations Management* 16(3):329–345.

[81] Saghafian S, Austin G, Traub SJ (2015) Operations research/management contributions to emergency department patient flow optimization: Review and research prospects. *IIE Transactions on Healthcare Systems Engineering* 5(2):101–123.

[82] Saghafian S, Hopp WJ, Iravani SMR, Cheng Y, Diermeier D (2018) Workload management in telemedical physician triage and other knowledge-based service systems. *Management Science* 64(11):5180–5197.

[83] Saghafian S, Imanirad R, Traub SJ (2019) Do physicians influence each other's performance? Evidence from the emergency department. *Working Paper,* Harvard University.

[84] Salway RJ, Valenzuela R, Shoenberger JM, Mallon WK, Viccellio A (2017) Emergency Department (ED) overcrowding: Evidence-based answers to frequently asked questions. *Revista Medica Clinica Las Condes* 28(2):213–219.

[85] Savage DW, Woolford DG, Weaver B, Wood D (2015) Developing emergency department physician shift schedules optimized to meet patient demand. *Canadian Journal of Emergency Medicine* 17(1):3–12.

[86] Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences* 95(11):5857–5864.

[87] Sherman HD (1984) Hospital efficiency measurement and evaluation: Empirical test of a new technique. *Medical Care* 22(10):922–938.

[88] Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science* 44(1):49–61.

[89] Simar L, Wilson PW (2002) Non-parametric tests of returns to scale. *European Journal of Operational Research* 139(1):115–132.

[90] Simar L, Wilson PW (2011) Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis* 36(1):33–53.

[91] Sinreich D, Jabali O, Dellaert NP (2012) Reducing emergency department waiting times by adjusting work shifts considering patient visits to multiple care providers. *IIE Transactions* 44(3):163–180.

[92] Smith CA, Varkey AB, Evans AT, Reilly BM (2004) Evaluating the performance of inpatient attending physicians: A new instrument for today's teaching hospitals. *Journal of General Internal Medicine* 19(7):766–771.

[93] Spaite DW, Bartholomeaux F, Guisto J, Lindberg E, Hull B, Eyherabide A, Lanyon S, Criss EA, Valenzuela TD, Conroy C (2002) Rapid process redesign in a university-based emergency department: Decreasing waiting time intervals and improving patient satisfaction. *Annals of Emergency Medicine* 39(2):168–177.

[94] Steinbach D, Tatsi E (2016) The effect of team production on individual performance. *Working Paper*, Goethe University Frankfurt, Department of Management and Microeconomics, and Stockholm University, Swedish Institute for Social Research (SOFI).

[95] Storfa AH, Wilson ML (2015) Physician productivity: Issues and controversies. *American Journal of Clinical Pathology* 143(1):6–9.

[96] Tan TF, Netessine S (2014) When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science* 60(6):1574–1593.

[97] The Becker's Hospital Review (2016) Hospitals: Is your ED's length of stay costing you millions? Retrieved September 20, 2018, https://www.beckershospitalreview.com/finance/hospitals-is-your-ed-s-length-of-stay-costing-you-millions.html.

[98] Traub SJ, Wood JP, Kelley J, Nestler DM, Chang YH, Saghafian S, Lipinski CA (2015) Emergency department rapid medical assessment: Overall effect and mechanistic considerations. *Journal of Emergency Medicine* 48(5):620–627.

[99] Traub SJ, Stewart CF, Didehban R, Bartley AC, Saghafian S, Smith VD, Silvers SM, LeCheminant R, Lipinski CA (2016) Emergency department rotational patient assignment. *Annals of Emergency Medicine* 67(2):206–215.

[100] Trusov M, Bodapati AV, Bucklin RE (2010) Determining influential users in internet social networks. *Journal of Marketing Research* 47(4):643–658.

[101] Tsugawa Y, Jha AK, Newhouse JP, Zaslavsky AM, Jena AB (2017) Variation in physician spending and association with patient outcomes. *JAMA Internal Medicine* 177(5):675–682.

[102] Varabyova Y, Schreyögg J (2013) International comparisons of the technical efficiency of the hospital sector: Panel data analysis of OECD countries using parametric and non-parametric approaches. *Health Policy* 112(1):70–79.

[103] Venkataraman S, Fredendall LD, Taaffe KM, Huynh N, Ritchie G (2018) An empirical examination of surgeon experience, surgeon rating, and costs in perioperative services. *Journal of Operations Management* 61(1):68–81.

[104] Wagner JM, Shimshak DG, Novak MA (2003) Advances in physician profiling: The use of DEA. *Socio-Economic Planning Sciences* 37(2):141–163.

[105] Waldinger F (2012) Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *Review of Economic Studies* 79(2):838–861.

[106] Wallace RB, Whitt W (2005) A staffing algorithm for call centers with skill-based routing. *Manufacturing and Service Operations Management* 7(4):276–294.

[107] Wang PP (1999) Sequencing and scheduling N customers for a stochastic server. *European Journal of Operational Research* 119(3):729–738.

[108] Wang X, Debo LG, Scheller-Wolf A, Smith SF (2010) Design and analysis of diagnostic service centers. *Management Science* 56(11):1873–1890.

[109] Ward AR, Armony M (2013) Blind fair routing in large-scale service systems with heterogeneous customers and servers. *Operations Research* 61(1):228–243.

[110] Woehr DJ, Arciniega LM, Poling TL (2013) Exploring the effects of value diversity on team effectiveness. *Journal of Business and Psychology* 28(1):107–121.

[111] Yankovic N, Green LV (2011) Identifying good nursing levels: A queuing approach. *Operations Research* 59(4):942–955.

[112] Zheng W, Sun H, Zhang P, Zhou G, Jin Q, Lu X (2018) A four-stage DEA-based efficiency evaluation of public hospitals in China after the implementation of new medical reforms. *PLoS One* 13(10):e0203780.

[113] Zoogah DB, Vora D, Richard OC, Peng MW (2011) Strategic alliance team diversity, coordination, and effectiveness. *The International Journal of Human Resource Management* 22(3):510–529.

Appendix A: Variable Importance Graphs

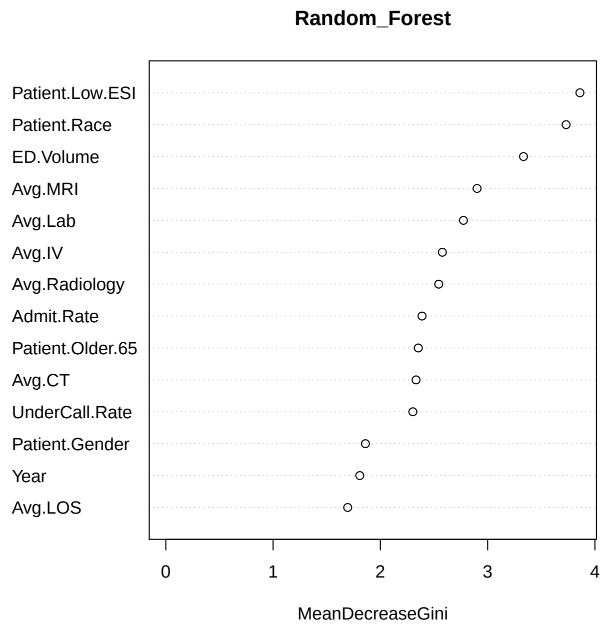


Figure A1: Variable Importance Graph - Random Forest Effectiveness Model

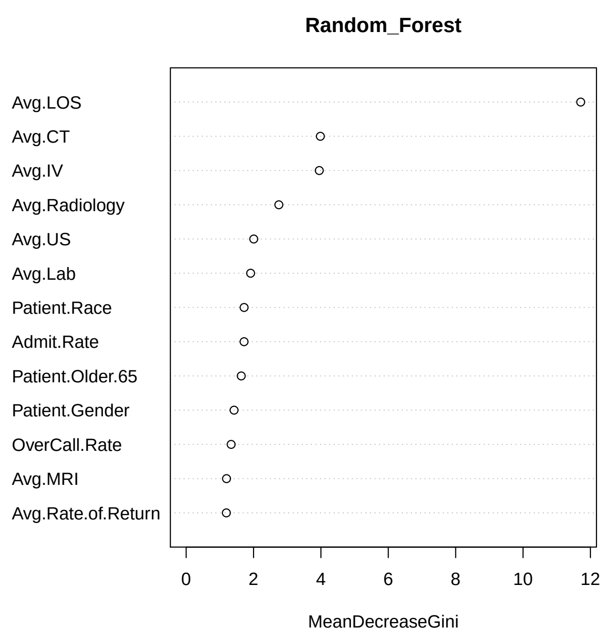


Figure A2: Variable Importance Graph - Random Forest Efficiency Model

Appendix B: Regression Results - Main Effect Estimates

Table B1: Regression Results - Faster Peer

	(1)	(2)
	LOS	72-Hour Rate of Return
Patient Age	0.0936***	0.0011
	(0.0222)	(0.0017)
Patient Gender	2.0878***	-0.0590
	(0.7093)	(0.0578)
Patient Race	-0.1318	0.1774
	(1.0056)	(0.1094)
Patient ESI	-1.3682	-0.1356***
	(1.0936)	(0.0518)
Patient Admission	19.8987***	-0.5681***
	(1.6904)	(0.0907)
ED Volume	9.9204***	-0.0021
	(0.5318)	(0.0035)
US Order Count	11.7711***	-0.1320
	(1.0573)	(0.1005)
MRI Order Count	32.7309***	0.0054
	(2.5395)	(0.1770)
Radiology Order Count	8.7060***	-0.3119***
	(0.9788)	(0.0494)
IV Med Fluid Order Count	9.0816***	0.0299
	(0.6090)	(0.0182)
Lab Order Count	1.9591***	-0.0239***
	(0.1219)	(0.0078)
CT Order Count	15.9243***	-0.1779**
	(0.9580)	(0.0664)
Faster Peer	5.2402***	0.0004
	(0.6709)	(0.0267)
Constant	127.782***	-2.1981***
	(8.4191)	(0.5344)
Observations	253,922	253,922
Time Fixed Effects	Yes	Yes
Physician Fixed Effects	Yes	Yes
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table B2: Regression Results - Slower Peer

	(1) LOS	(2) 72-Hour Rate of Return
Patient Age	0.0940***	0.0001
	(0.0204)	(0.0009)
Patient Gender	2.4324**	-0.0920*
	(1.1615)	(0.0478)
Patient Race	-1.3517	0.1357
	(0.9245)	(0.1020)
Patient ESI	-0.8567	-0.0914**
	(1.4441)	(0.0432)
Patient Admission	18.7836***	-0.4328***
	(3.3336)	(0.0888)
ED Volume	11.4024***	-0.0054
	(1.0729)	(0.0036)
US Order Count	10.3093***	-0.0928
	(1.5695)	(0.0613)
MRI Order Count	25.6152***	0.0013
	(4.2991)	(0.1497)
Radiology Order Count	6.7595***	-0.2866***
	(1.3282)	(0.0434)
IV Med Fluid Order Count	7.7752***	0.0174
	(0.8247)	(0.0163)
Lab Order Count	1.7228***	-0.0328***
	(0.2751)	(0.0059)
CT Order Count	13.1314***	-0.0656
	(1.8237)	(0.0585)
Slower Peer	-5.1070***	0.0170
	(0.6293)	(0.0263)
Constant	131.6160***	-2.0948***
	(6.4960)	(0.3138)
Observations	253,922	253,922
Time Fixed Effects	Yes	Yes
Physician Fixed Effects	Yes	Yes
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table B3: Regression Results - Higher-Quality Peer

	(1) LOS	(2) 72-Hour Rate of Return
Patient Age	0.1413***	-0.0011
	(0.0157)	(0.0011)
Patient Gender	2.1213***	-0.0653
	(0.5222)	(0.0481)
Patient Race	-2.0426*	0.1687**
	(0.9208)	(0.0729)
Patient ESI	1.1299	-0.0635*
	(0.8837)	(0.0380)
Patient Admission	20.6569***	-0.4326***
	(1.6434)	(0.0641)
ED Volume	10.2164***	-0.0045
	(0.4919)	(0.0036)
US Order Count	11.9380***	-0.1570**
	(1.0656)	(0.0734)
MRI Order Count	31.1662***	0.0300
	(2.2902)	(0.1374)
Radiology Order Count	8.3117***	-0.2646***
	(0.8279)	(0.0296)
IV Med Fluid Order Count	8.9931***	0.0130
	(0.5144)	(0.0128)
Lab Order Count	1.8106***	-0.0309***
	(0.1309)	(0.0062)
CT Order Count	15.1579***	-0.0643
	(1.1683)	(0.0457)
Higher-Quality Peer	-0.7553	0.0850*
	(0.6697)	(0.0459)
Constant	132.9968***	-2.3875***
	(7.9453)	(0.3245)
Observations	253,922	253,922
Time Fixed Effects	Yes	Yes
Physician Fixed Effects	Yes	Yes

Note: *p<0.05; **p<0.01; ***p<0.001

Table B4: Regression Results - Lower-Quality Peer

	(1)	(2)
	LOS	72-Hour Rate of Return
Patient Age	0.1145***	-0.0012
	(0.0190)	(0.0011)
Patient Gender	2.4443***	-0.0657**
	(0.7368)	(0.0335)
Patient Race	-1.4484**	0.1795**
	(0.7145)	(0.0794)
Patient ESI	0.9877	-0.0622
	(0.8922)	(0.0468)
Patient Admission	18.7906***	-0.3885***
	(2.5009)	(0.0571)
ED Volume	10.9392***	-0.0039
	(0.8478)	(0.0028)
US Order Count	11.2892***	-0.1262*
	(1.5698)	(0.0696)
MRI Order Count	27.5377***	0.0415
	(3.1634)	(0.1477)
Radiology Order Count	7.5608***	-0.2630***
	(1.1230)	(0.0488)
IV Med Fluid Order Count	8.2593***	0.0185
	(0.7901)	(0.0153)
Lab Order Count	1.7048***	-0.0306***
	(0.2115)	(0.0052)
CT Order Count	14.1751***	-0.0737*
	(1.6722)	(0.0419)
Lower-Quality Peer	0.6836	-0.2036***
	(0.8513)	(0.0533)
Constant	132.5669***	-2.1545***
	(7.4841)	(0.3136)
Observations	253,922	253,922
Time Fixed Effects	Yes	Yes
Physician Fixed Effects	Yes	Yes
Note:	*p<0.05; **p<0.01; ***p<0.001	

Appendix C: Covariate Balance Tables - Chapter 2

Table C1: Pre-Matching Covariate Balance - Faster Peer

	0	1	p	test SMD
Number of Observations	193140	60782		
ED Volume (mean (sd))	27.24 (14.96)	28.41 (13.07)	<0.001	0.084
Patient Age (mean (sd))	59.31 (20.64)	59.61 (20.53)	<0.001	0.014
Female Patient (mean (sd))	0.54 (0.50)	0.54 (0.50)	0.926	<0.001
White Patient (mean (sd))	0.91 (0.28)	0.91 (0.28)	0.017	0.009
Patient ESI (mean (sd))	2.96 (0.55)	2.96 (0.54)	0.441	0.003

Table C2: Pre-Matching Covariate Balance - Slower Peer

	0	1	p	test SMD
Number of Observations	148532	105390		
ED Volume (mean (sd))	28.16 (13.35)	26.94 (15.63)	<0.001	0.084
Patient Age (mean (sd))	59.53 (20.56)	59.26 (20.65)	0.001	0.013
Female Patient (mean (sd))	0.54 (0.50)	0.54 (0.50)	0.752	0.001
White Patient (mean (sd))	0.91 (0.28)	0.91 (0.28)	0.179	0.005
Patient ESI (mean (sd))	2.96 (0.54)	2.96 (0.55)	0.473	0.003

Table C3: Pre-Matching Covariate Balance - Higher-Quality Peer

	0	1	p	test SMD
Number of Observations	153461	100461		
ED Volume (mean (sd))	27.78 (14.87)	27.54 (13.16)	<0.001	0.017
Patient Age (mean (sd))	59.46 (20.60)	59.37 (20.60)	0.244	0.005
Female Patient (mean (sd))	0.54 (0.50)	0.54 (0.50)	0.759	0.001
White Patient (mean (sd))	0.91 (0.28)	0.91 (0.28)	0.517	0.003
Patient ESI (mean (sd))	2.96 (0.54)	2.96 (0.54)	0.760	0.001

Table C4: Pre-Matching Covariate Balance - Lower-Quality Peer

	0	1	p	test SMD
Number of Observations	155705	98217		
ED Volume (mean (sd))	27.57 (13.66)	27.91 (15.32)	<0.001	0.023
Patient Age (mean (sd))	59.38 (20.61)	59.51 (20.57)	0.102	0.007
Female Patient (mean (sd))	0.54 (0.50)	0.54 (0.50)	0.407	0.003
White Patient (mean (sd))	0.91 (0.28)	0.91 (0.28)	0.480	0.003
Patient ESI (mean (sd))	2.96 (0.54)	2.96 (0.54)	0.949	<0.001

Table C5: Post-Matching Covariate Balance - Faster Peer

	0	1	p	test SMD
Number of Observations	60782	60782		
ED Volume (mean (sd))	27.61 (12.62)	27.62 (12.63)	0.973	<0.001
Patient Age (mean (sd))	59.32 (20.57)	59.31 (20.57)	0.991	<0.001
Female Patient (mean (sd))	0.54 (0.50)	0.54 (0.50)	1.000	<0.001
White Patient (mean (sd))	0.92 (0.27)	0.92 (0.27)	1.000	<0.001
Patient ESI (mean (sd))	3.01 (0.54)	3.01 (0.54)	1.000	<0.001

Table C6: Post-Matching Covariate Balance - Slower Peer

	0	1	p	test SMD
Number of Observations	105390	105390		
ED Volume (mean (sd))	26.90 (13.09)	26.94 (15.63)	0.590	0.002
Patient Age (mean (sd))	59.29 (20.40)	59.26 (20.65)	0.805	0.001
Female Patient (mean (sd))	0.54 (0.50)	0.54 (0.50)	0.910	<0.001
White Patient (mean (sd))	0.91 (0.28)	0.91 (0.28)	0.907	0.001
Patient ESI (mean (sd))	2.96 (0.55)	2.96 (0.55)	0.682	0.002

Table C7: Post-Matching Covariate Balance - Higher-Quality Peer

	0	1	p	test SMD
Number of Observations	100461	100461		
ED Volume (mean (sd))	27.54 (13.01)	27.51 (12.89)	0.621	0.002
Patient Age (mean (sd))	59.45 (20.59)	59.37 (20.59)	0.356	0.004
Female Patient (mean (sd))	0.54 (0.50)	0.54 (0.50)	0.907	0.001
White Patient (mean (sd))	0.91 (0.28)	0.91 (0.28)	0.937	<0.001
Patient ESI (mean (sd))	2.96 (0.54)	2.96 (0.54)	0.155	0.006

Table C8: Post-Matching Covariate Balance - Lower-Quality Peer

	0	1	p	test SMD
Number of Observations	98217	98217		
ED Volume (mean (sd))	27.73 (13.78)	27.83 (14.01)	0.113	0.007
Patient Age (mean (sd))	59.44 (20.57)	59.51 (20.57)	0.468	0.003
Female Patient (mean (sd))	0.54 (0.50)	0.54 (0.50)	0.860	0.001
White Patient (mean (sd))	0.91 (0.28)	0.91 (0.28)	0.949	<0.001
Patient ESI (mean (sd))	2.96 (0.55)	2.96 (0.54)	0.346	0.004

Appendix D: Covariate Balance Tables - Chapter 3

Table D1: Pre-Matching Covariate Balance - Speed

	0	1	p	test SMD
Number of Observations	459	618		
Patient Age Shift 1 (mean (sd))	61.08 (3.56)	61.12 (3.60)	0.867	0.010
Patient Gender Shift 1 (mean (sd))	0.54 (0.07)	0.53 (0.09)	0.777	0.018
Patient Race Shift 1 (mean (sd))	0.92 (0.05)	0.92 (0.05)	0.603	0.032
Patient ESI Shift 1 (mean (sd))	2.94 (0.10)	2.94 (0.11)	0.587	0.034
ED Volume Shift 1 (mean (sd))	28.29 (7.51)	25.72 (6.94)	<0.001	0.355
Shift 1 Preference (mean (sd))	0.55 (0.04)	0.55 (0.03)	0.188	0.080
Patient Age Shift 2 (mean (sd))	57.00 (5.50)	56.82 (5.29)	0.577	0.034
Patient Gender Shift 2 (mean (sd))	0.56 (0.12)	0.54 (0.13)	0.063	0.115
Patient Race Shift 2 (mean (sd))	0.91 (0.07)	0.91 (0.07)	0.954	0.004
Patient ESI Shift 2 (mean (sd))	2.92 (0.13)	2.93 (0.14)	0.501	0.042
ED Volume Shift 2 (mean (sd))	22.29 (6.21)	20.74 (6.20)	<0.001	0.250
Shift 2 Preference (mean (sd))	0.25 (0.02)	0.25 (0.02)	0.339	0.059
Patient Age Shift 3 (mean (sd))	55.85 (12.30)	55.86 (13.66)	0.983	0.001
Patient Gender Shift 3 (mean (sd))	0.49 (0.31)	0.51 (0.32)	0.260	0.070
Patient Race Shift 3 (mean (sd))	0.90 (0.18)	0.91 (0.17)	0.192	0.080
Patient ESI Shift 3 (mean (sd))	2.92 (0.32)	2.92 (0.33)	0.898	0.008
ED Volume Shift 3 (mean (sd))	11.60 (7.71)	11.42 (9.98)	0.756	0.020
Shift 3 Preference (mean (sd))	0.23 (0.07)	0.23 (0.07)	0.460	0.046

Table D2: Pre-Matching Covariate Balance - Quality

	0	1	p	test SMD
Number of Observations	867	210		
Patient Age Shift 1 (mean (sd))	61.06 (3.60)	61.30 (3.49)	0.383	0.068
Patient Gender Shift 1 (mean (sd))	0.54 (0.08)	0.53 (0.08)	0.232	0.093
Patient Race Shift 1 (mean (sd))	0.92 (0.05)	0.91 (0.05)	0.033	0.164
Patient ESI Shift 1 (mean (sd))	2.94 (0.10)	2.95 (0.10)	0.654	0.034
ED Volume Shift 1 (mean (sd))	26.40 (7.04)	28.53 (8.08)	<0.001	0.280
Shift 1 Preference (mean (sd))	0.54 (0.03)	0.56 (0.04)	<0.001	0.527
Patient Age Shift 2 (mean (sd))	56.97 (5.27)	56.60 (5.82)	0.368	0.067
Patient Gender Shift 2 (mean (sd))	0.55 (0.12)	0.54 (0.13)	0.628	0.037
Patient Race Shift 2 (mean (sd))	0.91 (0.07)	0.90 (0.07)	0.147	0.111
Patient ESI Shift 2 (mean (sd))	2.92 (0.14)	2.92 (0.13)	0.954	0.005
ED Volume Shift 2 (mean (sd))	21.29 (6.12)	21.83 (6.76)	0.264	0.083
Shift 2 Preference (mean (sd))	0.25 (0.02)	0.24 (0.02)	0.005	0.201
Patient Age Shift 3 (mean (sd))	56.21 (13.03)	54.37 (13.27)	0.067	0.140
Patient Gender Shift 3 (mean (sd))	0.49 (0.31)	0.55 (0.33)	0.021	0.175
Patient Race Shift 3 (mean (sd))	0.91 (0.17)	0.90 (0.19)	0.464	0.055
Patient ESI Shift 3 (mean (sd))	2.91 (0.32)	2.93 (0.34)	0.469	0.055
ED Volume Shift 3 (mean (sd))	11.55 (9.44)	11.29 (7.41)	0.708	0.031
Shift 3 Preference (mean (sd))	0.23 (0.07)	0.22 (0.06)	0.123	0.123

Table D3: Pre-Matching Covariate Balance - Admission Rate

	0	1	p	test SMD
Number of Observations	483	594		
Patient Age Shift 1 (mean (sd))	61.20 (3.50)	61.02 (3.65)	0.431	0.048
Patient Gender Shift 1 (mean (sd))	0.53 (0.08)	0.54 (0.08)	0.298	0.064
Patient Race Shift 1 (mean (sd))	0.92 (0.05)	0.92 (0.05)	0.456	0.046
Patient ESI Shift 1 (mean (sd))	2.94 (0.10)	2.95 (0.10)	0.253	0.070
ED Volume Shift 1 (mean (sd))	26.56 (7.06)	27.03 (7.48)	0.292	0.065
Shift 1 Preference (mean (sd))	0.55 (0.03)	0.55 (0.04)	0.032	0.133
Patient Age Shift 2 (mean (sd))	56.91 (5.26)	56.88 (5.48)	0.928	0.006
Patient Gender Shift 2 (mean (sd))	0.55 (0.12)	0.55 (0.13)	0.770	0.018
Patient Race Shift 2 (mean (sd))	0.91 (0.07)	0.91 (0.07)	0.573	0.035
Patient ESI Shift 2 (mean (sd))	2.92 (0.14)	2.93 (0.14)	0.179	0.082
ED Volume Shift 2 (mean (sd))	21.27 (6.23)	21.50 (6.27)	0.549	0.037
Shift 2 Preference (mean (sd))	0.25 (0.02)	0.25 (0.02)	0.159	0.087
Patient Age Shift 3 (mean (sd))	55.93 (12.95)	55.80 (13.21)	0.874	0.010
Patient Gender Shift 3 (mean (sd))	0.50 (0.32)	0.51 (0.31)	0.842	0.012
Patient Race Shift 3 (mean (sd))	0.91 (0.17)	0.90 (0.19)	0.305	0.063
Patient ESI Shift 3 (mean (sd))	2.91 (0.33)	2.92 (0.33)	0.526	0.039
ED Volume Shift 3 (mean (sd))	11.68 (10.24)	11.35 (8.01)	0.548	0.036
Shift 3 Preference (mean (sd))	0.23 (0.06)	0.23 (0.07)	0.553	0.036

Table D4: Post-Matching Covariate Balance - Speed

	0	1	p	test SMD
Number of Observations	618	618		
Patient Age Shift 1 (mean (sd))	60.88 (3.10)	61.12 (3.60)	0.214	0.071
Patient Gender Shift 1 (mean (sd))	0.53 (0.06)	0.53 (0.09)	0.394	0.048
Patient Race Shift 1 (mean (sd))	0.92 (0.04)	0.92 (0.05)	0.456	0.042
Patient ESI Shift 1 (mean (sd))	2.95 (0.08)	2.94 (0.11)	0.139	0.084
ED Volume Shift 1 (mean (sd))	27.04 (6.23)	25.72 (6.94)	<0.001	0.200
Shift 1 Preference (mean (sd))	0.55 (0.03)	0.55 (0.03)	0.665	0.025
Patient Age Shift 2 (mean (sd))	56.71 (4.13)	56.82 (5.29)	0.678	0.024
Patient Gender Shift 2 (mean (sd))	0.55 (0.10)	0.54 (0.13)	0.134	0.085
Patient Race Shift 2 (mean (sd))	0.91 (0.06)	0.91 (0.07)	0.835	0.012
Patient ESI Shift 2 (mean (sd))	2.93 (0.10)	2.93 (0.14)	0.996	<0.001
ED Volume Shift 2 (mean (sd))	21.12 (4.84)	20.74 (6.20)	0.229	0.068
Shift 2 Preference (mean (sd))	0.25 (0.02)	0.25 (0.02)	0.630	0.027
Patient Age Shift 3 (mean (sd))	55.12 (10.60)	55.86 (13.66)	0.285	0.061
Patient Gender Shift 3 (mean (sd))	0.51 (0.26)	0.51 (0.32)	0.942	0.004
Patient Race Shift 3 (mean (sd))	0.91 (0.16)	0.91 (0.17)	0.957	0.003
Patient ESI Shift 3 (mean (sd))	2.93 (0.23)	2.92 (0.33)	0.654	0.026
ED Volume Shift 3 (mean (sd))	11.31 (6.83)	11.42 (9.98)	0.807	0.014
Shift 3 Preference (mean (sd))	0.22 (0.05)	0.23 (0.07)	0.104	0.093

Table D5: Post-Matching Covariate Balance - Quality

	0	1	p	test SMD
Number of Observations	210	210		
Patient Age Shift 1 (mean (sd))	60.99 (3.05)	61.30 (3.49)	0.346	0.092
Patient Gender Shift 1 (mean (sd))	0.53 (0.06)	0.53 (0.08)	0.759	0.030
Patient Race Shift 1 (mean (sd))	0.92 (0.04)	0.91 (0.05)	0.532	0.061
Patient ESI Shift 1 (mean (sd))	2.95 (0.09)	2.95 (0.10)	0.888	0.014
ED Volume Shift 1 (mean (sd))	27.69 (6.68)	28.53 (8.08)	0.251	0.112
Shift 1 Preference (mean (sd))	0.56 (0.03)	0.56 (0.04)	0.053	0.189
Patient Age Shift 2 (mean (sd))	56.48 (4.59)	56.60 (5.82)	0.815	0.023
Patient Gender Shift 2 (mean (sd))	0.55 (0.10)	0.54 (0.13)	0.380	0.086
Patient Race Shift 2 (mean (sd))	0.91 (0.06)	0.90 (0.07)	0.168	0.135
Patient ESI Shift 2 (mean (sd))	2.91 (0.11)	2.92 (0.13)	0.450	0.074
ED Volume Shift 2 (mean (sd))	22.21 (5.72)	21.83 (6.76)	0.529	0.062
Shift 2 Preference (mean (sd))	0.24 (0.02)	0.24 (0.02)	0.365	0.089
Patient Age Shift 3 (mean (sd))	55.43 (10.44)	54.37 (13.27)	0.364	0.089
Patient Gender Shift 3 (mean (sd))	0.53 (0.27)	0.55 (0.33)	0.430	0.077
Patient Race Shift 3 (mean (sd))	0.89 (0.17)	0.90 (0.19)	0.912	0.011
Patient ESI Shift 3 (mean (sd))	2.91 (0.27)	2.93 (0.34)	0.468	0.071
ED Volume Shift 3 (mean (sd))	12.06 (6.31)	11.29 (7.41)	0.249	0.113
Shift 3 Preference (mean (sd))	0.22 (0.05)	0.22 (0.06)	0.386	0.085

Table D6: Post-Matching Covariate Balance - Admission Rate

	0	1	p	test SMD
Number of Observations	594	594		
Patient Age Shift 1 (mean (sd))	60.97 (3.06)	61.02 (3.65)	0.778	0.016
Patient Gender Shift 1 (mean (sd))	0.53 (0.07)	0.54 (0.08)	0.187	0.077
Patient Race Shift 1 (mean (sd))	0.92 (0.04)	0.92 (0.05)	0.996	<0.001
Patient ESI Shift 1 (mean (sd))	2.94 (0.09)	2.95 (0.10)	0.482	0.041
ED Volume Shift 1 (mean (sd))	27.09 (6.17)	27.03 (7.48)	0.880	0.009
Shift 1 Preference (mean (sd))	0.55 (0.03)	0.55 (0.04)	0.021	0.135
Patient Age Shift 2 (mean (sd))	56.81 (4.36)	56.88 (5.48)	0.800	0.015
Patient Gender Shift 2 (mean (sd))	0.55 (0.10)	0.55 (0.13)	0.623	0.029
Patient Race Shift 2 (mean (sd))	0.91 (0.06)	0.91 (0.07)	0.279	0.063
Patient ESI Shift 2 (mean (sd))	2.93 (0.10)	2.93 (0.14)	0.968	0.002
ED Volume Shift 2 (mean (sd))	21.36 (5.50)	21.50 (6.27)	0.683	0.024
Shift 2 Preference (mean (sd))	0.25 (0.02)	0.25 (0.02)	0.880	0.009
Patient Age Shift 3 (mean (sd))	56.87 (10.75)	55.80 (13.21)	0.124	0.089
Patient Gender Shift 3 (mean (sd))	0.50 (0.28)	0.51 (0.31)	0.570	0.033
Patient Race Shift 3 (mean (sd))	0.91 (0.16)	0.90 (0.19)	0.456	0.043
Patient ESI Shift 3 (mean (sd))	2.91 (0.24)	2.92 (0.33)	0.510	0.038
ED Volume Shift 3 (mean (sd))	11.55 (6.28)	11.35 (8.01)	0.634	0.028
Shift 3 Preference (mean (sd))	0.23 (0.06)	0.23 (0.07)	0.686	0.023