Emergency Department Management:
Data Analytics for Improving Productivity and Patient Experience

by

Krista M. Foster

B.S., B.A., Youngstown State University, 2009

M.S., The Pennsylvania State University, 2011

Submitted to the Graduate Faculty of the

Joseph M. Katz Graduate School of Business in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

JOSEPH M. KATZ GRADUATE SCHOOL OF BUSINESS

This dissertation was presented

by

Krista M. Foster

It was defended on

June 26, 2020

and approved by

Jennifer Shang, Professor of Business Administration, Katz Graduate School of Business

Sunder Kekre, Professor of Operations Management, Tepper School of Business

Arvind Venkat, MD, National Director of Research, US Acute Care Solutions, Vice Chair for Research and Faculty Academic Affairs, Department of Emergency Medicine, Allegheny Health Network, and Professor of Emergency Medicine, Drexel University College of Medicine

G. G. Hegde, Associate Professor of Business Administration, Katz Graduate School of Business

Leon Valdes, Assistant Professor of Business Administration, Katz Graduate School of Business

Dissertation Chair: Jennifer Shang, Professor of Business Administration, Katz Graduate School of Business

Emergency Department Management:
Data Analytics for Improving Productivity and Patient Experience

Krista M. Foster, PhD

University of Pittsburgh, 2020

The onset of big data, typically defined by its volume, velocity, and variety, is transforming the healthcare industry. This research utilizes data corresponding to over 23 million emergency department (ED) visits between January 2010 and December 2017 which were treated by physicians and advanced practice providers from a large national emergency physician group. This group has provided ED services to health systems for several years, and each essay aims to address operational challenges faced by this group's management team.

The first essay focuses on physician performance. We question how to evaluate performance across multiple sites and work to understand the relationships between patient flow, patient complexity, and patient experience. Specifically, an evaluation system to assess physician performance across multiple facilities is proposed, the relationship between productivity and patient experience scores is explored, and the drivers of patient flow and complexity are simultaneously identified.

The second essay explores the relationship between physician performance and malpractice claims as we investigate whether physicians' practice patterns change after they are named in a malpractice lawsuit. Overall, the results of this analysis indicate that the likelihood of being named in a malpractice claim is largely a function of how long a physician has practiced. Furthermore, physician practice patterns remain consistent after a physician is sued, but patient experience

scores increase among sued physicians after the lawsuit is filed. Such insights are beneficial for management as they address the issue of medical malpractice claims.

The final essay takes a closer look at the relationship between advanced practice providers (APPs) and physicians. Can EDs better utilize APPs to reduce waiting times and improve patient flow? A systematic data-driven approach which incorporates descriptive, predictive, and prescriptive analyses is employed to provide recommendations for ED provider staffing practices.

Table of Contents

List of Tables

List of Figures

Preface

I would like to acknowledge the many people who have encouraged and supported me throughout the process of completing my dissertation and PhD. First, I would like to thank my adviser Jennifer Shang for both her personal and professional support. I am so appreciative of her mentorship and kindness and know that I would not have succeeded without her guidance and encouragement. At times, she had more confidence in me than I had in myself, and I will forever be grateful to her. I am also thankful for the guidance I received from the other members of my committee: Arvind Venkat provided countless hours of help in ensuring the practicality of all models included in my dissertation. Sunder Kekre shared his knowledge and expertise with me and helped me to position my dissertation essays in the operations management literature. G. G. Hegde was a constant source of encouragement and positivity, and Leon Valdes provided such helpful advice regarding the job market and transitioning to faculty. I am blessed to have such a supportive committee, and I value their continued advice and mentorship.

I would like to thank Carrie Woods, whose kindness has meant so much to me. From my first visit to campus, she has provided personal, professional, and emotional support, and she always goes above and beyond to help students. I am also grateful for to all of the occupants of Mervis 241 over the past several years for their camaraderie and for the advice from senior students.

Finally, I would not have been able to complete this without the unconditional love and support of my family, especially my parents, Josie and Gary Foster, who have always encouraged me. I love you, and I dedicate this dissertation to you.

1.0 Introduction

The onset of big data, typically defined by its volume, velocity, and variety, is transforming the healthcare industry. Massive investments in electronic health records and advances in technology such as smartphones and wearable medical devices have given rise to both structured and unstructured healthcare data, which are continuously changing. Such innovations have created numerous opportunities for both researchers and practitioners, and rich data is being collected and utilized by healthcare providers, pharmaceutical companies, and insurance companies. For instance, several organizations have launched big data initiatives in healthcare, ranging from efforts to improve medical treatments to personalized medicine (Nambiar et al. 2013). In many cases, the goal is patient-centric healthcare (Sonnati 2015). Rising healthcare costs have motivated other projects as well. Improvements in record-keeping have helped to reduce healthcare costs due to fraud, abuse, waste, and erroneous insurance claims (Srinivasan & Arunasalam 2013). As healthcare systems aim to balance the patient experience (or patient satisfaction) with costs and revenues, data from electronic health records and RFID tags have become invaluable. The increasing availability of data relating to both patients and physicians provides an opportunity to reevaluate the methods used to measure and assess physician performance and practice patterns.

This research was motivated by the common dilemmas facing executives in managing consolidated multi-facility emergency physician management networks (EPMNs). This group of emergency physicians (EPs) has provided emergency department (ED) services to both single- and multi-hospital health systems for several years. The increasing consolidation in staffing emergency departments is both an adaptation to the changing healthcare landscape and a mechanism to remain competitive in an era of shrinking profit margins in providing patient care.

Previously, single hospitals would either employ directly or contract with a small group of emergency physicians to provide patient care services in their ED. Natural growth might lead this hospital to partner with one or more local hospitals and therefore hire more emergency physicians or ask their contracting local group to do the same.

With rapid consolidation of hospitals into healthcare networks to achieve economies of scale, the nature of staffing EDs has changed (see Fig. 1). These changes have resulted in horizontal integration in terms of staffing as well as risk-pooling. Many healthcare systems today have multiple hospitals dispersed geographically (similar to a distributed supply chain), have varying profit margins, and are of different appeal to an already undersized (short supply) emergency physician workforce (Reiter et al. 2016). To address these challenges, healthcare systems often seek large EPMNs to manage their ED providers and assume the financial responsibility of their emergency physicians' compensation.

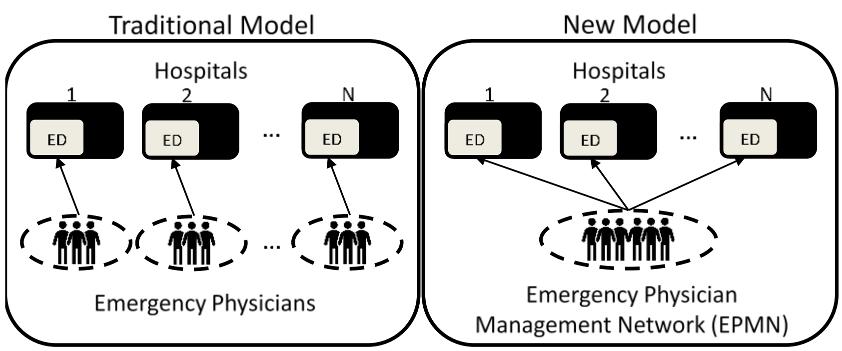


Figure 1 Comparison of ED Staffing Models

Multi-hospital ED consolidation through EPMNs helps aggregate resources and provides adequate capacity to meet this need. However, as the demand for physician services is growing faster than supply, EPMNs are constantly under pressure to maintain physician supply to meet patient demand. The data used for this research was collected by a physician-owned EPMN that served about six million patients per year at over 170 sites in 21 states as of 2017 and continues to grow. The EPMN under study has surveyed and documented emergency medicine (EM) trends and observed noticeable increases in the number of EDs managed by physician networks. Specifically, nearly 40% of all EDs in the US are currently managed by EPMNs. These complex networks constantly strive to balance the demands of three key stakeholders: the company, the facilities and their patients, and the physicians. While the company aims to acquire new contracts with facilities and attract physicians, every facility must be staffed, and physicians desire to be paid competitively. As EPMNs normally contract with multi-facility healthcare systems, management constantly faces challenges when staffing EDs in hospitals of various performance levels (e.g., productivity and patient experience), while under the pressure to ensure all physicians are evaluated and compensated fairly. With continuing healthcare reform efforts and the current shortage of EPs, the EPMN under study is constantly seeking methods to incentivize and retain current physicians, attract new physicians, and engage in continuous improvement to enhance ED performance.

Each of the essays herein rely on data from this large physician-owned EPMN which is described in detail in Section 2. Section 3 addresses physician performance as we ask how to evaluate performance across multiple sites and work to understand the relationships between patient flow, patient complexity, and patient experience (Foster et al. 2018). Specifically, an evaluation system to assess physician performance across multiple facilities is proposed, the relationship between productivity and patient experience scores is explored, and the drivers of patient flow and complexity are simultaneously identified. A secondary result of this analysis reveals that the support of advanced practice providers (APPs) such as physician assistants has a direct positive impact on patient flow.

Section 4 focuses on the relationship between physician performance and malpractice claims (Carlson et al. 2018) and investigates whether physicians' practice patterns change after they are named in a malpractice lawsuit (Carlson et al. 2020). Overall, the results of this analysis indicate that the likelihood of being named in a malpractice claim is largely a function of how long a physician has practiced. Furthermore, physician practice patterns remain consistent after a physician is sued, but patient experience scores increase among sued physicians after the lawsuit is filed. Such insights are beneficial for management as they address the issue of medical malpractice claims.

Section 5 takes a closer look at the relationship between APPs and physicians. Can EDs better utilize APPs to reduce waiting times and improve patient flow? A systematic data-driven approach which incorporates descriptive, predictive, and prescriptive analyses is employed to provide recommendations for ED provider staffing practices. Finally, we discuss directions for future research.

2.0 Data Description

The data used for all analyses are proprietary data maintained by an EPMN with which we have collaborated. This company contracts with hospitals and healthcare systems across the United States to manage emergency departments and staff emergency medicine providers. They maintain data corresponding to ED visits, physicians' demographics, physicians' clinical hours, patient experience surveys, and malpractice lawsuits. Visit characteristics, including Current Procedural Terminology Evaluation and Management (CPT E & M) codes and relative value units (RVUs) generated, were abstracted by trained billing specialists. During this period, billing specialists were required to have or acquire relevant certification(s) between their second and third employment year, with ongoing training, auditing, and external evaluation. The group also maintains a demographic and credentialing database of all physicians. Physicians' clinical hours were tracked electronically (Tangier; Sparks, MD), while patient experience data (Press Ganey Associates Inc., South Bend, IN) were linked to physicians monthly. This physician group also maintained its own risk-retention program that recorded all malpractice claims during the study period. Because hospital contracts can change over time, the number of facilities, and thus the number of physicians, varied from month to month. Tables 1 – 3 provide characteristics of these data between January 2010 and December 2017. The complete database contains detailed data corresponding to over 23 million ED visits between January 2010 and December 2017. In each of the essays herein, we define a subset of the available data to use for the analysis.

Table 1 Facility Characteristics by Year

		Year							
		2010	2011	2012	2013	2014	2015	2016	2017
Yearly Facility Counts		59	63	59	65	73	80	123	151
Full Year		54	51	52	48	53	61	73	115
Partial Year		5	12	7	17	20	19	50	36
by Annual Visits (Full Year Only)	<20,000	9	5	6	6	8	12	18	39
	20,000-39,999	17	18	18	17	18	22	26	40
	40,000-59,999	19	17	17	12	14	14	16	19
	60,000-79,999	6	8	5	7	9	10	10	13
	80,0000+	3	3	6	6	4	3	3	4
by Facility Type	Hospital ED	Not recorded prior to 2014				60	63	87	100
	Other					13	17	36	51
by EM Residency	Yes	9	9	10	10	9	9	9	9
	No	50	54	49	55	64	71	114	142
by Teaching Hospital	Yes	12	12	13	13	14	14	14	14
	No	47	51	46	52	59	66	109	137
by Trauma Level	1	2	2	2	2	3	3	4	5
	2	7	8	9	9	10	10	12	16
	3	7	7	7	7	4	2	4	5
	4	2	2	1	1	1	2	1	1
	None	41	44	40	46	55	63	102	124
by State	AZ	3	4	3	3	3	3	1	0
	CA	10	10	7	4	4	4	2	2
	CO	0	0	0	0	0	0	13	18
	CT	3	5	5	5	5	5	6	8
	FL	0	0	0	0	0	0	10	10
	HI	1	1	1	1	2	2	2	2
	IL	4	5	5	5	5	5	6	8
	KS	0	0	0	0	0	0	0	1
	KY	0	0	0	0	0	1	1	1
	MD	0	0	0	0	0	0	14	18
	MI	0	0	2	2	2	2	2	2
	NC	11	11	11	10	11	10	14	14
	NH	0	0	0	1	1	1	1	2
	NV	7	7	7	7	3	3	3	7
	NY	4	4	4	5	5	5	6	5
	OH	9	8	7	11	11	18	17	23
	OK	1	1	1	2	3	3	4	4
	PA	4	5	5	7	16	16	19	22
	RI	0	0	0	1	1	1	1	1
	VA	0	0	0	0	0	0	0	2
	WV	2	2	1	1	1	1	1	1

Table 2 Provider Characteristics by Year

		Year							
		2010	2011	2012	2013	2014	2015	2016	2017
Physician	Physicians	775	865	874	863	897	1,001	1,412	1,752
	Clinical Shifts	99,018	105,794	110,668	105,195	111,455	114,683	145,811	178,398
	Gender Male	542	606	601	582	597	674	948	1,159
	(%)	(69.9%)	(70.1%)	(68.8%)	(67.4%)	(66.6%)	(67.3%)	(67.1%)	(66.1%)
	Race White	609	674	689	681	705	779	944	1,053
	(%)	(78.6%)	(77.9%)	(78.8%)	(78.9%)	(78.6%)	(77.8%)	(66.9%)	(60.1%)
APP	APPs	263	316	342	376	421	521	841	1,077
	Clinical Shifts	24,651	29,213	33,039	35,803	48,577	50,333	82,374	112,646
	Gender Male	APP demographics not recorded prior to 2014				146	172	290	369
	(%)					(34.7%)	(33.0%)	(34.5%)	(34.3%)
	Race White					353	436	623	724
	(%)					(83.8%)	(83.7%)	(74.1%)	(67.2%)
	Physician Assistants					336	405	636	799
	(%)					(79.8%)	(77.7%)	(75.6%)	(74.2%)

Table 3 Visit Characteristics by Year

		Year							
		2010	2011	2012	2013	2014	2015	2016	2017
Number of Visits		2,266,635	2,480,776	2,580,737	2,438,166	2,626,216	2,765,378	3,446,818	4,459,566
% of Visits by Gender	Male	1,002,799	1,099,066	1,140,692	1,080,927	1,165,190	1,230,730	1,527,758	1,985,573
	Female	1,263,823	1,381,703	1,440,021	1,357,228	1,460,859	1,534,481	1,918,106	2,472,684
	Other	13	7	24	11	167	167	954	1,309
% of Visits by Disposition	Admitted	378,266	417,356	453,757	426,729	454,535	475,077	584,500	776,402
	Discharged	1,740,160	1,904,561	1,982,728	1,881,429	2,027,765	2,142,421	2,650,452	3,402,832
	Against Medical Advice	21,557	24,783	29,595	24,586	24,820	24,295	34,303	49,067
	Left Without Treatment	51,898	56,906	54,183	46,248	52,745	58,465	70,530	80,353
	Transferred	27,633	31,077	31,337	33,249	38,052	39,428	45,765	62,609
	ED Death/ DOA	3,465	3,572	3,742	3,511	3,671	3,699	4,826	6,009
	Other/ Unknown	43,656	42,521	25,395	22,414	24,628	21,993	56,442	82,294
% of Visits by Provider Type	Physician	Not recorded prior to 2014				2,166,356	2,214,064	2,660,844	3,199,941
	APP					406,237	505,025	741,258	1,136,434
	Both					46,815	46,263	44,270	43,942
	None					6,808	26	446	79,249
Median Patient Age		35	36	36	37	37	37	38	40
Median Patient Length of Stay (hours)		2.60	2.70	2.75	2.77	2.77	2.72	2.78	2.92
Median Patient Length of Stay (Admitted)		4.67	4.87	4.95	5.05	4.97	4.87	4.80	4.90
Median Patient Length of Stay (Discharged)		2.32	2.40	2.40	2.42	2.42	2.38	2.47	2.57
Median Patient RVUs		3.21	3.40	3.37	3.37	3.30	3.33	3.32	3.32

3.0 Understanding Variability in Physician Practice Patterns: Key Performance Indicators in Emergency Physician Management Networks

3.1 Motivation

> *"The director of an emergency medicine group is struggling. The group of emergency physicians has provided emergency department (ED) services to a five-hospital health system for the past ten years. Given new legislation, the health system is seeking opportunities to maximize ED productivity and minimize costs. Simultaneously, the emergency physicians in the group are in revolt. They feel that based on the particular hospitals in which they work, their productivity-based compensation varies drastically – driven by factors out of their control such as location, services available at each hospital, and the scheduling idiosyncrasies of each hospital. The emergency physicians also feel that the health system is imposing unrealistic productivity and patient experience goals upon the group without any objective basis for measuring their performance against peers. The director must address the demands from the health system to show where productivity can be improved and from the physicians to measure their productivity objectively. Now the director is at risk for seeing physicians leave, losing the contract with the health system or both. Is there a better way?"*

EPMNs often contract with multi-facility healthcare systems, and the need to staff EDs in hospitals of various performance levels (e.g., productivity and patient experience) introduces challenges for management. Despite varying levels of profitability under different hospitals, the administrators need to hire quality physicians to work, especially in less appealing ED facilities and locations. They also need to ensure physicians are fairly assessed and compensated. Herein lies a typical value chain challenge of maintaining a sufficient level of resources (quality physicians) and efficiently managing the ED's operations where care must be available at all times to meet patient demand under budget constraints. There is an earnest need to explore this evolving industry and identify effective ways to deliver care effectively and efficiently.

In order to maintain contracts with healthcare systems, efficiently manage EDs, and ensure physicians' continued dedication to their practice, EPMNs must focus on both clinical and operational performance. Clinical performance (e.g. clinical outcomes, unexpected return visits, medico-legal risk) depends heavily on the quality of the physicians hired and available risk management and clinical education programs. Therefore, clinical outcomes are largely subsumed by hiring practices, which are beyond the scope of this research. However, operational metrics are important to maintaining health system contracts and physician investment in the EPMN, and this is an area that is largely unexplored in the emergency medicine and related operations literature. How to fairly and effectively evaluate and incentivize EPs while improving productivity and patient experience has become a pressing issue for the executives. Thus, management is in search of a method that would equitably assess physicians within the EPMN based on their productivity (objective score) and patient experience (subjective score) in order to balance performance and customer satisfaction, both of which are key to maintaining health system contracts for ED services. Their relative standings among physicians in the network will help management to make justifiable decisions on performance-based compensation and training requirements.

As variation in productivity has implications on ED patient flow and waiting times, objective measurement of EP performance is important, particularly when it comes to its consequent impact on ED performance. Researchers studying EP productivity, however, have mostly focused on a single facility. In this research, we address the emerging trend that emergency physicians work in multiple facilities. We conduct a multi-center, multi-year and multi-physician study to investigate EP operational performance in the EPMN setting to draw lessons on how to measure and enhance physician productivity and patient experience. For such a study, we make use of EP profiles, daily schedules, monthly patient experience surveys, patients' visit details, and patients' insurance information to better understand the performance of physicians in this network using big data analytics.

EPs encounter a number of work conditions that affect their performances. These circumstances make the revenue generated by an EP less relevant in performance evaluation. For instance, if patients are uninsured, they cannot pay for treatment, and different insurances pay different amounts for similar services, completely out of the treating physician's control. Moreover, an EP's output flow rate (Patients/hr) is highly dependent on patient acuity and chronic medical conditions, the patient arrival rate, and the ED's capacity. As an example, an elderly adult patient with chest pain and several comorbidities will likely take longer to treat than a young adult with an ankle injury who is otherwise healthy, as the former is much more serious and complex. It would be unfair to penalize physicians who treat complex patients for their resultant lower patient flow rates, or equivalently, to reward physicians who are assigned to simpler cases by relying solely upon the volume-based Patients/hr measure. Relative value units (RVUs) are thus often used to assess emergency physician performance as they mirror the time and supplies/devices needed from the healthcare workers/facility to care for the patient as well as the cognitive expertise and potential medico-legal risk (Venkat et al, 2015, Appendix A.1). RVUs reflect the revenue potential for a particular visit, while the actual revenue generated depends on the specific insurance, location and type of hospital. Conventionally, RVUs/hr is a marker for revenue potential and is a proxy for physician productivity, capturing the yield from both volume and complexity of a physician's efforts. However, one should note that RVUs/hr is not a pure measure of productivity nor is it an absolute measure of revenue.

To date, comparison of EP performance within such a large network has not been available due to the technical difficulties involved (e.g., demand heterogeneity in different facilities, shift variation, and varying availability of diagnostic equipment). RVUs/hr is driven by complexity (RVUs/Patient) and patient flow (Patients/hr) at the facility level, both of which affect an EP's performance. Thus, significant disparity may be found in terms of facility and physician performance. The variation across facilities thereby limits the efficacy of the conventional RVUs/hr measure, making it necessary to take into account these differences while developing fair performance metrics to assess physicians within the EPMN.

In order to overcome the aforementioned difficulties when evaluating EP productivity, we propose four new indices for assessing physician performance relative to peers: *revenue potential index, patient volume index*, *patient complexity index,* and *patient experience index*. We employ the revenue potential index and patient experience index within a large EPMN to differentiate the high performers from those physicians lagging behind on each metric, resulting in a 2-by-2 graph. Then, we segment physicians into clusters to uncover possible physician characteristics affecting their performance. We empirically verify that physician productivity is a function of the proposed complexity and volume indices and subsequently identify drivers of the two proposed indices such that management can help physicians target improvements in both dimensions. Our study highlights that the use of big data analytics to manage complex and large-scale physician groups has major potential for developing and deploying new metrics that are sophisticated, yet relatively straightforward to implement, and offers operational insights for volume and complexity. The proposed metrics are transparent and take into account facility-specific differences. These metrics can be linked to performance-based pay, making them attractive to physicians and therefore mitigating the obstacles management faces in recruiting and assigning physicians to facilities.

3.2 Related Literature

Service outputs can be classified into two components: quantity-oriented (volume) and quality-oriented (process and outcome) (Grönroos and Ojasalo, 2004). In EDs, patients (heterogeneous customers) arrive unannounced and expect high quality care (output) within a short time. The idea of service systems with heterogeneous customers is not new (Armony & Ward 2010; Ward & Armony 2013); and most of these systems focus on throughput (volume). Similarly, in EDs, operational efficiency and patient experience are important to all stakeholders (e.g., patients, physicians, and administrators). Facilities and physicians are judged on patient experience via survey data, such as those from the Press Ganey© (PG) survey (see Appendix A.1). While patient experience has become a focal point of healthcare providers, physician productivity remains crucial from a resource management perspective. Management's ultimate goal is to concurrently achieve high productivity and excellent patient experience. In this research, we aim to address these two aspects of performance while making use of the recent explosion of big data in the healthcare sector.

The impact of big data and analytics within the healthcare industry has been widely recognized by healthcare professionals and beyond. Murdoch and Detsky (2013) outlined ways in which big data may be used to improve the quality and efficiency of healthcare delivery, from better knowledge dissemination to personalized patient care. Raghupathi and Raghupathi (2014) cited specific examples in their review of big data in healthcare, including earlier predictions of sudden increases in flu-related emergency room visits. McKinsey & Company also weighed in on big data's place in the healthcare sector (Kayyali et al. 2013). Both Ellaway et al. (2014) and Moskowitz et al. (2015) discuss the impact of big data on education and training for clinicians and other healthcare professionals. Furthermore, Obermeyer and Emanuel (2016) recommend the development of algorithms to make the best use of healthcare data. Most recently, Ba and Nault (2017) identified healthcare IT that incorporates OM methods as an emerging research area. All of these researchers agree that the availability of big data, including physician work histories and patient records, is changing the management of facilities, physicians, and patients. EDs and physician management networks are no exception. For example, the American College of Emergency Physicians (ACEP) has introduced the Clinical Emergency Data Registry (CEDR), which is designed to collect data from emergency physicians across the US to measure healthcare quality and to identify practice patterns, trends and outcomes in emergency care (ACEP 2016). The ultimate goal of CEDR is to inform emergency physicians and eventually improve the overall quality of emergency care, suggesting that emergency medicine clinicians are embracing data-driven knowledge and decision-making.

Several operations management researchers have also focused on problems in the healthcare sector, leveraging data to motivate research questions and validate models and theories. The single-facility healthcare operations literature abounds, and EDs have been the center of several studies (Powell et al. 2012, Kc 2013; Song et al. 2015; Lee et al. 2015). For instance, Venkat et al. (2015) used ED data to gain insights into its management and revenue generation. However, operations researchers who have studied multiple facilities mainly focus on differences in hospitals and compare their efficiencies (Theokary and Ren 2011; Bhargava & Mishra, 2014; Blank & Eggink, 2014; Büchner et al. 2016). Angst et al. (2011) studied how U.S. hospitals convert existing medical technologies into integrated information technology. No research hitherto has examined the dynamics between physicians and patients, and compared individual practitioners' performances across the multiple EDs each physician serves. Also, no researchers to our knowledge have explored the implications of multi-facility physician management networks.

Different from the emergency medicine literature's physician studies (e.g. Brennan et al. 2007; Johnson et al. 2008; Clinkscales et al. 2016), we look into physician performance with the aim to efficiently manage and incentivize physicians within an EPMN that must staff many EDs and attract large numbers of practicing EPs. Our research differs significantly from both the operations and EM literatures as we consider two key dimensions of physician performance: patient experience and productivity (the latter comprising volume and complexity). Through big data, we are able to offer practice-based evidence to develop insights for multi-facility physician management networks instead of the conventional approach. The proposed metrics take into account physician efforts and peer effects. Our setting is unique as the physicians under study work at multiple sites under the management of an organization, whose goal is to ensure appropriate staffing across multiple facilities under contract for emergency services.

Our research benefits EPMNs, an increasingly prevalent EM practice model. Unlike the frequently used data from the Centers for Medicare and Medicaid Services (CMS), all of the physicians in our data are managed by an EPMN, and the proprietary data uniquely links patient, physician, hospital, revenue, and insurance information. To our knowledge, this is the first research to study physician productivity on this scale and with this level of detail. We contribute to the literature by addressing the trend in which physician work assignments include varying hours across multiple sites and shifts.

By developing relative physician performance indices, we help management understand the impact of multiple-facility employee assignments on both relative productivity and patient experience. The visual display of each physician's two-dimensional scores is conducive to employee development, as it allows management to readily assess physician performance.

3.3 Theoretical Framework

Our research framework is motivated by management's need to explicitly and equitably differentiate strong and weak performers within the network, so as to provide justifiable bonuses or prescribe remedial actions. Through identifying driving forces contributing to physician performance differences, we offer the EPMN management implications for continuous improvement.

We match the 10,615,879 patient visits to physicians and facilities to develop indices that assess physicians within the network (Fig. 2). Although physicians may treat many patients across multiple facilities, the indices quantify performance within the network using a single value. Such metrics allow management to easily identify the best performers to reward with performance-based bonuses, whereas the poor performers would require intervention in the form of training. To gain managerial insights for effective network resource management, we conduct cluster analysis to sort physicians into distinctive groups.

Justification for the proposed indices, including a stylized example to illustrate their benefits, is detailed in §3.4. We then use the indices to develop statistical models and benchmark physician performance (Fig. 3). Our research relates Physician's Revenue Generation Potential with Patient Volume and Patient Complexity, while controlling for patients' medical and demographic information as well as physician and facility characteristics.

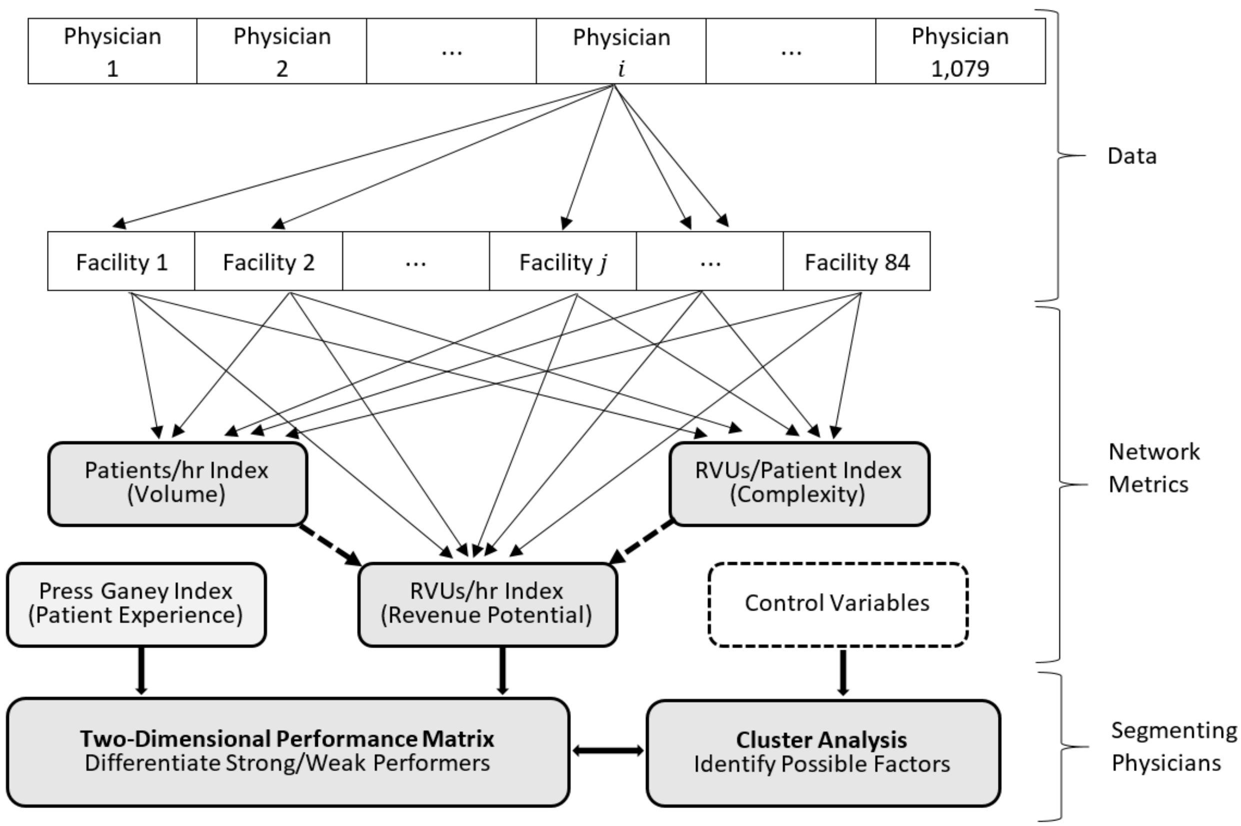


Figure 2 Developing Network Metrics for Physician Performance

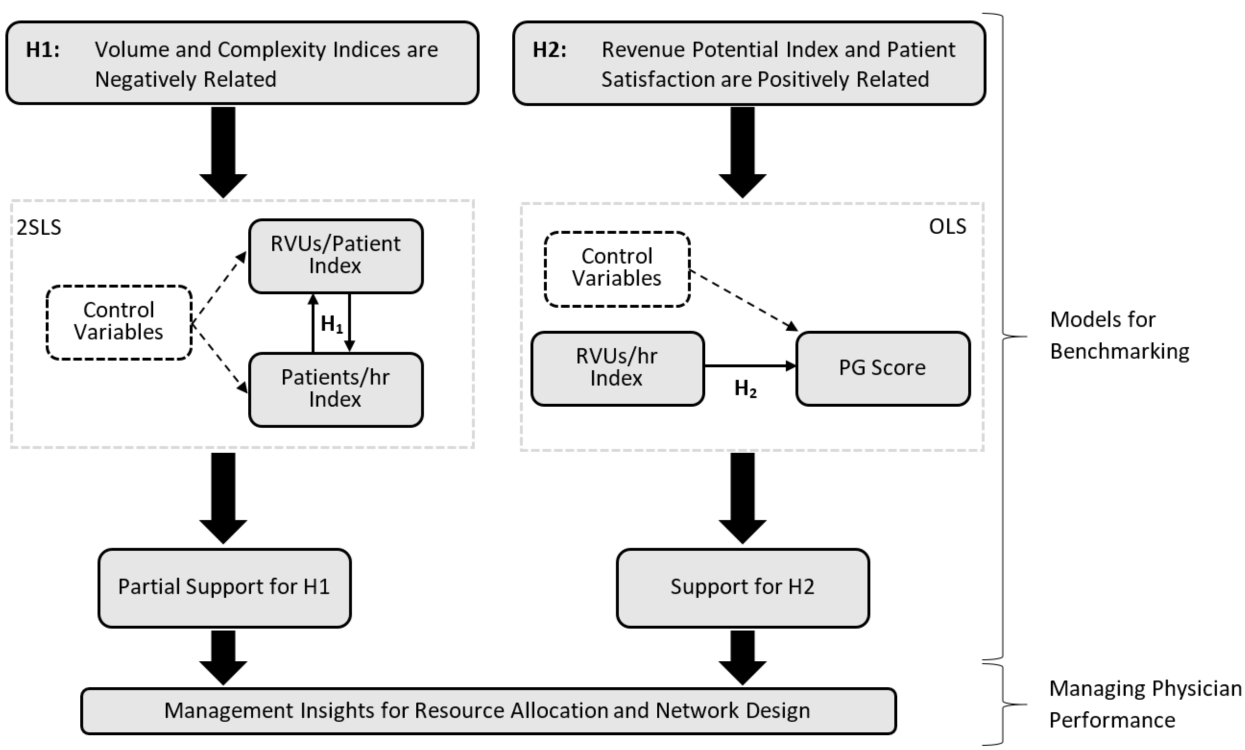


Figure 3 Research Outline

Using the performance indices, we develop a two-stage least squares model to simultaneously reveal the underlying drivers of the volume and complexity indices. Subsequently, we link Physician's Revenue Generation Potential to Patient Experience to better understand how different aspects of performance interact and to empirically test if tradeoffs exist for physicians at the network level. By using these metrics and statistical models, we gain managerial insights for an EPMN.

3.3.1 Factors Impacting Physician Performance

We develop integrated relative performance indices to assess the EPMN's physicians in §3.4, addressing management's need for a fair and coherent network performance metric. In this section, we develop models that relate these indices to physician, patient, and facility traits to better understand the underlying drivers.

3.3.1.1 Identifying the Drivers of Performance Indices and Linking the Indices

In order to identify factors driving physician performance, we employ two-stage least squares (2SLS) regression to simultaneously model the proposed relative performance metrics, RVUs/Patient Index and Patients/hr Index, by Eqs. (1) – (2). The model is estimated using log-transformed values of RVUs/Patient Index and Patients/hr Index to address the nonlinear relationship characterized by diminishing returns.

$$\ln(RVUs/Patient\ Index) = \alpha_{11} + \beta_{11} \ln(Patients/hr\ Index) + \boldsymbol{\gamma_1}^{\boldsymbol{T}} Controls + \varepsilon_1 \tag{1}$$

$$\ln(Patients/hr\ Index) = \alpha_{21} + \beta_{21} \ln(RVUs/Patient\ Index) \tag{2}$$

$$+\boldsymbol{\gamma_2}^{\boldsymbol{T}} Controls + \varepsilon_2$$

While patient complexity and patient flow are related (RVUs/Patient × Patients/hr = RVUs/hr) on a system-wide level, these have not been studied at the physician-level. For example, Eitel et al (2010) found that at the ED-level, the criticality of patients is associated with flow. This relationship may also hold at the physician-level. Based on management's desire to measure physician operational performance on two dimensions (patient satisfaction and productivity), it is important to fully understand both metrics and their drivers (patient complexity and patient flow). As patients with high complexity require more physician time to treat, we posit that an inverse relationship exists between RVUs/Patient Index and Patients/hr Index (see the left-side of Fig. 3). Hypothesis 1 tests such a relationship.

H1. *There exists a negative relationship between the number of patients that a physician treats per hour relative to peers (Patients/hr Index) and the relative number of RVUs the physician generates per patient (RVUs/Patient Index). That is,* $\beta_{11} < 0$ *and* $\beta_{21} < 0$.

Although H1 seems intuitive, the unpredictable circumstances of emergency care may counter-intuitively disrupt the presumed inverse relationship between complexity and volume. In emergency medicine, there are numerous examples where high RVU patients are treated quite efficiently in the ED. The ensuing examples justify the importance of testing H1. Consider a patient with a severe allergic reaction. The patient will be rapidly treated, observed, and commonly discharged, but the visit will likely result in high RVUs due to the critical consequences if this condition is not treated appropriately. In other cases, an EP may be constrained by the environment, such as the specific number and type of arriving patients or the availability of beds for admitted patients. For instance, there may be a high number of patients requiring admission to the hospital during flu season, but most of these patients do not require the resources of an intensive care unit

(ICU). In this case, high RVU, critically ill patients may rapidly be placed in ICUs, while less complex patients requiring admission may wait hours for an inpatient bed. Such circumstances make the exploration of H1 highly relevant to the study of productivity in the emergency department.

3.3.1.2 Linking RVUs/hr Index to Patient Experience Index in the Network

After identifying the drivers of physician performance in Eqs. (1)–(2), we now explore the extent to which physician performance is related to patient experience as measured by the Press Ganey© (PG) patient experience index. We present an OLS model to describe the relationship between RVUs/hr Index and PG Index while controlling for patient and physician level differences in Eq.(3).

$$PG\ Index = \alpha_{31} + \beta_{31}\,(RVUs/Hour\ Index) + \boldsymbol{\gamma_3}^{\boldsymbol{T}} Controls + \varepsilon_3 \qquad (3)$$

As RVUs/hr consists of two components, Patients/hr and RVUs/Patient, which are presumed to be the primary drivers of RVUs/hr Index, we study how each of these components influence patient experience. First, we consider Patients/hr, which reflects the logistics of providing patient services. If an ED visit is extended due to a long wait time or prolonged service, the patient may blame the physician, leading to dissatisfaction. This frustration would be reflected in physician PG survey responses. Such phenomena have been studied in the EM literature. For example, Handel et al (2014) observed that patients with low door-to-room times gave higher experience scores, and Pines et al (2008) found a negative relationship between overcrowding (longer wait times, prolonged treatment times) and patient experience assessment. Furthermore, Hwang et al. (2015) found that implementation of a fast track significantly increased patient experience. Thus, we expect Patients/hr to be positively associated with patient experience.

Alternatively, Boudreaux et al. (2004) found that the strongest predictor of ED patient satisfaction is how satisfied the patient is with interpersonal interactions with the ED staff. We conjecture that patients may view visits with higher RVUs (more tests and resources) as a sign that physicians care and are sincere (positive interactions). Patients may perceive increased numbers of examinations, procedures, or therapeutic interventions, leading to higher RVUs, as signs that they have been taken seriously, and thus RVUs/Patient may reflect a physician's individualized focus toward patients. Thus, RVUs/Patient will also be positively correlated with patient satisfaction, even though RVUs/Patient and Patients/hr are inversely related.

Together, we anticipate that as a physician's productivity (RVUs/hr=RVUs/Patients×Patients/Hr) increases, patients will perceive that the physician is both highly skilled and focused, leading to higher PG percentile rank scores (hereafter abbreviated as PG scores), resulting in a higher PG Index. Boudreaux et al (2006) further provided the theoretical foundation that patient satisfaction is dependent on physician performance. Therefore, as a physician's ability to speed up processing or handle more complex patients is associated with high RVUs/hr Index, we posit that physicians with high RVUs/hr Index will exhibit better PG scores relative to their peers (see the right-side of Fig. 3).

H2. *There exists a positive relationship between the number of RVUs that a physician generates per hour relative to peers (RVUs/hr Index) and the relative patient experience score (PG Index). That is,* $\beta_{31} > 0$.

3.4 Data and Variables

The data under study were collected from one of the largest EM providers in the US. These data are unique as they are from a large private EPMN, which includes numerous physicians across 84 facilities in 14 US states from 2010-2014. The percentage of facilities in each state is shown in Fig. 4. Our data initially contained 90 EM facilities, 1,434 physicians, 622,716 clinical shifts, and 11,060,222 patient visits, comprising visit records, physician profiles, healthcare provider schedule logs, and facility profiles from January 2010 to June 2014. Visit characteristics, including Current Procedural Terminology Evaluation and Management (CPT E/M) codes and relative value units (RVUs) generated, were abstracted by trained coders. The coders need to acquire relevant certification(s) between their second and third employment year, with ongoing training, auditing, and external evaluation. The EPMN also maintains a demographic and credentialing database of all physicians. Physicians' clinical hours were tracked electronically using physician scheduling software (e.g. Tangier; Sparks, MD). Patient experience data (Press Ganey Associates Inc., South Bend, IN) were linked to physicians monthly.

These data are stored in various large databases, four of which are used in this research. The digitally recorded databases include a variety of data types (numeric, factors, strings, dates), and as might be expected, not all data are initially stored in the correct format (e.g. numeric displayed as string). Similarly, date-time information for visits and shifts do not follow a standard format and required careful conversion. In addition, several measures (e.g., patient length of stay and physician shift length) need to be computed from the database, and missing data needed to be removed. During each stage of the analysis, we applied inclusion/exclusion criteria and verified the real-world validity of descriptive values by consulting with our clinical collaborator.

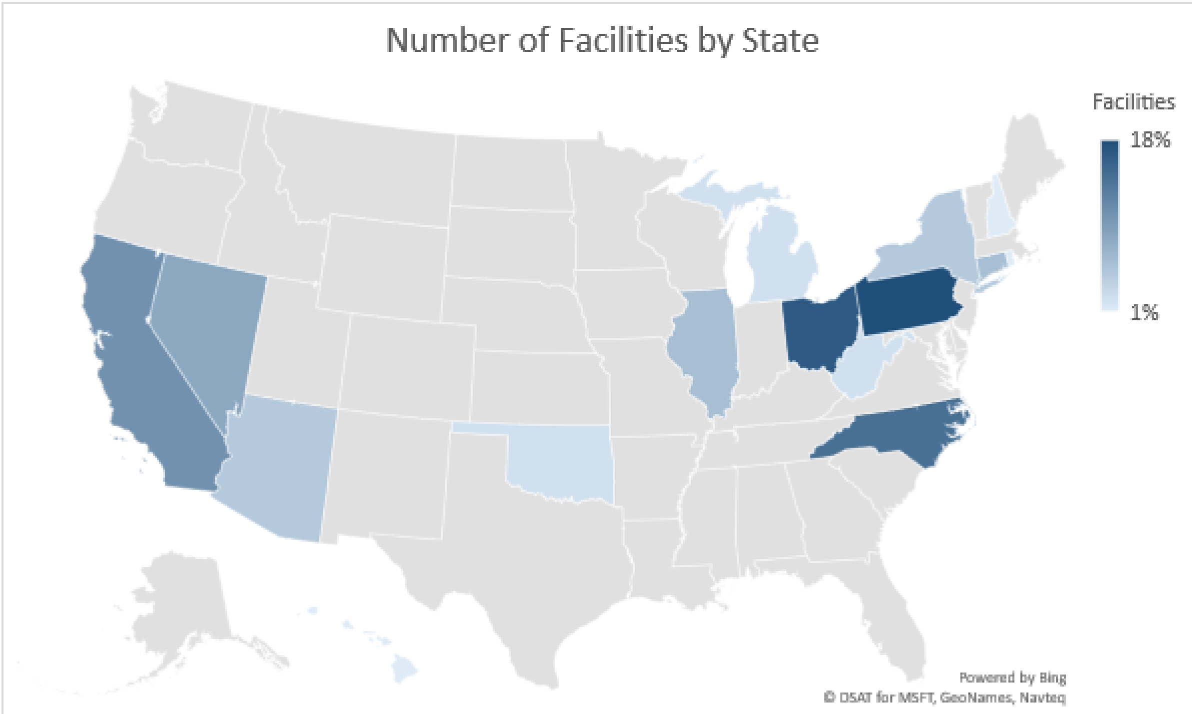


Figure 4 Locations of included EDs

In the initial stages, we ensure that these four databases are correctly linked using common fields. We then use the matched data to construct shift-level and month-level data, which are combined to obtain a record for each physician in each facility. This procedure involves the calculation of several aggregate variables, such as the proportion of hours worked during each shift and the Advanced Practice Provider (APP) (e.g., physician assistant or nurse practitioner) support ratio. These two measures present challenges as physician and APP schedules constantly change, and each shift needed to be broken down into hourly segments to determine these ratios. The initial task was demanding, but we are able to automate the procedure for future use by developing user-defined functions. The final step was to use the derived data to compute the indices and independent variables for each physician. The resulting dataset was used for statistical analyses herein.

Six newly acquired facilities and 355 physicians with short work histories (<500 patient visits) in the dataset were excluded. Thus, the final data for this study include 84 EM facilities, 1,079 physicians, and 10,615,879 patient visits with a total of 54 attributes. The facilities, which vary in yearly volume and capabilities (e.g., trauma designation or academic), are primarily located in urban and suburban regions, with only five EDs (6%) serving rural communities. Due to the range of locations and facility types, some EDs are profit-centers, while others are cost-centers, but the EPMN is still responsible for maintaining physician staffing levels at all facilities. Table 4 includes summary statistics for the final attributes we compiled for each physician and used for our analysis, including facility characteristics, patient visit records, the merged information from physicians' and patients' interactions, and physician's demographic and professional attributes. The data used for the analyses herein were aggregated over the entire study period, resulting in one record per physician (n = 1,079). However, the same methodology could be applied using monthly data.

The patient visit records include the hospital, the date and time of patient arrival at and departure from the ED, age, gender, the attending physician, disease codes (ICD-9), and the discharge disposition (admitted, discharged, transferred, elopement, left without being seen, or died) for each patient visit. Additionally, information regarding the proportion of patients with each payment source (Commercial Insurance, Medicare, Medicaid, or Self-Pay) and the RVUs associated with the visit are provided. We grouped ICD-9 codes into three categories (Group 1: Circulatory, Respiratory, Digestive, and Genitourinary; Group 2: Symptoms, Signs & Ill-Defined Conditions; and Group 3: Injury & Poisoning) that reflect the most prominent diagnostic groups attributed to ED patients. The patient level variable definitions are detailed in Appendix Table 1.

Table 4 Summary Statistics (10,615,879 patient visits of n=1,079 physicians at 84 facilities)

Level	Categorical Variable	Count (%)
Facility Characteristics	≤20000 Yearly Visits	23 (27.4%)
	20000-40000 Yearly Visits	27 (32.1%)
	40000-60000 Yearly Visits	22 (26.2 %)
	60000-80000 Yearly Visits	8 (9.5%)
	80000-10000 Yearly Visits	3 (3.6%)
	>100000 Yearly Visits	1 (1.2%)
	Teaching Facility	15 (17.9%)
	EM Residency Training Site	12 (14.3%)

Level	Continuous Variable	Mean	Std. Deviation
Visits	Average Patient Age	40.406	9.060
	% Male Patients	44.274	3.130
	% ICD9 Group 1	36.620	5.515
	% ICD9 Group 2	36.539	8.801
	% ICD9 Group 3	23.404	5.300
	% Admitted	17.673	9.980
	Commercial Index	1.004	0.063
	Medicaid Index	0.992	0.184
	Medicare Index	1.018	0.312
	Self-Insurance Index	0.991	0.109
Physician	Physician Age	45.870	9.994
	# Facilities Worked	2.456	2.294
	% 6AM-3PM Hours	36.434	14.337
	% 3PM-12AM Hours	47.467	11.609
	% 12AM-6AM Hours	16.086	14.998
	Coding Com/ Patient	0.507	0.484
	Physician PG Scores	54.241	24.773
	Physician PG Index	0.986	0.443
	APP Support Ratio (%)	22.772	10.515
	RVUs/hr	9.623	2.198
	RVUs/Patient Index	1.005	0.089
	Patients/hr Index	0.995	0.235
	RVUs/hr Index	0.991	0.188

Level	Categorical Variable	Percent
Physician Characteristics	Male	69.045
	White	78.221
	Primary Pediatric Practice	4.912
	Efficiency Training	29.935
	Patient Satisfaction Training	34.198

Notes: See Tables A1 and A2 in Appendix 1 for detailed variable definitions.
Indexed measures are computed by Eqs. (7)-(10) and (A1.1).

Each attending physician has a profile, which denotes demographic information, date of residency completion, and if explicitly trained on efficiency and patient experience by the EPMN. In addition, the healthcare provider work logs contain data corresponding to physicians and APPs.

For each shift worked, these include start and end times, the facility, whether a physician or APP, the provider's monthly PG patient experience percentile rank score, and the facility's monthly PG percentile rank score. Note that PG scores for physicians are reported on a monthly basis for each facility, so we use weighted averages to compute one physician PG score per physician at each facility, weighting by the number of hours worked during each month (Eq.(4)). Within each facility j, the total number of months that physician i has worked is $S_{i,j}$. The total number of hours physician i worked during month l at facility j is denoted as $Hours_{i,j,l}$, and the corresponding physician PG score is $PG_{i,j,l}$. See Appendix Table 2 for definitions of all physician-level variables.

$$Physician\ PG\ Score_{i,j} = \frac{\sum_{l=1}^{S_{i,j}} PG_{i,j,l} \times Hours_{i,j,l}}{\sum_{l=1}^{S_{i,j}} Hours_{i,j,l}} \tag{4}$$

3.4.1 Physician Performance Measures & Indices Development

The EPMN seeks to explicably and impartially assess all physicians within the network. Management wants to reward the strongest performers, while improving the performance of the weakest performers. Physician performance is based on two dimensions: patient experience and physician productivity. The EPMN administers patient experience surveys (Press Ganey Associates, Inc.) and reports each physician's monthly percentile score for each facility worked. The EPMN then uses these PG scores to assess patient satisfaction.

RVUs/hr are currently used by the EPMN to measure physician productivity at each facility the physician works. However, the goal of the network management is not to simply compare physicians in a single facility, but how to motivate physicians to accept the assignments deemed necessary by the EPMN. Thus, both metrics must be modified to reflect each physician's performance within the entire network. Such metrics should objectively reflect a physician's relative standing in the EPMN, regardless of the number and location of the facilities at which the physician works or other facility characteristics. In the following, we discuss why the conventional simple averages of RVUs/hr and PG scores would be unfair to physicians working in multiple facilities and the need to propose an alternative.

3.4.1.1 Absolute Measure of Physician Performance

Let F_i denote the total number of facilities in which physician i has worked. Within each facility j, the total number of patients visiting physician i is $V_{i,j}$ and the total number of hours physician i worked is $Hours_{i,j}$. The RVUs incurred by physician i's patient k in facility j is denoted as $RVU_{i,j,k}$. Thus, the total number of RVUs generated by physician i in facility j during the 54-month period are given by $Total\ RVU_{i,j} = \sum_{k=1}^{V_{i,j}} RVU_{i,j,k}$. The RVU/hr performance measure for physician i in facility j can be expressed as

$$RVUs/hr_{i,j} = \frac{Total\ RVUs_{i,j}}{Hours_{i,j}} \tag{5}$$

Patients/hr and RVUs/Patient can be computed for physician i in facility j similarly.

For a physician working at multiple facilities within the EPMN, RVUs/hr = (sum of RVUs in the network)/(sum of hours worked in the network); network Patients/hr and RVUs/Patient can be obtained similarly. Yet, these measures are facility-dependent. Facility demand affects service times, which directly influences the number of patients a physician cares for hourly (Kc and Terwiesch, 2009). Thus, comparing physicians across facilities using these metrics would be unfair, but RVUs/hr performance targets set by individual facilities are inadequate for evaluating EP productivity in network settings. Our data indicate the average EP within the EPMN works at

2.46 facilities, and EP groups often contract with healthcare systems with multiple EDs. These EDs have varying capabilities, patient catchment areas and staffing models. Moreover, EPs do not control patient arrivals and work various clinical schedules to cover facilities 24/7. The high degree of variability in patient volume and acuity leads to fluctuations in RVUs/hr, making the conventional approach invalid and necessitating more sophisticated methods to measure EP productivity.

3.4.1.2 Proposed Relative Indices for Physicians in Network

The new performance indices proposed here are better suited for concurrently evaluating all physicians working in a large EPMN, where a physician may be assigned to multiple facilities. Eq.(5) quantifies physician i's performance in facility j. We define facility j's average performance as:

$$Average\ RVUs/hr_j = \frac{\sum_{i=1}^{N} \sum_{k=1}^{V_{i,j}} RVUs_{i,j,k}}{\sum_{i=1}^{N} Hours_{i,j}} \tag{6}$$

Taking the ratio Eq.(5)/Eq.(6), i.e., $\frac{RVUs/hr_{i,j}}{Average\ RVUs/hr_j}$, we know how the physician performs relative to peers in the specific facility, j. Next, we weight the ratio by the hours worked in each facility to derive a composite rating across all facilities (Eq.(7)). Indexing this metric acknowledges the inability of EPs to control for demand, capacity, and competencies in the facilities worked.

$$RVUs/hr\ Index_i = \frac{\sum_{j=1}^{F_i} \left[\frac{RVUs/hr_{i,j}}{Average\ RVUs/hr_j} \times Hours_{i,j} \right]}{\sum_{j=1}^{F_i} Hours_{i,j}} \tag{7}$$

We define the volume index (Eq.(8)) and the complexity index (Eq.(9)) similarly, using hours in a facility and number of patients treated in a facility, respectively, as the weights:

$$Patients/hr\ Index_i = \frac{\sum_{j=1}^{F_i}\left[\frac{Patients/hr_{i,j}}{Average\ Patients/hr_j} \times Hours_{i,j}\right]}{\sum_{j=1}^{F_i} Hours_{i,j}} \tag{8}$$

$$RVUs/Patient\ Index_i = \frac{\sum_{j=1}^{F_i}\left[\frac{RVUs/Patient_{i,j}}{Average\ RVUs/Patient_j} \times V_{i,j}\right]}{\sum_{j=1}^{F_i} V_{i,j}} \tag{9}$$

The patient experience index (Eq.(10)) is also computed using hours worked in a facility as the weights:

$$PG\ Index_i = \frac{\sum_{j=1}^{F_i}\left[\frac{PG\ Score_{i,j}}{Average\ PG\ Score_j} \times Hours_{i,j}\right]}{\sum_{j=1}^{F_i} Hours_{i,j}} \tag{10}$$

Note that an index of 100% indicates that a physician is performing on par with peers in the network; 110% implies 10% better performance than peers; while 85% suggests that performance is 15% inferior to peers.

Advantages of the proposed indices are discussed in Appendix A.3, where we compare Eqs. (7)–(9) with alternate metrics. Appendix A.3 also details the mathematical rationale for supporting the indices we proposed.

3.4.1.3 The Need for the New Physician Network Performance Indices

Normalizing by the facility averages removes the effects of scale (facility average) and addresses the effects of relative hours worked at various facilities in the EPMN. Such a method is fair since facility assignment is often beyond physicians' control, given the need of the EPMN to staff all contracted facilities and the competitive nature of obtaining more healthcare facilities to contract with the EPMN. Table 5 gives a stylized example, which resembles information found in the data, to highlight the need and advantage of using the proposed indices (Eq.(7)). Physician A worked in two facilities (Facilities 1 and 2), while Physician B only worked in Facility 2. Facility 1 has higher average RVUs/hr (14 vs. 4).

Table 5 The Need of Using Performance Indices – An Example

					RVUs/hr				
Phys.	Fac.	RVUs	Patients	Hours	Fac.	Phys.	RVUs/ hr Ratio	RVUs/hr	RVUs/hr Index
A	1	9000	3600	1200	14.0	7.5	0.54	6.0	**69.6%**
	2	3000	900	800	4.0	3.8	0.94		
B	2	9000	2500	2200	4.0	4.1	1.02	4.1	**102.3%**

Since Physician B only works at one facility, calculating RVUs/hr is straightforward, but facility differences complicate the calculation for Physician A. First, we examine the absolute measures, which is the ratio of Total RVUs to Total Hours. In that case, Physician A [6.0=(9000+3000)/(1200+800)] generates more RVUs/hr than Physician B [4.1=9000/2200] as shown in Table 5. Based on the absolute RVU/hr measure, Physician A is more productive (6.0) than Physician B (4.1). We then compare physicians on the relative indices, detail their derivations below, and contrast this with the absolute measures.

We first calculate the ratios for each facility where the physician works. Physician A's RVUs/hr in Facility 1 is 7.5 (=9000/1200), while Facility 1's average RVU/hr is 14. Thus, Physician A is a below average performer in Facility 1, with a ratio of 0.54 (=7.5/14) indicating that performance is 46% below peers. Physician A also appears to be slightly below average at Facility 2 (0.94). Using Eq. (7), the index is:

RVUs/hr Index: 69.6% = (0.54*1200+0.94*800)/(1200+800)

The indices in Table 5 show that Physician A is below average with regard to RVUs/hr (69.6%) and Physician B is above average with regard to RVUs/hr (102.3%). The absolute measure and the indexed measure give different conclusions about Physicians A and B's performances. By reflecting the relative performance at facilities in which a physician works and weighing performance on the time spent in each facility, the index proposed in Eq.(7) neutralizes the scale (facility size) bias. Thus, the proposed index better reflects the true performance when the network administrators assign facilities. This demonstrates the importance of adopting the relative index for assessing physician performance within large networks where physicians work at multiple facilities.

3.4.1.4 Comparing Network Physician Performance Using Proposed Index

We find a strong positive correlation between RVU/hr Index and absolute RVUs/hr ($r=0.7687$, p-value < 0.001), confirming that a substantial portion of revenue potential (RVUs/hr) can be explained by the proposed index. It captures the "intrinsic" ability of a physician to potentially generate revenue across facilities since it weights relative performance by hours at the various facilities. The unexplained portion may be driven by patient characteristics. Thus, the index is an equitable measure of physician performance.

To provide a more comprehensive view and offer appropriate "carrots and sticks" for reward and remedy, we jointly consider both productivity and patient satisfaction by plotting the new RVU/hr Index (x-axis) and PG Index (y-axis) on the xy-plane. Depending on which quadrant a physician falls, management can easily identify the need for improvement with regard to productivity, patient satisfaction, or both, or if the physician should be rewarded for outstanding performance on both dimensions (Fig. 5).

Fig. 5 displays physicians' performances, where each dot represents one physician. The horizontal and vertical lines signify average performance on each dimension, and thus divide physicians into four groups (quadrants). Management may choose cutoff values within each quadrant to further differentiate physicians within each group, e.g., awarding those who perform one standard deviation above the mean (center point) on both dimensions. Additionally, they may choose to weight the two measures differently depending on specific performance goals. The "best" physicians (26%) who deliver above average performance in both productivity and patient satisfaction, i.e., RVUs/hr Index ≥ 1 and PG Index ≥ 1, are located in Quadrant I and may be used for benchmarking. Physicians who fall in Quadrant III (30%) are below average performers on productivity and patient experience, and require immediate attention from management. Quadrant II (26%) shows those short on productivity, while Quadrant IV (18%) depicts those with low patient experience; both II and IV are candidates for remedial training depending on the specific threshold set by management. It may be necessary to further address ways to balance patient experience and productivity in Quadrants II and IV.

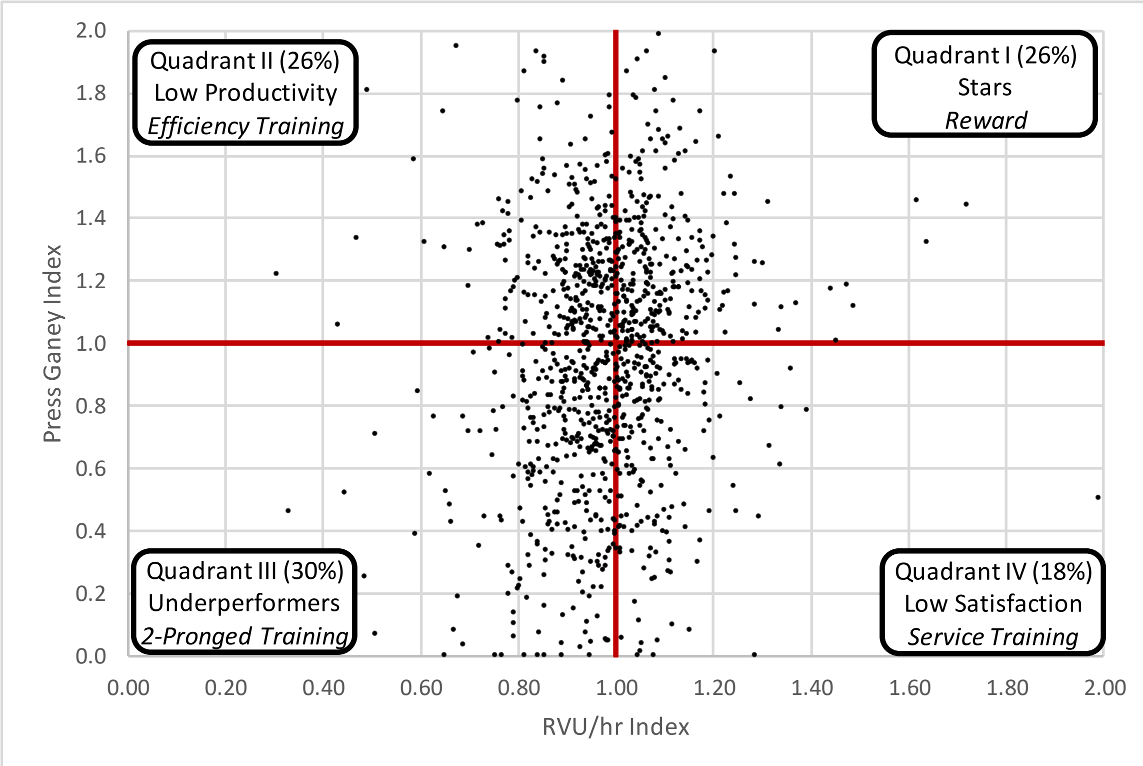


Figure 5 Integrating Productivity and Patient Experience: Insights for Rewards and Improvement Training

3.5 Cluster Analysis to Manage Physician Segments

We have proposed a fairer, systematic, logical and comprehensible performance metric to objectively evaluate physicians working in an EPMN in §3.4. To identify individual traits that may affect physician performance, we group physicians based on non-performance attributes and examine the disparities among clusters. Such information can be used to guide and reward different groups of physicians; and more efficiently and effectively allocate physicians to facilities to meet the needs of the EPMN.

Due to known differences between pediatric and general EDs, we treat pediatric physicians as a separate group. Since pediatric physicians only account for 5% of the physicians in our data, we have excluded physicians from the cluster analysis if the average age of their patients was less than

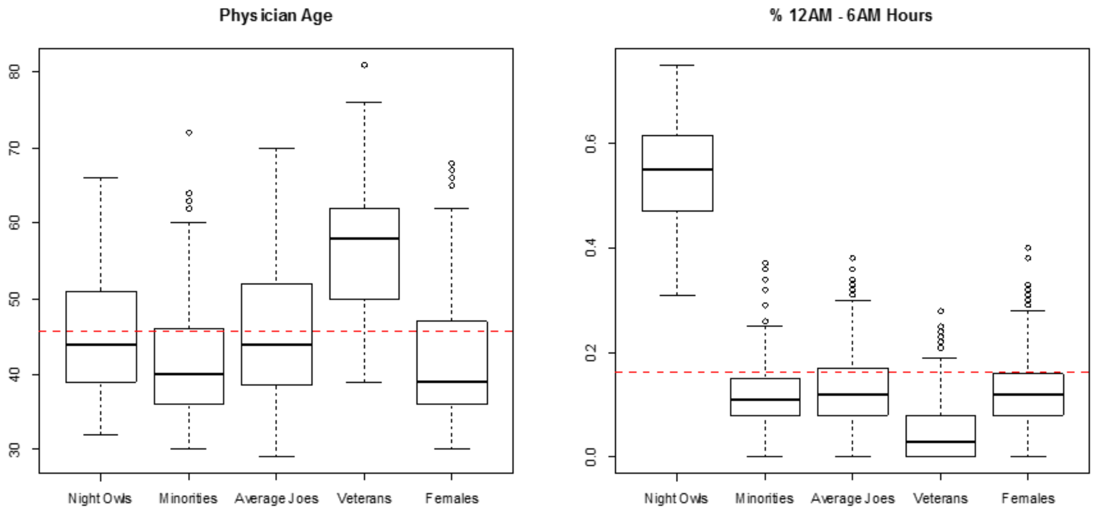
18. Using the variables in Table 6, we group the 1,026 non-pediatric EPs into five fundamentally different clusters using K-means.

Table 6 Summary Statistics of Variables used in Cluster Analysis

	Average Patient Age ≥ 18 (N=1026)	
Variable	Mean	Standard Deviation
Physician Age	45.749	9.967
# Facilities Worked	2.518	2.334
% 6AM-3PM Hours	36.724	14.257
% 3PM-12PM Hours	47.139	11.342
% 12PM-6AM Hours	16.126	15.176
% Admitted	18.052	10.018
APP Support Ratio (%)	23.268	10.226

Variable	Count	Percent
Male	715	69.688
White	812	79.142

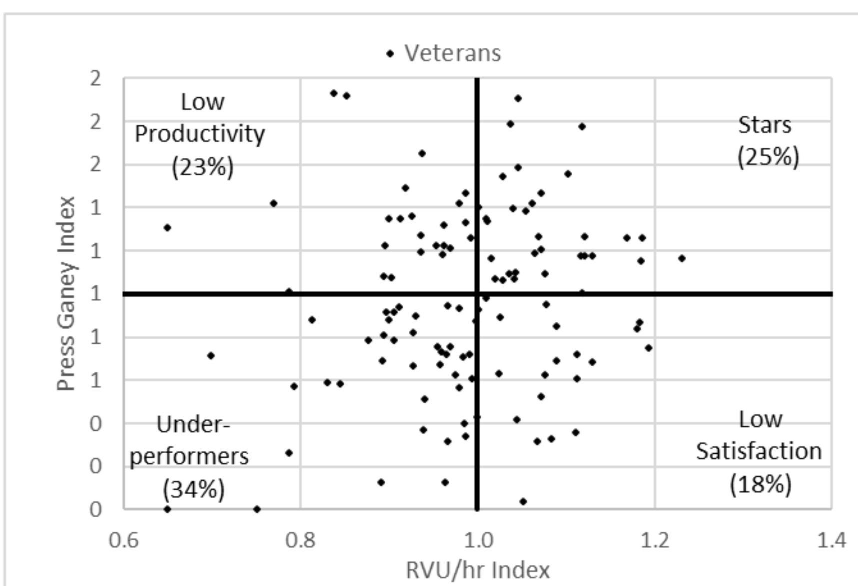
We have colloquially named clusters based on dominant characteristics (e.g., Night Owl, Veteran, etc.) Figs. 6(a) – (b) graphically display differences among the five clusters. The dashed line represents the overall average, while each boxplot shows the distribution of the variable on the y-axis. The plot provides the (1) minimum, (2) 1st quartile (25th percentile), (3) median (50th percentile), (4) 3rd quartile (75th percentile), and (5) maximum. Figs. 6(a) – (b) shows that Veteran physicians are older and tend to work fewer nights, while Night Owls work a disproportionately high number of night shifts relative to others. Minority and Female physicians are generally younger, consistent with medical school trends.



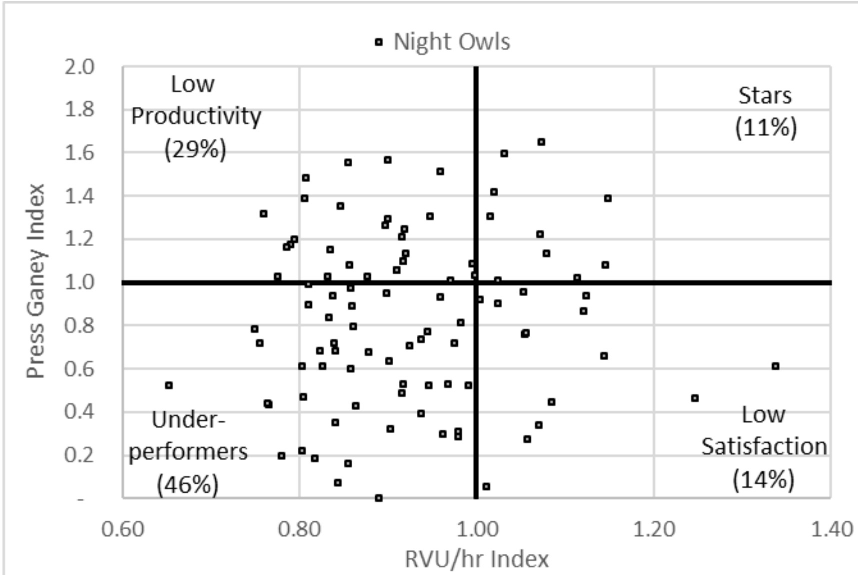
(a) **(b)**

Figure 6 Distributions of age and night-shift by clusters

By zooming in on Fig. 5 and examining the five clusters separately, we can link the cluster membership with performance relative to the entire physician population. Fig. 7(a) shows a relatively even distribution of Veterans across the four quadrants, while Fig. 7(b) suggests that the majority of Night Owls have low RVUs/hr indices, with many also earning lower PG scores. One-way ANOVA has confirmed that the cluster differences in PG scores and RVUs/hr Index are statistically significant.



(a) 25% of Veterans (Quadrant I) achieve above average performance on both dimensions



(b) 75% (Quadrants II & III) of Night Owls earn below average RVUs/hr Index

Figure 7 RVUs/hr Index (Productivity) versus PG Index (Patient Experience) by Cluster

Thus far, we have provided a structural, evidence-based approach to segment EPs within the network. The clusters suggest that EPs may have different priorities at different stages in their personal and professional lives. For instance, younger physicians (i.e., many female and minority physicians) may have different family and financial concerns than veteran doctors (Darcy et al. 2012; Dyrbye et al. 2013). Like most professions, the composition of EPs encompasses varying career ambitions and personal responsibilities. It is thus vital to understand such distinctions when matching physicians with divergent priorities to different shifts and multiple facilities. Specifically, given the need to maintain continuous and universal availability of EDs, managers should consider physician characteristics when pursuing operational productivity and greater patient experience. There are legal reasons why we do not recommend linking compensations with these clusters. By law, employers cannot compensate based on age, gender, or race. Unfortunately, the cluster analysis reveals that physicians are naturally grouped by these factors. Thus, these clusters only help to develop insights for our study, instead of serving as the basis for compensation.

The clustering results lead us to reflect on current practices. Namely, management must appreciate the dynamics of doctors' professional and personal development and recognize that physicians will respond distinctively to incentives. Measuring performance collectively for all clusters or offering identical incentives to all physicians may not be effective. For example, comparing Night Owls with Veterans who work fewer night shifts could be unfair due to differences in patient composition. In the following section, we use the cluster differences to motivate our selection of control variables.

3.6 Empirical Results

The indices in Eqs. (7) – (10) neutralize the impact of exogenous patient demand and allow for fairer comparisons among physicians operating in an EPMN. In this section, we first model RVUs/hr Index to demonstrate how the four indices are linked. Then, we identify the drivers of these indices, i.e., to understand how patient, physician, and facility factors impact physicians' relative indices. Finally, we analyze the relationship between the revenue potential index and PG patient experience index to discern if physicians with high revenue potential sacrifice patient experience scores.

3.6.1 Drivers of Relative Indices: A Simultaneous Equations Model

Through a log-log model (not presented here), we empirically show that the product of the volume and complexity indices explains 99.86% of the variability in RVUs/hr Index. Namely, ln(RVUs/hr Index) = 1.00*ln(Patients/hr Index) + 0.99*ln(RVUs/Patient Index). While this product does not necessarily result in the RVUs/hr Index mathematically, the strength of this empirical relationship led us to further examine the indices' relationships. Appendix A.4 proves that the multiplicative equivalence holds when the average flow rates among facilities are equal. As the facility flow rates in our dataset are quite comparable for a given physician, this explains the strong empirical result. Specifically, the average facility flow rate deviation for a physician is 0.3, corresponding to a mean absolute percentage difference (MAPD) of 8.6%. Note that Physician i's MAPD is computed by $MAPD_i = \frac{\sum_{j=1}^{F_i} \frac{|Patients/hr_j - Average\ Patients/hr|}{Average\ Patients/hr}}{F_i}$, with F_i denoting the total number of facilities in which physician i has worked.

The coefficients for the volume and complexity indices are both approximately equal to positive one and highly significant, and these effects are synergistic. As both indices strongly influence the RVUs/hr Index, physicians can enhance their RVUs/hr Index through improving volume or complexity performance (or both). Subsequently, we explore how exogenous factors affect complexity and volume indices, and thus indirectly drive revenue potential, using a system of simultaneous equations. In addition to the control variables identified through clustering, we chose to include physician characteristics and variables that change shift-by-shift, e.g. patient population, type of shift (day, afternoon, or night). While geographic location, demographics around the facility, the number of beds, etc. are important factors, these are facility related variables, which are indirectly accounted for in our indices.

The first and second stage estimates from the two-stage least squares (2SLS) model (Eqs.(1)–(2)) are summarized in Tables 7(a) and 7(b), respectively, and we have confirmed the robustness of these results using bootstrap standard errors (see Appendix A.5). We utilize the 2SLS regression procedure to simultaneously estimate the two equations because it accounts for correlations between endogenous variables (volume and complexity indices); and between endogenous variables and the 2^{nd}-stage errors. We confirm that this system is identified based on the rank condition, which is a necessary and sufficient condition for identification. More specifically, each of the control variables excluded from one equation appears in the other equation. Eq.(1) in Table 7(b) corresponds to the complexity index and comprises patient characteristics (e.g., Patient Age, % Male Patients, % Admitted, and ICD-9 Code groups). These controls are not included in Eq.(2) for the volume index because physicians do not schedule or select their patients. In contrast, physician characteristics (e.g., age, gender, and race) and coding communications per patient are only included in Eq.(2). Other exogenous variables (insurance types, shifts worked, and the APP support ratio) overlap between the two equations, signifying both their direct and indirect effects on the volume and complexity indices.

The coefficient for ln(Patients/hr Index) in Eq.(1) is significant and positive (β=0.1841), which is inconsistent with our first hypothesis (H1). Conversely, the coefficient for ln(RVUs/Patient Index) in Eq.(2) is significant and negative (β=-0.8224), suggesting an inverse relationship and providing support for H1. Thus, the simultaneous equations model partially supports H1. Tables 7(a) – (b) indicate that many factors influence the volume and complexity indices, and the relationships between the exogenous variables and the indices involve both direct and indirect effects. This demonstrates the extent to which the volume and complexity indices are intertwined with each other and with exogenous factors.

We have thus far analyzed the factors influencing the volume and complexity indices and how these two indices drive RVUs/hr Index, but does the need to increase physician productivity in the EPMN result in reduced patient experience? We next examine how revenue potential (efficiency) affects patient experience.

Table 7 2SLS Model: Relationship between Volume and Complexity Indices (N=1,079)

(a) First Stage Regression Estimates (Exogenous Variables Only)

		Y=ln(RVUs/Patient Index)			Y=ln(Patients/hr Index)		
		First Stage Regression					
Level	Variable	Coefficient		[Standard Error]	Coefficient		[Standard Error]
Physician	(Constant)	-0.4735		[0.1040]	0.4202		[0.3025]
	ln(RVUs/Patient Index)						
	ln(Patients/hr Index)						
	Peds Indicator (0 = General, 1 = Peds)	0.4418	***	[0.0610]	-0.4752	**	[0.1775]
	Physician Age	-0.0001		[0.0004]	0.0008		[0.0012]
	Physician Male (0 = Female, 1 = Male)	-0.0022		[0.0038]	0.0365	**	[0.0112]
	Physician White (0 = Non-white, 1 = White)	0.0031		[0.0042]	0.0303	*	[0.0123]
	% 12AM - 6AM Hours	0.0701	***	[0.0180]	-0.1119	*	[0.0523]
	% 6AM - 3PM Hours	-0.0322		[0.0176]	0.2241	***	[0.0512]
	Coding Com/Patient	-0.0097	**	[0.0037]	-0.0402	***	[0.0107]
	APP Support Ratio	0.1468		[0.0748]	0.4859	*	[0.2174]
	APP Support Ratio*Physician Age	-0.0025		[0.0016]	-0.0090	*	[0.0046]
Patient	Average Patient Age	0.0289	***	[0.0039]	-0.0257	*	[0.0113]
	(Average Patient Age)^2	-0.0003	***	[0.0000]	0.0003	*	[0.0001]
	% Male Patients	-0.0232		[0.0863]	0.1888		[0.2510]
	Commercial Index	0.0032		[0.0294]	-0.2592	**	[0.0855]
	Medicaid Index	-0.1902	***	[0.0129]	0.1274	**	[0.0375]
	Medicare Index	0.0371	***	[0.0066]	-0.1053	***	[0.0193]
	Self-Pay Index	-0.2229	***	[0.0202]	0.3605	***	[0.0589]
	% ICD9 Group 1	0.3313	***	[0.0394]	-0.2913	*	[0.1147]
	% ICD9 Group 2	0.2155	***	[0.0306]	-0.1530		[0.0890]
	% ICD9 Group 3	0.0582		[0.0495]	-0.2163		[0.1440]
	% Admitted Patients	0.1255	***	[0.0233]	-0.1552	*	[0.0678]
Adjusted R^2		64.64%			27.51%		

(b) Second Stage Regression Estimates

		Eq. (1): ln(RVUs/Patient Index)		Eq. (2): ln(Patients/hr Index)	
		Second Stage Regression			
Level	Variable	Coefficient	[Standard Error]	Coefficient	[Standard Error]
Physician	(Constant)	-0.5648	[0.1258]	0.0351	[0.1626]
	ln(RVUs/Patient Index)			-0.8224 ***	[0.1391]
	ln(Patients/hr Index)	0.1841 **	[0.0659]		
	Peds Indicator (0 = General, 1 = Peds)	0.5384 ***	[0.0809]		
	Physician Age			0.0002	[0.0012]
	Physician Male (0 = Female, 1 = Male)			0.0367 **	[0.0109]
	Physician White (0 = Non-white, 1 = White)			0.0337 **	[0.0120]
	% 12AM - 6AM Hours	0.0839 ***	[0.0228]	-0.0191	[0.0504]
	% 6AM - 3PM Hours	-0.0811 **	[0.0268]	0.2205 ***	[0.0488]
	Coding Com/Patient			-0.0457 ***	[0.0104]
	APP Support Ratio	0.0170	[0.0232]	0.5794 **	[0.2132]
	APP Support Ratio*Physician Age			-0.0098 *	[0.0044]
Patient	Average Patient Age	0.0345 ***	[0.0050]		
	(Average Patient Age)^2	-0.0004 ***	[0.0001]		
	% Male Patients	-0.0768	[0.1064]		
	Commercial Index	0.0476	[0.0395]	-0.2756 **	[0.0824]
	Medicaid Index	-0.2124 ***	[0.0183]	-0.0485	[0.0497]
	Medicare Index	0.0564 ***	[0.0107]	-0.0650 **	[0.0202]
	Self-Pay Index	-0.2929 ***	[0.0357]	0.1876 **	[0.0720]
	% ICD9 Group 1	0.3697 ***	[0.0519]		
	% ICD9 Group 2	0.2598 ***	[0.0368]		
	% ICD9 Group 3	0.0865	[0.0598]		
	% Admitted Patients	0.1629 ***	[0.0302]		

Independent variables defined in Tables A1 & A2

3.6.2 Drivers of PG Score

CMS (2015) has recently introduced Value-Based Purchasing (VBP) programs. These quality initiatives financially motivate hospitals to improve upon current operations, as reimbursement now depends on patient outcomes and patient experience. The PG survey is a widely employed tool for gauging perceptions among discharged patients. The survey includes questions related to the individual physician and the facility, and both physician and facility PG scores (percentiles) are reported monthly. In this section, we examine the linkage between the proposed revenue potential index and patient experience scores. Specifically, we consider whether physicians trade off the RVUs/hr Index for the PG Index. For this, we employ OLS regression and model the physician PG Index as a function of the RVUs/hr Index and exogenous control variables (Table 8). The specific variables chosen for the model are based on the recommendations of the clinicians and administrative staff of the EPMN under study.

Table 8 PG Index Model Results

Dependent Variable: PG Index[1]

Level	Variable	Coefficient		[Standard Error]
Physician	(Constant)	0.8027		[0.2200]
	RVUs/hr Index	0.3234	***	[0.0719]
	Physician Age	-0.0062	***	[0.0014]
	Physician Male (0 = Female, 1 = Male)	0.0480		[0.0291]
	Physician White (0 = Non-white, 1 = White)	0.1421	***	[0.0323]
	% 12AM - 6AM Hours	-0.5652	***	[0.1027]
	% 3PM - 12AM Hours	-0.3506	**	[0.1308]
	Efficiency Flag (0 = Not Complete, 1 = Complete)	0.0651	*	[0.0303]
	Patient Satisfaction Flag (0 = Not Complete, 1 = Complete)	-0.0772	**	[0.0292]
Patient	Average Patient Age	-0.0052	***	[0.0015]
	% ICD9 Group 1	-0.1543		[0.2867]
	% ICD9 Group 2	0.7095	**	[0.2150]
	% ICD9 Group 3	1.1821	***	[0.3447]
	R^2	10.87%		
	Adjusted R^2	9.85%		

***p < 0.001, **p < 0.01, *p < 0.05; two-tailed tests
[1]Refer to Eq. (10) to derive the PG Index
Independent variables defined in Tables A1 & A2

The results in Table 8 show that almost 10% of the variability in the PG Index is explained by our model, and most predictors are statistically significant. Specifically, we observe a positive and significant coefficient for RVUs/hr Index (β=0.3234) that supports our second hypothesis (H2).

Other significant factors in Table 8 suggest that PG Index is further influenced by physician-specific characteristics, such as work schedules as well as the average patient age. For example, physicians who work more night shifts receive significantly lower scores relative to their peers, possibly due to the acuteness of patients' conditions, stress experienced during night visits, and lower staffing levels at night. Working more hours between 3pm and 12am is also associated with lower relative scores. Furthermore, physicians seeing older patients receive lower scores relative to their peers on average. As physicians with lower PG scores are often required to complete the Patient Satisfaction training, we find a negative relationship between training completion and PG Index. Due to this bias and because training completion dates are not reported, we must be cautious when interpreting such results. We recommend that future data collection in the EPMN include the dates and reasons for training.

Table 8 suggests that a significant relationship exists between the RVUs/hr Index and PG Index. However, we speculate that unobserved factors play important roles in explaining relative PG scores and hence the relatively low explanatory power of the model.

The results of these hypotheses are important because this is how physicians get paid and how they are evaluated. It also helps to have objective metrics to compare healthcare providers so management can make better provider evaluation and incentive decisions. The primary driver of RVUs/hour is better training in multi-tasking to manage multiple patients. The primary driver of PG scores is to reduce length of stay for discharged patients (see Pines et al. 2017), and there are clearly proven techniques for improving communication to increase PG. Thus, the proposed metrics allow for such evaluations while taking into account factors largely outside of an individual physician's control. The indices provide insight into what is fixed and what needs to be adjusted.

3.7 Managerial Insights, Limitations, and Conclusions

3.7.1 Managerial Insights

3.7.1.1 Benchmarking Physician Performance and Continuous Improvement

Our findings have practical implications for EDs. Indexing physician metrics such as Patients/hr, RVUs/Patient, RVUs/hr, and PG Scores to facility-based averages mitigates the exogenous factors that affect physician performance. The proposed indices provide simple and intuitive benchmarks for objective evaluation of physicians. These indices facilitate physician segmentation into high and low performers, which initiates new management processes to incentivize and train physicians. Management may opt to customize the evaluation and ranking process by providing weights for each dimension of performance. Moreover, resulting clusters highlight distinct differences among physicians and provide further managerial insights.

Healthcare administrators are constantly looking for best practices to benchmark performance and improve processes. EDs are no exception, as evidenced by the Emergency Department Benchmarking Alliance (Wiler et al. 2015). However, benchmarking practices typically occur at the facility-level. With our study, benchmarking at the physician level both within a facility and across facilities is possible. That is, by identifying characteristics of the highest performing physicians in terms of RVUs/hr Index, other physicians can recognize factors under their control and strive for first-rate performance. Furthermore, the proposed indices provide objective measures of the network performance of physicians given that demand is beyond their control. The new indices capture physician performance relative to their peers and are highly correlated with the revenue potential of the physician in the EPMN. These indices neutralize the exogenous demand effects while capturing relative effort compared to peers. From the perspective of the administrators, the index approach is fair, simple, intuitive, and effective.

Our statistical models acknowledge that the same productivity level may not be attainable for all physicians at all facilities. We recommend that managers incorporate relative indices and provide adjustment factors when evaluating physicians. Exogenous demand and other facility-level factors impact the indices, and thus our 2SLS regression model (Table 7) can be used to adjust for various physician-specific and patient-specific factors to support more equitable comparisons of physicians' RVUs/hr performance. At a minimum, our facility-adjusted indices reduce the exogenous variability inherent to the respective absolute measures, and ED management could use our model to set performance standards. These results are helpful in tracking continuous improvement of physicians over time as well. Recognizing these results could lead to the development of new training programs or instituting mandatory completion of existing training for certain physicians. The clustering results reinforce our case for defining performance objectives (and incentives) pertaining to distinct physician segments. The clusters provide guidance for management to set physician-specific targets, schedule and allocate EPs, and implement differential physician "care and recruitment" strategies.

3.7.1.2 Management and Physician Benefits

While EDs typically cannot control the type or number of patients arriving for treatment during a given shift, they should recognize that patient experience scores are generally lower for physicians working mostly night shifts. In addition, physicians with more APP support tend to achieve higher volume indices, demonstrating that they treat more patients per hour. The proposed indexing approach is valuable for accurate assessment of physician productivity and could lead to more equitable compensation across physicians as the proposed performance indices are generalizable and applicable to various compensation models, such as fee-for-service, capitated, or value-based reimbursement. While physicians may favor daytime shifts for physiological reasons, our results indicate that performance differs significantly between day and night shifts. Physicians who work more night hours are somewhat penalized in both PG scores and RVUs/hr. Thus, management should take note of physicians willing to work nights in their facility assignment and in their compensation decisions as our empirical results provide compelling evidence and rationale for differentiating recompense. This information may even translate into scheduling decisions. For instance, a facility might stipulate a minimum proportion of night hours for all physicians in order to address the night shift effect. As physicians age, working nights becomes physically more demanding; compensation adjustments may need to follow to allow for fewer night shifts.

EPs may also gain from understanding our results. Since RVUs/hr is used to evaluate the physicians, they can benefit from recognizing the factors that affect productivity, especially those factors within their control (e.g., shift preference). This knowledge may help physicians to improve their productivity by controlling its drivers. It may also aid them in employment decisions. Physician involvement in shift choices offers them the ability to take into account both personal and financial concerns to make educated decisions that best fit their needs. All stakeholders benefit from physician productivity improvements, and understanding and tracking productivity is a critical step toward improving ED operations.

We have had discussions with the organization in question on piloting the use of these indices in particular health systems to prospectively assess their values. Currently, EPMN management employs arbitrary benchmarking that does not adjust for facility capabilities, physicians' career stages, and patient differences across EDs and shifts. Testing and subsequent implementation of these the transparent and easy to implement indices with follow-up assessment is our planned next step.

3.7.2 Limitations

We have taken into account several control variables in our study of EPs, but other factors could also affect physician performance. Capacity variables such as the number of ED beds, nursing support, and the availability of diagnostic tools in a facility can influence processing rates as well. One limitation of our study is the inability to account for these effects. Moreover, we did not consider the impact of information technology. For example, some facilities have recently implemented electronic health record systems, which temporarily slow down operations and negatively impact ED performance (Ward et al. 2014a; Ward et al. 2014b). While we examined productivity, providers and hospitals use several other clinical metrics to assess physician performance, such as patient outcomes and benchmark goals put forth by national organizations such as CMS. We focused on the operational performance metrics and have yet to link them to these other markers.

From a modeling perspective, we aggregated data over the entire study period (54 months) to analyze overall EP performance. However, this data may be modeled at a monthly level. Performing a cross-sectional analysis limits our ability to explore learning effects over time. Studying the data longitudinally would allow us to examine the influence of exogenous shocks on demand. For instance, outbreaks of Zika virus, Ebola virus, and various influenzas impact ED demand in prominent facilities, but we do not view the effects of these exogenous shocks when data is aggregated over many months. Future research could study EP performance on a more granular level, including quasi-experiments to expose causal relationships.

Furthermore, the way in which PG scores are reported raises some concerns about their usage. A physician's scores at any time are based on the PG surveys completed by previous patients at a given facility. Therefore, PG scores are not reported until a number of patients from that facility submit their surveys (possibly requiring several weeks' worth of data to reach an adequate sample). As with any survey, there may also be response bias, and PG percentile scores could fluctuate significantly from month to month.

Finally, implementation of the proposed indices for benchmarking may not be easy. Our discussions with physicians indicate some opposition to management by numbers alone. Thus, physician resistance, actionable items resulting from benchmarking, and understanding the meaning of these indices by a healthcare audience will likely present some challenges.

3.7.3 Conclusions

Our study is unique in several ways. First, we use big data to conduct a multi-facility, multi-year study of physician performance within a large EPMN, which is highly relevant as these physician networks continue to expand, especially in emergency medicine. We develop performance metrics that adjust for facility-level differences and allow for objectively comparing physicians who work at multiple facilities across a large and diverse EPMN. Our proposed indices overcome the deficiencies of existing physician performance measures that are only appropriate in the single-facility case and make equitable comparisons of physicians within the network possible.

We empirically demonstrate the value of the proposed indices in evaluating physicians within this large network. We subsequently use cluster analysis to identify physician segments with similar characteristics, which helps management to recognize physicians' priorities and to better understand which factors are driving physician performance. We verify that the proposed volume and complexity indices explain a substantial portion of variation in revenue potential productivity across physicians. The 2SLS model (Table 7) simultaneously examines the linkage between the volume and complexity indices and their drivers. Finally, we explore the relationship between physician productivity and patient experience (Table 8).

The 24/7 nature of EDs, the high incidence of burnout and the overall shortage of EPs relative to patient need make it necessary to effectively manage EPs across their entire careers. This research is highly relevant to the EM field as large physician management groups attempt to balance professional career satisfaction and the needs of large health systems for ED services. As consolidation occurs throughout healthcare in general and EM in particular, more physician management networks are emerging. The need for more robust data-driven measurements of network physicians persists, as these performance metrics are crucial for maintaining market position and bargaining power. Refined tools to measure and evaluate productivity carry lessons for assessing and managing the operational efficiency of the healthcare system where the ED stands at an important nexus.

While we have focused on ED-specific issues and problems associated with management of large EPMNs, the research framework in Fig. 2 – 3 may be adapted for application to other settings. For example, regional managers of retail stores or restaurant chains overseeing various facilities may face disparities in performance across different sites due to site-specific attributes. By identifying dimensions on which performance will be judged and indexing those measures relative to site averages, organizations can develop an equitable method to assess an employee's relative performance across sites. Thus, the proposed research framework (Fig. 2 – 3), comprising the indexing system, performance matrix (quadrants), clustering, and driver identification, is generalizable and can be broadly applied after making industry-specific or company-specific adaptations. Our model is thus valuable for performance enhancement and employee development in data-intensive business settings.

4.0 Emergency Physician Practice Patterns and Malpractice Claims

4.1 Background and Motivation

Emergency medicine is a specialty with high malpractice risk because of the undifferentiated patient population and limited time and resources to manage acutely ill and injured individuals. Emergency physicians are likely to be involved in malpractice claims; more than 75% of emergency physicians will be named in a malpractice claim at some point in their career (Jena et al. 2011). On average, physicians spend 50.7 months of their career involved in litigation (Seabury et al. 2013). To help reduce risk, roughly 9 in 10 physicians report overusing or over-ordering tests or procedures, termed *defensive medicine* (Bishop et al. 2010, Mello et al. 2010).

A malpractice claim can negatively affect a provider through anxiety, depression, and even thoughts of suicide, referred to as *medical malpractice stress syndrome* (Sanbar & Firestone 2007). Additionally, being sued may affect how physicians practice, for example, by leading them to order more tests and treatments for the purpose of avoiding future litigation or changing care patterns in other ways. However, no study outside of obstetrics has evaluated how being named in a malpractice claim changes physicians' practice patterns, including whether they practice more defensively.

Defensive medicine and the price of the system to administer the medical liability process costs an estimated $56 billion per year in the United States (Mello et al. 2010). These estimates are derived indirectly from studies of how physicians perceive they may order tests differently due to malpractice fear. High costs of defensive medicine are used to justify tort reforms: the changing of rules around how and when physicians can be named in a claim and how claims are adjudicated. Yet, the literature on practice changes after tort reform is mixed (Gandhi et al. 2006, Karcz et al. 1990, Brown et al. 2010, Pines et al. 2009, Pines et al. 2010, Studdert et al. 2016). Some studies find a degree of practice change at the community-level after tort reform while others find no such effect (Waxman et al. 2014, Farmer et al. 2018, Gimm 2010, Currie & MacLeod 2006, Moghtaderi et al. 2019).

Emergency physicians are ideal subjects for the study of this question. They perceive themselves at high risk for malpractice claims because they care for an undifferentiated patient population, lack a previous relationship with patients, and have both limited time and resources (Carrier et al. 2010). In addition, there are few barriers for emergency physicians to practice defensively compared with those commonly encountered in other specialties: there is no insurance preauthorization, and emergency physicians have a high degree of autonomy in clinical decisions. Finally, they treat large numbers of diverse patients, which could allow the assessment of moderate changes in clinical practice patterns.

We study how commonly measured markers of provider practice were affected after physicians were named in a malpractice lawsuit. For example, if they order more tests after being sued, this should affect relative value units (RVUs) per visit; if they are more likely to admit marginal patients (those where the decision of whether or not to admit is a function of physician judgement rather than obvious physiologic criteria) to the hospital, this should affect hospital admission rates; and if they spend more time with each patient, this should affect RVUs per hour, discharge length of stay, and assessed patient experience. Our objective is to evaluate whether emergency physicians' clinical practice patterns change after being named in a malpractice claim—and if so, how—focusing on common, aggregate markers of emergency physician practice, including care intensity (RVUs per visit and hospital admission rate), care speed (length of stay and RVUs per hour), and how patients assess their experiences with their emergency physician (Press Ganey percentile rank). Since previous research showed that these practice patterns are not associated with the probability of being named in a malpractice claim (Carlson et al. 2018), it is important to understand if physicians perceive that changing practice patterns may mitigate the risk of a lawsuit. Such perceptions could have widespread financial implications and may warrant interventions by management.

4.2 Data and Variables

In order to assess changes in physician practice patterns after being named in a malpractice claim, we conduct a retrospective study using data from a national emergency physician management network (EPMN), with data from 59 emergency departments (EDs) in 11 states between January 2010 and December 2015. The EPMN under study provided the data, but did not control the research question, analyses, or decision to publish results. This group maintained its own risk-retention program (a privately owned entity in which policyholders are also owners) during the study period and captured all malpractice claims against physicians. We use a difference-in-differences approach to compare physicians who were named in a malpractice claim during the study period to matched controls who were not named in malpractice claims between 2010 and 2015. Difference-in-differences methods are commonly used to estimate causal effects of external shocks (Lechner 2011): here, the shock from being named in malpractice claim. A "malpractice claim" was defined as any filed malpractice lawsuit during the study period regardless of case disposition. Because simply being named in a malpractice claim induces significant stress for the provider, we elected to examine behavior change after being named in a claim rather than after the verdict (Sanbar & Firestone 2007). Also, because claim duration can vary substantially and named physicians may acquire information about claims at various points in the lawsuit in ways we could not observe, we examined changes in clinical practice patterns relative to the claim filing date.

Fig. 8 provides a flowchart for how the sample for the main analysis and sensitivity analyses were determined. Our initial sample included ED data from 105 facilities which were staffed by the EPMN and corresponded to 1,558 emergency physicians, 14,671,102 ED visits, and 628,942 ED clinical shifts from January 2010 through December 2015. Ninety-nine of these emergency physicians were named in 107 lawsuits. Physicians named in a malpractice suit were identified from a database of malpractice claims maintained by the ED group from June 2010 through May 2014.

Next, we exclude outlying or extreme visits/shifts and shifts that are likely to be administrative in nature as follows. We exclude visits resulting in greater than 20 relative value units (RVUs), clinical shifts $\leq$ 4 hours (likely administrative based on typical ED staffing patterns), clinical shifts with average RVUs/hour $\geq$ 30, and clinical shifts with average patients/hour $\geq$ 10. This excludes 2.1% of visits and 6.3% of clinical shifts.

We require at least 4 months of data within this ED group for each named and control physician preceding the lawsuit date for inclusion in this study, based on prior research indicating that this is a reasonable period of time during which an emergency physician adapts to the local ED environment and the physician's practice pattern stabilizes (Carlson et al. 2018). This minimum pre-lawsuit period also provides a minimum period in which we can assess whether pre-treatment trends are parallel for named and control physicians, a core assumption of the difference-in-differences methodology.

All Data: January 2010 – December 2015		
1,558 Emergency Physicians (99 Named, 107 Claims)	14,671,102 Emergency Department (ED) Visits	628,942 ED Clinical Shifts

Exclusion Criteria:			
Criteria (applied sequentially)	**Physicians Removed**	**Visits Removed**	**Shifts Removed**
1. Visits with > 20 RVUs	0	8,496	0
2. Shifts ≤ 4 hours	25 (0 Named)	219,481	38,433
3. Shifts with RVUs/Hour ≥ 30	0	66,094	743
4. Shifts with Visits/Hour ≥ 10	0	13,318	141
Total Removed	25 Physicians (0 Named, 0 Claims)	307,389 Visits (2.1%)	39,317 Shifts (6.3%)

Match Named Physicians to Controls in Same Facility-Month using Propensity-Matching, ≥ 4 Months Data Both Pre- and Post-Claim Date for Inclusion of Named and Control Physicians

Data for All Visits Study Analysis, January 2010 – December 2015		
205 Physicians (65 Named, 69 Claims)	1,674,734 Visits	75,464 Shifts

Data for Body System/Clinical Issue, Failure to Diagnose, and Non-Failure to Diagnose Study Analyses, January 2010-December 2015			
Visits Within Same Body System/Clinical Issue of Malpractice Claim	198 Physicians (63 Named, 67 Claims)	1,611,852 Visits	72,863 Shifts
Claim Allegation - Failure to Diagnose	140 Physicians (42 Named, 46 Claims)	1,130,544 Visits	51,278 Shifts
Claim Allegation – Non-Failure to Diagnose	75 Physicians (23 Named, 23 Claims)	574,212 Visits	25,982 Shifts

Figure 8 Flow Diagram for Sample Definition

We also require at least four months of data following the lawsuit date for named and control physicians; thus, a minimum total of eight months of continuous data for each named physician. The minimum post-lawsuit period provides a minimum period during which we can assess whether any post-lawsuit effects appear gradually over time. These requirements for minimum observation time and suitable controls limit the named physician sample to 65 physicians, named in 69 lawsuits, and the potential control physician sample to 387.

4.2.1 Propensity Score Matching

We then match the remaining named (treated) physicians to at most three control physicians who practiced in the same facility in the same month using propensity-score matching on the pre-treatment period. Two-month averages were computed for practice variables (RVUs per hour, RVUs per visit, visit length for discharged patients, admit rate, total monthly patients, and Press Ganey (PG) percentile ranks) in Months -1 and -2 and in Months -3 and -4 for each physician. Variables used in matching include these two-month averages and physician demographics (physician age, physician race, physician gender, emergency medicine board certification, and years since residency completion). Since two of the named physicians did not have PG percentile ranks in Months -4 through -1, the matches for those two physicians were based on physician demographics and the two-month averages of RVUs per hour, RVUs per visit, visit length for discharged patients, admit rate, and total monthly patients. R's *optmatch* package was used to match treated physicians to controls working in the same facility. Matching was done with replacement, and we specified a caliper width of 2, resulting in 140 control physicians after propensity matching. Table 9 summarizes the differences between the treatment and the control physicians to which they were matched.

Since treated physicians often worked at more than one facility in the same month, with greater than 20% of clinical hours at a second ED in 22.2% of physician-months, we included data from all facilities in which a treated physician worked. In the case of multiple facilities worked at in a month, we matched the named physician's visits and shifts at facility 1 to control physicians' shifts and visits at facility 1 and matched the named physician's visits and shifts at facility 2 to control physicians' shifts and visits at facility 2, etc. If control physicians worked at more than one facility in a given month, we used only their shifts at the same facility as the named physicians for comparison. While this resulted in multiple observations per physician per month, it better reflects the practice patterns of emergency physicians and accounts for variations among the facilities. By controlling for a physician-by-facility fixed effect in the regression models, we essentially treated each physician-facility pair as separate for comparisons in outcome measures pre- and post-claim.

Table 9 Comparison of Treatment and Control Physicians

Difference (Control - Treatment)		
Variable	**Mean**	**SD**
Physician Age	0.01	11.75
Race White (1/0)	-0.02	0.45
Race Black (1/0)	0.00	0.00
Race Other (1/0)	0.02	0.45
Gender (1/0)	0.01	0.61
Board Certified (1/0)	-0.02	0.29
Years Since Residency Completion	0.38	11.89
PG Score (Months -1 & -2)	3.31	41.13
RVU/Patient (Months -1 & -2)	-0.03	0.88
LOS Discharged (Months -1 & -2)	0.01	1.04
Admit Rate (Months -1 & -2)	0.00	0.10
RVUs/Hour (Months -1 & -2)	-0.28	2.57
Average Number of Patients per Month (Months -1 & -2)	-6.73	152.75
PG Score (Months-3 & -4)	3.72	37.85
RVUs/Patient (Months-3 & -4)	-0.02	0.92
LOS Discharged (Months-3 & -4)	0.07	0.93
Admit Rate (Months-3 & -4)	0.00	0.10
RVU/Hour (Months-3 & -4)	-0.30	2.27
Average Number of Patients per Month (Months-3 & -4)	-3.04	143.93

While eight months of data were required for inclusion, we tracked physicians' practice patterns from 12 months preceding the filing date through 23 months following the lawsuit filing date. We chose this timeframe because we observe a substantial drop in the number of named physicians before and after these time bounds. Table 10 shows sample size by event month. For the four physicians who were named in multiple (two) claims during the study period, we treated each claim as a separate event, but we assessed robustness if only the first claim for these physicians was studied and found the results were comparable to those reported.

Table 10 Physician Counts by Month for Difference-in-Differences Analysis

	All Visits		Body System/Clinical Issue Analysis		Failure to Diagnose		Non-Failure to Diagnose	
Month Number (0 = Filing Date)	Named Physicians	Control Physicians	Named Physicians	Control Physicians	Named Physicians	Control Physicians	Named Physicians	Control Physicians
-12	53	87	51	84	32	52	21	40
-11	53	89	51	86	33	54	20	40
-10	53	96	51	93	33	61	20	40
-9	56	108	54	105	35	72	21	43
-8	57	113	55	110	36	77	21	43
-7	62	124	60	120	39	83	23	49
-6	63	129	61	124	40	86	23	50
-5	63	128	61	124	41	88	22	48
-4	65	137	63	132	42	95	23	52
-3	64	133	62	129	42	94	22	47
-2	65	131	63	127	42	89	23	52
-1	65	139	63	134	42	97	23	52
0	65	137	63	132	42	96	23	51
1	65	139	63	134	42	97	23	52
2	65	134	63	128	42	95	23	49
3	64	131	62	125	41	90	23	51
4	63	130	61	125	40	91	23	49
5	63	132	61	127	40	93	23	49
6	64	127	62	121	41	89	23	48
7	59	119	57	114	38	82	21	45
8	59	118	57	113	38	82	21	44
9	60	111	58	106	38	77	22	43
10	57	105	55	100	36	72	21	41
11	57	101	55	96	36	71	21	38
12	56	95	54	90	35	65	21	38
13	54	89	52	84	33	59	21	38
14	54	91	52	86	34	61	20	38
15	53	89	51	84	34	60	19	36
16	52	85	50	80	33	57	19	35
17	50	80	48	76	32	53	18	32
18	48	78	46	74	30	52	18	31
19	48	79	46	75	30	51	18	32
20	47	73	45	69	30	50	17	27
21	45	68	43	64	29	50	16	22
22	43	64	41	60	28	46	15	22
23	45	67	43	62	29	43	16	27
~~24~~	~~15~~	~~23~~	~~15~~	~~23~~	~~9~~	~~17~~	~~6~~	~~7~~

All periods are measured in event time, relative to the lawsuit filing date, except as specified below for Press Ganey. For each claim, the 30 days prior and the filing date are treated as month 0 and the 90 days prior and the filing date are treated as quarter 0. For example, if the filing date was June 26, 2013, month zero for that physician would begin on May 27, 2013 and end on June 26, 2013, and quarter 0 would run from March 28, 2013 through June 26, 2013.

4.2.2 Summary Statistics

A total of 1,674,734 ED visits involving 205 emergency physicians (65 named in 69 malpractice claims and 140 matched controls) at 59 EDs in 11 US states from 2010 to 2015 met inclusion criteria (Fig. 8). Table 11 shows that the most prevalent claims related to a neurologic condition (n=17, 24.6%), the majority alleged a failure to diagnose (n=46, 66.7%), and the most common filing state was California (n=21, 30.4%). Of 69 claims, 47 resulted in voluntary dismissal of the named emergency physician by the plaintiff, 17 were settled, and five were tried, with two plaintiff victories. Named and control emergency physicians were balanced on pre-exposure demographic and operational factors after propensity-score matching (Table 9). We did not observe differential dropping out from the sample of named physicians after their being named in a malpractice claim.

Table 11 Summary Statistics for Physician and Facility Characteristics

	All Physicians	*Named Physicians (N=65)*	*Matched Control Physicians (N=140)*	*Standardized difference in two groups (95% CI)*
Physician Characteristics				
Mean Age at First Included Physician-Month	44.3 (9.6)	45.4 (9.5)	43.7 (9.7)	0.18 (-0.12,0.47)
Mean Years since Residency Completion at First Included Physician-Month	11.4 (9.4)	12.5 (9.6)	10.9 (9.4)	0.18 (-0.12, 0.47)
% Male	143 (69.8%)	49 (75.4%)	94 (67.1%)	0.18 (-0.11, 0.48)
% Board Certified in Emergency Medicine	192 (93.7%)	61 (93.8%)	131 (93.6%)	0.01 (-0.28, 0.31)
Operational Characteristics (for included months under study)				
Total ED Visits as Attending Physician of Record	1,674,734	565,823	1,108,911	
Per-physician Mean Monthly ED Visits as Attending Physician of Record	289.6 (153.5)	292.5 (131.7)	288.4 (162.2)	0.03 (-0.02,0.08)
Total Physician-Months	6,168	2,074	4,094	
Mean Number of Physician-Months Per Physician	30.1 (10.4)	31.9 (7.4)	29.2 (11.4)	0.28 (-0.02, 0.57)
ED Characteristics (same for named and control physicians, by construction)				
ED Mean Annual Visit Volume	43,220 (20,540)			
Mean Annual ED Admission Rate	16.4% (7.1%)			
Percentage with Trauma Designation (Level 1-4)	17 (28.8%)			
Percentage in Academic Hospitals	9 (15.3%)			
Percentage with Emergency Medicine Residency Program	8 (13.6%)			
Malpractice Claim Characteristics (n=69)				
Primary Body System/Clinical Issue of Malpractice Claim (%) [#]				
Blood/Lymphatic	1 (1.4%)			
Cardiovascular	8 (11.6%)			
ENT	1 (1.4%)			
Endocrine	2 (2.9%)			
Eye	1 (1.4%)			
Gastrointestinal	9 (13.0%)			
Genitourinary	3 (4.3%)			
Skin/Wound	5 (7.2%)			
Neurologic	17 (24.6%)			
OB-GYN	5 (7.2%)			
Orthopedic	7 (10.1%)			
Psychiatric	3 (4.3%)			
Respiratory	6 (8.7%)			
Non-Organ-System Based Clinical Issue (Medical Battery)	1 (1.4%)			
Malpractice Claim Allegation (%) [#]				
Failure to Diagnose	46 (66.7%)			
Non-Failure to Diagnose	23 (33.3%)			
Claim Disposition (%)				
Physician Voluntarily Dismissed from Claim	47 (68.1%)			
Out-of-Court Settlement	17 (24.6%)			
Trial - Defense Verdict	3 (4.3%)			
Trial - Plaintiff Verdict	2 (2.9%)			
State (%)				
CA	21 (30.4%)			
CT	5 (7.2%)			
HI	2 (2.9%)			
IL	5 (7.2%)			
NC	7 (10.1%)			
NV	6 (8.7%)			
NY	2 (2.9%)			
OH	15 (21.7%)			
OK	2 (2.9%)			
PA	2 (2.9%)			
WV	2 (2.9%)			

[#] Based on the internal classification by the risk management department of this national emergency physician group and review by collaborating physicians.

4.2.3 Outcome Measures

For each outcome measure, we studied all visits within the same ED with the named and control physicians as the attending physician of record. We used only within-ED comparisons because emergency medical practice is constrained by the capabilities of the ED and affiliated hospital (e.g., trauma center, stroke center, available consultation services). Because treated physicians often worked at more than one facility in the same month (greater than 20% of clinical hours at a second ED in 22% of physician-months in our sample), we included data from all facilities where a treated physician worked. If a named physician worked at multiple facilities in a single month, we matched the named physician's visits and shifts at facility 1 to control physicians' shifts and visits at facility 1 and matched the named physician's visits and shifts at facility 2 to control physicians' shifts and visits at facility 2, etc. Although this resulted in multiple observations per physician per month, we thought it better reflected the practice patterns of emergency physicians and accounted for variations among the facilities. We used physician-by-facility fixed effects in all regression models, allowing us to evaluate changes in practice patterns for the same physician in the same facility. Because previous work has not shown an association between working at multiple facilities and the likelihood of being named in a malpractice claim, we did not separately study named physicians who worked in multiple facilities (Carlson et al. 2018).

Since many malpractice claims against emergency physicians involve issues related to failure to diagnose and initiate treatment (Daniels et al. 2017, Colaco et al. 2015, DePasse et al. 2017), we also separately studied claims alleging failure to diagnose (46 claims) and other allegations (23 claims). We conducted this secondary analysis because of the focal importance in emergency medicine of early diagnosis and intervention for acute conditions and the judgment of the emergency physicians on the research team that a failure-to-diagnose claim might affect physician behavior more significantly than other claims. We also studied post-claim visits involving the same body system or clinical condition as the relevant malpractice claim as a secondary analysis. This was based on the potential that knowledge of the previous claim could affect behavior for these types of future visits more strongly than others. For the body system and clinical condition analysis, per-shift outcomes (RVUs/hour and Press Ganey percentile rank) were not applicable.

We selected outcome measures that were observable in our data and could plausibly be affected by practice changes after a malpractice claim. Specifically, we studied five outcome measures, including two measures of care intensity (hospital admission rate [%] and RVUs per visit), two measures of care speed (RVUs per hour and length of stay in hours for discharged patients), and one measure of subjective patient experience (monthly physician Press Ganey percentile rank). The first four operational measures were computed from visit- and shift-level data. Press Ganey percentile ranks were reported monthly. We lagged physician monthly Press Ganey percentile ranks by two months, according to empirical analysis of the average time needed for survey return and processing. That is, the Press Ganey percentile rank that we matched to month 0 in our data set was recorded in month 2 in the raw data. For example, if the filing date for a named physician was June 26, 2013, we used the Press Ganey percentile rank that the ED group received in August 2013 for month 0 and the rank for September 2013 for month 1, so that the Press Ganey rank from a given month would correspond more closely to visits from month 0.

4.3 Methods and Hypotheses

4.3.1 Methods

We used several graphical and regression approaches to determine whether named physicians' practice behavior changed after they were sued for malpractice. These include "simple difference-in-differences" regressions, which assume a onetime change in outcome, occurring immediately after the shock. Additionally, distributed lag regressions allow for the gradual emergence of a treatment effect. We conduct all analyses in event time relative to the filing date for each named physician.

First, we performed simple difference in difference (DiD) regressions that assume the effect of being named in a malpractice claim occurs immediately following the filing date. This DiD model is specified in Eq. (11). For simplicity, one equation is shown, but we run five separate regressions, one for each of the outcome variables discussed in the previous section.

$$Y_{ift_1} = \beta_{0_1} + \beta_{1_1}\,\text{Named}_i + \beta_{2_1}\,\text{AfterClaim}_{ift} + \beta_{3_1}\left(\text{Named}_i \times \text{AfterClaim}_{ift}\right) + \beta_{4_1}(i \times f) + \varepsilon_{ift_1} \quad (11)$$

Here Y is the outcome, i indexes physicians (for physicians who received two claims each, these are treated as separate events), f indexes facilities, and t indexes month relative to the filing date. Therefore, Y_{ift_1} represents the outcome measure for physician i at facility f in month t. $Named_i$ is a dummy variable which indicates whether physician i was been named in a malpractice claim during the study period. $AfterClaim_{ift}$ is a dummy variable which indicates if time t is after the relevant filing date. Therefore, β_{3_1}, the coefficient for the $\text{Named}_i \times \text{AfterClaim}_{ift}$ interaction, s the DiD estimate that represents the change in the outcome variable after being named in a malpractice claim relative to control physicians. We include the interaction between physician and facility fixed effects (i*f) to control for unobserved, time-constant factors specific to a physician-facility pair, and standard errors were clustered on facility. This regression model does not include patient covariates (e.g., age, gender, insurance type) because there should not be systematic differences in the characteristics of patients treated by named versus control physicians within the same facility at the same time, given the nature of emergency department care, where providers have no ability to predict or select which patients to see and by law and clinical practice must see any patient who presents to the ED. We similarly did not include physician-level covariates in the regression equations because named and control physicians were matched on these covariates in the pre-treatment period and also based on previous work suggesting no likely impact in this regard (Carlson et al. 2018).

We also plot leads and lags graphs showing pre- and post-treatment trends by quarter. These graphs provide evidence on whether the pre-treatment trends are parallel as well as evidence on when any apparent treatment effect appears, relative to the lawsuit filing date. While the parallel trends assumption cannot be directly tested, plausibility can be assessed during the pre-shock period with leads-and-lags graphs, which provide estimates of the treatment effect in each period, both before the shock (when there should be no treatment effect) and afterward. The graphs are evaluated visually for evidence of nonparallel trends.

To investigate whether pre-treatment trends differ between treatment and control physicians, we use a leads and lags model in event time, treating the 90 days prior and the filing date as time zero for each named physician, and track forward up to 23 months and backward up to 12 months in the same manner. As we included data for months 22 and 23, we averaged the data over these months to determine the data for quarter 8, and the data for quarter -4 was determined by month -12. The monthly records were assigned to the appropriate quarter as follows:

Month	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Quarter	-4	-3			-2			-1			0			1			2			3			4			5			6			7			8	

Despite the uneven quarter lengths at the extremes of our study period, we compared the quarterly leads and lags graphs to monthly graphs (not shown) and found that they yield equivalent results. Again, we include a physician-by-facility fixed effect and use cluster-robust standard errors, clustered on facility. Eq. (12) specifies the model used for quarterly leads and lags models.

$$Y_{ift_2} = \beta_{0_2} + \beta_{1_2}\,\text{Named}_i + \sum_{k=-4}^{8}[\beta_{2_2}^{k}\,k + \beta_{3_2}^{k}(\text{Named}_i \times k)] + \beta_{4_2}(i \times f) + \varepsilon_{ift_2} \quad (12)$$

Here, k indexes “event time” relative to the filing date (quarterly). We include 4 leads (one year) and 8 lags (23 months) in our specification. Time zero is treated as the reference category, and each $\beta_{2_2}^{k}$ estimates the quarter *k* effect within the control group relative to time zero. Each $\beta_{3_2}^{k}$ coefficient estimates the DiD for quarter *k* relative to time zero (plotted in Fig. 9 for the analysis which includes all visits). Again, this equation is used for each of the five outcome measures.

Finally, we use distributed lags to test for changes in emergency physician practice patterns after a malpractice claim. While the leads-and-lags models estimate quarterly (or monthly) coefficients, relative to a base year, the distributed lag model estimates quarterly incremental changes relative to a pre-claim mean. Eq. (13) specifies the distributed lags model with six quarterly lags.

$$Y_{ift_3} = \beta_{0_3} + \beta_{1_3}\,\text{Named}_i + \sum_{k=0}^{6}[\beta_{2_3}^{k}\,k + \beta_{3_3}^{k}(\text{Named}_i \times \text{lag}_k)] + \beta_{4_3}(i \times f) + \varepsilon_{ift_3} \quad (13)$$

Here the first treatment lag (lag_1) equals 1 in the quarter of the malpractice claim and subsequent quarters; lag_k turns on in the k^{th} quarter after the claim and stays on. Thus, the coefficient on $\text{Named}_i \times \text{lag}_1$ (i.e., $\beta^1_{3_3}$) estimates the effect of the lawsuit in the claim quarter, and the coefficients $\beta^k_{3_3}$ ($k = 2,3,\ldots,6$) estimate the additional effect in each subsequent quarter. The overall treatment effect is thus the sum of these coefficients.

4.3.2 Hypotheses

We chose the aforementioned outcome measures because they are commonly used, aggregate measures of emergency physician performance, which we considered likely to be influenced by a change in clinical practice resulting from being named in a malpractice claim. For example, a natural response might be to slow down and practice more deliberately, which should be detectable as fewer RVUs per hour and longer discharge length of stay. Emergency physicians may decrease their threshold for ordering diagnostic tests after being named in a claim, and ordering more tests should lead to increased RVUs per visit. Providers also may be more cautious in deciding which patients to admit, resulting in higher hospital admission rates. These practices correspond to practicing more defensively after a malpractice claim and lead us to the following hypotheses:

H3. *Physicians practice more defensively after being named in a malpractice claim. That is,*

- ***a.*** $\beta_{3_1} < 0$ *when Y measures RVUs per hour.*
- ***b.*** $\beta_{3_1} > 0$ *when Y measures length of stay for discharged patients.*
- ***c.*** $\beta_{3_1} > 0$ *when Y measures admission rate.*
- ***d.*** $\beta_{3_1} > 0$ *when Y measures RVUs per visit.*

In addition, emergency physicians may change the ways in which they interact with patients, according to assumptions about malpractice risk, which could affect patients' rating of their experiences. For instance, physicians may alter the ways in which they interact with patients if they feel that a communication issue contributed to being named in a malpractice claim. Improving communication (either the content or the delivery of that content) might be seen as a way to reduce the risk of a lawsuit. While we cannot directly test for changes in physicians' soft skills or bedside manner, we believe that such changes would be reflected in patient experience scores. Consequently, we hypothesize that patient experience scores will increase after a physician is named in a malpractice claim.

H4. *Physicians improve communication with patients after being named in a malpractice claim, and thus patient experience scores increase.* $\beta_{3_1} > 0$ *when Y measures patient experience.*

These hypotheses are tested in the following section.

4.4 Analyses and Results

According to the leads and lags graphs (see Fig. 9), pretreatment trends were reasonably parallel for named versus control physicians for all outcomes, which supports the appropriateness of the core difference-in-differences assumption of parallel trends. The results of the simple DiD regressions suggest that after being named in a malpractice claim, emergency physicians had improved patient experience scores relative to control physicians, which supports hypothesis H4. Specifically, the average monthly physician Press Ganey percentile rank increased by 6.52 (95% confidence interval [CI] 0.67 to 12.38). Mean monthly Press Ganey percentile ranks increased for named physicians in the first quarter after the filing date and remained elevated during the two-year treatment period (Fig. 9).

Outcomes for care intensity and care speed showed no significant change for named versus control physicians: RVUs per visit (–0.02; 95% CI –0.10 to 0.05), RVUs per hour (–0.07; 95% CI –0.59 to 0.46), visit length of stay for discharged patients (–0.01; 95% CI –0.10 to 0.09), and hospital admission rate (–0.008; 95% CI –0.018 to 0.002). That is, consistent changes in these outcomes were not observed, and we do not find support for hypotheses H3(a)-(d). Complete results are detailed in Table 12 (Panel A) and Fig. 9.

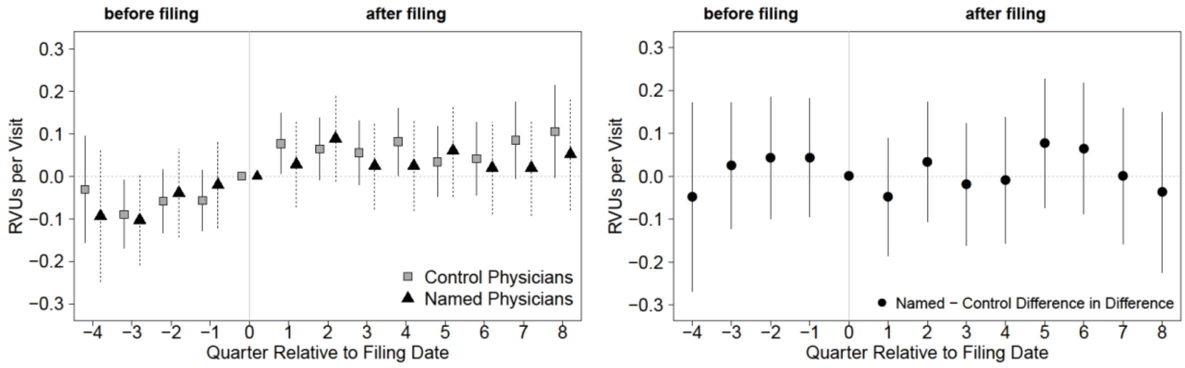
We then perform secondary analyses on subsets of the data. First, we consider the claims alleging failure to diagnose (46 claims) and other allegations (23 claims) separately to assess whether a failure-to-diagnose claim might affect physician behavior more significantly than other claims. The increase in average monthly Press Ganey percentile ranks for named physicians in the subset of 46 failure-to-diagnose claims, relative to controls, was 10.52 percentile ranks (95% CI 3.72 to 17.32). This increase in scores began shortly after filing (Fig. 10) and continued to increase in the post-claim period. This was confirmed in the distributed lag analyses (Table 13 [Panel A]). As with the all-visits analysis, other outcomes within the subset of failure-to-diagnose claims were similar between named and control physicians (Fig. 10, Table 12 [Panel B] and Table 13 [Panel B]). For the 23 non-failure-to-diagnose malpractice claims, there was no evidence of a change in either monthly Press Ganey percentile ranks or other outcomes (Fig. 11, Table 12 [Panel C] and Table 13 [Panel C]).

We then study post-claim visits involving only the same body system or clinical condition as the relevant malpractice claim to determine if behaviors for these types of visits are influenced more strongly than others. Note that for the body system and clinical condition analysis, per-shift outcomes (RVUs/hour and Press Ganey percentile rank) were not applicable. Prior to study analysis, collaborating emergency physicians classified ICD-9 codes into body system and clinical condition groups using diagnosis codes within the major disease categories of the ICD-9 Tabular Index (for visits from January 2010 through September 2015). Emergency medical practice also often involves ruling out acute conditions rather than definitive diagnosis. Therefore, we treated ICD-9 diagnosis codes within symptoms, signs, and ill-defined conditions (780-799) and injury and poisoning (800-999) as non- exclusive – these codes could be placed in more than one body system/clinical condition category based on physician authors consensus. ICD-10 codes came into application in October 2015. Only three named physicians had visits during the ICD-10 time period (October-December 2015). For these named physicians, the physician authors reviewed their visits' primary ICD-10 codes and classified into body system or clinical condition. The classification of ICD-9 and ICD-10 codes is shown in Appendix Table 6.

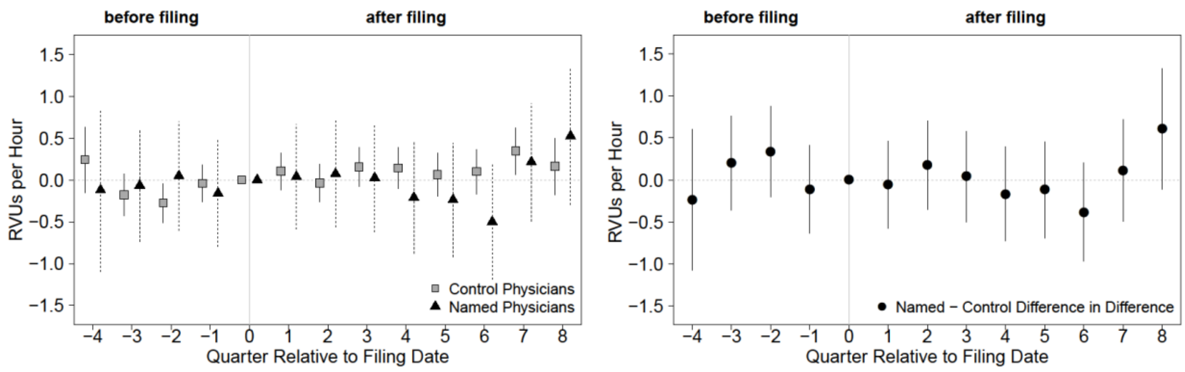
The EPMN which provided the data for this study classifies malpractice claims by body system or clinical issue involved as well as by malpractice allegation (failure-to-diagnose and non-failure-to-diagnose categories). Prior to study analysis, the physician authors reviewed the ED group's body system/clinical issue classification for each malpractice claim and re-adjusted three claims, for which there was physician authors consensus that the presenting condition of the patient was clearly in another body system, based on the malpractice claim details: one claim was moved from Neurologic to Psychiatric; one from OB-GYN to Integument/Wounds; and one from Neurologic to Cardiovascular.

When analyzed within the same body system or clinical condition as the malpractice claim (Fig. 12, Table 12 [Panel D] and Table 13 [Panel D]), outcomes were similar between named and control physicians: RVUs per visit (0.03; 95% CI –0.08 to 0.14), visit length (–0.02; 95% CI –0.14 to 0.10), and hospital admission rate (–0.011; 95% CI –0.027 to 0.006). In this secondary analysis, we could not assess changes in RVUs per hour or monthly Press Ganey percentile ranks as those outcomes rely on all of a physician's visits.

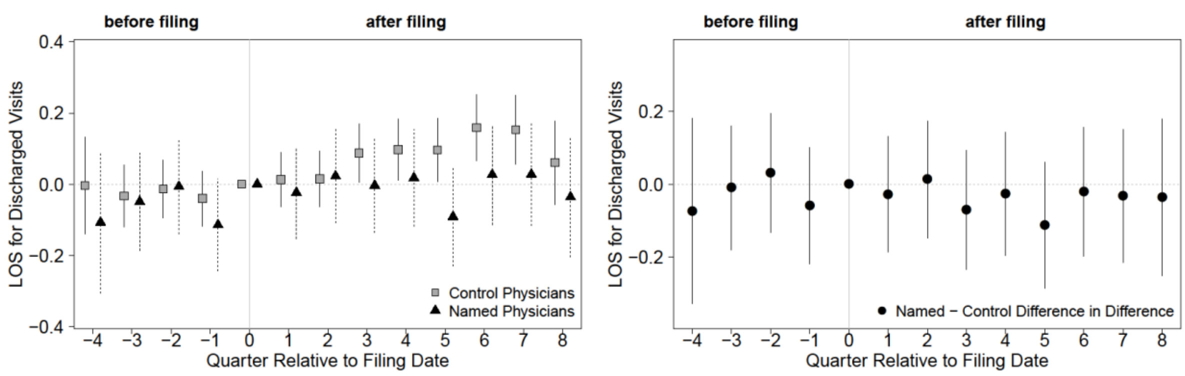
Based on the primary and secondary analyses, we do not find any evidence that physicians practice more defensively after a malpractice claim with regard to RVUs per hour, discharge length of stay, RVUs per visit, or admission rates. That is, there is no support for hypotheses H3(a)-(d). However, the data suggest that patient experience scores do increase after a physician is named in a malpractice claim, supporting hypothesis H4. While we cannot prove the reason for the increase in patient experience scores (e.g., changes in communication or demeanor), these results suggest a difference in patients' perceptions of the physicians.



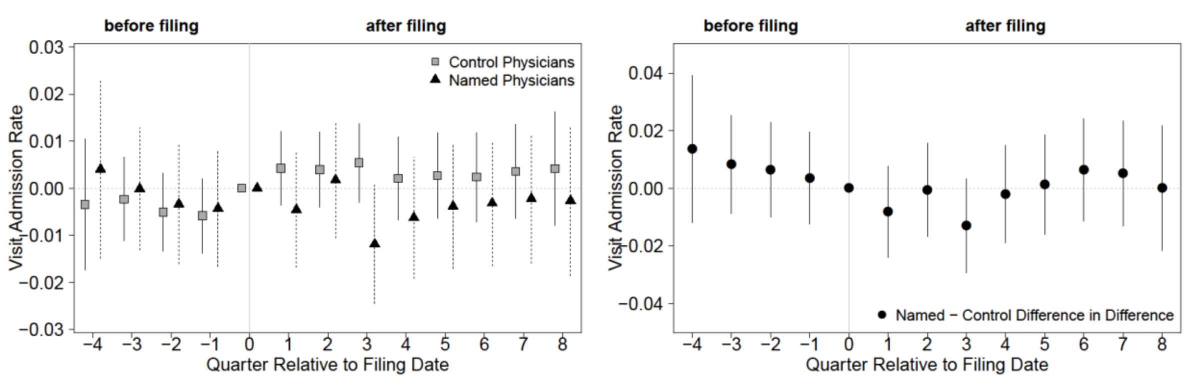
(a) RVUs/visit



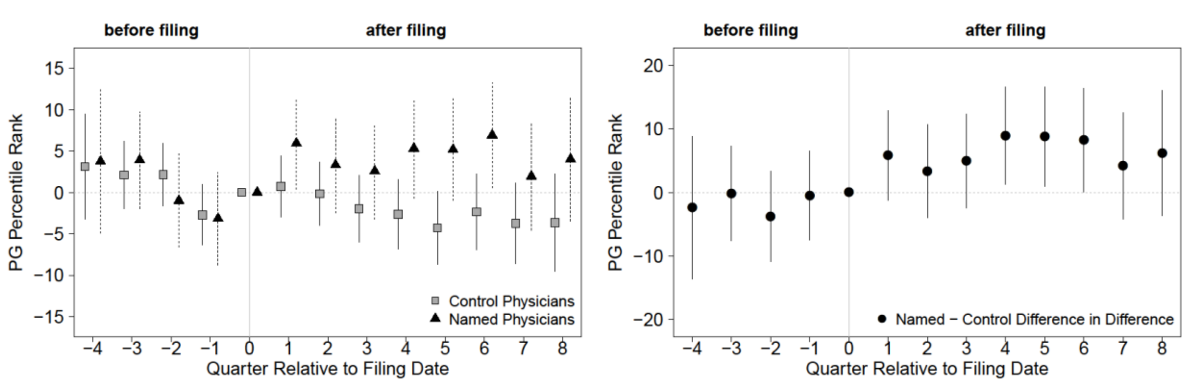
(b) RVUs/hour



(c) Visit Length of Stay (LOS) for Discharged Patients (hours)



(d) Proportion of Visits Resulting in Hospital Admission



(e) Physician Press Ganey Percentile Rank

Figure 9 Leads and Lags Figures for Outcome Measures: All Visits

Table 12 Difference-in-differences (DiD) Analysis: Named versus Control Physicians

Panel A. All Visits

	Named Physicians (N=65)		*Control Physicians (N=140)*			
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
Relative Value Units (RVUs)/Visit	3.68 (0.81)	3.72 (0.85)	3.65 (0.77)	3.73 (0.79)	0.01 (-0.11, 0.14)	-0.02 (-0.10, 0.05) SE = 0.04
RVUs/Hour	10.03 (3.62)	10.07 (5.18)	9.56 (2.70)	9.52 (2.77)	0.52 (-0.20, 1.24)	-0.07 (-0.59, 0.46) SE = 0.26
Visit Length (hrs.)	2.67 (1.18)	2.73 (1.19)	2.62 (1.08)	2.67 (1.11)	0.06 (-0.12, 0.23)	-0.01 (-0.10, 0.09) SE = 0.05
Proportion of Visits Resulting in Hospital Admission	0.188 (0.109)	0.182 (0.116)	0.186 (0.093)	0.191 (0.097)	-0.005 (-0.02, 0.01)	-0.008 (-0.018, 0.002) SE = 0.005
Monthly Physician Press Ganey Percentile Rank	51.61 (36.88)	57.97 (36.13)	54.91 (37.05)	55.42 (37.35)	0.37 (-3.53, 4.27)	6.52 (0.67, 12.38) SE = 2.92
Physician months	731	1,379	1,602	2,704		
Visits	207,502	358,321	428,846	680,065		

Panel B. Failure to Diagnose Claims: All Visits

	Named Physicians (N=42)		*Control Physicians (N=98)*			
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
RVUs/Visit	3.60 (0.70)	3.71 (0.79)	3.60 (0.69)	3.70 (0.74)	0.004 (-0.10, 0.11)	0.01 (-0.07, 0.08) SE = 0.04
RVUs/Hour	9.88 (3.02)	9.49 (2.87)	9.84 (2.88)	9.67 (2.87)	-0.10 (-0.64, 0.44)	-0.39 (-0.88, 0.11) SE = 0.25
Visit Length (hrs.)	2.58 (0.91)	2.69 (0.98)	2.69 (0.99)	2.73 (1.03)	-0.07 (-0.20, 0.06)	0.07 (-0.02, 0.15) SE = 0.04
Proportion of Visits Resulting in Hospital Admission	0.179 (0.099)	0.179 (0.109)	0.188 (0.090)	0.193 (0.095)	-0.012 (-0.026, 0.002)	-0.002 (-0.013, 0.009) SE = 0.006
Monthly Physician Press Ganey Percentile Rank	49.36 (36.42)	60.22 (34.97)	51.40 (37.78)	52.98 (38.05)	3.87 (-1.49, 9.22)	10.52 (3.72, 17.32) SE = 3.37
Physician months	470	896	1,069	1,846		
Visits	141,685	230,895	294,428	462,536		

Panel C. Non-Failure to Diagnose Claims: All Visits

	Named Physicians (N=23)		*Control Physicians (N=52)*			
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
RVUs/Visit	3.85 (0.97)	3.76 (0.96)	3.77 (0.87)	3.77 (0.84)	0.02 (-0.24, 0.28)	-0.03 (-0.21, 0.14) SE = 0.08
RVUs/Hour	10.32 (4.57)	11.21 (7.83)	9.39 (2.28)	9.51 (2.51)	1.43 (-0.35, 3.22)	0.54 (-0.64, 1.73) SE = 0.57
Visit Length (hrs.)	2.83 (1.57)	2.82 (1.51)	2.64 (1.16)	2.68 (1.19)	0.16 (-0.26, 0.58)	-0.12 (-0.38, 0.13) SE = 0.12
Proportion of Visits Resulting in Hospital Admission	0.207 (0.123)	0.189 (0.130)	0.184 (0.091)	0.190 (0.094)	0.008 (-0.16, 0.21)	-0.018 (-0.037, 0.0001) SE = 0.009
Monthly Physician Press Ganey Percentile Rank	55.84 (37.43)	53.35 (38.02)	60.34 (34.99)	58.80 (35.15)	-5.14 (-12.41, 2.13)	-0.27 (-8.80, 8.27) SE = 4.00
Physician months	261	483	571	968		
Visits	65,817	127,426	142,055	238,914		

Panel D. Visits Involving Same Body System/Clinical Issue as Malpractice Claim

	Named Physicians (N=63)[+]		*Control Physicians(N=135)*			
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
RVUs/Visit	3.72 (0.98)	3.83 (1.00)	3.74 (0.97)	3.88 (0.99)	-0.03 (-0.18, 0.11)	0.03 (-0.08, 0.14) SE = 0.05
Visit Length (hrs.)	3.05 (1.99)	3.10 (1.70)	3.03 (1.52)	3.09 (1.62)	0.01 (-0.17, 0.20)	-0.02 (-0.14, 0.10) SE = 0.06
Proportion of Visits Resulting in Hospital Admission	0.179 (0.177)	0.177 (0.175)	0.176 (0.177)	0.198 (0.190)	-0.01 (-0.04, 0.02)	-0.011 (-0.027, 0.006) SE = 0.008
Physician months	712	1,341	1,562	2,616		
Visits	202,503	347,133	412,249	649,967		

[+]Two named physicians were excluded from this sub-analysis, one because the malpractice claim was for a non-body system-based issue (medical battery), and one who had no subsequent visits corresponding to the body system or clinical issue of the malpractice claim.

Table 13 Distributed Lags Analysis

Panel A. Distributed Lags for All Visits (Quarterly)

	Relative Value Units (RVUs)/Visit	RVUs/Hour	Discharge Visit Length	Proportion of Visits Resulting in Hospital Admission	Physician Press Ganey Percentile Rank
Malpractice Claim Month and After	-0.07 [0.07]	-0.04 [0.24]	0.09 [0.07]	-0.008 [0.007]	-0.66 [3.98]
Quarter 1 and After	-0.01 [0.07]	-0.06 [0.19]	-0.10 [0.07]	-0.005 [0.008]	7.44 [4.47]
Quarter 2 and After	0.08 [0.05]	0.18 [0.28]	0.04 [0.06]	0.008 [0.007]	-2.47 [3.86]
Quarter 3 and After	-0.05 [0.06]	-0.16 [0.29]	-0.08 [0.05]	-0.012 [0.006]	1.63 [4.15]
Quarter 4 and After	0.01 [0.05]	-0.19 [0.23]	0.05 [0.07]	0.011 [0.006]	3.95 [4.47]
Quarter 5 and After	0.09 [0.06]	0.05 [0.18]	-0.09 [0.07]	0.003 [0.006]	-0.16 [3.77]
Quarter 6 and After	-0.06 [0.07]	0.18 [0.36]	0.08 [0.05]	0.003 [0.007]	-2.59 [4.37]
Sum	-0.01 [0.07]	-0.02 [0.53]	-0.01 [0.08]	-0.001 [0.009]	7.13 [5.39]
Observations (months)	9,186	9,186	9,186	9,186	7,610
Named Physician months	2,937	2,937	2,937	2,937	2,309
Control Physician months	6,249	6,249	6,249	6,249	5,301
Named Physicians	65	65	65	65	64
Control Physicians	140	140	140	140	138
R^2	0.23	0.30	0.49	0.36	0.21

Panel B. Distributed Lags for 46 Failure to Diagnose Claims: All Visits (Quarterly)

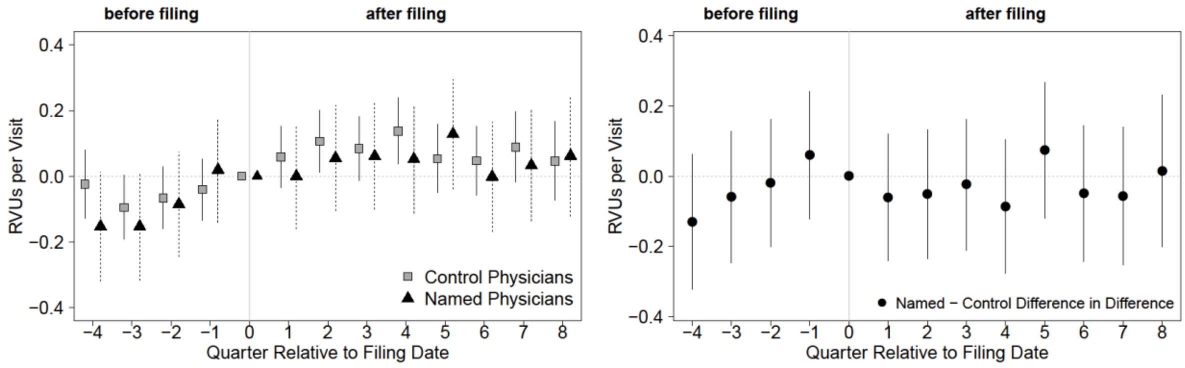
	Relative Value Units (RVUs)/Visit	RVUs/Hour	Discharge Visit Length	Proportion of Visits Resulting in Hospital Admission	Physician Press Ganey Percentile Rank
Malpractice Claim Month and After	-0.01 [0.09]	-0.25 [0.21]	0.03 [0.08]	-0.010 [0.009]	3.06 [5.19]
Quarter 1 and After	-0.05 [0.08]	0.14 [0.23]	-0.04 [0.08]	-0.001 [0.007]	6.77 [5.98]
Quarter 2 and After	0.02 [0.07]	-0.08 [0.25]	0.07 [0.07]	0.009 [0.008]	-0.34 [4.61]
Quarter 3 and After	0.03 [0.08]	0.09 [0.21]	-0.01 [0.07]	-0.018 [0.009]	-0.13 [4.60]
Quarter 4 and After	-0.04 [0.06]	-0.63 [0.29]	-0.01 [0.09]	0.016 [0.007]	3.72 [6.41]
Quarter 5 and After	0.17 [0.10]	0.03 [0.24]	0.00 [0.07]	0.014 [0.008]	-7.54 [6.50]
Quarter 6 and After	-0.09 [0.10]	0.14 [0.23]	0.15 [0.08]	0.004 [0.01]	11.03 [4.45]
Sum	0.04 [0.06]	-0.57 [0.41]	0.18 [0.07]	0.013 [0.010]	16.57 [5.56]
Observations (months)	6,382	6,382	6,382	6,382	5,381
Named Physician months	1,946	1,946	1,946	1,946	1,536
Control Physician months	4,436	4,436	4,436	4,436	3,845
Named Physicians	42	42	42	42	42
Control Physicians	98	98	98	98	98
R^2	0.27	0.35	0.54	0.41	0.20

Panel C. Distributed Lags for Visits Involving Same Body System or Clinical Issue as Malpractice Claim (Quarterly)

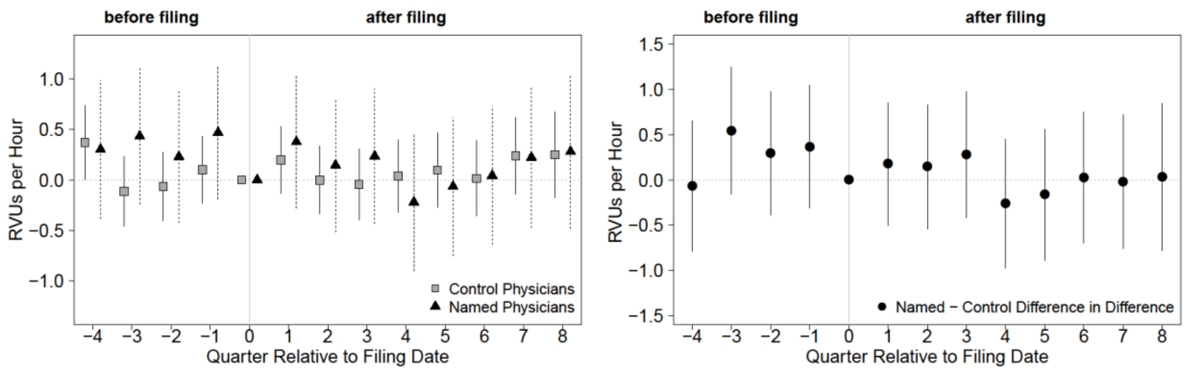
	Relative Value Units (RVUs)/Visit	Discharge Visit Length	Proportion of Visits Resulting in Hospital Admission
Malpractice Claim Month and After	0.03 [0.08]	0.04 [0.12]	0.007 [0.016]
Quarter 1 and After	-0.02 [0.10]	0.00 [0.16]	-0.014 [0.019]
Quarter 2 and After	0.05 [0.08]	-0.11 [0.14]	-0.019 [0.011]
Quarter 3 and After	-0.10 [0.07]	-0.07 [0.13]	-0.003 [0.013]
Quarter 4 and After	0.16 [0.09]	0.19 [0.13]	0.032 [0.017]
Quarter 5 and After	-0.07 [0.08]	-0.06 [0.14]	-0.003 [0.019]
Quarter 6 and After	-0.08 [0.08]	-0.01 [0.12]	-0.015 [0.017]
Sum	-0.03 [0.09]	-0.02 [0.11]	-0.014 [0.013]
Observations (months)	9,139	9,035	9,139
Named Physician months	2,683	2,648	2,683
Control Physician months	6,456	6,387	6,456
Named Physicians	63	63	63
Control Physicians	135	135	135
R^2	0.44	0.45	0.52

Panel D. Distributed Lags for 23 Non-Failure to Diagnose Claims: All Visits (Quarterly)

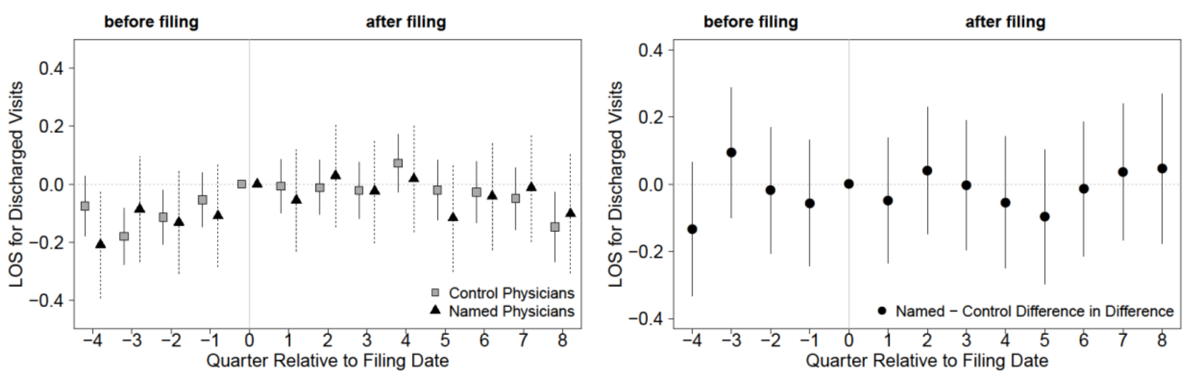
	Relative Value Units (RVUs)/Visit	RVUs/Hour	Discharge Visit Length	Proportion of Visits Resulting in Hospital Admission	Physician Press Ganey Percentile Rank
Malpractice Claim Month and After	-0.20 [0.11]	0.33 [0.62]	0.30 [0.12]	-0.006 [0.013]	-6.23 [5.45]
Quarter 1 and After	0.12 [0.18]	-0.41 [0.26]	-0.29 [0.14]	-0.014 [0.016]	6.60 [5.35]
Quarter 2 and After	0.20 [0.12]	0.69 [0.52]	-0.05 [0.10]	0.012 [0.01]	-7.51 [6.28]
Quarter 3 and After	-0.17 [0.12]	-0.52 [0.80]	-0.19 [0.09]	-0.008 [0.009]	7.87 [7.23]
Quarter 4 and After	0.07 [0.11]	0.57 [0.32]	0.19 [0.12]	0.001 [0.011]	8.03 [6.52]
Quarter 5 and After	-0.03 [0.10]	-0.08 [0.33]	-0.22 [0.19]	-0.011 [0.008]	9.73 [5.33]
Quarter 6 and After	-0.07 [0.13]	0.57 [0.92]	-0.10 [0.07]	-0.001 [0.009]	-29.93 [7.80]
Sum	-0.08 [0.17]	1.15 [1.30]	-0.35 [0.19]	-0.026 [0.018]	-11.45 [8.18]
Observations (months)	3,587	3,587	3,587	3,587	2,988
Named Physician months	991	991	991	991	773
Control Physician months	2,596	2,596	2,596	2,596	2,215
Named Physicians	23	23	23	23	22
Control Physicians	52	52	52	52	50
R^2	0.17	0.28	0.42	0.28	0.25



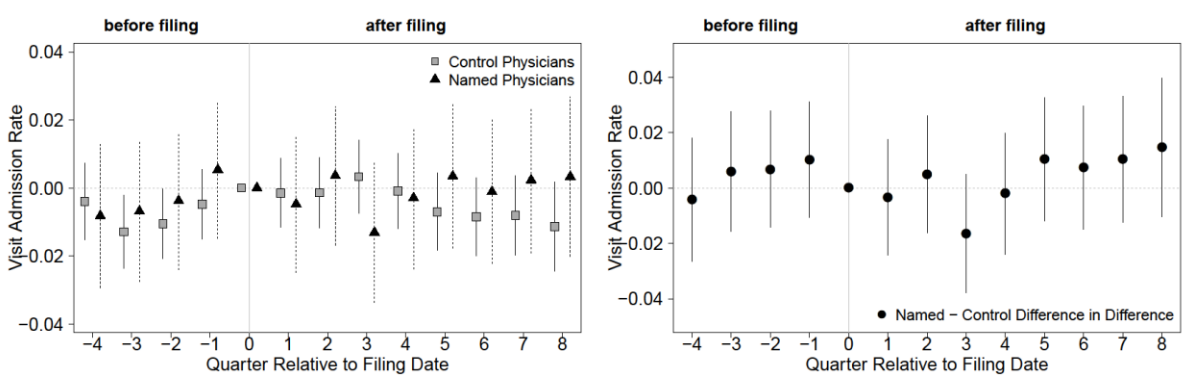
(a) RVUs/visit



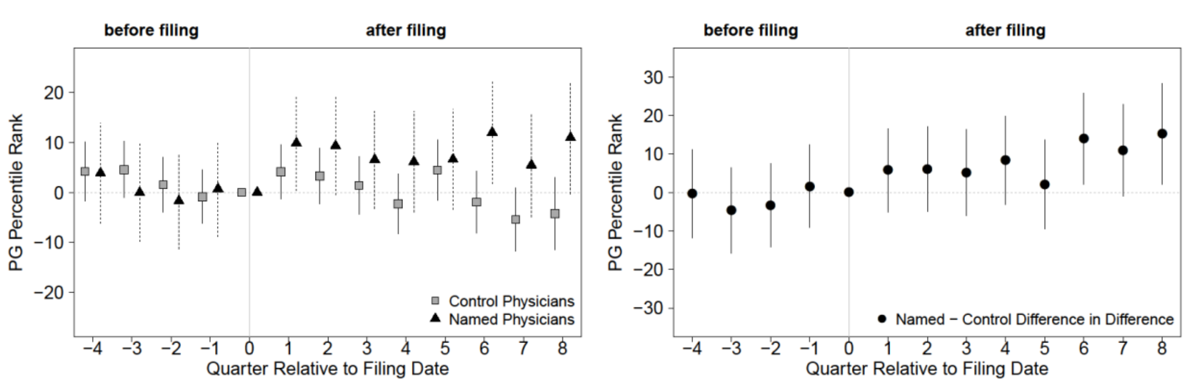
(b) RVUs/hour



(c) Visit Length of Stay for Discharged Patients (hours)

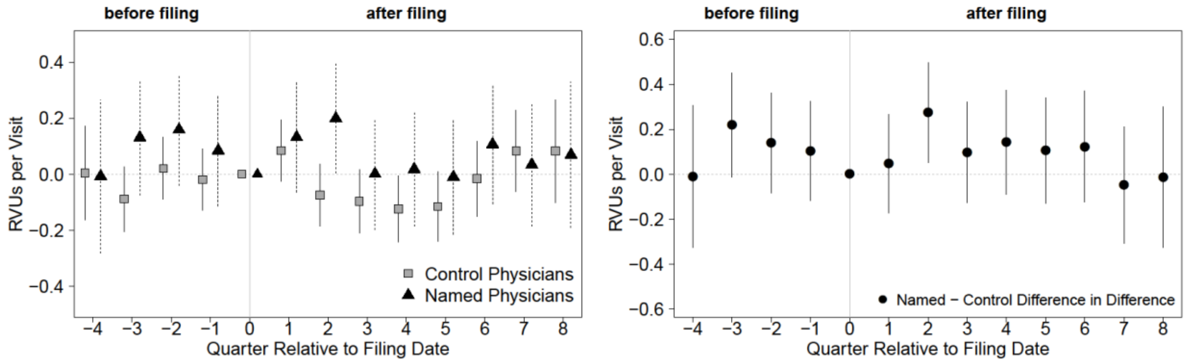


(d) Proportion of Visits Resulting in Hospital Admission

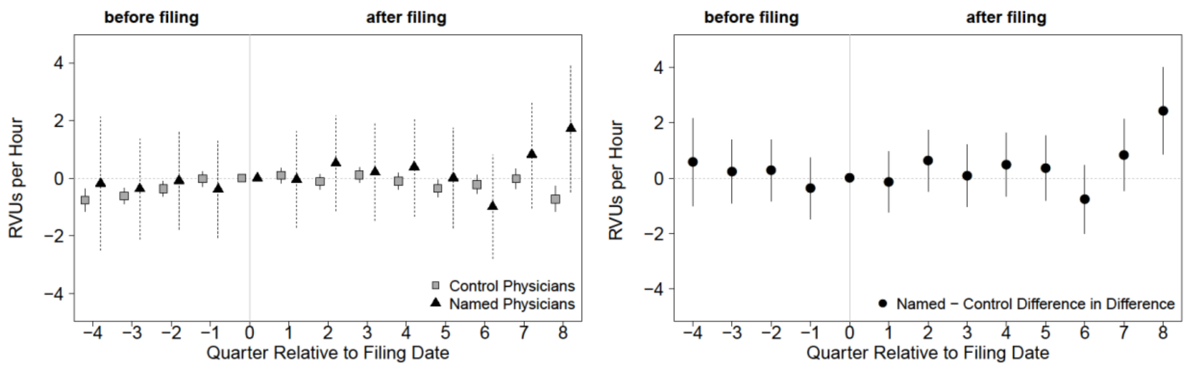


(e) Physician Press Ganey Percentile Rank

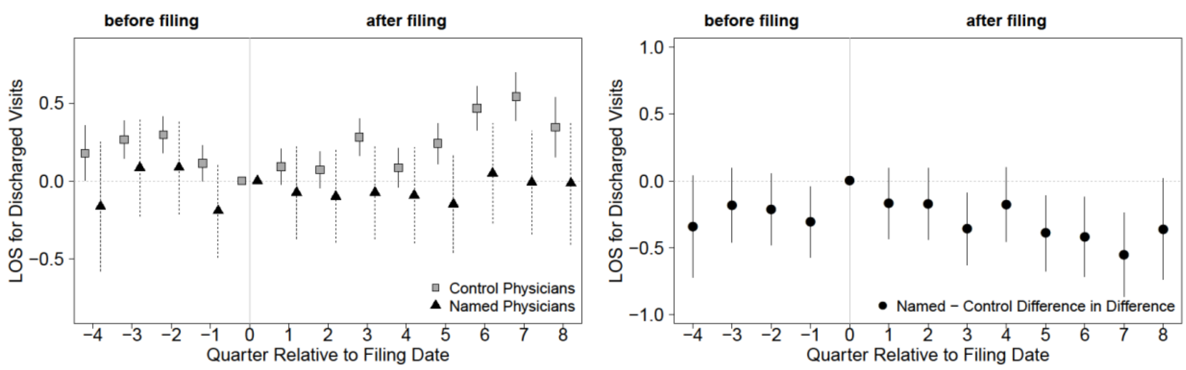
Figure 10 Leads and Lags Comparing Named Physicians in 46 Failure to Diagnose Claims Versus Their Control Physicians for Outcome Measures in All Included ED Visits.



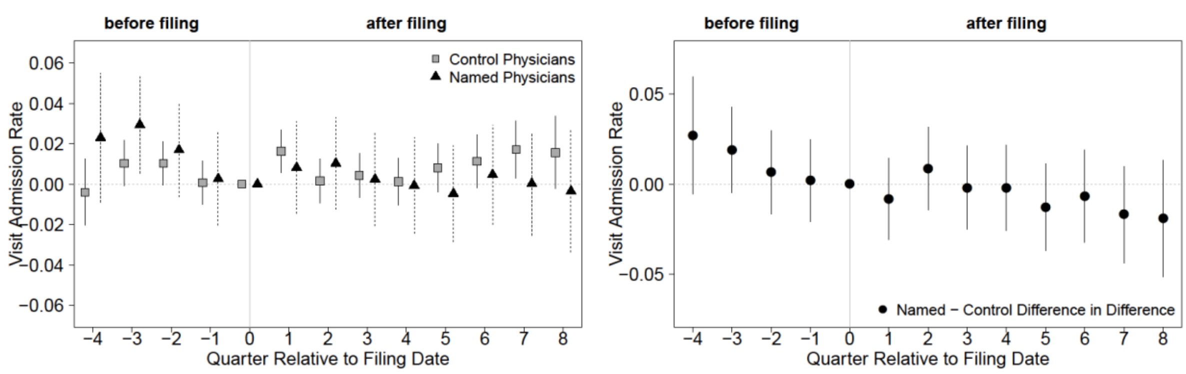
(a) **RVUs/visit**



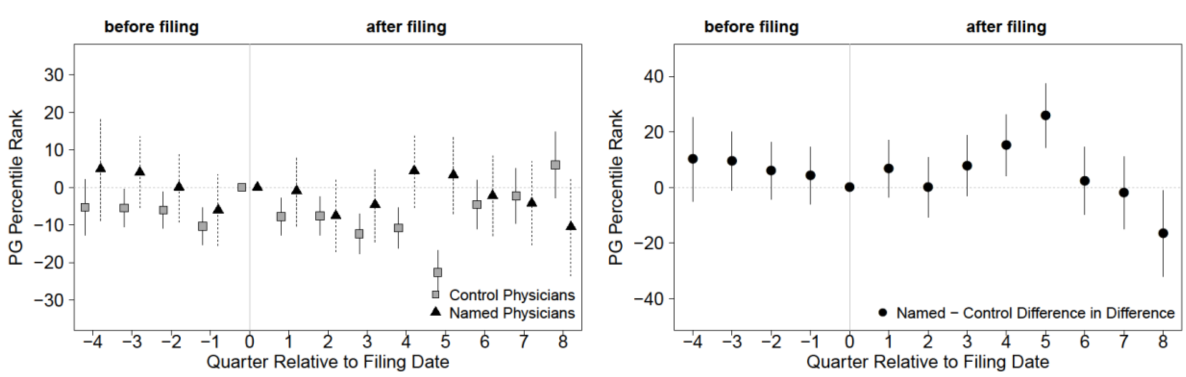
(b) RVUs/hour



(c) Visit Length of Stay for Discharged Patients (hours)

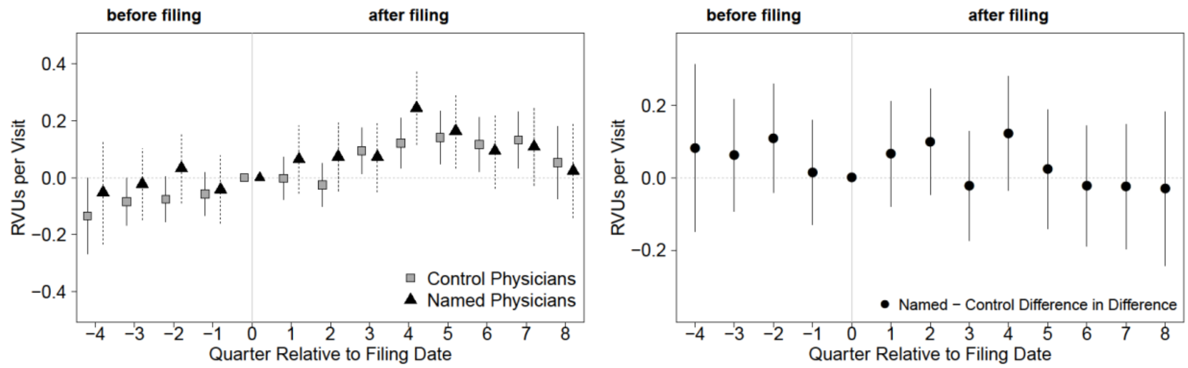


(d) Proportion of Visits Resulting in Hospital Admission

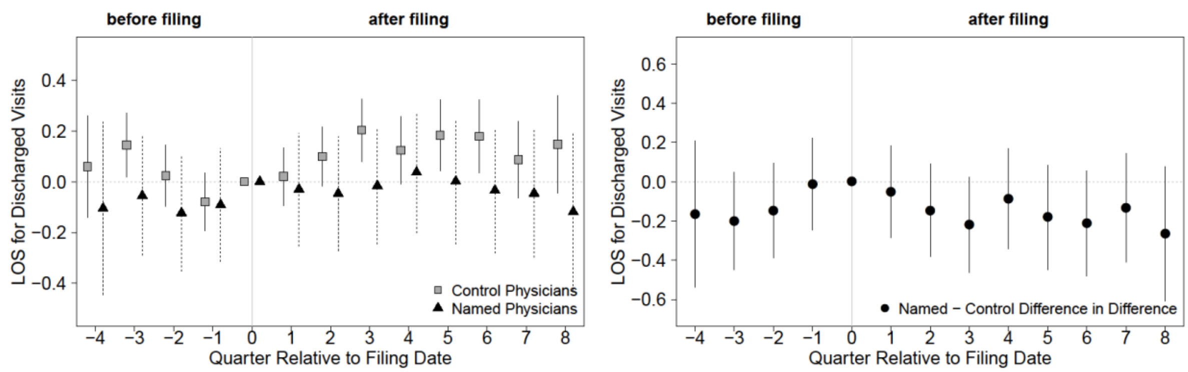


(a) Physician Press Ganey Percentile Rank

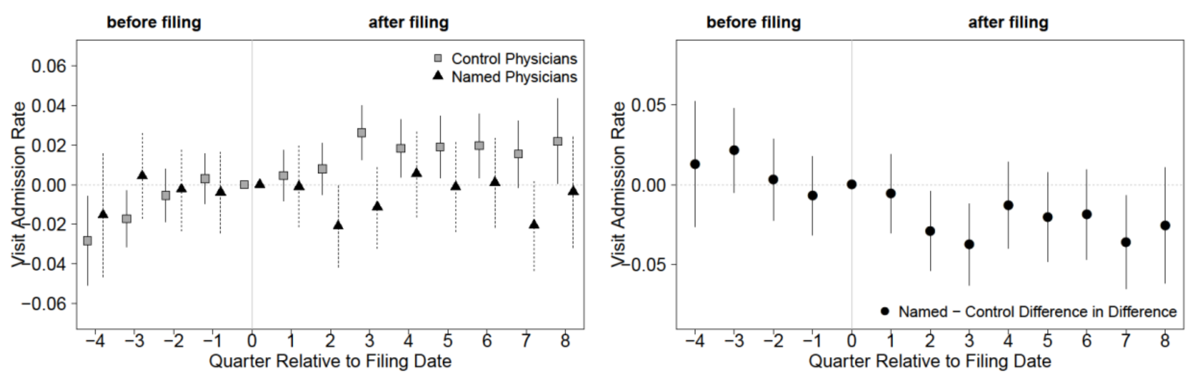
Figure 11 Leads and Lags Comparing Named Physicians in 23 Non-Failure to Diagnose Claims Versus Their Control Physicians for Outcome Measures in All Included ED Visits



(a) RVUs/visit



(b) Visit Length of Stay (LOS) for Discharged Patients (hours)



(c) Proportion of Visits Resulting in Hospital Admission

Figure 12 Leads and Lags for Outcome Measures: Body System/Clinical Issue Specific Visits

4.4.1 Robustness Checks

We performed three additional sensitivity analyses. First, three states in the data set have pre-notification laws, in which the provider is notified of an impending lawsuit before the lawsuit filing date (Hawaii: indirect process, no specific period (Hawaii Revised Statutes § 671); California: notice required three months before filing (California Code of Civil Procedure § 364); and West Virginia, notice required one month before filing (West Virginia Code § 55-7B)). In all other states in this sample, the date the ED group received notice of the claim corresponded to the claim filing date. We performed a sensitivity analysis to account for these laws, by removing two claims from Hawaii and adjusting the event date to three months earlier for California and one month earlier for West Virginia. The results after this adjustment were also unchanged from those reported. We report results from this robustness check in Appendix Table 7.

An additional sensitivity analysis was performed in which all visits by all potential control physicians were included, without propensity-score matching. Control physicians in this analysis were those not named in a malpractice claim during the entire study period and who were working in the same EDs as named physicians regardless of matching on pre-claim characteristics. Control physicians were matched to named physicians on a monthly basis (using the same start and end dates as the control physicians) and required to have at least four months of data prior to the comparison month, but without any specification on post-monthly comparison presence in the dataset. Similar to the main analysis, separate records were used for each facility worked during a given month such that named physicians in a given month were analyzed against controls from each facility at which they worked during that month. The broader inclusion criteria allowed a larger sample in this analysis (Appendix Tables 8 – 9 and Appendix Fig. 3 – 5). The results show similar trends to the main analysis with propensity-score matched controls.

Finally, we studied the possibility of difference by gender (Appendix Table 10). Our data hint that there may be baseline differences in PG scores by emergency physician gender, although the results are not statistically significant. However, the power for this comparison was limited because there were only 16 named female physicians. Given the limited number of female physicians involved in a lawsuit in our dataset, additional study will be required to better understand the relationship between provider gender and practice changes after being named in a malpractice suit.

4.5 Conclusions

4.5.1 Limitations

Our work has several limitations. Although we were able to assess aggregate measures of clinical practice (RVUs/visit, RVUs/hour, discharge visit length of stay, hospital admission percentage, and physician Press Ganey percentile rank), we lacked more granular data and therefore could not examine specific clinical actions that may have changed after a claim, such as ordering specific laboratory or advanced imaging tests. However, greater overall testing should result in higher RVUs per visit, which was not observed (Medical Decision Making and the Marshfield Clinic Scoring Tool FAQ). There may be other measures of practice that may change after a malpractice claim (e.g., the number of consultations, number of referrals, or changes in documentation and communication style) and may be more difficult to observe and require further study.

It may be that many emergency physicians at baseline practice a high level of defensive medicine. This could make it more challenging to observe differences in practice patterns, even with a larger sample size. Although Press Ganey scores by emergency physician gender did not reach significance, our data hint that there may be baseline differences. Given the limited number of female physicians involved in a lawsuit in our data set, additional study will be required to better understand the relationship between provider gender and practice changes after a physician is named in a malpractice lawsuit.

We studied only physicians who were ED attending physicians of record. How other members of the health care team, such as advanced practice providers or trainees, were affected by a malpractice claim was not measured. In any observational study, there could be unmeasured differences between treating and control persons, although our use of a difference-in-differences analysis, with facility×physician fixed effects, combined with propensity-score matching and an exact match on facility, should address many potential differences. We obtained similar results without matching, suggesting that our results are unlikely to be sensitive to our choice of a particular matching approach.

Our finding of a significant post-lawsuit relative increase in Press Ganey percentiles should be evaluated, taking into account the limitations of the Press Ganey metric, including month-to-month variability for individual physicians (Pines et al. 2018). However, despite these limitations, Press Ganey is a commonly used measure. We also did not study changes in patient outcomes, such as rates of missed diagnoses or ED revisits. The assessment of practice changes for future visits involving the same body system or clinical condition was limited by the smaller number of future visits involving specific conditions. A key aspect of ED practice is a focus on ruling out acute conditions rather than making definitive diagnoses. Therefore, our results may not generalize to other specialties.

Our sample of greater than 1.6 million ED visits is large and leads to reasonably tight confidence bounds, but our sample of claims was small and from a single ED group, which employs principally board-certified emergency physicians (ie, those with maximum qualifications). Based on a .05 significance level, our analyses had sufficient power to detect changes of approximately 0.08 RVUs per visit (a 2% change from baseline), 0.5 RVUs per hour (a 5% change from baseline), 0.10-hour visit length for discharged patients (a 2% change from baseline), a 1.2% change in hospital admission rate (a 6% change from baseline), and 5.7 Press Ganey percentile ranks. Thus, despite the limited number of malpractice claims, our study was sufficiently powered to find differences we would consider clinically meaningful.

This ED group employs predominantly board-certified emergency physicians. How these results generalize to non–board-certified emergency physicians or to physicians with different baseline practice patterns will require additional study. Although some emergency physician groups may mandate care bundles to lower risk in how providers manage specific common complaints (e.g., chest pain, headache), none were used across the EPMN during the study period. How use of mandated care bundles may influence behavior surrounding a malpractice claim is unknown and deserving of further investigation. Few claims went to trial, limiting the ability to study these separately. We also do not know the claim history of emergency physicians included in this study before the study period or joining this emergency medicine group.

4.5.2 Discussion

In this study, we found evidence of improved assessed patient experience after emergency physicians were named in a malpractice claim, but no significant changes in aggregate measures of care intensity and speed of care. Patient experience ratings increased for named physicians promptly after the malpractice claim filing date and were most prominent in cases in which the malpractice claim alleged a failure to diagnose. Failure-to-diagnose claims (e.g., a missed diagnosis of acute myocardial infarction, stroke, or meningitis) are different from other claims for emergency physicians; for example, adverse effects of treatment (e.g., allergic reaction to administered medication). Failure-to-diagnose claims tend to occur according to the cognitive decision making of a single emergency physician: despite the emergency physician's having evaluated the patient to rule out serious conditions (it is hoped), the contention of the claim is often that the physician "missed" something.

Being accused of missing something might engender self-doubt about personal competence and introspection about whether one's clinical practice approach may need to change. However, although we observed improvements in assessed patient experience, potentially related to improved communication as perceived by the patient, there was no evidence of a change toward more defensive practice, such as admitting more patients or performing more tests (which likely would have been evident in RVUs/visit or ED length of stay). There is an association between patient complaints and perceived lack of physician empathy and malpractice claim risk (Hickson et al. 1994, Hickson et al. 2002, Cydulka et al. 2011). Emergency physicians may be aware of this association and alter their approach to interacting with patients after a malpractice claim, perhaps through improved communication, which may involve providing patients with a better understanding of risks and benefits in testing or other clinical decisions. In an environment in which many physicians perceive themselves to be time pressured (Nugus et al. 2011), Press Ganey score changes could reflect that physicians are attempting to improve patient experience through communication. However, the degree of change on average was moderate (6.5 Press Ganey percentile ranks overall and 10.5 for failure-to-diagnose claims). The association between Press Ganey ranks and other measures of care quality is unknown, and previous work has suggested that the Press Ganey scores of individual physicians can fluctuate (Pines et al. 2018). Further study is needed to address whether these changes are clinically meaningful to patients and physicians. In addition, future studies may examine changes in specific behaviors such as bedside communication, education, or other factors that might have led to the observed improvement in assessed patient experience after a malpractice claim.

These findings suggest that efforts to reduce the number of malpractice claims may not meaningfully affect defensive medical practices among emergency physicians. This is consistent with literature finding little change in ED practice for clinical decisions attributable to defensive medicine after tort reform (Waxman et al. 2014). However, our findings that malpractice claims led to improved patient experience may be relevant for efforts to support open patient-physician communication, such as apology statutes. Apology statutes might allow improved communication to take place before and perhaps without costly litigation, if coupled with direct patient compensation for preventable adverse events.

In conclusion, emergency physicians named in a malpractice claim have higher assessed patient experience post-claim relative to matched control physicians, but care intensity and speed did not significantly change.

5.0 Balancing Emergency Physicians and Advanced Practice Providers to Meet Patient Demand

5.1 Motivation

Consider a local emergency department (ED). In the early morning hours, (3am to 7am), only one physician and one advanced practice provider (APP) were scheduled to work, but the ED experienced higher patient volumes than usual. This created a backlog of patients waiting for care. During the 7am to 11am shift, the ED is staffed with three physicians and two physician assistants. An elderly female patient arrives at 10am complaining of abdominal pain. The patient overhears that some patients have already been waiting for more than two hours. Furthermore, she knows from previous experiences that arriving patients with more urgent needs will be treated first. With no other options for care, the patient will wait almost three hours before she is eventually seen by a physician, or she may choose to leave the ED before being treatment.

Unfortunately, this situation is not unique, and long ED waiting times continue to be a problem. The American College of Emergency Physicians (ACEP) cites four reasons for long ED waiting times: (1) long boarding times, (2) waiting for specialists, (3) mass casualty events, natural disasters, and local disease outbreaks, and (4) staffing (ACEP 2009). Boarding time refers to the time between the decision to admit a patient to the hospital (as inpatient) and the patient's departure from the ED area, and reducing boarding times involves operational changes at the hospital level. Similarly, the additional waiting time attributed to specialist consultations cannot be addressed within the ED alone. Mass casualty events, natural disasters, and local disease outbreaks are largely unpredictable, but EDs typically have disaster plans for such events. For these reasons, we focus on addressing changes in ED provider staffing as a method to reduce patient waiting times.

Traditionally, emergency physicians diagnose and treat patients within the ED, but many EDs also employ APPs such as physician assistants and nurse practitioners. This has become common practice as the current shortage of emergency physicians continues (Reiter et al. 2016). These APPs are licensed to diagnose and treat patients, prescribe medications, and order tests, but there are limitations to the conditions they can treat without a physician's involvement. For example, APPs are less likely to perform invasive procedures, such as inserting chest tubes (Nyberg et al. 2010). APPs are commonly required to practice under a physician's supervision and thus, are not perfect substitutes for physicians. However, many of their responsibilities overlap. While APPs hold advanced degrees (master's or doctorate) and obtain professional licensure and certification for the states in which they practice, their education time is shorter than that of physicians, and APPs do not require residency training after graduation as physicians do. While Kraus et al. (2018) report that formal postgraduate training programs to provide emergency medicine physician assistants with the skills and knowledge to work in an ED do exist, the qualifications for physician assistant licensure are (1) graduation from an accredited physician assistant program and (2) passage of the Physician Assistant National Certification Examination (Cawley & Hooker 2013). Once licensed, the scope of practice may vary according to training, experience, facility policy, and state law, but physician assistants have the authority to prescribe medications in all states (Cawley & Hooker 2013). For these reasons, physician assistant and nurse practitioner programs are viewed as attractive alternatives for students interested in medicine, and, consequently, the number of licensed APPs has been increasing in recent years (Gaudio & Borensztein 2018, Fraher et al. 2016, Hooker et al. 2016). Furthermore, APPs are much less costly to employ than physicians, with national averages indicating that emergency physician salaries can be more than three times that of APPs working in emergency departments (Katz 2017, AANP 2018, AAPA 2018). Still, our analyses of a national physician-owned emergency medicine group suggest that EDs vary in the ways they staff and utilize APPs.

Given the shortage of emergency physicians, an increase in the number of APPs, the ability of APPs to treat some ED patients, and the lower cost of APPs, we surmise that APPs may be instrumental in reducing patient waiting times in the ED. Carter and Chochinov (2007) previously suggested that APPs could reduce waiting times and called for the medical community to explore the use of APPs in EDs, but Larkin and Hooker (2010) later indicated that ED patients prefer to see a physician. Similarly, Dill et al. (2013) found that half of the patients they surveyed prefer that their primary care providers are physicians. However, when these patients were presented scenarios in which they could see an APP sooner than a physician, the majority of respondents indicated that they would choose to see an APP instead of waiting (Dill et al. 2013). Similarly, Doan et al. (2012) found that patients tended to favor a physician assistant in a number of scenarios in which the waiting time for a physician was longer for a physician than for a physician assistant. Furthermore, the proportion of patients preferring to see a physician assistant slightly decreased as the difference in waiting times between the two types of providers decreased (Doan et al. 2012). Thus, there is evidence suggesting that patients may adjust their provider preferences in favor of shorter waiting times. Additionally, studies have shown no differences between physicians and APPs with regard to diagnostic accuracy (van der Linden et al. 2010) or patient satisfaction (Vallejo et al. 2015, Jeanmonod et al. 2013), suggesting that these providers are highly capable and that the patients who do end up seeing an APP are just as satisfied with their care. Additionally, our data suggest that at least 50% of patients arriving in any ED are classified as levels 3 through

5 on the Emergency Severity Index (ESI) scale, which typically indicates non-life-threatening or lower risk symptoms. APP involvement varies widely by facility, with some facilities that do not employ APPs to treat ED patients and others that rely on APPs to treat over half of ED patients. Thus, it is apparent that APPs have the necessary training and qualifications to treat many ED patients.

While it seems clear that APPs have the potential to affect ED waiting times, there is not a standard practice for staffing APPs in EDs. In fact, ACEP (2013) provides guidelines regarding the roles of APPs in EDs, but those guidelines indicate that both the scope of practice and degree of supervision may vary based on state laws and regulations, guidelines developed by ED medical directors, and supervising physicians. We have observed various staffing patterns among EDs with similar characteristics and comparable annual patient volumes. Published case studies such as Sturmann et al. (1990) are not generalizable to other facilities, and other proposed scheduling tools (Myers et al. 2014) focus specifically on Level 1 trauma centers, the hospitals which, according to the American Trauma Society (ATS, 2019), are “capable of providing total care for every aspect of injury – from prevention through rehabilitation.” Such hospitals often have additional providers such as medical students and residents, which makes their situation different from smaller non-academic facilities with fewer capabilities. For this reason, we have chosen to focus our research on medium-sized facilities (40,000 to 60,000 patient visits per year) which do not have a Level 1 trauma designation. These facilities tend to have fewer resources, yet they must still be prepared to treat any patient that arrives in the ED.

In this research, we propose a systematic data-driven approach to determine the number of APPs and physicians to staff in order to reduce patient waiting times and length of stay. Our approach incorporates a variety of methodologies, including data mining, simulation, statistics, and machine learning. Our proposed methods take into account the realities of EDs such as known trends in ED arrival rates, the stochastic nature of arrivals, and staffing limitations.

Next, we discuss the existing literature related to our problem. We then outline our research framework. Our proposed models are presented, and the data are described. We detail our analyses and results. Finally, we discuss our conclusions, including managerial insights and directions for future work.

5.2 Related Literature

Initially, APPs were introduced to EDs in order to meet the growing demand for emergency care, and the proportion of APPs treating patients has continued to rise in recent years (Brown et al. 2012). For example, Barata et al (2015) called for the use of nurse practitioners and physician assistants in lower-acuity settings during peak hours to improve the flow and care of patients in a pediatric ED. Another study showed that adding a physician assistant as a triage liaison provider, who assesses the patient before an ED treatment room is available, is beneficial with regard to both median length of stay and the proportion of patients who leave without treatment (Nestler et al. 2012). While the medical community tends to agree that APPs play an important role in improving ED operations, the specific roles of APPs can vary. We first examine the literature to find commonalities in the ways in which APPs are utilized in EDs.

Generally, patients seen by APPs are of lower acuity (Brown et al. 2012), and research has indicated that APPs treat more low acuity patients per hour and generate more relative value units (RVUs) per hour than those generated by emergency medicine resident physicians, while patient satisfaction scores remain similar for the two types of providers (Jeanmonnod et al. 2013). RVUs reflect the time and supplies/devices needed from the healthcare workers/facility to care for the patient. In a high acuity setting, APPs tend to treat more patients per hour than emergency medicine (EM) residents, but they generate fewer RVUs per patient (Hamden et al. 2014). This indicates either that APPs do not document as thoroughly or that APPs tend to treat less sick patients. Silberman et al. (2012) found that APPs treated more patients per hour and generated more RVUs per hour when staffing a low acuity area compared to a high acuity area. Regardless of the setting, it seems clear that APPs typically treat lower acuity patients than their physician counterparts.

However, hospital characteristics and location may also affect how APPs work. For example, Sawyer and Ginde (2014) found that physician assistants who practiced in rural areas tended to report a broader scope of practice, greater autonomy, and lower access to physician supervision than their urban counterparts. Additionally, Nelson et al. (2016) studied rural EDs in Washington state and found that the EDs located in more remote areas were more likely to staff APPs as the sole providers, with physicians providing backup. These differences suggest that rural emergency departments should be treated differently than those in urban or suburban areas.

While Jeanmonnod et al. (2013) noted similarities in patient satisfaction scores, Pavlik et al (2017) used 72-hour recidivism as an outcome measure and concluded that physician assistants and emergency physicians are similar in their management of pediatric patients 6 years or younger. Such studies reflect the ability of APPs to treat patients while also maintaining quality of care. Furthermore, Horowitz et al. (2009) studied hospital-level performance on ED wait time and visit length and found that less than half of hospitals consistently achieved recommended wait times and visit length, potentially contributing to an increase in adverse events and an reduction in the quality of care. These results highlight the potential for improvements in ED operations, especially with regard to providers.

Next, we present our research framework, which combines multiple data mining methods and simulation to describe the current state of an ED, to predict the effects of changes to the current system, and to prescribe an optimal staffing policy for ED providers.

5.3 Research Framework

Before addressing our research questions, we must fully understand the current state of EDs. We use data from a national EM management network to understand how EDs utilize both physicians and APPs. We consider the observed staffing levels of both provider types, differences in the types of patient they treat, and differences in operational measures between the two types of providers (e.g., patients per hour, patient wait times, patient length of stay, and the proportion of patients admitted to the hospital). We use descriptive statistics and data visualization to gain insights about ED providers and the patients they treat as well as how they have changed over time. The descriptive analyses demonstrate that there is no standard way in which APPs have been utilized, and even changes over time have been inconsistent. Most of these analyses are based on eight EDs for which complete data for 2014 through 2017 are available. We then use data from these eight EDs to predict ED patient arrivals, the assignment of patients to ED providers, and the length of stay for discharged patients. The results from the data-driven predictive analyses are incorporated into a simulation model for an ED in a medium-sized metropolitan area. After replicating the current state of the ED, we use a Taguchi L-27 experimental design (Krishnaiah &

Shahabudeen 2012) to simulate both controllable and uncontrollable changes to the ED. Taguchi designs are a special case of fractional factorial designs which are used to identify the optimum process parameters, and the L-27 design allows us to include up to 13 three-level factors in a 27-run experiment with an orthogonal design matrix. Using patient LOS as an outcome measure, we analyze the experimental results to identify the optimal staffing levels for physicians and APPs in the ED. This framework is depicted in Fig. 13.

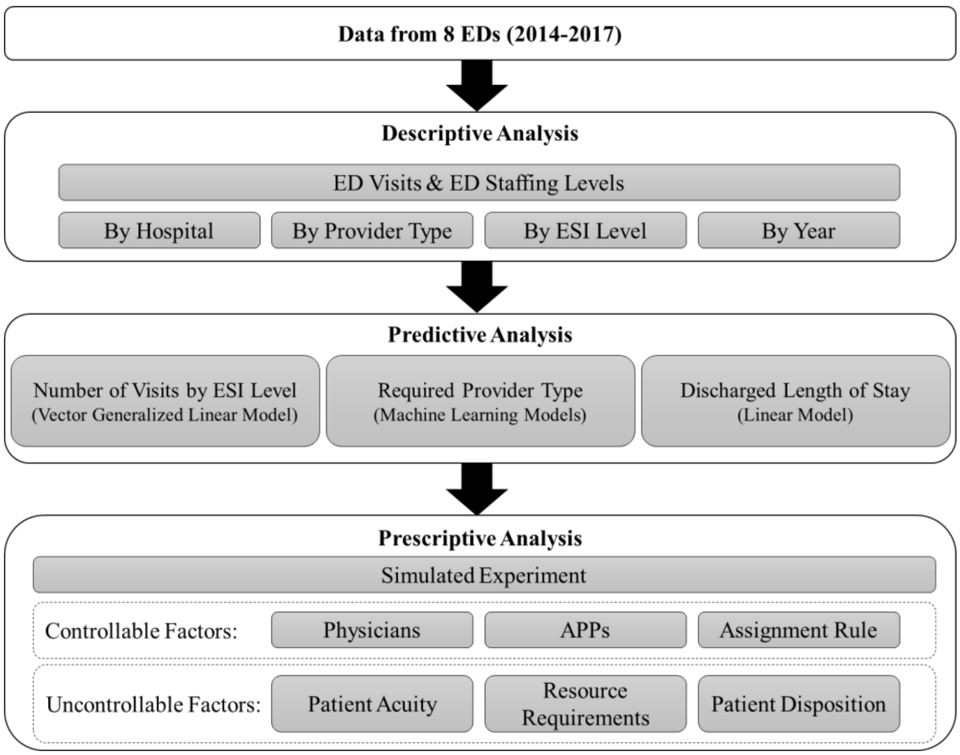


Figure 13 Research Framework

Figure 13 lists the three dependent variables for which we fit the predictive models. First, we consider how the use of APPs affects the time that patients spend in the ED. We specifically consider the length of stay for discharged patients because the length of stay for admitted patients is influenced by factors beyond the ED's control such as the availability of inpatient beds. We utilize ordinary least squares (OLS) regression to test hypotheses regarding the use of APPs in

EDs. The dependent variable is log-transformed discharge visit length (i.e., LOS). Independent variables include an APP indicator variable (1 if provider is an APP, 0 otherwise), the number of physicians working at a patient's arrival time, the number of APPs working at a patient's arrival time, and the patient's ESI Level. We also control for the patient's age, gender, and insurance (payer) as well as the ED census, arrival time (categorized by four-hour half-shifts), and the facility in which they arrive. This allows us to estimate the difference in the length of stay for discharged patients between patients treated by physicians and those treated by APPs, while controlling for patient and facility factors. We can also estimate the effects of adding the two different types of providers. In our first specification, we group ESI levels into three groups as previously discussed: levels 1 and 2, level 3, or levels 4 and 5. An alternate specification excluded levels 1 and 2 since those patients are unlikely to be treated by an APP alone. The model is specified in Eq. (12).

$$\begin{aligned}
\ln(LOS) = {} & \beta_0 + \beta_1 \text{ APP Indicator } + \beta_2 \text{ Physician Count} + \beta_3 \text{ APP Count} \\
& + \beta_4 \text{ Physician Count} \times \text{APP Indicator} + \beta_5 \text{ APP Count} \times \text{APP Indicator} \\
& + \sum_{k=2}^{3} \beta_{6_k} \text{ ESI Level}_\text{k} + \sum_{k=2}^{3} \beta_{7_k} \text{ ESI Level}_\text{k} \times \text{APP Indicator} \\
& + \sum_{k=2}^{3} \beta_{8_k} \text{ ESI Level}_\text{k} \times \text{Physician Count} + \sum_{k=2}^{3} \beta_{9_k} \text{ ESI Level}_\text{k} \times \text{APP Count} \\
& + \sum_{k=2}^{11} \beta_{10_k} \text{ Age Group}_\text{k} + \beta_{11} \text{ Male} + \sum_{k=2}^{5} \beta_{12_k} \text{ Payer}_\text{k} + \beta_{13} \text{ ED Census} \\
& + \sum_{k=2}^{6} \beta_{14_k} \text{ Half Shift}_\text{k} + \sum_{k=2}^{6} \beta_{15_k} \text{ Half Shift}_\text{k} \times \text{APP Indicator} \\
& + \sum_{k=2}^{6} \beta_{16_k} \text{ Half Shift}_\text{k} \times \text{Physician Count} + \sum_{k=2}^{6} \beta_{17_k} \text{ Half Shift}_\text{k} \times \text{APP Count} \\
& + \sum_{k=2}^{8} \beta_{18_k} \text{ Facility}_\text{k} + \varepsilon
\end{aligned} \tag{12}$$

APPs generally treat lower acuity patients in EDs. Thus, we expect patients treated by APPs to require fewer resources in the ED and have a shorter length of stay than those treated by physicians. APPs tend to have fewer responsibilities than physicians and treat fewer patients per hour than physicians. One possible reason for this may be that APPs experience fewer interruptions while treating a patient. Thus, we hypothesize that after controlling for patient characteristics and the current state of the ED, the discharge visit length for visits treated by APPs will be shorter than for visits treated by physicians:

H5. *The average discharge visit length for visits treated by APPs will be shorter than the average discharge length for visits treated by physicians. That is,* $\beta_1 < 0$.

It is logical that the length of stay should decrease when more providers are available (either physicians or APPs). Thus, we hypothesize that as the number of physicians increases, the average discharge visit length will decrease, holding all else constant. Similarly, the average discharge visit length will decrease as the number of APPs increases while controlling for other factors. APPs who treat lower acuity patients enable physicians to focus their skills on higher acuity patients as well as other responsibilities. This should result in a reduction in the average length of stay for all discharged patients, regardless of the provider type. However, due to known differences in training between physicians and APPs, we expect that adding a physician will have a greater impact on the length of stay for discharged patients compared to adding an APP.

H6. *There exists a negative relationship between the number of providers and the length of stay (LOS) for discharged visits, and the effect of adding a physician will be greater in magnitude than that of adding an APP. Namely,*

a. There exists a negative relationship between the number of physicians and the length of stay (LOS) for discharged visits. That is, $\beta_2 < 0$.

b. *There exists a negative relationship between the number of APPs and the LOS for discharged visits. That is,* $\beta_3 < 0$.

c. *The relationship between the number of physicians and the LOS for discharged visits will be stronger (i.e., more negative) than that between APPs and LOS. That is,* $\beta_2 < \beta_3 < 0$.

We expect that with more physicians working, all patients will benefit because each additional physician increases the ED's capacity. Consequently, we expect that the gap in length of stay between patients treated by physicians and patients treated by APPs will narrow if more physicians are working. That is, the interaction terms, $\text{APP Indicator} \times \text{Physician Count}$ will be positive.

H7. *As the number of physicians increases, the average effect of being treated by an APP on discharge LOS will be smaller in magnitude (i.e., less negative). That is, while the coefficient for APP Indicator is negative* ($\beta_1 < 0$)*, the coefficient for the interaction term APP Indicator × Physician Count will be positive* ($\beta_4 > 0$).

Furthermore, while we expect an overall negative effect of the APP count on discharge visit length, we anticipate that this effect will be most negative for the lowest acuity patients. Since EDs prioritize the most severe (or highest acuity) patients, lower acuity patients are most likely to experience long wait times and thus longer lengths of stay when waiting to see a physician. The availability of APPs to treat lower acuity patients should result in shorter visit lengths. Consequently, we hypothesize that the $\text{APP Count} \times \text{ESI Level}$ interaction term will be most negative for the ESI level 4 and 5 group:

H8. *The negative relationship between the APPs count and the length of stay (LOS) for discharged visits will become stronger as ESI Level increases (patient's condition is less severe). That is,* $\beta_{8_k} < 0$ *for all k, and* β_{8_k} *decreases as k increases.*

Finally, we expect that the effects of both adding APPs and being treated by an APP will differ by time of day. In particular, EDs experience a different patient population during the night shifts. For example, fewer patients typically arrive during the night shifts, and the patients tend to be sicker. In that case, there should be less need for APPs if patients that arrive during those hours and APPs may not be equipped to handle many of the more severe conditions. Therefore, we suspect that the effect of adding an APP will be smaller during the night (11pm to 3am and 3am to 7am) compared to the reference shift (7am to 11am), and we hypothesize that the $\text{APP Count} \times \text{Shift 11pm to 3am}$ (and the $\text{APP Count} \times \text{Shift 3am to 7am}$) interaction term will be positive:

H9. *The negative relationship between the APPs count and the LOS for discharged visits will become weaker (i.e., less negative) for patients arriving during the night shifts. That is, the coefficients for the corresponding* $APP\ Count \times Shift$ *interaction terms will be positive:* $\beta_{17_k} > 0$ *for k=5 and k=6.*

We utilize the insights derived from the descriptive and predictive analyses to design a simulated experiment for a single ED. To better inform the experimental settings, we estimate the number of hourly visits using a vector generalized linear model. Visits are categorized based on Emergency Severity Index (ESI) level, requiring the use of multivariate methods to simultaneously estimate the number of hourly visits of each level, and in this case, the dependent variables are modeled as a multivariate Poisson distribution. Independent variables include the facility, month of the year, day of the week, time of day (represented by four hour half-shift increments), indicators for holidays and the day after a holiday, and the normal high and low temperatures for each given date and facility location. We decided to include federal holidays in which most businesses are closed as well as the day after each of those holidays after discussions with clinicians. Physicians have observed that patient volumes and patient characteristics differ on those days. Similar to weekends, holidays represent time off, and patients will typically put off going to the ED until the following day if possible. Then, the day after a holiday is similar to a Monday, and higher patient volumes are observed. For the purposes of this analysis, we identified New Year's Day, Memorial Day, Independence Day, Labor Day, Thanksgiving, and Christmas as holidays. The model specification is provided in Eq. (13), where $\boldsymbol{\eta_i} = (\eta_{1(i)} \quad \eta_{2(i)} \quad \eta_{3(i)})^T$ denotes the number of hourly arrivals during time period i which are classified as ESI level 1 or 2, ESI level 3, and ESI level 4 or 5, respectively. We chose to group ESI levels in this way after discussion with clinicians suggested that both ESI levels 1 and 5 are rare. Additionally, it would be highly unusual for an APP to treat an ESI level 1 or 2 patient without a physician as those represent the highest acuity patients, and many hospitals operate with a "Fast Track" and route ESI level 4 and 5 patients, the lowest acuity patients, to an APP. Each coefficient $\boldsymbol{\beta_k}$ is a column vector, and dummy variables are used to signify month of the year, day of the week, and half shifts (7am – 11am, 11am – 3pm, 3pm – 7pm, 7pm – 11pm, 11pm – 3am, and 3am – 7am). Additionally, we include dummy variables to indicate federal holidays in which most businesses are closed, as well as the day after each of those holidays. Finally, we took into account normal weather conditions for a given date since both health conditions and accidental injuries may be triggered by weather conditions, such as extreme heat or cold, and associated seasonal activities. Specifically, we utilize daily minimum and maximum temperature data collected by the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (NCEI). The arrival rate for category *i* severity can be expressed as

$$\boldsymbol{\eta_i} = \boldsymbol{\beta_1}\text{Facility}_i + \sum_{j=2}^{12} \boldsymbol{\beta_j}\,\text{Month}_i + \sum_{j=13}^{18} \boldsymbol{\beta_j}\,\text{Day}_i + \boldsymbol{\beta_{19}}\text{Holiday}_i \tag{13}$$

$$+\boldsymbol{\beta_{20}}\text{Day After Holiday}_i + \sum_{j=21}^{25} \boldsymbol{\beta_j}\,\text{Half Shift}_i + \boldsymbol{\beta_{26}}\text{MinTemp}_i + \boldsymbol{\beta_{27}}\text{MaxTemp}_i$$

An ED may be able to make better staffing decisions regarding the total number of providers if reasonable predictions are available for the number of patients of each severity type that will arrive during a given four-hour period of time (or half-shift). However, additional information is necessary to better decide on provider type. If patient classes (or ESI levels) and other visit characteristics could be used to predict which visits require a physician and which could be treated by an APP, management could improve decision-making. We implement various machine learning methods to classify visits by required provider type and assess these models. Independent variables include patient characteristics (ESI level, patient age group, patient gender, and payer information [e.g. Medicare, Medicaid, private insurance provider, etc.]) as well as the current state of the ED (ED census at the patient's arrival time, average ESI level of other patients in ED upon the patient's arrival, and the ratio of APPs to physicians working in ED upon a patient's arrival), the half shift in which a patient arrived (i.e., one of the six four-hour time blocks previously described; first half shift is 7am to 11am), and the facility. We include the half shift in which a patient arrives in order to further account for both different patient characteristics and different staffing patterns throughout the day. While a number of models were tested, the logistic regression model is specified in Eq. 14, where π is the probability that a patient is assigned to an APP.

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_{k=2}^{3} \beta_{1_k} \text{ ESI Level}_{\mathrm{k}} + \sum_{k=2}^{11} \beta_{2_k} \text{ Age Group}_{\mathrm{k}} + \beta_3 \text{ Male} + \sum_{k=2}^{5} \beta_{4_k} \text{ Payer}_{\mathrm{k}} \quad (14)$$

$$+ \beta_5 \text{ ED Census} + \beta_6 \text{ Average Severity of Other Patients}$$

$$+ \beta_6 \text{ APP to Physician Ratio} + \sum_{k=2}^{6} \beta_{7_k} \text{ Half Shift}_{\mathrm{k}} + \sum_{k=2}^{8} \beta_{8_k} \text{ Facility}_{\mathrm{k}} + \varepsilon$$

By adjusting both controllable and uncontrollable factors in a simulated experiment for a single ED, we estimate the effects of changes on visit length. Controllable factors include staffing decisions and the assignment of providers to patients, while uncontrollable factors include the number of ED visits and patient characteristics. In the next section, we detail these analyses and results.

5.4 Data and Analysis

Recall that our data are from a large physician-owned emergency medicine management organization. This company contracts with hospitals and healthcare systems across the United States to manage emergency departments and staff emergency medicine providers. The initial database contained detailed data corresponding to over 11.5 million ED visits at 125 EDs between August 2011 and December 2017. We excluded all EDs that terminated their relationship with the organization before 2017, and further excluded EDs with a large number of missing data fields as they didn't collect the necessary data during the study period. For each year in which an ED is included in our sample, we include the following fields: emergency severity index (ESI) of visits, the primary provider's type (physician or APP), and time stamps for the ED arrival times, times patients are first seen by a provider, and ED departure times. The resulting database for our analysis is comprised of 4,870,794 ED visits at 49 EDs between 2014 and 2017, with the number of EDs varying by year.

5.4.1 Descriptive Analysis

Table 14 shows how the sample of hospitals evolved over time.

Table 14 Description of ED Sample by Year

		Year			
		2014	2015	2016	2017
Yearly Hospital Counts		8	19	33	49
by Annual Visits	<20,000 visits	1	2	4	6
	20,000-39,999 visits	3	6	12	20
	40,000-59,999 visits	2	5	9	12
	60,000-79,999 visits	2	6	8	10
	80,0000+ visits	0	0	0	1
by EM Residency	Yes	1	4	4	4
	No	7	15	29	45
by Teaching Hospital	Yes	2	6	7	7
	No	6	13	26	42
by Trauma Level	1	0	1	1	1
	2	1	2	3	3
	3	1	2	2	2
	4	0	0	1	1
	None	6	14	26	42
by State	CA	1	1	1	1
	CT	1	1	1	2
	FL	0	0	0	8
	HI	1	2	2	2
	IL	2	2	2	2
	KY	0	0	1	1
	MD	0	0	0	4
	NC	0	1	6	7
	NH	0	1	1	1
	NY	1	1	2	2
	OH	2	4	7	9
	OK	0	2	2	2
	PA	0	4	8	8

Due to the nature of emergency medicine, the characteristics of patient visits vary. Table 15 describes the visit data. In total, 1,476 emergency medicine providers (894 physicians and 582 APPs) treated the patients in the sample.

Table 16 shows a comparison of visit characteristics between patients treated by a physician and patients treated by an APP. Overall, APPs are most likely to treat visits classified as ESI level 4 and 5; at least 67% (e.g., 61.9% + 5.9% = 67.8% in 2017) of visits treated by APPs each year are classified as ESI level 4 or 5 (Table 16). Physicians are most likely to treat visits classified as ESI level 3 with at least 50% of visits assigned to physicians classified as ESI level 3 each year (Table 16). The majority of patients are eventually discharged from the ED, regardless of provider type, but the admission rate among APPs is significantly lower than the admission rate among physicians (difference in proportions = 0.19, p-value < 2.2×10^{-16}). Furthermore, the observed median LOS among discharged patients treated by APPs is about one hour shorter than discharged patients treated by physicians in the same facility during the same year (median difference in LOS discharged = 0.98 hours, p-value < 2.2×10^{-16}). These data indicate that APPs play an important role in treating non-life-threatening conditions and lower risk patients who arrive in an ED.

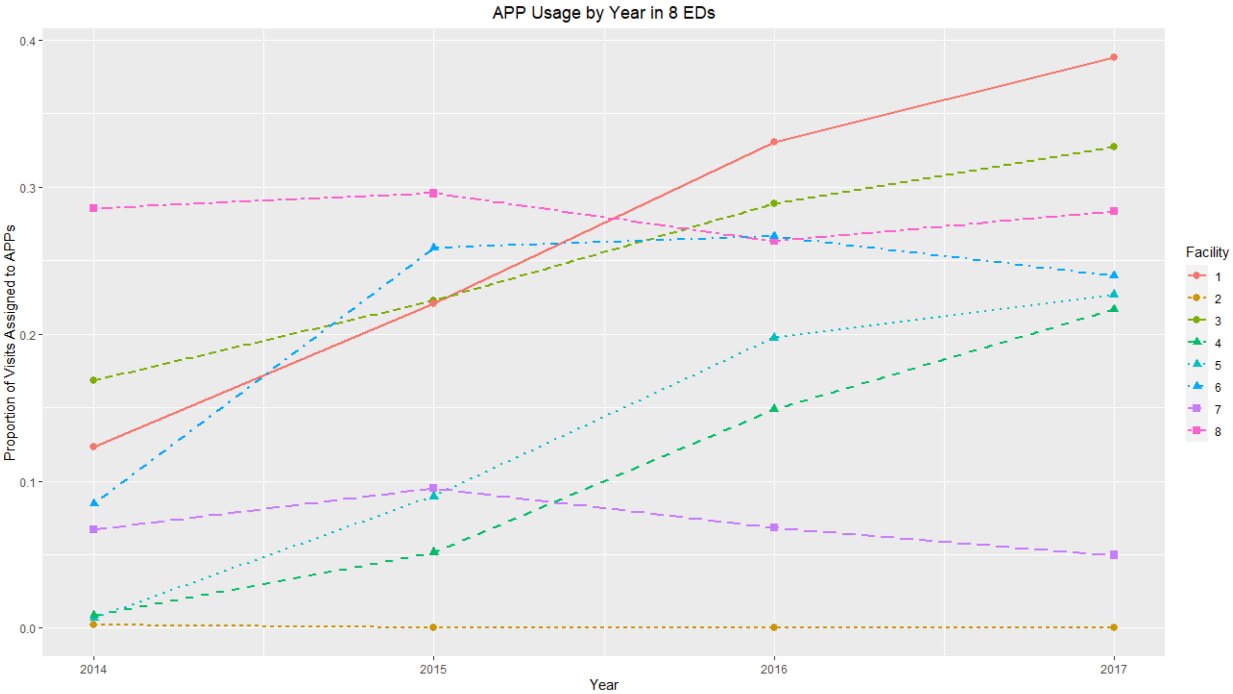
We also consider how the staffing of APPs and the work assigned to APPs differs by ED and how these have changed over time. Fig. 14 depicts the proportion of visits treated by APPs each year from 2014 to 2017 for eight EDs in which we have four continuous years of data. It is evident that APP usage varies by both ED and year. For example, some EDs have increased APP usage each year, while others have opted not to utilize APPs to treat patients independently of physicians. When we take into account the visit's ESI level, it becomes clear that when EDs do assign APPs to patients, the visits assigned to their care are typically classified as ESI levels 4 or 5 (Fig. 14).

Table 15 Description of ED Visits by Year

		Year				
		2014	2015	2016	2017	Total
Number of Visits		381,858	901,048	1,458,409	2,129,479	4,870,794
% of Yearly Visits by ESI Level	1	0.8	0.9	0.8	0.8	0.8
	2	11.6	13.7	13.8	15.7	14.4
	3	44.7	45.7	46.1	49.1	47.2
	4	36.3	34.4	34.8	31.3	33.3
	5	6.6	5.3	4.5	3.0	4.2
	Not Specified	0.0	0.0	0.0	0.0	0.0
% of Visits by Gender	Male	44.5	43.8	43.4	43.6	43.7
	Female	55.5	56.2	56.6	56.3	56.3
	Other	0.0	0.0	0.0	0.0	0.0
% of Visits by Disposition	Admitted	15.6	18.4	17.2	19.5	18.3
	Discharged	80.4	77.8	78.6	76.2	77.5
	Against Medical Advice	1.0	0.9	0.9	1.1	1.0
	Left Without Treatment	0.8	1.2	1.4	1.4	1.3
	Transferred	1.9	1.4	1.5	1.4	1.5
	ED Death	0.1	0.1	0.1	0.1	0.1
	DOA	0.0	0.0	0.0	0.0	0.0
	Other	0.0	0.2	0.3	0.3	0.2
% of Visits by Provider Type	Physician	86.5	80.2	77.4	69.0	74.9
	APP	11.0	17.9	21.2	28.3	22.9
	Both	2.4	1.9	1.4	1.0	1.4
	None	0.1	0.0	0.0	1.8	0.8
Median Patient Age		39.0	38.0	39.0	41.0	40.0
Median Patient Waiting Time (hours)		0.17	0.15	0.17	0.17	0.17
Median Patient Length of Stay (hours)		2.9	2.9	2.9	3.0	2.9
Median Patient Length of Stay (Admitted)		5.8	4.9	4.6	4.8	4.8
Median Patient Length of Stay (Discharged)		2.5	2.5	2.5	2.6	2.6
Median Patient RVUs		3.3	3.3	3.3	3.3	3.3

Table 16 Description of ED Visits by Year and Provider Type

		Year									
		2014		2015		2016		2017		Total	
	Provider Type	P	A	P	A	P	A	P	A	P	A
Number of Visits		330,225	42,028	722,476	161,557	1,128,627	308,599	1,468,376	602,637	3,649,704	1,114,821
% of Yearly Visits by ESI Level	1	0.9	0.0	1.1	0.0	1.0	0.0	1.1	0.0	1.1	0.0
	2	13.2	0.5	16.8	0.5	17.5	0.8	20.5	3.9	18.2	2.4
	3	50.3	6.4	54.7	7.5	56.0	10.5	57.7	28.2	55.9	19.5
	4	29.4	82.3	23.9	78.9	22.7	78.1	18.9	61.9	22.0	69.6
	5	6.1	10.8	3.5	13.0	2.9	10.5	1.9	5.9	2.9	8.4
	Not Specified	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% of Visits by Gender	Male	43.9	46.1	43.2	45.5	42.8	44.7	43.5	43.7	43.3	44.3
	Female	56.1	53.9	56.8	54.5	57.2	55.3	56.5	56.3	56.7	55.7
	Other	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% of Visits by Disposition	Admitted	17.9	0.7	22.5	0.8	21.7	1.4	25.1	6.0	22.9	3.8
	Discharged	77.6	98.6	73.0	98.0	73.3	97.4	69.5	92.2	72.1	94.7
	Against Medical Advice	1.1	0.5	1.0	0.6	1.0	0.5	1.2	0.8	1.1	0.7
	Left Without Treatment	0.9	0.1	1.4	0.4	1.7	0.3	1.9	0.5	1.6	0.4
	Transferred	2.2	0.1	1.6	0.1	1.9	0.2	1.9	0.2	1.9	0.2
	ED Death	0.2	0.0	0.1	0.0	0.2	0.0	0.2	0.0	0.2	0.0
	DOA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Other	0.0	0.0	0.2	0.0	0.3	0.2	0.3	0.2	0.3	0.2
Median Patient Age		41	30	41	30	43	30	46	32	44	31
Median Patient Waiting Time		0.17	0.22	0.15	0.18	0.16	0.22	0.15	0.22	0.15	0.22
Median Patient Length of Stay (hours)		3.1	2.0	3.3	1.8	3.2	1.9	3.5	2.1	3.3	2.0
Median Patient Length of Stay (Admitted)		5.8	5.5	4.9	4.4	4.6	4.5	4.8	5.1	4.8	5.0
Median Patient Length of Stay (Discharged)		2.6	2.0	2.8	1.8	2.9	1.8	3.0	2.0	2.9	1.9
Median Patient RVUs		3.3	3.0	3.3	3.0	3.3	3.3	4.9	3.3	3.6	3.3



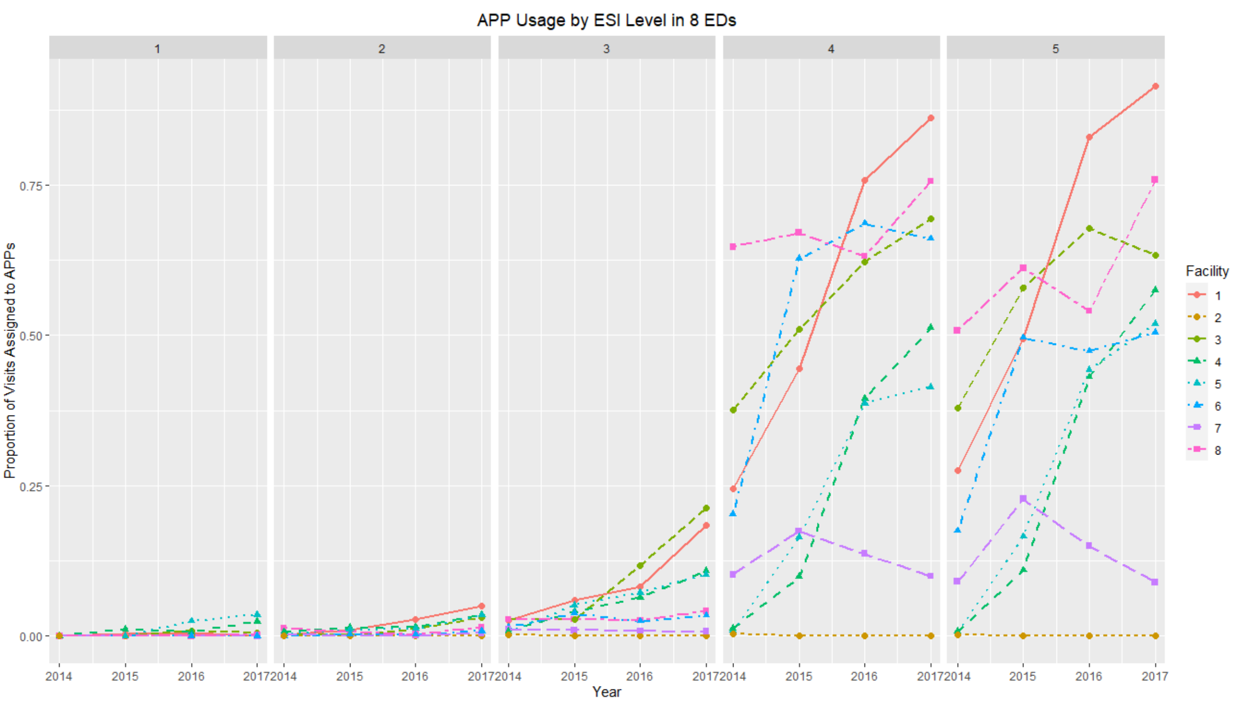


Figure 14 APP Usage by ESI Level in Eight EDs

Fig. 15 shows the distribution of APP usage by both ESI level and year for the entire sample. Because the sample size is increasing with each year, the variability in the proportion of visits assigned to APPs tends to increase over time, but we also see that the median level of APP usage also tends to increase over time. Furthermore, we see that ESI levels 4 and 5 are most likely to be treated by APPs.

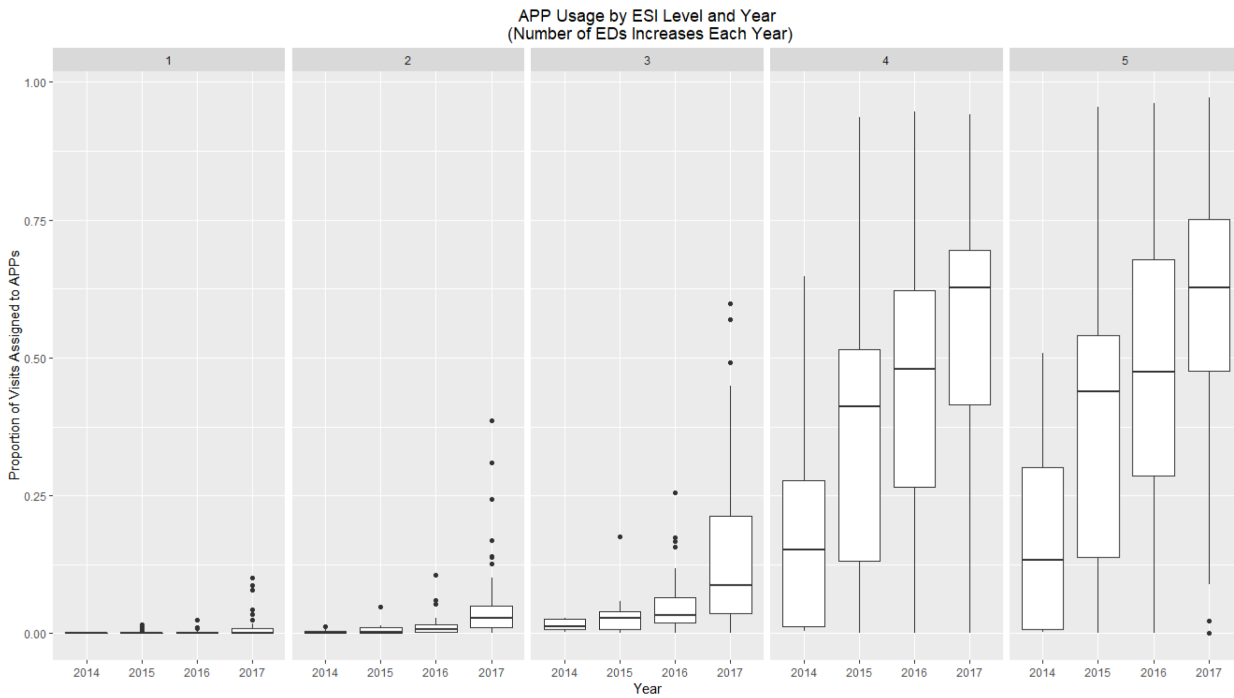


Figure 15 APP Usage by ESI Level and Year

We suspected that the change in the proportion of visits assigned to APPs was partially due to changes in staffing decisions over time. Fig. 16 plots the median staffing levels of physicians and APPs by the hour of the week (Sunday 12am = hour 0) for each year for a single non-academic ED which sees 40,000 to 60,000 visits annually. Over time, this ED has made some changes to its staffing patterns with respect to both physicians and APPs. For example, the median number of physicians working on Mondays at 4pm has decreased from four to two, a 50% reduction in the number of physicians. At the same time, the median number of APPs working on Mondays at 4pm has increased from two to three. After examining this graph, we considered the proportion of hours during each year in which this ED staffed more physicians than APPs. Fig. 17 shows that this proportion decreased each year between 2014 (95.8%) and 2017 (31.7%), and the proportion of visits treated by physicians also steadily declined during the same period. Furthermore, we examined the minimum, median, and maximum number of patient arrivals by hour of the week (Fig. 18) and ESI level (Fig. 19) to determine if the patient population was changing from year to year, but found that the arrival patterns remained consistent over time. Plotting the proportion of ED arrivals classified as ESI levels 4 and 5 and the proportion of ED visits treated by APPs in Fig. 20, we can see that the proportion of visits assigned to APPs has increased since 2014, while the proportion of visits classified as ESI levels 4 and 5 has remained relatively stable.

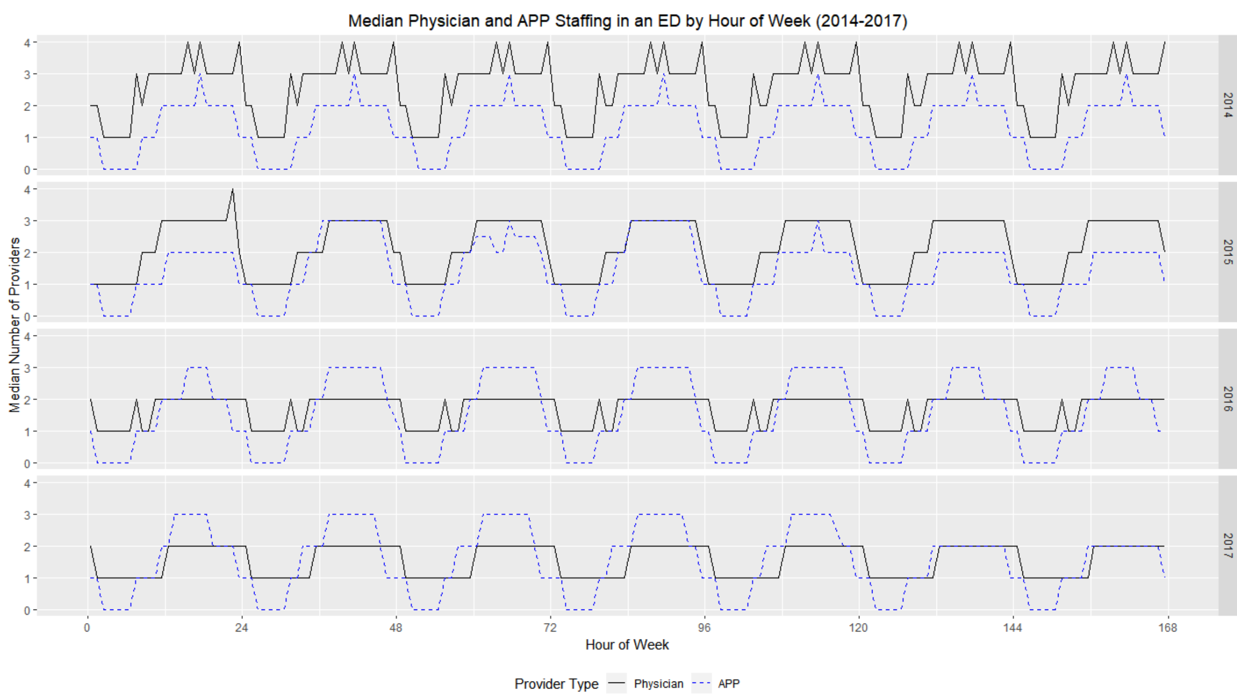


Figure 16 Median Physician and APP Staffing in an ED by Hour of Week (2014-2017)

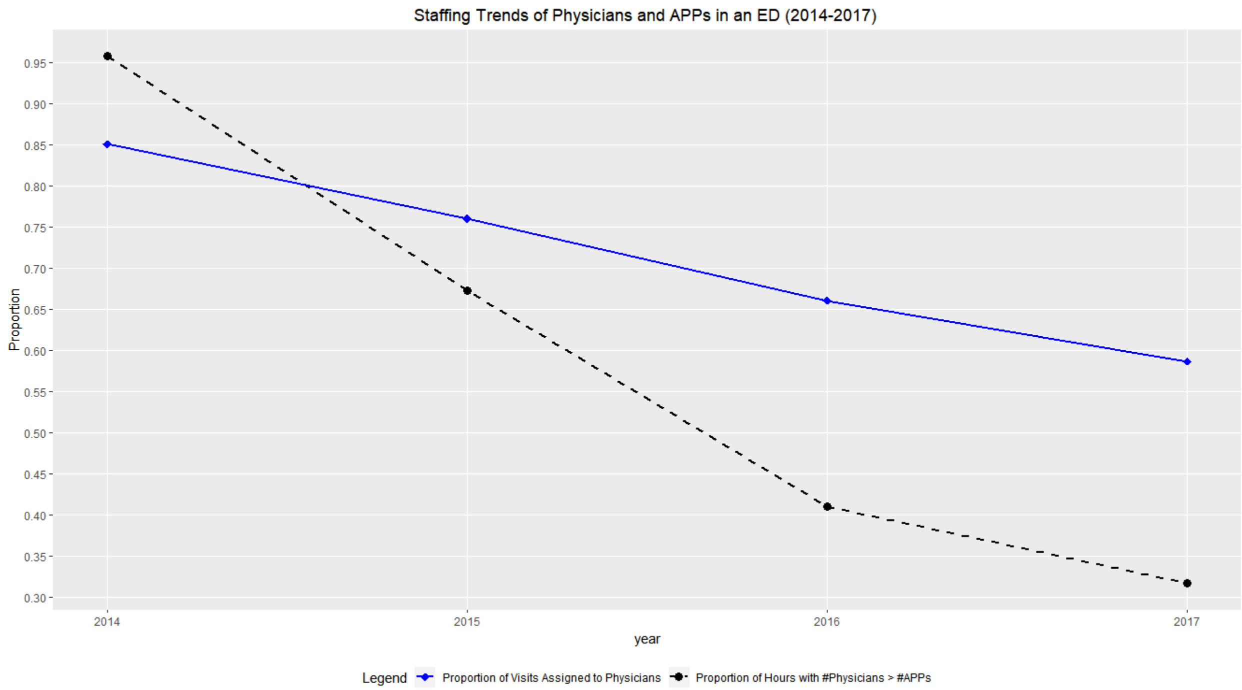


Figure 17 Physician and APP Staffing Trends (2014-2017)

Figure 18 Number of ED Patient Arrivals by Hour of Week

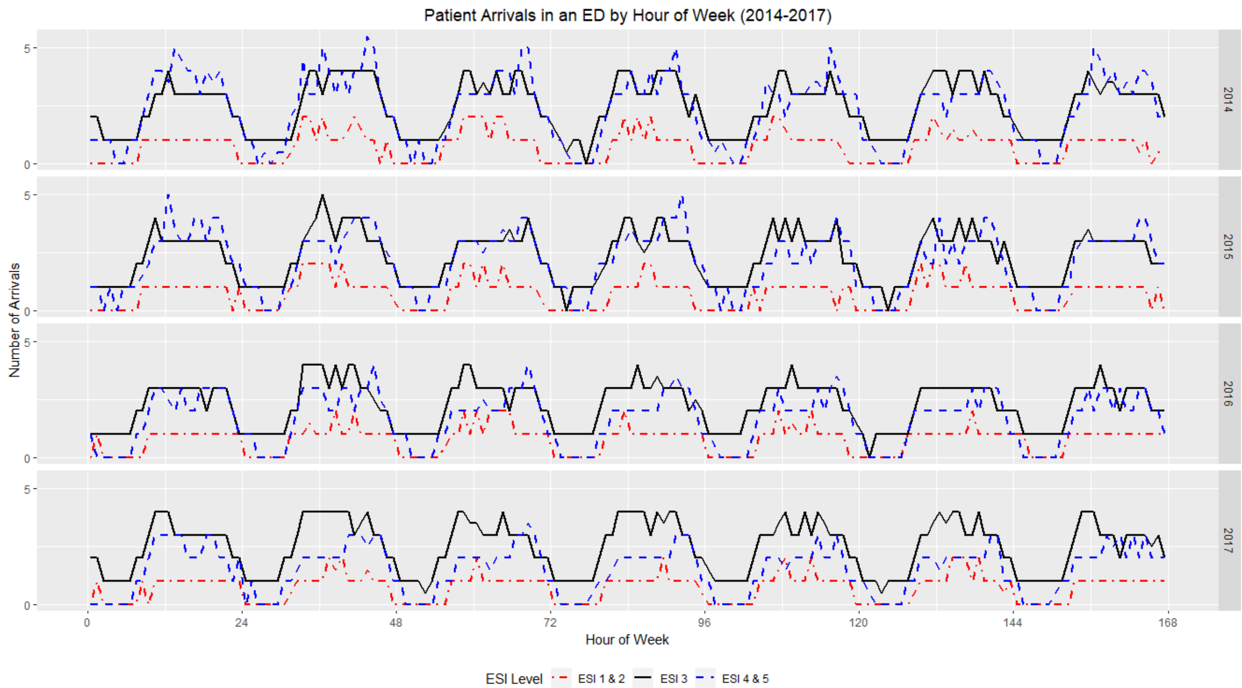


Figure 19 Number of ED Patient Arrivals by ESI Level and Hour of Week

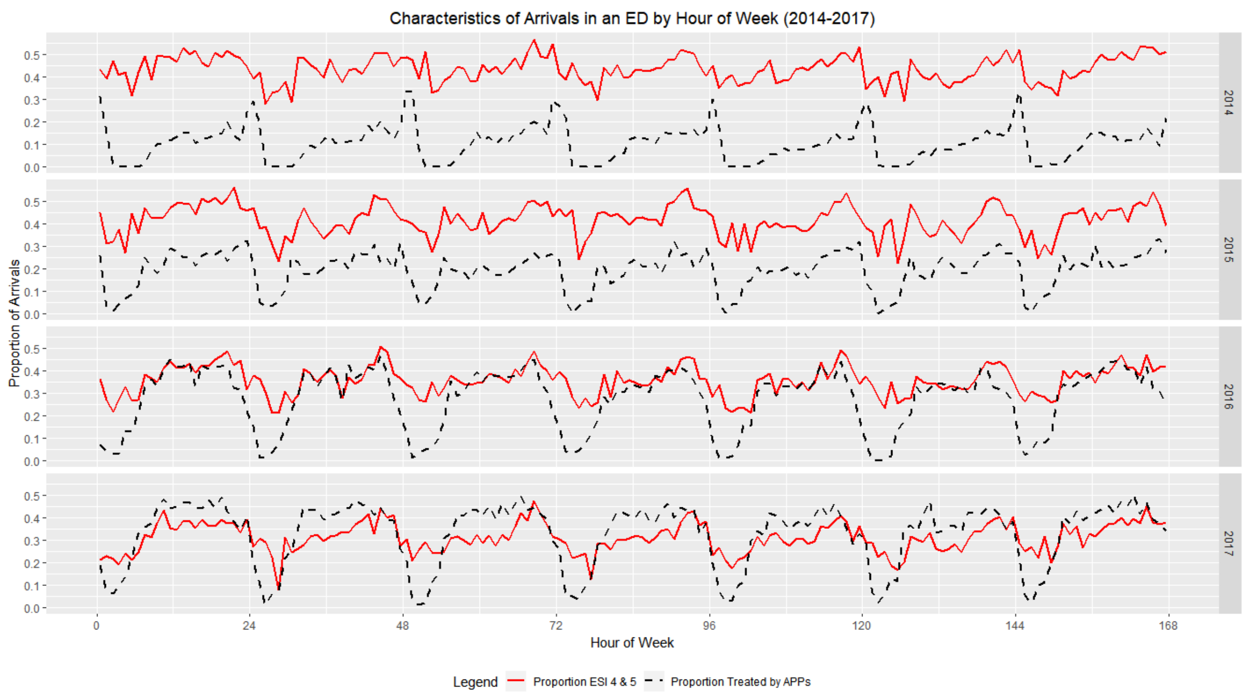


Figure 20 Characteristics of ED Patient Arrivals by Hour of Week

It is clear that this ED's staffing patterns and scheduling practices have evolved over time, but we have not yet examined how this may affect operational characteristics such as waiting times and patient length of stay (LOS) for discharged patients. Before controlling for visit level factors, we find that average and median LOS for discharged patients were both significantly lower in 2014 compared to each of the subsequent years. Similarly, the average and median waiting times for discharged patients were lowest in 2014. This observation may be due to a number of reasons including changes in the patient population, changes to provider staffing, or unobservable changes to the way the facility operates (e.g. nursing staff or policy changes). We consider both median LOS and median wait times for discharged patients by hour of the week in Fig. 21 and 22. In addition to the overall median LOS, we consider the medians for ESI level 3 visits and ESI level 4 visits separately as these two groups cover over 80% of the discharged patients treated each year. The graphs indicate that there are time trends in both LOS and wait times. First, we notice that within each day, the median LOS is typically highest for patients arriving overnight (specifically, between 1am and 4am), with some variability by year. We do note that this result could be due to a lack of resources to safely discharge patients until the morning. Then, we observe that the median waiting times tend to be higher during afternoons and evenings, and the variability in waiting times appears to have increased between 2014 and 2017. This is especially true for ESI level 3 visits, which, we previously observed, are predominantly treated by physicians. To investigate these time trends, we separated the ESI level 3 visits by provider, but since there are hours when APPs are not present in the ED and thus not treating patients, we limited the sample further to visits between noon and midnight.

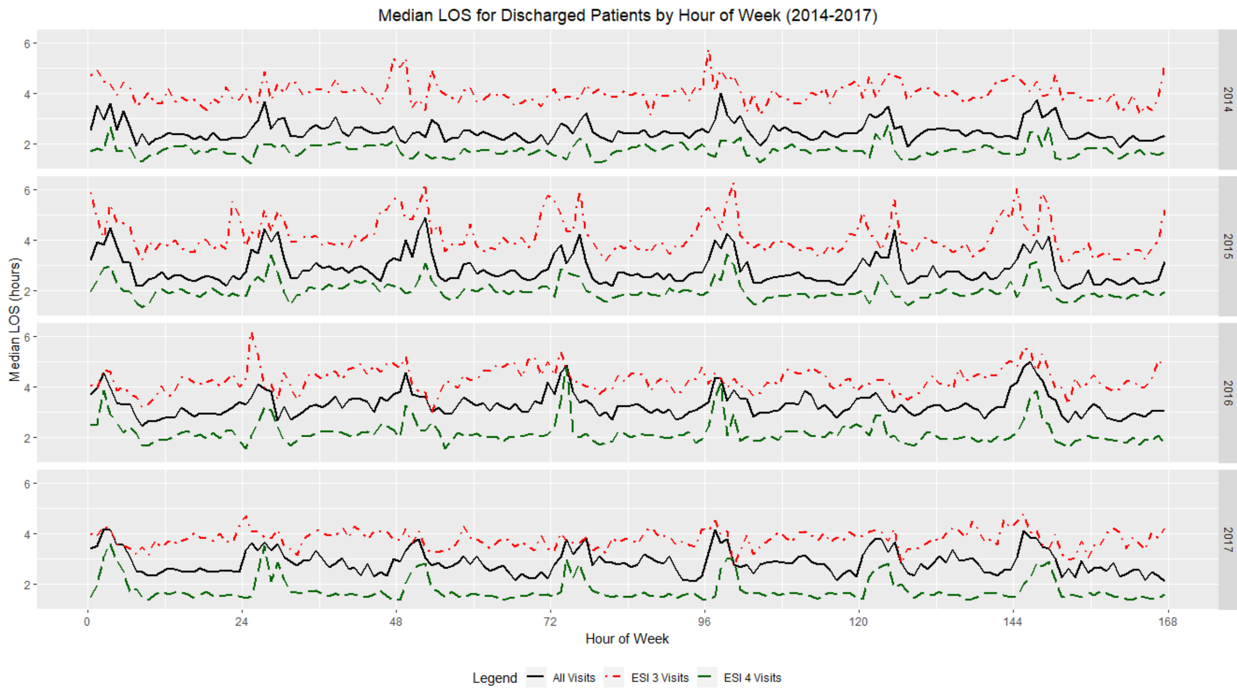


Figure 21 Median Visit Length for Discharged ED Patients by Hour of Week

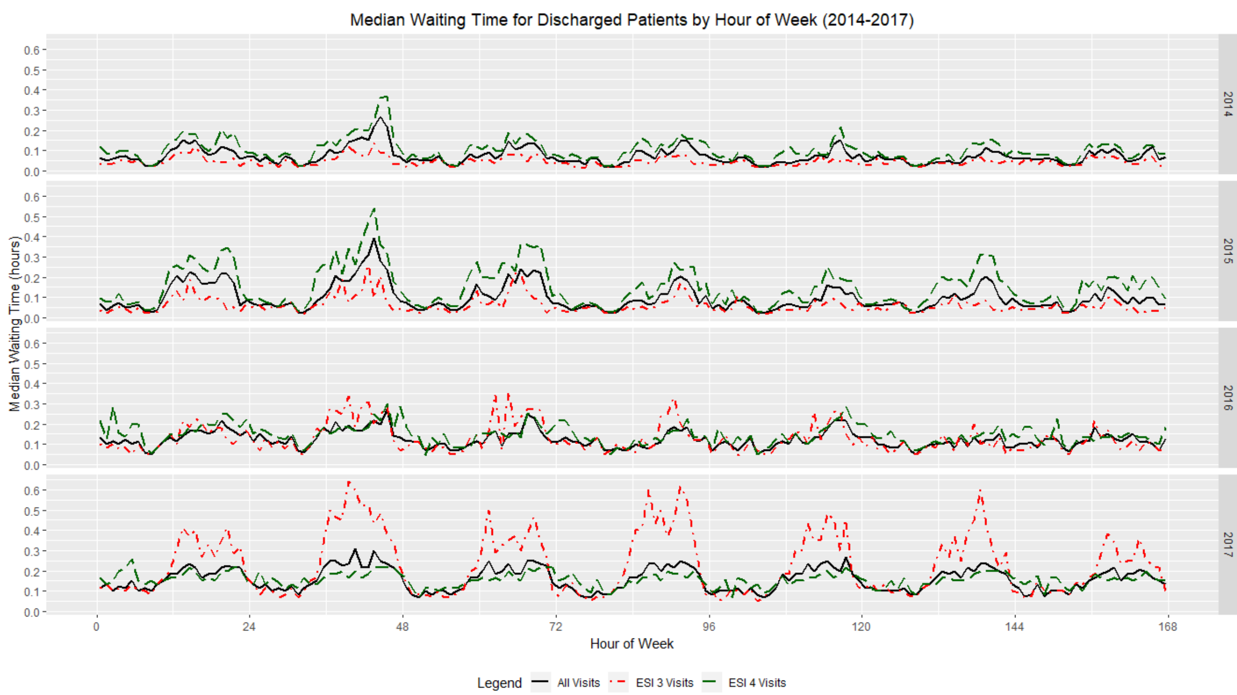


Figure 22 Median Waiting Time for Discharged ED Patients by Hour of Week

After controlling for patient characteristics (age, gender, and ESI level), temporal factors (day of the week and hour of the day), and the operational state of the ED during the hour of arrival (number of new arrivals, the average ESI level of new arrivals, the number of patients who arrived in previous periods, the average ESI level of patients arriving in previous periods, and the total number of providers working), we still find that the average LOS and average waiting times for discharged patients were lowest in 2014. However, we do find that, after controlling for these same factors, patients treated by physicians who are eventually discharged only wait slightly longer than those treated by APPs on average (difference in means = 0.01 hours, p-value $=6.77 \times 10^{-8}$), but their LOS is 0.97 hours longer on average (p-value $< 2.2 \times 10^{-16}$). We acknowledge that this result may be caused by underlying patient comorbidities that are not captured by the ESI Level alone.

While we have shown one ED which has made staffing and scheduling changes over time, other EDs have maintained similar scheduling practices throughout the study period. For example, one non-academic ED, which treats 60,000 to 80,000 visits annually and is a Level 2 trauma center, has continued to utilize only physicians to treat patients. Another ED that sees 20,000 to 40,000 visits per year has consistently relied on similar numbers of physicians (1 to 2) and APPs (0 to 2) to treat patients during each of the four years observed, but this ED also has an emergency medicine residency program, which likely provides additional resources. While we note differences in the characteristics of the observed facilities, these observations demonstrate that there is not a standard way in which EDs utilize APPs to treat patients. In fact, ACEP (2013) acknowledges that multiple staffing models that utilize APPs exist and states that ED medical directors are responsible for choosing "the most appropriate staffing model to achieve operational efficiency, while maintaining clinical quality." Thus, we aim to provide a framework to systematically determine a more effective staffing pattern.

5.4.2 Predictive Analysis

In order to better inform decisions, we consider predictive models to predict visit characteristics based on patient, physician, facility, and temporal factors. We utilize emergency severity index (ESI) levels to differentiate visits into three categories. The highest acuity patients are assigned to ESI levels 1 or 2, while the lowest acuity patients are assigned to ESI levels 4 or 5. For each of the following predictive models, data from 2014 through 2016 are used to train the model, and the test set consists of the 2017 data.

We first use visit characteristics as well as provider type and the current state of the ED to predict the length of stay for discharged patients. We limit this analysis to discharged patients because the length of stay for admitted is influenced by factors beyond the ED's control such as the availability of inpatient beds. Table 17 summarizes the results of six regression models, and Table 18 reports the results when ESI Levels 1 and 2 are excluded since it is rare for an APP to treat those patients. In all twelve cases, we observe a negative and statistically significant coefficient for Provider Type APP, suggesting that visit length is shorter for visits treated by APPs when compared to those treated by physicians when controlling for patient and ED related factors. This provides strong supports for H5. Namely, the coefficient from the initial model in Table 17 (Model A) suggests that discharged patients who are treated by APPs have shorter ED stays than those treated by physicians on average ($\beta_1 = -0.0767$). Exponentiating this value, we estimate that the average LOS for discharged patients treated by APPs is 0.93 times the LOS for discharged patients treated by physicians, after controlling for both patient and ED factors. When we limit the analysis to ESI Levels 1 through 3, the result is equivalent ($\beta_1 = -0.0774, \exp(\beta_1) = 0.9255$). While we note that this observation may result from unobserved differences in the patients between the two provider types (e.g., APPs may be assigned to "easier" within each ESI level), we have controlled for differences in patient demographics, severity, time of arrival, and the current ED conditions.

Next, we test our hypothesis H6 which considers the effects of adding providers. In each of the six models reported in Table 17, the coefficients for Physician Count and APP Count are both negative and statistically significant. This suggests support for H6 (a) and (b). All six models reported in Table 18, also show that the coefficient for APP Count is negative and statistically significant, further supporting H6 (b). However, the full model reported in Table 18 (Model F) suggests that after controlling for patient characteristics, time of day, and current ED conditions, the Physician Count coefficient is not significantly different from zero. That is, adding a physician may not effectively reduce the LOS for an ESI level 3 patient who arrives between 7am and 11am and is treated by a physician (our base case), but significant interaction terms suggest the effects of adding a physician are dependent on patient severity and time of arrival. Thus, our models largely support H6 (a) and (b) and suggest that adding a provider (either a physician or APP) is associated with shorter LOS for discharged patients. The estimates provided in column A of Table 17 suggest that the average LOS is reduced by a factor of 0.93 [$= \exp(\beta_2) = \exp(-0.0721)$] for each additional physician, holding all else constant, while the average LOS is reduced by a factor of 0.98 [$= \exp(\beta_3) = \exp(-0.0244)$] for each additional APP. This result is similar when the analysis is limited to lower acuity patients (Table 18: $\beta_2 = -0.0762$ and $\beta_3 = -0.0762$). Additionally, the 95% confidence intervals for β_2 and β_3 suggest that the effect of adding a physician is stronger than the effect of adding an APP. Thus, we also find support for H6 (c).

In the second model (Table 17, Model B), we test how the effect of being treated by an APP may change with additional providers. The interactions, APP Indicator × Physician Count and APP Indicator × APP Count are both positive and statistically significant ($\beta_4 = 0.0049$ and $\beta_5 = 0.0393$). This supports the idea that the effect of being treated by an APP diminishes as the number of providers increases, our hypothesis, H7. Alternatively, these positive interaction terms mean that the marginal effect on LOS of adding a provider (either an APP or a physician) is greater in magnitude (i.e., more negative) for patients who are treated by physicians compared to those treated by APPs. These results also hold when the analysis is limited lower acuity patients (Table 18, Model B: $\beta_4 = 0.0056$ and $\beta_5 = 0.0376$).

The third and fourth models (Table 17, Models C and D) test whether or not our previous results differ based on patient severity (ESI Level). In Model C, we see that the effect of being treated by an APP is lower in magnitude for ESI Level 4 and 5 patients compared to ESI Level 1 and 2 patients ($\beta_{7_3} = 0.2129$), but we know from the descriptive analysis that it's rare for an APP to treat an ESI Level 1 or 2 patient. Thus, we refer to Table 18 for the comparison between Level 3 and Level 4 or 5 and find that the effect of being treated by an APP is actually stronger (i.e., more negative) among ESI Level 4 and 5 patients ($\beta_{7_3} = 0.1723$). In the previous section, we presented descriptive statistics that indicated that several EDs have been increasing their usage of APPs in the past several years, especially with regard to ESI Level 4 and 5 patients, and this result confirms that such policy changes are associated with shorter discharge LOS.

We further consider how the number of providers impacts different patients in Model D. First we consider the interaction between Physician Count and ESI Level. The positive and significant coefficient for Physician Count × ESI Level 3 ($\beta_{8_2} = 0.0443$) suggests that the impact of adding a physician is greater for ESI Level 1 and 2 patients than for ESI Level 3 patients. That is, ESI Level 1 and 2 patients will see a larger reduction in discharged LOS, after controlling for the patient's age, gender, and insurance, the number of APPs working, the ED census, the half-shift of arrival, and the facility. The negative and significant coefficient for Physician Count × ESI Level 4 or 5 (β_{8_3} =-0.0522) suggests that the lowest acuity patients see the greatest benefit from additional physicians. When we limit the analysis to ESI Levels 3 through 5, it is clear that ESI Level 4 and 5 patients experience a greater reduction in LOS when a physician is added in comparison to Level 3 patients ($\beta_{8_3} = -0.0966$).

We next consider the interaction between APP Count and ESI Level. The coefficient for APP Count × ESI Level 3 is not significantly different from zero (β_{9_3} = 0.0003), which indicates the effect of adding an APP is similar for ESI Level 1 and 2 patients and ESI Level 3 patients. The positive and significant coefficient for APP Count × ESI Level 4 or 5 (β_{9_3} = 0.0413) suggests that the lowest acuity patients see the least benefit from additional APPs. This result holds when co ESI Levels 3 through 5 (β_{9_3} = 0.0410), and thus we do not find support for H8.

Finally, we consider differences based on a patient's arrival time. We observe that the difference in average discharge LOS between patients treated by APPs and patients treated by physicians was greatest between the hours of 11pm and 3am ($\beta_{15_5} = -0.1296$). The smallest difference between the two groups was observed between 3am and 7am (β_{15_6} = 0.2969), with all four-hour time periods differing significantly from the reference group (7am to 11am). Similar results are seen in Table 18 when the most acute patients are excluded from the analysis. We also observe that the effects of adding providers differ by time of day. Namely, adding a physician is associated with significantly shorter LOS between the hours of 11am and 11pm when compared to the reference group (7am to 11am), while the most beneficial times to add an APP are between 7am and 11am or 7pm and 11pm. The results are similar when limited to lower acuity patients. However, the results do not fully support H9 as the negative relationship between the APP count and the LOS for discharged visits becomes weaker for patients arriving between 11pm and 3am ($\beta_{17_5} = 0.0322$), but the coefficient for the 3am to 7am time period ($\beta_{17_6} = 0.0015$) is not statistically significant. This is also true of the results presented in Table 18 ($\beta_{17_5} = 0.0356$ and $\beta_{17_6} = -0.0114$).

These results highlight that ED staffing decisions should depend on the time of day as well as the expected patient population. If visit lengths are too long and management is considering adding a provider, physicians will generally be more effective than APPs. However, physicians are also a more costly resource, so management should carefully consider the costs and benefits. Furthermore, the benefits of adding providers vary by both patient severity and time of day. For instance, our analyses indicate that physicians will have the biggest impact on LOS between 11am and 11pm, while APPs have the greatest potential between 7am and 11am and 7pm and 11pm. This suggests that it makes most sense to add a physician between the hours of 11am and 7pm, one of the busiest time periods of the day for most EDs, but the decision is less clear for other time periods.

Table 17 Predicting Discharge Visit Length

Dependent Variable: ln(LOS) (Discharge Visits Only)	A			B			C			D			E			F		
Variable	Coeff		Std Error	Coeff		Std Error	Coeff		Std Error	Coeff		Std Error	Coeff		Std Error	Coeff		Std Error
(Constant)	1.1890	***	0.0041	1.2100	***	0.0042	1.1930	***	0.0041	1.1956	***	0.0066	1.1550	***	0.0061	1.1500	***	0.0080
APP Indicator	-0.0767	***	0.0018	-0.1480	***	0.0044	-0.2653	***	0.0213	-0.0670	***	0.0018	-0.1236	***	0.0038	-0.4636	***	0.0217
Physician Count	-0.0721	***	0.0010	-0.0730	***	0.0011	-0.0720	***	0.0010	-0.0637	***	0.0023	-0.0465	***	0.0025	-0.0338	***	0.0032
APP Count	-0.0244	***	0.0009	-0.0373	***	0.0010	-0.0241	***	0.0009	-0.0458	***	0.0018	-0.0345	***	0.0027	-0.0486	***	0.0031
APP Indicator × Physician Count				0.0049	**	0.0015										0.0623	***	0.0019
APP Indicator × APP Count				0.0393	***	0.0011										0.0196	***	0.0015
ESI Level 1 or 2 (reference)																		
ESI Level 3	-0.2258	***	0.0024	-0.2265	***	0.0024	-0.2231	***	0.0024	-0.3292	***	0.0063	-0.2253	***	0.0024	-0.3102	***	0.0063
ESI Level 4 or 5	-0.8207	***	0.0025	-0.8231	***	0.0025	-0.8333	***	0.0025	-0.7494	***	0.0063	-0.8208	***	0.0025	-0.6796	***	0.0065
APP Indicator × ESI Level 3							0.0395		0.0218							0.0485	*	0.0217
APP Indicator × ESI Level 4 or 5							0.2129	***	0.0213							0.2209	***	0.0213
Physician Count × ESI Level 3										0.0443	***	0.0024				0.0388	***	0.0024
Physician Count × ESI Level 4 or 5										-0.0522	***	0.0023				-0.0812	***	0.0025
APP Count × ESI Level 3										0.0003		0.0018				-0.0002		0.0018
APP Count × ESI Level 4 or 5										0.0413	***	0.0018				0.0259	***	0.0020
APP Indicator × Shift 11am-3pm													0.0722	***	0.0047	0.0432	***	0.0050
APP Indicator × Shift 3pm-7pm													0.0845	***	0.0046	0.0435	***	0.0051
APP Indicator × Shift 7pm-11pm													0.0578	***	0.0049	0.0192	***	0.0053
APP Indicator × Shift 11pm-3am													-0.1296	***	0.0072	-0.1439	***	0.0072
APP Indicator × Shift 3am-7am													0.2969	***	0.0188	0.3239	***	0.0188
Physician Count × Shift 11am-3pm													-0.0329	***	0.0027	-0.0377	***	0.0027
Physician Count × Shift 3pm-7pm													-0.0358	***	0.0026	-0.0388	***	0.0026
Physician Count × Shift 7pm-11pm													-0.0312	***	0.0026	-0.0364	***	0.0026
Physician Count × Shift 11pm-3am													0.0004		0.0036	-0.0002		0.0035
Physician Count × Shift 3am-7am													0.1020	***	0.0055	0.1077	***	0.0054
APP Count × Shift 11am-3pm													0.0105	***	0.0027	0.0064	*	0.0027
APP Count × Shift 3pm-7pm													0.0169	***	0.0026	0.0118	***	0.0026
APP Count × Shift 7pm-11pm													-0.0006		0.0027	-0.0048	.	0.0027
APP Count × Shift 11pm-3am													0.0322	***	0.0035	0.0335	***	0.0035
APP Count × Shift 3am-7am													0.0015		0.0154	0.0081		0.0153
0 to 3 months	-0.0961	***	0.0049	-0.0953	***	0.0049	-0.0943	***	0.0049	-0.0993	***	0.0049	-0.0965	***	0.0049	-0.0962	***	0.0049
4 to 36 months	-0.0961	***	0.0036	-0.0963	***	0.0036	-0.0949	***	0.0036	-0.0980	***	0.0036	-0.0963	***	0.0036	-0.0964	***	0.0036
3 to 8 years	-0.0971	***	0.0028	-0.0970	***	0.0028	-0.0965	***	0.0028	-0.0982	***	0.0028	-0.0973	***	0.0028	-0.0975	***	0.0028
9 to 17 years	-0.0348	***	0.0024	-0.0345	***	0.0024	-0.0349	***	0.0024	-0.0334	***	0.0024	-0.0354	***	0.0024	-0.0335	***	0.0024
18 to 34 years (reference)																		
35 to 44 years	0.0505	***	0.0019	0.0504	***	0.0019	0.0501	***	0.0019	0.0506	***	0.0019	0.0505	***	0.0019	0.0500	***	0.0019
45 to 54 years	0.0881	***	0.0020	0.0880	***	0.0020	0.0873	***	0.0020	0.0881	***	0.0020	0.0880	***	0.0020	0.0870	***	0.0019
55 to 64 years	0.1061	***	0.0022	0.1061	***	0.0022	0.1051	***	0.0022	0.1061	***	0.0022	0.1060	***	0.0022	0.1046	***	0.0022
65 to 74 years	0.0965	***	0.0032	0.0962	***	0.0032	0.0954	***	0.0032	0.0968	***	0.0032	0.0966	***	0.0032	0.0954	***	0.0032
75 to 84 years	0.1296	***	0.0037	0.1293	***	0.0037	0.1282	***	0.0037	0.1300	***	0.0037	0.1300	***	0.0037	0.1284	***	0.0037
85 years plus	0.1825	***	0.0043	0.1818	***	0.0043	0.1809	***	0.0043	0.1815	***	0.0042	0.1828	***	0.0043	0.1793	***	0.0042
Male Indicator	-0.0372	***	0.0012	-0.0377	***	0.0012	-0.0375	***	0.0012	-0.0377	***	0.0012	-0.0374	***	0.0012	-0.0383	***	0.0012
Commercial Insurance (reference)																		
Medicaid	0.0016		0.0016	0.0019		0.0016	0.0017		0.0016	0.0002		0.0016	0.0017		0.0016	0.0010		0.0016
Medicare	0.0648	***	0.0025	0.0650	***	0.0025	0.0645	***	0.0025	0.0635	***	0.0025	0.0649	***	0.0025	0.0635	***	0.0024
Self-Pay	-0.0215	***	0.0025	-0.0038		0.0022	-0.0031		0.0022	-0.0054	*	0.0022	-0.0036		0.0022	-0.0053	*	0.0022
Other	-0.0033		0.0022	-0.0212	***	0.0025	-0.0214	***	0.0025	-0.0180	***	0.0025	-0.0224	***	0.0025	-0.0192	***	0.0025
ED Census	0.0133	***	0.0001	0.0133	***	0.0001	0.0133	***	0.0001	0.0134	***	0.0001	0.0136	***	0.0001	0.0136	***	0.0001
Shift 7am-11am (reference)																		
Shift 11am-3pm	0.0351	***	0.0022	0.0381	***	0.0022	0.0345	***	0.0022	0.0332	***	0.0022	0.0747	***	0.0060	0.0925	***	0.0060
Shift 3pm-7pm	0.0506	***	0.0024	0.0539	***	0.0024	0.0495	***	0.0024	0.0486	***	0.0024	0.0843	***	0.0059	0.1007	***	0.0059
Shift 7pm-11pm	0.1010	***	0.0023	0.1022	***	0.0023	0.0991	***	0.0023	0.0981	***	0.0023	0.1556	***	0.0058	0.1716	***	0.0058
Shift 11pm-3am	0.1152	***	0.0024	0.1126	***	0.0024	0.1147	***	0.0024	0.1159	***	0.0024	0.1044	***	0.0068	0.1041	***	0.0068
Shift 3am-7am	0.0708	***	0.0030	0.0583	***	0.0030	0.0712	***	0.0030	0.0743	***	0.0030	-0.0795	***	0.0086	-0.0886	***	0.0086
Facility 1 (reference)																		
Facility 2	-0.1631	***	0.0025	-0.1693	***	0.0025	-0.1605	***	0.0025	-0.1605	***	0.0025	-0.1592	***	0.0025	-0.1629	***	0.0025
Facility 3	0.1137	***	0.0030	0.1102	***	0.0030	0.1147	***	0.0030	0.1154	***	0.0030	0.1123	***	0.0030	0.1144	***	0.0030
Facility 4	0.0371	***	0.0029	0.0315	***	0.0029	0.0400	***	0.0029	0.0441	***	0.0029	0.0434	***	0.0029	0.0442	***	0.0029
Facility 5	-0.0493	***	0.0029	-0.0516	***	0.0029	-0.0463	***	0.0029	-0.0418	***	0.0029	-0.0436	***	0.0029	-0.0421	***	0.0029
Facility 6	-0.3245	***	0.0031	-0.3269	***	0.0031	-0.3268	***	0.0031	-0.3283	***	0.0031	-0.3332	***	0.0032	-0.3412	***	0.0032
Facility 7	-0.0205	***	0.0041	-0.0344	***	0.0041	-0.0175	***	0.0041	-0.0255	***	0.0041	-0.0187	***	0.0041	-0.0463	***	0.0042
Facility 8	0.0517	***	0.0022	0.0471	***	0.0022	0.0469	***	0.0022	0.0503	***	0.0022	0.0455	***	0.0023	0.0299	***	0.0023
R^2	35.63%			35.73%			35.73%			36.08%			35.82%			36.50%		
Adjusted R^2	35.63%			35.73%			35.73%			36.08%			35.82%			36.50%		

***p < 0.001, **p < 0.01, *p < 0.05; two-tailed tests

Table 18 Predicting Discharge Visit Length (ESI Levels 1 and 2 Excluded)

Dependent Variable: ln(LOS) (Discharge Visits Only)

Variable	A			B			C			D			E			F		
	Coeff		Std Error	Coeff		Std Error	Coeff		Std Error	Coeff		Std Error	Coeff		Std Error	Coeff		Std Error
(Constant)	0.9435	***	0.0037	0.9650	***	0.0039	0.9498	***	0.0037	0.8465	***	0.0040	0.9066	***	0.0059	0.8159	***	0.0060
APP Indicator	-0.0774	***	0.0018	-0.1473	***	0.0044	-0.2267	***	0.0046	-0.0665	***	0.0018	-0.1223	***	0.0038	-0.4159	***	0.0066
Physician Count	-0.0762	***	0.0011	-0.0773	***	0.0012	-0.0762	***	0.0011	-0.0225	***	0.0013	-0.0491	***	0.0025	0.0036		0.0026
APP Count	-0.0262	***	0.0009	-0.0396	***	0.0010	-0.0259	***	0.0009	-0.0494	***	0.0011	-0.0352	***	0.0027	-0.0505	***	0.0028
APP Indicator × Physician Count				0.0056	***	0.0015										0.0637	***	0.0019
APP Indicator × APP Count				0.0376	***	0.0011										0.0191	***	0.0014
ESI Level 3 (reference)																		
ESI Level 4 or 5	-0.5933	***	0.0014	-0.5952	***	0.0014	-0.6081	***	0.0015	-0.4187	***	0.0032	-0.5940	***	0.0014	-0.3674	***	0.0034
APP Indicator × ESI Level 4 or 5							0.1722	***	0.0049							0.1708	***	0.0050
Physician Count × ESI Level 4 or 5										-0.0966	***	0.0013				-0.1204	***	0.0015
APP Count × ESI Level 4 or 5										0.0410	***	0.0011				0.0267	***	0.0013
APP Indicator × Shift 11am-3pm													0.0693	***	0.0047	0.0401	***	0.0050
APP Indicator × Shift 3pm-7pm													0.0827	***	0.0047	0.0417	***	0.0051
APP Indicator × Shift 7pm-11pm													0.0556	***	0.0049	0.0180	***	0.0053
APP Indicator × Shift 11pm-3am													-0.1291	***	0.0072	-0.1434	***	0.0072
APP Indicator × Shift 3am-7am													0.3028	***	0.0187	0.3298	***	0.0186
Physician Count × Shift 11am-3pm													-0.0342	***	0.0027	-0.0395	***	0.0027
Physician Count × Shift 3pm-7pm													-0.0383	***	0.0027	-0.0416	***	0.0027
Physician Count × Shift 7pm-11pm													-0.0328	***	0.0027	-0.0385	***	0.0027
Physician Count × Shift 11pm-3am													-0.0002		0.0037	-0.0009		0.0037
Physician Count × Shift 3am-7am													0.1063	***	0.0057	0.1127	***	0.0056
APP Count × Shift 11am-3pm													0.0093	***	0.0028	0.0045		0.0028
APP Count × Shift 3pm-7pm													0.0159	***	0.0027	0.0098	***	0.0027
APP Count × Shift 7pm-11pm													-0.0014		0.0028	-0.0065	*	0.0028
APP Count × Shift 11pm-3am													0.0356	***	0.0037	0.0366	***	0.0036
APP Count × Shift 3am-7am													-0.0114		0.0159	-0.0046		0.0158
0 to 3 months	-0.0828	***	0.0049	-0.0819	***	0.0049	-0.0811	***	0.0049	-0.0859	***	0.0049	-0.0832	***	0.0049	-0.0828	***	0.0049
4 to 36 months	-0.0826	***	0.0036	-0.0828	***	0.0036	-0.0815	***	0.0036	-0.0844	***	0.0036	-0.0828	***	0.0036	-0.0829	***	0.0036
3 to 8 years	-0.0877	***	0.0028	-0.0877	***	0.0028	-0.0873	***	0.0028	-0.0889	***	0.0028	-0.0879	***	0.0028	-0.0882	***	0.0028
9 to 17 years	-0.0362	***	0.0024	-0.0360	***	0.0024	-0.0363	***	0.0024	-0.0350	***	0.0024	-0.0368	***	0.0024	-0.0350	***	0.0024
18 to 34 years (reference)																		
35 to 44 years	0.0523	***	0.0020	0.0521	***	0.0020	0.0518	***	0.0020	0.0523	***	0.0020	0.0523	***	0.0020	0.0516	***	0.0019
45 to 54 years	0.0890	***	0.0020	0.0888	***	0.0020	0.0882	***	0.0020	0.0889	***	0.0020	0.0890	***	0.0020	0.0878	***	0.0020
55 to 64 years	0.1104	***	0.0023	0.1103	***	0.0023	0.1093	***	0.0023	0.1101	***	0.0023	0.1103	***	0.0023	0.1086	***	0.0023
65 to 74 years	0.1125	***	0.0034	0.1121	***	0.0034	0.1112	***	0.0034	0.1127	***	0.0033	0.1126	***	0.0033	0.1113	***	0.0033
75 to 84 years	0.1515	***	0.0039	0.1512	***	0.0039	0.1500		0.0039	0.1520	***	0.0039	0.1521	***	0.0039	0.1504	***	0.0039
85 years plus	0.2046	***	0.0045	0.2040	***	0.0045	0.2028	***	0.0045	0.2036	***	0.0044	0.2051	***	0.0045	0.2013	***	0.0044
Male Indicator	-0.0416	***	0.0012	-0.0421	***	0.0012	-0.0419	***	0.0012	-0.0422	***	0.0012	-0.0419	***	0.0012	-0.0428	***	0.0012
Commercial Insurance (reference)																		
Medicaid	-0.0090	***	0.0017	-0.0087	***	0.0017	-0.0088	***	0.0017	-0.0105	***	0.0016	-0.0089	***	0.0017	-0.0096	***	0.0016
Medicare	0.0546	***	0.0026	0.0547	***	0.0026	0.0544	***	0.0026	0.0529	***	0.0026	0.0547	***	0.0026	0.0528	***	0.0026
Self-Pay	-0.0106	***	0.0023	-0.0110	***	0.0023	-0.0103	***	0.0023	-0.0128	***	0.0023	-0.0108	***	0.0023	-0.0125	***	0.0023
Other	-0.0233	***	0.0025	-0.0232	***	0.0025	-0.0233	***	0.0025	-0.0198	***	0.0025	-0.0243	***	0.0025	-0.0212	***	0.0025
ED Census	0.0141	***	0.0001	0.0141	***	0.0001	0.0141	***	0.0001	0.0141	***	0.0001	0.0143	***	0.0001	0.0144	***	0.0001
Shift 7am-11am (reference)																		
Shift 11am-3pm	0.0368	***	0.0023	0.0401	***	0.0023	0.0364	***	0.0023	0.0348	***	0.0023	0.0785	***	0.0061	0.0980	***	0.0061
Shift 3pm-7pm	0.0512	***	0.0024	0.0546	***	0.0024	0.0501	***	0.0024	0.0490	***	0.0024	0.0894	***	0.0060	0.1075	***	0.0061
Shift 7pm-11pm	0.1029	***	0.0023	0.1042	***	0.0023	0.1011	***	0.0023	0.0998	***	0.0023	0.1596	***	0.0059	0.1775	***	0.0060
Shift 11pm-3am	0.1154	***	0.0025	0.1128	***	0.0025	0.1149	***	0.0025	0.1164	***	0.0025	0.1046	***	0.0070	0.1046	***	0.0069
Shift 3am-7am	0.0700	***	0.0031	0.0570	***	0.0031	0.0705	***	0.0031	0.0742	***	0.0030	-0.0836	***	0.0089	-0.0935	***	0.0088
Facility 1 (reference)																		
Facility 2	0.0020		0.0042	-0.0128	*	0.0042	0.0047		0.0042	-0.0032		0.0041	0.0043		0.0042	-0.0246	***	0.0042
Facility 3	0.0692	***	0.0023	0.0641	***	0.0023	0.0643	***	0.0023	0.0678	***	0.0023	0.0624	***	0.0024	0.0456	***	0.0024
Facility 4	-0.1499	***	0.0025	-0.1564	***	0.0025	-0.1476	***	0.0025	-0.1469	***	0.0025	-0.1456	***	0.0025	-0.1499	***	0.0025
Facility 5	0.1433	***	0.0031	0.1396	***	0.0031	0.1444	***	0.0031	0.1454	***	0.0031	0.1422	***	0.0031	0.1447	***	0.0031
Facility 6	0.0646	***	0.0030	0.0587	***	0.0030	0.0675	***	0.0030	0.0728	***	0.0030	0.0715	***	0.0031	0.0722	***	0.0031
Facility 7	-0.0356	***	0.0029	-0.0383	***	0.0030	-0.0328	***	0.0029	-0.0270	***	0.0029	-0.0295	***	0.0030	-0.0277	***	0.0030
Facility 8	-0.3382	***	0.0033	-0.3404	***	0.0033	-0.3405	***	0.0033	-0.3414	***	0.0033	-0.3461	***	0.0033	-0.3537	***	0.0033
R^2		33.52%			33.62%			33.62%			34.02%			33.73%			34.49%	
Adjusted R^2		33.52%			33.62%			33.62%			34.02%			33.73%			34.49%	

***p < 0.001, **p < 0.01, *p < 0.05; two-tailed tests

Next, we fit a model to predict the numbers of arrivals of each ESI level during a given time period. Due to our previous observations and the low probability of Level 1 and Level 5 patients, we grouped ESI levels 1 and 2 together, and we also combined ESI Levels 4 and 5. In order to simultaneously estimate the number of hourly visits of each level, we use a vector generalized linear model (Eq. (13)). Since the number of arrivals during a given hour is a count, we model these counts using a Poisson distribution. Control variables include the facility, month of the year, day of the week, time of day (half-shift), indicators for holidays and the day after a holiday, and the normal high and low temperatures for each given date and facility location. The mean absolute deviation (MAD) and mean squared error (MSE) are comparable for the training and test sets (Table 19). The test set performance suggests that the predictions made using historic data are off by about one patient on average. Furthermore, we find that 61.42% of predicted values (across all three severity levels) are off by less than one patient arrival, and only 5.02% of predicted values are off by more than three patient arrivals. Thus, that historic data provides a reasonable estimate of the number of hourly arrivals of each type during a given half-shift. By combining this historic information about patient severity and predictions regarding the appropriate type of provider, management can make more informed decisions regarding provider staffing.

Finally, we develop a model to classify visits based on the type of provider that each visit will require. For this analysis, we limit the sample to patients who arrive when at least one physician and at least one APP are working. This ensures that either provider type was an option when the patient was assigned to a provider. We predict the provider type based on the patient's age, gender, ESI level, and payer information (e.g. Medicare or a commercial insurance provider). We control for the facility, the average ESI level of patients currently in the ED as well as the ratio of APPs to physicians to get a sense of the ED's current state. Several machine learning approaches were tested to determine the method that most accurately classifies visits by provider type. Table 20 summarizes the performance of eight classification models. We trained each model using data from 2014 through 2016 (the training set) and implemented 10-fold cross validation. Summary statistics corresponding to the reported accuracy (proportion of correct predictions) of each model are shown. Several of these models perform similarly, suggesting that there are clear patterns in the ways in which patients are assigned to the two provider types.

Table 19 Predicting Number of Patients per Hour by ESI Level

		ESI Level		
		1 & 2	3	4 & 5
	(Constant)	-0.2047	0.7900	0.8636
Month of Year	January	0.0904	0.1061	0.1165
	February	0.0465	0.0896	0.1100
	March	0.0785	0.1109	0.0937
	April	0.0567	0.0646	0.0587
	May	0.0172	0.0449	0.0572
	June	-0.0136	0.0149	0.0245
	August	0.0067	0.0173	-0.0173
	September	0.0339	0.0291	0.0020
	October	0.0481	-0.0057	-0.0400
	November	0.0286	0.0175	-0.0463
	December	0.0911	0.0605	0.0047
Day of Week	Monday	0.1312	0.1246	0.0555
	Tuesday	0.0877	0.0842	-0.0183
	Wednesday	0.0755	0.0623	-0.0327
	Thursday	0.0478	0.0357	-0.0697
	Friday	0.0223	0.0379	-0.1101
	Saturday	-0.0582	-0.0228	-0.0659
	Holiday	-0.1124	-0.1112	-0.0113
	Day After Holiday	0.0128	0.0798	0.1134
Half Shift	Shift 11am-3pm	0.3524	0.2836	0.2975
	Shift 3pm-7pm	0.2971	0.2424	0.3359
	Shift 7pm-11pm	0.0837	0.0634	0.2186
	Shift 11pm-3am	-0.5520	-0.6222	-0.7272
	Shift 3am-7am	-0.9111	-0.9855	-1.3491
	Minimum Temperature (normal, based on historic data)	0.0009	0.0007	0.0016
	Maximum Temperature (normal, based on historic data)	0.0022	0.0015	0.0012
Facility	Facility 2	-1.7717	0.0961	0.0496
	Facility 3	-0.8301	-0.4572	-0.6840
	Facility 4	-0.2143	-0.1652	-0.7270
	Facility 5	-1.4861	-0.0554	-0.5191
	Facility 6	0.3538	0.3326	0.1783
	Facility 7	-1.9719	-1.1857	-0.9150
	Facility 8	0.2988	0.3582	0.3385
Training Set (2014 – 2016)	Mean Squared Error (MSE)	1.0468	5.0395	5.5270
	Mean Absolute Deviation (MAD)	0.7187	1.7397	1.7947
Test Set (2017)	Mean Squared Error (MSE)	1.1525	4.5651	4.4106
	Mean Absolute Deviation (MAD)	0.7561	1.6571	1.6377

Table 20 Predicting Provider Type

Method	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Linear Discriminant Analysis	83.13%	83.24%	83.31%	83.31%	83.35%	83.51%
Logistic Regression	82.97%	83.27%	83.34%	83.31%	83.41%	83.52%
K Nearest Neighbor	81.38%	81.69%	81.92%	81.91%	82.17%	82.39%
Naïve Bayes	80.49%	82.23%	83.03%	82.74%	83.48%	84.03%
Cart	82.88%	83.07%	83.42%	83.47%	83.93%	84.05%
C-5.0	83.85%	84.03%	84.15%	84.21%	84.37%	84.83%
Bagged Cart	78.44%	78.57%	78.77%	78.74%	78.84%	79.08%
Stochastic Gradient Boosting	83.73%	84.00%	84.17%	84.18%	84.39%	84.60%

For ease of interpretability, the results of the logistic regression model are provided here (Table 21 and Fig. 23). These results confirm previous observations about the likelihood of ESI Level 4 and 5 patients being treated by APPs. While the odds of being treated by an APP increase with ESI Level, the odds of an ESI Level 4 or 5 patient being treated by an APP are more than 150 times the odds of an ESI Level 1 or 2 patient being treated by an APP. As the average ESI Level within the ED increases, it indicates less sick patients on average. Consequently, the logistic regression model suggests that the odds of being treated by an APP increase as the average ESI level increases (i.e., for lower acuity patients). This may be due to an increased availability of physicians to treat all patients. Additionally, as more APPs are working, the APP ratio will increase, and thus the odds of a patient being treated by an APP increase.

Together, these models provide insights about ED visits, assignments of the patients to different provider types, and the factors that affect the length of stay for discharged patients. Specifically, we find that the number and severity of visits remains relatively stable over time, validating management's use of historic trends in staffing physicians and APPs. As expected, the assignment of providers to these visits is strongly related to patient characteristics, including ESI level. We controlled for patient characteristics, patient arrival times, and the ED conditions upon arrival to assess how the use of APPs impacts LOS for discharged patients. We noted that patients treated by APPs experienced shorter LOS and that adding providers (either physicians or APPs) were associated with shorter LOS. Additionally, the benefit of adding a physician was greater than adding an APP as evidenced by larger reductions in LOS. In either case, the patients treated by physicians saw a greater impact when an additional provider was added. Furthermore, patients classified as ESI Level 4 or 5 benefitted most from being treated by APPs and also profited most from the addition of a physician. However, this group of low acuity patients observed minimal benefit from the addition of an APP. Time of day is important in predicting LOS, and differences throughout the day affect how beneficial APPs are. Next, we incorporate these insights into a simulation model to replicate the current operations of an ED and then run a simulated experiment in which we assess how changes to both controllable and uncontrollable factors affect visit lengths.

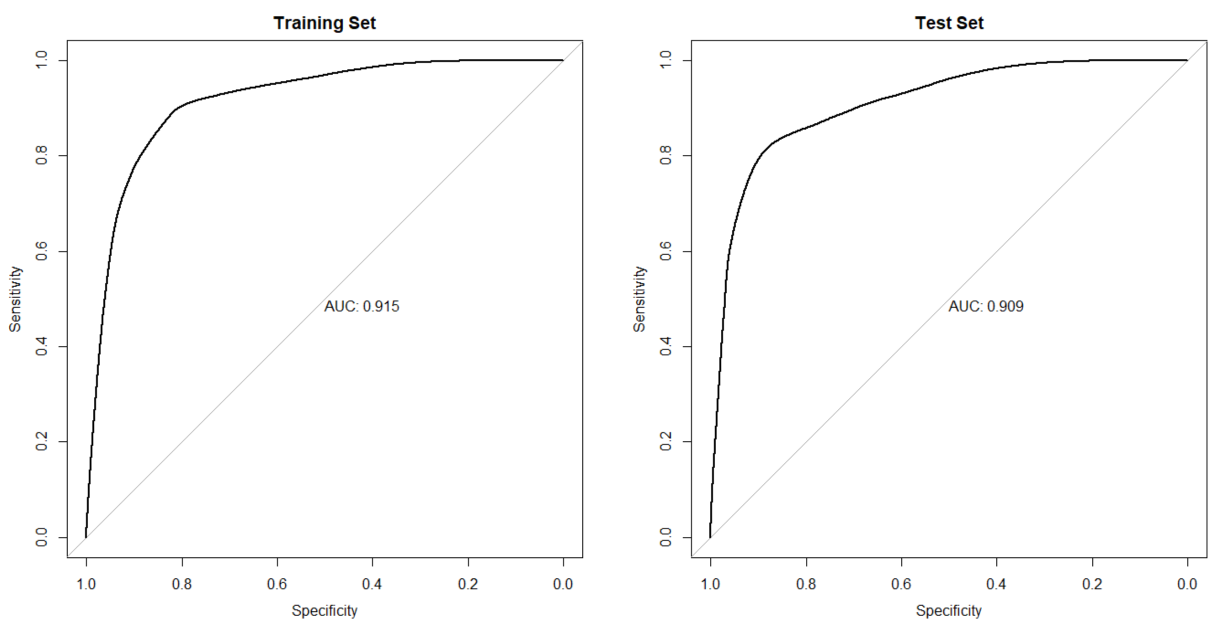


Figure 23 ROC Curve: Logistic Regression Model for Provider Type (APP = 1 or Physician = 0)

Table 21 Predicting Provider Type: Logistic Regression Results

Dependent Variable: Prob(Provider Type = APP)					
	Variable	exp(Coefficient)	Coefficient		Standard Error
	(Constant)	0.0036	-5.6133	***	0.0519
ESI Level	ESI Level 1 or 2	Reference Group			
	ESI Level 3	4.9982	1.6091	***	0.0344
	ESI Level 4 or 5	109.9395	4.6999	***	0.0337
Age	0 to 3 months	0.5624	-0.5756	***	0.0305
	4 to 36 months	0.9700	-0.0305		0.0217
	3 to 8 years	1.0290	0.0286		0.0165
	9 to 17 years	1.0976	0.0932	***	0.0148
	18 to 34 years	Reference Group			
	35 to 44 years	0.9364	-0.0658	***	0.0128
	45 to 54 years	0.8458	-0.1675	***	0.0136
	55 to 64 years	0.7427	-0.2974	***	0.0161
	65 to 74 years	0.6510	-0.4292	***	0.0247
	75 to 84 years	0.4739	-0.7469	***	0.0314
	85 years plus	0.3696	-0.9953	***	0.0397
Gender	Male Indicator	1.0053	0.0053		0.0082
Payer	Commercial Insurance	Reference Group			
	Medicaid	0.9186	-0.0849	***	0.0115
	Medicare	0.8099	-0.2109	***	0.0184
	Self-Pay	0.9170	-0.0867	***	0.0148
	Other	1.5598	0.4445	***	0.0163
	ED Census	0.9956	-0.0045	***	0.0005
	Average Severity of Other Patients	1.2634	0.2338	***	0.0090
	APP to Physician Ratio	2.2704	0.8200	***	0.0107
Half Shift	Shift 7am-11am	Reference Group			
	Shift 11am-3pm	0.8782	-0.1298	***	0.0143
	Shift 3pm-7pm	0.7775	-0.2517	***	0.0148
	Shift 7pm-11pm	0.6390	-0.4479	***	0.0150
	Shift 11pm-3am	0.3215	-1.1348	***	0.0200
	Shift 3am-7am	0.0489	-3.0179	***	0.2194
Facility*	Facility 1	Reference Group			
	Facility 2	0.0034	-5.6881	***	0.0870
	Facility 3	2.5217	0.9249	***	0.0176
	Facility 4	0.4393	-0.8225	***	0.0179
	Facility 5	0.3919	-0.9367	***	0.0196
	Facility 7	1.5979	0.4687	***	0.0353
	Facility 8	1.9998	0.6930	***	0.0115

*Note that Facility 6 is excluded because no visits were attributed to APPs working independently

5.4.3 Prescriptive Analysis

In order to prescribe rules for making staffing decisions, we design a simulated experiment. We begin by simulating the current state of one ED in our sample. After replicating the ED's current state, we utilize a Taguchi L-27 design to study the effects of changes to the ED's current staffing practices, which are controllable, as well as uncontrollable changes to the patient population. Staffing variables include the number of physicians working during each shift (day, afternoon, and night), the APP to physician ratio during each shift, and the rules for assigning providers to patients. The uncontrollable patient variables include patient complexity (expected value of ESI level), the proportion of patients requiring labs, testing, and both labs and testing, and the proportion of patients who are admitted to the hospital. Table 22 details the levels for each of the variables in our experiment, and Table 23 lists each of the 27 experimental runs. For each variable, the medium (or M) setting represents the observed data or baseline setting. These settings were largely identified through the previously presented descriptive analysis. Patients arriving in the ED are assigned to one of the five ESI levels using a discrete probability distribution that reflects the actual data. Six different distributions are used and dependent on the time of arrival. As with previous analyses, we use four-hour blocks of time, or half-shifts. The three rules used for assignment of patients to providers are as follows:

(1) ESI Level 1, 2, and 3 always assigned to physicians, ESI Levels 4 and 5 assigned to first available provider.
(2) ESI Levels 1 and 2 always assigned to physicians, all others assigned to first available provider.
(3) ESI Levels 1 and 2 always assigned to physicians, ESI Levels 4 and 5 always assigned to APPs, ESI Level 3 assigned to first available provider.

Rule 2 is viewed as the base case because many emergency departments already operate in this way. APPs typically do not have the necessary training and expertise to independently treat Level

1 or 2 patients, so those cases are always assigned to physicians, but other patients may be assigned to either provider type. Rule 1 uses the same idea but assumes that Level 3 patients should only be seen by physicians. Finally, Rule 3 represents an ED with a "Fast Track" which immediately routes the lowest acuity patients to APPs in an attempt to reduce wait times and ED crowding. In each of these cases, the highest priority patient (i.e. the patient with the lowest ESI level) is always served first, and patients within the same ESI level are treated on a first come first serve (FCFS) basis. The simulation model also accounts for the possibility that a more severe patient who arrives in an ED may take precedence over the provider's current patient. That is, a physician may temporarily leave a lower acuity patient to treat a new high acuity patient. Additionally, an APP may call on a physician for assistance and vice versa, resulting in patients treated by both provider types. Resource requirements (e.g., testing or labs) for each visit are based on the historic data, and service times are randomly assigned based on both patient and ED factors.

The results of the experimental runs are described in Table 24. Using these simulated results, we construct a model to determine directions for improvement. For each factor, we code the High, Medium, and Low settings as 1, 0, and -1, respectively. These results (Table 25) suggest that increasing the number of physicians during the day shift (7am to 3pm) is associated with significantly fewer patients in the ED on average and a lower utilization rate for physicians. Furthermore, increasing the number of physicians during the afternoon shift (3pm to 11pm) is associated with shorter visit lengths and fewer patients in the ED on average as well as a lower utilization rate for physicians. While the goal is not to have idle physicians, a small reduction in utilization could free up physicians for other responsibilities or allow them to spend more time interacting with patients. Adding physicians during the night shift did not produce statistically significant differences from the baseline. On the other hand, increasing the number of APPs working per physician (or APP ratio) during a shift resulted in significantly shorter visit lengths on average, but minimal changes were observed with respect to the number of patients in the ED and the physician utilization rate. These results further emphasize that adding physicians will have a greater influence than adding APPs. However, increasing the number of APPs is still shown to be beneficial, so further analysis will be necessary to determine the exact tradeoff between a physician and an APP for a specific ED. While these regression results suggest potential directions for improvements, further experimentation is recommended on a facility basis to determine the specific settings that result in improved ED operations while also meeting the facility's financial and personnel constraints.

Table 22 Description of Factor Levels for Simulated Experiment

Controllable Factors		**Levels**		
Number of Physicians	day			
Number of Physicians	afternoon	1	2	3
Number of Physicians	night			
APP to Physician Ratio	day			
APP to Physician Ratio	afternoon	L (ratio $\leq$ 0.5)	M (0.67 $\leq$ ratio $\leq$ 1.33)	H (1.5 $\leq$ ratio $\leq$ 2.5)
APP to Physician Ratio	night			
Assignment Rule (Patients to Providers)		Rule 1	Rule 2	Rule 3

Uncontrollable Factors	**Levels**		
Patient Complexity (Expected Value for ESI Level)	L = 3.11	M = 3.15	H = 3.19
Proportion of Patients Requiring Labs (No testing)	L = 0.39	M = 0.43	H = 0.47
Proportion of Patients Requiring Testing (No labs)	L = 0.01	M = 0.02	H = 0.03
Proportion of Patients Requiring Both Labs & Testing	L = 0.14	M = 0.16	H = 0.18
Proportion of Patients Admitted	L = 0.11	M = 0.14	H = 0.17

Table 23 Experimental Design for Simulation

Run	Physicians (D)	Physicians (A)	Physicians (N)	APP Ratio (D)	APP Ratio (A)	APP Ratio (N)	Assignment Rule	Patient Complexity (Expected Value)	Labs	Testing	Labs + Testing	Admitted
1	1	2	2	M	L	L	Rule 1	3.15	M	M	H	H
2	2	3	1	M	H	L	Rule 2	3.15	H	L	L	M
3	3	2	1	H	M	L	Rule 3	3.19	M	L	L	H
4	2	2	3	L	H	L	Rule 2	3.11	M	H	M	H
5	1	2	2	M	H	H	Rule 3	3.11	L	L	M	M
6	2	2	3	L	L	M	Rule 3	3.15	H	L	H	L
7	2	1	2	H	M	H	Rule 1	3.15	H	L	M	H
8	2	1	2	H	L	M	Rule 3	3.11	M	H	L	M
9	1	1	1	L	H	H	Rule 3	3.19	H	H	H	H
10	3	1	3	M	H	M	Rule 1	3.19	M	L	H	M
11	1	2	2	M	M	M	Rule 2	3.19	H	H	L	L
12	1	1	1	L	M	M	Rule 2	3.15	M	M	M	M
13	3	1	3	M	M	L	Rule 3	3.15	L	H	M	L
14	3	3	2	L	M	L	Rule 3	3.11	H	M	H	M
15	1	1	1	L	L	L	Rule 1	3.11	L	L	L	L
16	1	3	3	H	L	L	Rule 1	3.19	H	H	M	M
17	3	3	2	L	L	H	Rule 2	3.19	M	L	M	L
18	3	2	1	H	L	H	Rule 2	3.15	L	H	H	M
19	3	1	3	M	L	H	Rule 2	3.11	H	M	L	H
20	1	3	3	H	H	H	Rule 3	3.15	M	M	L	L
21	2	1	2	H	H	L	Rule 2	3.19	L	M	H	L
22	1	3	3	H	M	M	Rule 2	3.11	L	L	H	H
23	2	3	1	M	L	M	Rule 3	3.19	L	M	M	H
24	3	3	2	L	H	M	Rule 1	3.15	L	H	L	H
25	2	3	1	M	M	H	Rule 1	3.11	M	H	H	L
26	2	2	3	L	M	H	Rule 1	3.19	L	M	L	M
27	3	2	1	H	H	M	Rule 1	3.11	H	M	M	L

Table 24 Results of Simulated Experiment (Average of 10 Runs; Each Run = 52 Weeks)

Run	Patients in ED	Time in ED (hours)	Utilization of APPs (%)	Utilization of Physicians (%)
1	20.2704	3.6960	21.5992	71.1442
2	8.5794	1.5591	36.7139	45.8602
3	11.9494	2.1529	44.4162	70.4786
4	8.6754	1.5764	34.2779	47.6667
5	10.1847	1.8505	56.0331	56.4701
6	59.1205	10.7393	27.2186	93.5016
7	11.2345	2.0429	27.1698	60.8932
8	25.6181	4.6551	42.5391	75.4981
9	11.8892	2.1601	44.4162	70.3158
10	10.8860	1.9831	26.6400	59.5376
11	10.5257	1.9165	53.5287	54.4810
12	60.8423	14.3317	61.9248	99.9552
13	22.8029	4.1434	36.6944	67.8703
14	19.1905	3.4839	35.0343	64.5407
15	144.0000	2.5049	0.7865	99.8234
16	14.3249	2.5804	17.7152	66.6142
17	8.2320	1.4918	38.3743	43.7849
18	8.6903	1.5738	37.1236	46.1110
19	11.5751	2.1021	26.4259	55.6060
20	9.7193	1.7692	54.6787	52.5943
21	18.1689	3.3032	49.4564	82.0670
22	9.7067	1.7664	50.1257	50.8049
23	8.5764	1.5603	52.6464	50.6623
24	8.4910	1.5331	13.4880	51.7366
25	8.8865	1.6073	21.2448	51.5164
26	8.2332	1.4859	23.6186	50.0788
27	9.0661	1.6408	21.3161	51.7605

Table 25 Analysis of Simulated Experiment (Average of 10 Runs; Each Run = 52 Weeks)

	Variable	ln(Visit Length)		ln(Patients in ED)		Utilization of Physicians	
Controls	(Intercept)	0.8773	***	2.6612	***	62.6435	***
	Physicians (Day)	-0.1318		-0.2464	*	-6.1543	*
	Physicians (Afternoon)	-0.2845	*	-0.3992	***	-10.7473	**
	Physicians (Night)	0.0407		-0.0758		-2.3449	
	APP Ratio (Day)	-0.1334		-0.2492	*	-3.5879	
	APP Ratio (Afternoon)	-0.1927		-0.3234	**	-4.7076	
	APP Ratio (Night)	-0.1981		-0.3284	**	-7.1497	*
	Assignment Rule	0.1880		0.0567		2.157	
	Labs	0.1081		-0.0223		0.4416	
	Testing	-0.0193		-0.1489		-2.7414	
	Labs+Testing	0.1456		0.0144		1.8546	
	Admitted	-0.1279		-0.2581	*	-3.7828	
	Patient Complexity	-0.0445		-0.1742		-0.3148	
	R^2	58.11%		84.01%		73.79%	
	Adjusted R^2	22.21%		70.30%		51.32%	

***p < 0.001, **p < 0.01, *p < 0.05; two-tailed tests

5.5 Conclusions & Managerial Insights

We analyze historic data to gain a better understanding of how APPs are utilized in EDs and the potential benefits of APPs from the ED patient perspective, mainly focusing on the visit length for discharged patients since it has been linked to patient satisfaction previously. We apply data from eight hospitals over a four-year period (2014-2017) to gain a better understanding of how APPs are currently utilized in EDs. Through descriptive analyses and data visualization, we observe that current practices vary. While patients arriving in one ED were treated exclusively by physicians (as of 2017), low acuity patients arriving in other EDs were likely to be treated by an APP, and the number of patients being treated by APPs increased each year during the study period. We then developed predictive models to better understand how providers are assigned to patients across these EDs and how that assignment as well as staffing levels relate to visit length.

In these predictive models, we consider the effects of being treated by an APP and adding more APPs to an ED, all else constant. We also compare the addition of an APP to the addition of a physician. We note that adding a physician has a larger impact on LOS for discharged patients, presumably because emergency physicians have more comprehensive education and training and more specialized knowledge. However, we also note the benefits gained from adding an APP after accounting for a number of patient and ED-related factors. Namely, with each additional APP, we observe a shorter average LOS amongst all discharged patients, regardless of whether or not they were treated by an APP. In fact, the difference in LOS was most pronounced among the patients that were treated by physicians, suggesting that APPs are freeing up physicians to more quickly care for sicker patients. The results of our predictive analyses make a case for the ED “Fast Track” in which lower acuity patients are treated as a separate population and almost always see an APP. Using the results of the descriptive and predictive analyses, we develop a simulation model to experiment with potential process changes within the ED. Additional experimentation on a facility basis could help to determine the specific settings that result in improved ED operations while taking into account the facility’s financial and personnel constraints.

While the use of observational data limits our ability to infer causality, we worked to address such limitations. Since APPs are typically used during high volume times, we include the current ED census and patient arrival time as control variables. We also run our analyses excluding the highest acuity patients as they are unlikely to be seen by an APP and may skew the results. Still, this is a limitation of the analyses. Additionally, this work has focused on the patient’s perspective, and more specifically, on patient length of stay. The provider perspective is also important, and we acknowledge that we haven't compared provider productivity or actual patient outcomes for physicians and APPs. Ideally, further research may also incorporate patient outcomes such as patient experience scores or returns to the ED within 72 hours.

Our findings have practical implications for EDs. While many EDs already utilize APPs, the staffing decisions are not made systematically. We have confirmed that APPs provide benefits within the ED and simulated an experiment for a single ED to determine directions for improvement. Management may opt to simulate a series of experiments which are feasible given the available staff, budget, and other constraints, applying response surface methodology to continuously find directions for improvement. Once a desirable setting is confirmed through simulation, the ED may experiment with adding APPs to shifts. Thus, this proposed research framework, comprising the descriptive analyses, data visualization, predictive models, and simulated experiments, has provided insights about the usage of APPs in EDs. These insights are valuable for increasing patient throughput in EDs and may ultimately be incorporated to improve the patient experience and patient outcomes.

6.0 Summary & Future Work

Healthcare operations is a research area that continues to evolve, and the increasing availability of data provides opportunities for more informed decision making. In each of these three essays, we utilized data corresponding to millions of visits at various EDs staffed by a large national emergency physician group. Using these data, we addressed issues related to managing emergency physicians and APPs. Future research may expand upon each of these essays.

For example, we developed metrics to evaluate performance across multiple sites and work to understand the relationships between patient flow, patient complexity, and patient experience in the first essay. In the future, we may further assess physician productivity by analyzing changes within a provider's shift. For instance, do physicians' practice patterns differ between the beginning, middle, and end of a shift? Do those differences depend on the length of the shift (e.g., eight hours versus 10 hours versus 12 hours)? These are important questions that may provide insights regarding the optimal shift length in different emergency department settings.

In the second essay, we investigated what happens when a physician is accused of medical malpractice, an event that can be both financially and emotionally taxing. As part of this analysis, we examined whether or not a physician's practice patterns changed specifically when patients presented symptoms related to the same body system as the patient in the recent malpractice claim against that physician. Future research may center on disease- or symptom-specific research questions, comparing how physicians and facilities differ in their treatments for certain conditions. Most analyses included within this dissertation utilize all of a providers' visits within a specified time frame. While incorporating all visits is helpful when making generalizations, there is potential for future research related to specific medical conditions, procedures, or treatments. By isolating a more focused sample, we may be able to uncover treatment patterns or trends related to a given disease or medical condition. Such research would be conducive to better assessing quality of care.

The final essay builds on a secondary result from the first essay which indicated that the support of APPs such as physician assistants has a direct positive impact on patient flow. We took a closer look at the relationship between APPs and physicians and ED staffing practices and plan to continue to build on this research moving forward. For instance, in this work, we mainly focus on when and where APPs could be added, but more work is required to better understand the substitutability of APPs and physician and when a physician could be replaced by one or more APPs. In addition to using the historic data to make such staffing decisions, there is potential for improvements to short-term demand predictions, which could provide ED management with better information to improve resource allocation and make short-term scheduling adjustments. For example, social media data (e.g., Twitter) may provide insights about the surrounding community with posts about accidents, mass-casualty events, and even disease outbreaks. Monitoring such data could allow EDs to better inform decisions in real-time. For instance, an ED may be able to call in additional providers before an influx of patients. Such improvements in short-term demand forecasts are valuable for ED management.

Research related to managing EDs in terms of either operational metrics (e.g. length of stay, RVUs, or admission rates) or quality of care (e.g., patient outcomes or adherence to standards) continue to be valuable for healthcare administrators and physicians. Additionally, such research has the potential for downstream patient benefits. EDs represent an important facet of the healthcare industry, and as the group continues to grow and collect additional data, new research questions constantly arise.

Appendix A Development of Key Performance Indicators

This appendix supports the analysis and models presented in Section 3.

Appendix A.1 Relative Value Units and Press Ganey Scores

Management in the EPMN under study needs not only to maintain their existing work force and retain current physicians, but also to appeal to new physicians by offering attractive compensation. Currently, some physicians feel they are disadvantaged due to the types of hospitals (with varying demand and capability) where they are assigned to work and the sorts of patient these facilities draw. To mitigate such concerns, management constantly looks for new performance metrics to ensure fairness and efficiency in operations. The choice of indices is directly motivated by management's need to focus on these operational outcomes. To focus on the operational outcomes, we use the conventional metrics (i.e. productivity and patient experience assessment) as references.

Typically, an emergency physician's productivity is measured by RVUs per hour, where RVU stands for relative value unit, a standard resource-based measure. The RVUs for a visit are assigned by trained medical coders. More specifically, RVUs are widely used as a measure for physician services billing and determining physician compensation. While the charges and revenues associated with a particular patient visit are highly variable based on specific laboratory, diagnostic imaging, and treatment modalities used, RVUs provide a standardized measure and indicate the amount of effort expended. Readers are referred to Venkat et al. (2015) for more details.

DOCTORS	very poor 1	poor 2	fair 3	good 4	very good 5
1. Courtesy of the doctor	○	○	○	○	○
2. Degree to which the doctor took the time to listen to you	○	○	○	○	○
3. Doctor's concern to keep you informed about your treatment	○	○	○	○	○
4. Doctor's concern for your comfort while treating you	○	○	○	○	○

Appendix Figure 1 Press Ganey (PG) survey questions

The "Physician PG score" for each clinician in a facility is calculated using the survey results from the previous 30 days of service. The mean score is then ranked and converted to a percentile.

The organization has formal courses for physicians to improve their performance with respect to patients per hour and PG physician percentile rank. Both courses are 2-3 days in length and provide practical techniques that physicians can apply in their interactions with individual patients and managing the flow of patients in the ED. It is important to note that efficiency and patient experience training for emergency physicians and APPs is not simply a matter of remediation. Based on site circumstances, the EPMN may offer or mandate such training for all emergency physicians and APPs regardless of performance along these metrics. These courses offer continuing education to healthcare providers and means of improving operational performance. Therefore, currently physicians of varying performance levels are receiving the training.

Appendix A.2 Additional Tables and Variable Definitions

Similar to Eqs. (7) – (10), we propose Eq.(A1) to index the proportion of insurance types handled by each physician. It reflects whether the physician is treating more or less patients of a certain insurance type than peers. Additionally, physician *i*'s APP Support Ratio is the number of combined APP hours during all shifts divided by the total number of physician and APP hours during these shifts.

$$Insurance\ m\ Index_i = \frac{\sum_{j=1}^{F_i}\left[\frac{\%Insurance\ Type\ m_{\ i,j}}{Average\ \%Insurance\ Type\ m_j} \times V_{i,j}\right]}{\sum_{j=1}^{F_i} V_{i,j}} \quad (A.1)$$

Appendix Table 1 Definitions of Patient Level Variables

Variable	Definition
Average Patient Age$_i$	=Average age of patients treated by a physician across all visits in all facilities
Peds Indicator$_i$	$=\begin{cases} 1, \text{ if } \textit{Average Patient Age}_k < 18 \text{ years} \\ 0, \text{ otherwise} \end{cases}$
Coding Com/Patient$_i$	=Average number of communications with physician per patient visit due to unclear or potentially missing physician documentation
Admit Ratio$_i$	=Proportion of a physician's patients that were admitted
Commercial Insurance Ratio$_i$	=Proportion of a physician's patients with commercial insurance
Medicare Ratio$_i$	=Proportion of a physician's patients with Medicare
Medicaid Ratio$_i$	=Proportion of a physician's patients with Medicaid
Self-Pay Ratio$_i$	=Proportion of a physician's patients without insurance
Male Ratio$_i$	=Proportion of male patients
ICD-9 Group 1 Ratio$_i$	=Proportion of Circulatory, Respiratory, Digestive, and Genitourinary visits (390-629)
ICD-9 Group 2 Ratio$_i$	=Proportion of visits due to Symptoms, Signs & Ill-defined Conditions (780-799)
ICD-9 Group 3 Ratio$_i$	=Proportion of Injury & Poisoning visits (800-999)

Appendix Table 2 Definitions of Physician Level Variables

Variable	Definition
# Facilities Worked$_i$	=Number of facilities in which a physician worked
Physician Age$_i$	=Age of Physician
Physician Gender$_i$	$=\begin{cases} 1, \text{ if Male} \\ 0, \text{ otherwise} \end{cases}$
Physician Race$_i$	$=\begin{cases} 1, \text{ if White} \\ 0, \text{ otherwise} \end{cases}$
Efficiency Flag$_i$	$=\begin{cases} 1, \text{ if EPMN-administered Efficiency Training Completed} \\ 0, \text{ otherwise} \end{cases}$
Satisfaction Flag$_i$	$=\begin{cases} 1, \text{ if EPMN-administered Patient Satisfaction Training Completed} \\ 0, \text{ otherwise} \end{cases}$
6am-3pm Ratio$_i$	=Proportion of a physician's hours worked between 6am and 3pm
3pm-12am Ratio$_i$	=Proportion of a physician's hours worked between 3pm and 12am
12am-6am Ratio$_i$	=Proportion of a physician's hours worked between 12am and 6am
APP Support Ratio$_i$	=Proportion of provider hours worked during a physician's shifts which are attributed to APPs

Appendix A.3 Aggregate Measure: An Alternative Index

In addition to the proposed relative indices presented in the paper, we also considered an alternate approach: aggregate measures. The difference is subtle and important. We first describe the aggregate measures of physician performance and then discuss why we ultimately choose to employ the relative index measures.

Each physician may be compared to the aggregated average performance of all facilities at which he or she works. Therefore, we consider an aggregate measure that accounts for facility-level differences by comparing a physician's actual RVUs (or Patients) to his or her expected RVUs (or Patients) if he or she performed at the facility average within each facility. Recall that $Average\ RVUs/hr_j = \frac{\sum_{i=1}^{N}\sum_{k=1}^{V_{i,j}} RVUs_{i,j,k}}{\sum_{i=1}^{N} Hours_{i,j}}$, and other facility averages can be computed similarly. Subsequently, we compute the physician's aggregate measures relative to facilities' averages in the EPMN in Eqs. (A.2) – (A.4).

$$Aggregate\ RVUs/hr_i = \frac{\sum_{j=1}^{F_i}\sum_{k=1}^{V_{i,j}} RVUs_{i,j,k}}{\sum_{j=1}^{F_i}(Average\ RVUs/hr_j \times Hours_{i,j})} \tag{A.2}$$

$$Aggregrate\ Patients/hr_i = \frac{\sum_{j=1}^{F_i} V_{i,j}}{\sum_{j=1}^{F_i}(Average\ Patients/hr_j \times Hours_{i,j})} \tag{A.3}$$

$$Aggregate\ RVUs/Patient_i = \frac{\sum_{j=1}^{F_i}\sum_{k=1}^{V_{i,j}} RVUs_{i,j,k}}{\sum_{j=1}^{F_i}(Average\ RVUs/Patient_j \times V_{i,j})} \tag{A.4}$$

Appendix Table 3 presents a stylized example in which Physician A worked in two facilities (Facilities 1 and 2), while Physician B only worked in Facility 2. Facility 1 has higher average Patients/hr (4 vs.1) and RVUs/hr (14 vs. 4). Since Physician B only works at one facility, calculating RVUs/hr is straightforward, but facility differences complicate the calculation for Physician A.

First, we examine the absolute measures, which is the ratio of Total RVUs to Total Hours. In that case, Physician A [6.0=(9000+3000)/(1200+800)] generates more RVUs/hr than Physician B [4.1=9000/2200] as shown in Appendix Table 3(a). Then we compare the physicians based on the ratio of actual RVUs earned to expected RVUs earned at all facilities using Eq. (A.2), which yields the numbers in Appendix Table 3(b). Note that the aggregate measure captures the net effect of facility size and hours worked. Finally, we compare physicians on the relative indices in Appendix Table 3(c). By reflecting the relative performance at facilities in which the physician works and weighing his or her performance on the time he or she spent or patients he or she saw in each facility, the proposed indices in Appendix Table 3(c) neutralize the scale (facility size) bias.

Note that the measures in Appendix Tables 3(b) and 3(c) are comparable when we do not observe heterogeneity across the facilities in which a physician works, but they give different conclusions about Physicians A and B's performances than Appendix Table 3(a).

Appendix Table 3 Comparison of Performance Metrics

Phys.	Fac.	RVUs	Patients	Hours	RVUs/hr Fac.	RVUs/hr Phys.	Patients/hr Fac.	Patients/hr Phys.	RVUs/Patient Fac.	RVUs/Patient Phys.	RVUs/ hr Ratio	Patients/ hr Ratio	RVUs/ Patient Ratio
A	1	9000	3600	1200	14.0	7.5	4.0	3.0	3.5	2.5	0.54	0.75	0.71
	2	3000	900	800	4.0	3.8	1.0	1.1	4.0	3.3	0.94	1.13	0.83
B	2	9,000	2,500	2,200	4.0	4.1	1.0	1.1	4.0	3.6	1.02	1.14	0.90

	(a) Total Physician Absolute Measures		
Physician	RVUs/hr	Patients/hr	RVUs/Patient
A	6.0	2.3	2.7
B	4.1	1.1	3.6

	(b) Aggregate Measures based on Facilities		
Physician	RVUs/hr	Patients/hr	RVUs/Patient
A	60.0%	80.4%	74.1%
B	102.3%	113.6%	90.0%

Computed by Eqs. (S.1) – (S.3)

	(c) Proposed Relative Index Measures		
Physician	RVUs/hr	Patients/hr	RVUs/Patient
A	**69.6%**	**90.0%**	**73.8%**
B	**102.3%**	**113.6%**	**90.0%**

Computed by Eqs. (5) – (7)

Eqs. (A.2)–(A.3) are better than the absolute measures because they normalize performance relative to expected performance at facilities, allowing for comparisons within the network. However, we contend that Eqs. (A.2)–(A.3) are inferior to the relative index measures in Eqs. (7)–(9). Our reasoning follows from the analysis of the two-facility case.

Recall that the proposed relative index first compares each physician's performance in a given facility to that facility's average and then weights the relative performance by the ratio of hours worked in that facility. In the two-facility case, that becomes: $\frac{RVUs/hr_1}{Average\ RVUs/hr_1}X + \frac{RVUs/hr_2}{Average\ RVUs/hr_2}(1-X)$, where $X =$ the proportion of hours worked in Facility 1, $RVUs/hr_j$ = the physician's RVUs/hr in Facility j, and $Average\ RVUs/hr_j$ = the average RVUs/hr for all physicians working in Facility j. This expression does not reward or penalize physicians for specific facility assignments, and thus, we contend that it is appropriate when physicians are assigned to multiple facilities. Alternatively, the aggregate measure simply compares each physician's aggregate performance to his or her expected performance based on all facilities in which he or she works. In the two-facility case, that is $\frac{RVUs/hr_1 \times X + RVUs/hr_2 \times (1-X)}{Average\ RVUs/hr_1 \times X + Average\ RVUs/hr_2 \times (1-X)}$. The aggregate measure has the potential to inappropriately penalize physicians who work a high proportion of hours at higher performing facilities or conversely reward physicians who work many hours at low-performing facilities. Thus, we reason that the proposed relative index is more suitable when facility assignment is beyond the physicians' control as is the case in our data.

In order to fully compare the effects of using each of the two measures discussed, we compare the mathematical implications of each for the two-facility case. We differentiate both the aggregate and relative index measures with respect to the proportion of hours worked in Facility 1. Recall that X = the proportion of hours worked in Facility 1, $RVUs/hr_j$ = the physician's RVUs/hr in Facility j, and $Average\ RVUs/hr_j$ = the average RVUs/hr for all physicians working in Facility j.

We first consider the proposed relative index: $\frac{RVUs/hr_1}{Average\ RVUs/hr_1} X + \frac{RVUs/hr_2}{Average\ RVUs/hr_2}(1-X)$. Differentiating with respect to X, we can determine the reward (or penalty) for working a larger proportion of hours at Facility 1 under the assumption that the "effort" at both facilities remains constant:

$$\frac{\partial}{\partial x}\left(\frac{RVUs/hr_1}{Average\ RVUs/hr_1} X + \frac{RVUs/hr_2}{Average\ RVUs/hr_2}(1-X)\right)$$
$$= \frac{RVUs/hr_1}{Average\ RVUs/hr_1} - \frac{RVUs/hr_2}{Average\ RVUs/hr_2}.$$

The derivative is positive if the physician's relative performance is higher for Facility 1 and negative if the physician's relative performance is higher for Facility 2. This indicates that there is always incentive to work more hours at the facility where relative performance is highest. It should also follow that there is more incentive to improve performance at the facility in which the most hours are worked.

We then consider the aggregate measure: $\frac{RVUs/hr_1 \times X + RVUs/hr_2 \times (1-X)}{Average\ RVUs/hr_1 \times X + Average\ RVUs/hr_2 \times (1-X)}$. Again, differentiating with respect to X, we can determine the reward (or penalty) for working a larger proportion of hours at Facility 1 under the assumption that the "effort" at both facilities remains constant:

$$\frac{\partial}{\partial x}\left(\frac{RVUs/hr_1 \times X + RVUs/hr_2 \times (1-X)}{Average\ RVUs/hr_1 \times X + Average\ RVUs/hr_2 \times (1-X)}\right)$$

$$= \frac{\left(\frac{RVUs/hr_1}{Average\ RVUs/hr_1} - \frac{RVUs/hr_2}{Average\ RVUs/hr_2}\right) \times Average\ RVUs/hr_1 \times Average\ RVUs/hr_2}{[Average\ RVUs/hr_1 \times X + Average\ RVUs/hr_2 \times (1-X)]^2}.$$

The numerator will be positive if the relative performance is higher at Facility 1 than at Facility 2, and it is a function of the difference in "effort" between the two facilities. The denominator is always positive, but it is a function of both X and the facility averages, so the incentives are not as clear. Under the assumption that $Average\ RVUs/hr_1 > Average\ RVUs/hr_2$, the denominator is increasing in X, but if $Average\ RVUs/hr_1 < Average\ RVUs/hr_2$, the denominator is decreasing in X. This means that, holding all else constant, the derivative is decreasing in X when $Average\ RVUs/hr_1 > Average\ RVUs/hr_2$ but increasing in X when $Average\ RVUs/hr_1 < Average\ RVUs/hr_2$.

Consequently, there is always an incentive to work more hours at the facility where the relative performance is better (from the numerator), but the incremental reward from increasing the proportion of hours at that facility is dependent on both the current proportion of hours worked and the difference between the two facilities' averages (from the denominator). We reason that our proposed relative index is preferable to the aggregate measure because these incentives should be universal and should not depend on the specific facilities in which a physician works. The very fact that the incentive based on the aggregate measure depends on the facility averages poses a potential problem. A summary companion of the two indices is given in Appendix Table 4.

Appendix Table 4 Summary and Contrast of Two Metrics

Proposed Index Eqs. (7) – (9)	Alternate Index Eqs. (A.2) – (A.4)
Index calculated by first computing physician performance relative to a facility and then averaging across facilities, weighted by hours worked.	Index calculated by first weighting facility averages and physician averages by hours worked and then computing a ratio for relative performance, i.e., physician average/facility average
Fairer and impartial index (prevents gaming)	Incremental gains of physician depend on facility average.
If network administration assigns facilities, this measure controls for exogenous facility demand and is appropriate.	If physician self-selects facilities, then alternate index is appropriate. It captures yield from aggregate demand across facilities.

Appendix A.4 Mathematical Relationship among Proposed Indices

Recall the mathematical definitions of each of the proposed indices from Eqs. (7) – (9), and consider the product of RVUs/Patient Index (Eq. 8) and Patients/hr Index (Eq. 9).

$RVUs/Patient\ Index_i \times Patients/hr\ Index_i$

$$= \frac{\sum_{j=1}^{F_i}\left[\frac{RVUs/Patient_{i,j}}{Average\ RVUs/Patient_j} \times V_{i,j}\right]}{\sum_{j=1}^{F_i} V_{i,j}} \times \frac{\sum_{j=1}^{F_i}\left[\frac{Patients/hr_{i,j}}{Average\ Patients/hr_j} \times Hours_{i,j}\right]}{\sum_{j=1}^{F_i} Hours_{i,j}}$$

$$= \frac{\sum_{j=1}^{F_i}\left[\frac{RVUs_{i,j}/V_{i,j}}{\sum_{i=1}^{N}\sum_{k=1}^{V_{i,j}} RVUs_{i,j,k}/\sum_{i=1}^{N} V_{i,j}} \times V_{i,j}\right]}{\sum_{j=1}^{F_i} V_{i,j}}$$

$$\times \frac{\sum_{j=1}^{F_i}\left[\frac{V_{i,j}/Hours_{i,j}}{\sum_{i=1}^{N} V_{i,j}/\sum_{i=1}^{N} Hours_{i,j}} \times Hours_{i,j}\right]}{\sum_{j=1}^{F_i} Hours_{i,j}}$$

$$= \frac{\sum_{j=1}^{F_i}\left[\frac{RVUs_{i,j}}{\sum_{i=1}^{N}\sum_{k=1}^{V_{i,j}} RVUs_{i,j,k}/\sum_{i=1}^{N} V_{i,j}}\right]}{\sum_{j=1}^{F_i} V_{i,j}} \times \frac{\sum_{j=1}^{F_i}\left[\frac{V_{i,j}}{\sum_{i=1}^{N} V_{i,j}/\sum_{i=1}^{N} Hours_{i,j}}\right]}{\sum_{j=1}^{F_i} Hours_{i,j}}$$

If Average Patients/hr is equivalent across all facilities, the above can be reduced to:

$$= \frac{\sum_{j=1}^{F_i}\left[\frac{RVUs_{i,j}}{\sum_{i=1}^{N}\sum_{k=1}^{V_{i,j}} RVUs_{i,j,k}/\sum_{i=1}^{N} V_{i,j}}\right]}{\sum_{j=1}^{F_i} V_{i,j}} \times \frac{\left[\frac{\sum_{j=1}^{F_i} V_{i,j}}{Average\ Patients/hr}\right]}{\sum_{j=1}^{F_i} Hours_{i,j}}$$

$$= \frac{\sum_{j=1}^{F_i}\left[\frac{RVUs_{i,j}}{\sum_{i=1}^{N}\sum_{k=1}^{V_{i,j}} RVUs_{i,j,k}/\sum_{i=1}^{N} V_{i,j}}\right]}{\sum_{j=1}^{F_i} Hours_{i,j}} \times \left[\frac{1}{Average\ Patients/hr}\right]$$

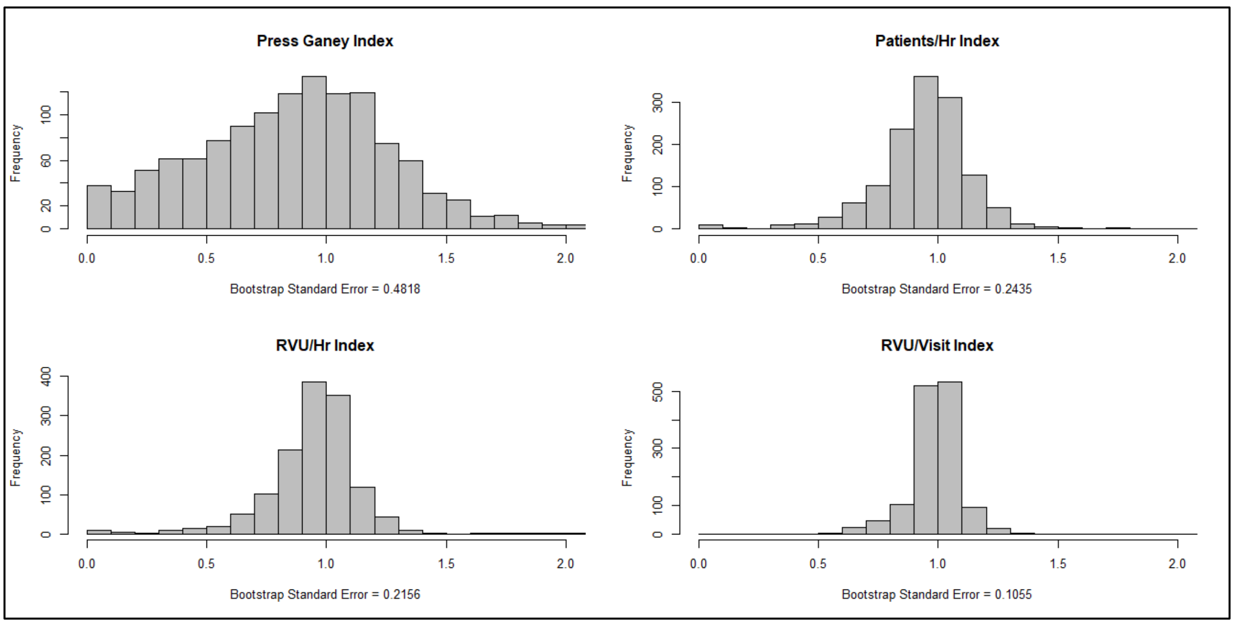
$$= \frac{\sum_{j=1}^{F_i}\left[\frac{RVUs_{i,j}}{\sum_{i=1}^{N}\sum_{k=1}^{V_{i,j}} RVUs_{i,j,k}/\sum_{i=1}^{N} Hours_{i,j}}\right]}{\sum_{j=1}^{F_i} Hours_{i,j}}$$

$$= RVUs/hr\ Index_i$$

Therefore, this proves that the product of RVUs/Patient Index and Patients/hr Index is equal to RVUs/hr Index when all of a physician’s facilities have the same Average Patients/hr.

Appendix A.5 Bootstrapped Standard Errors

We used bootstrapping to resample from the data. Using the bootstrap samples, we recalculate the four indices and their standard errors. We have plotted the distributions of the indices and indicate the standard error for each index in Appendix Fig. 2.



Appendix Figure 2 Sampling Distributions of Four Indices

While the bootstrap standard errors suggest overlap among the performance of physicians, such performance measures have been studied over time. Carlson et al (2017) suggest that month-to-month performance stabilizes after four months of practice. The variability among physicians still exceeds the variability over time for a single physician.

Furthermore, we computed the 2SLS second-stage estimates using bootstrapping. The results are consistent with the original second-stage results. Differences in significance only relate to four out of 27 control variables. Note that these differences do not affect the hypotheses, and the conclusions associated with our hypotheses remain the same.

Appendix Table 5 Second Stage Regression Estimates (bootstrapped standard errors)

		Eq. (1): ln(RVUs/Patient Index)		Eq. (2): ln(Patients/hr Index)	
Second Stage Regression (Standard Errors based on 800 Bootstrap Samples)					
Level	Variable	Coefficient	[Standard Error]	Coefficient	[Standard Error]
Physician	(Constant)	-0.5648	[0.2103]	0.0351	[0.2768]
	ln(RVUs/Patient Index)			-0.8224 ***	[0.1971]
	ln(Patients/hr Index)	0.1841 **	[0.0705]		
	Peds Indicator (0 = General, 1 = Peds)	0.5384 ***	[0.0996]		
	Physician Age			0.0002	[0.0016]
	Physician Male (0 = Female, 1 = Male)			0.0367 ***	[0.0091]
	Physician White (0 = Non-white, 1 = White)			0.0337 **	[0.0126]
	% 12AM - 6AM Hours	0.0839 ***	[0.0298]	-0.0191	[0.0725]
	% 6AM - 3PM Hours	-0.0811 ***	[0.0288]	0.2205 ***	[0.0609]
	Coding Com/Patient			-0.0457 ***	[0.0099]
	APP Support Ratio	0.0170	[0.0222]	0.5794	[0.3225]
	APP Support Ratio*Physician Age			-0.0098 *	[0.0073]
Patient	Average Patient Age	0.0345 ***	[0.0064]		
	(Average Patient Age)^2	-0.0004 ***	[0.0001]		
	% Male Patients	-0.0768	[0.1378]		
	Commercial Index	0.0476	[0.0524]	-0.2756 *	[0.1394]
	Medicaid Index	-0.2124 ***	[0.0581]	-0.0485	[0.0747]
	Medicare Index	0.0564	[0.0762]	-0.0650	[0.0368]
	Self-Pay Index	-0.2929 ***	[0.0591]	0.1876	[0.1507]
	% ICD9 Group 1	0.3697 ***	[0.0555]		
	% ICD9 Group 2	0.2598 ***	[0.0446]		
	% ICD9 Group 3	0.0865	[0.0870]		
	% Admitted Patients	0.1629 ***	[0.0342]		

Independent variables defined in Tables A1 & A2

Appendix B Malpractice Claim Analysis: Additional Details and Robustness Checks

This appendix supports the analysis and models presented in Section 4.

Appendix Table 6 Body System or Clinical Issue Classification of Primary Diagnosis ICD-9 and ICD-10 Codes as Relates to Malpractice Claim Subject

Body System	Primary ICD-9 Codes*	Primary ICD-10 Codes (if applicable)**
Neurologic	320-379 (all inclusive), 720-724 (all inclusive), 780 (all inclusive), 781 (all inclusive), 784.0, 784.3, 784.5, 784.6, 799.2, 800-801 (all inclusive), 803-806 (all inclusive), 850-854 (all inclusive), 907 (all inclusive), 950-957 (all inclusive), 986	G40.909, G44.209, G44.219, G44.319, G45.9, G51.9, G56.01, G70.00, G89.4, G91.9, G93.9, I62.01, I63.9, M48.02, M51.26, M51.36, M54.16, M54.2, M54.41, M54.42, M54.5, M54.6, M54.9, M62.830, R26.2, R41.0, R42, R44.0, R44.3, R51, R53.1, R53.83, R56.00, R56.9, S06.340A, S09.90XA, S12.190A, S22.079A, S32.019A
Gastrointestinal	520-579 (all inclusive), 783 (all inclusive), 787 (all inclusive), 789 (all inclusive), 793.4, 793.6	B96.81, C24.9, K02.9, K04.7, K08.8, K12.0, K13.0, K21.0, K28.1, K29.00, K29.90, K31.84, K35.3, K35.80, K40.21, K40.90, K52.9, K56.41, K56.60, K56.69, K57.32, K57.92, K57.93, K59.00, K59.8, K61.1, K62.5, K62.89, K63.1, K64.8, K74.60, K80.20, K80.50, K81.0, K85.8, K85.9, K92.0, K92.1, K92.2, K94.23, R10.0, R10.10, R10.11, R10.12, R10.13, R10.30, R10.31, R10.32, R10.33, R10.84, R10.9, R11.0, R11.10, R11.2, R13.10, R14.0, R17, R18.8, R19.5, R19.7, R63.4, Z20.09, Z43.1, Z43.3
Cardiovascular	390-459 (all inclusive), 780.2, 780.4, 785.0-785.3, 785.9, 786.5 (all inclusive), 786.7, 793.2, 796.2, 796.3	NA
Obstetrics & Gynecology	090-099 (all inclusive), 614-616 (all inclusive), 617-629 (all inclusive), 630-679 (all inclusive), 760-779 (all inclusive), 789 (all inclusive)	NA
Orthopedics	710-739 (all inclusive), 793.7, 805-848 (all inclusive), 880-897 (all inclusive), 905, 912-917 (all inclusive), 923-924 (all inclusive), 927-928 (all inclusive)	R25.2, S42.301A, S70.11XA, S16.1XXA, M54.5, S63.502A, S14.136A, M79.605, S93.602A, S42.002A, S62.631A, S63.501A, S32.9XXA, S32.059A, S61.431A, M25.462, S83.92XA, S93.402A
Respiratory	460-519 (all inclusive), 786 (all inclusive except 786.5), 793.1, 793.2, 794.2	NA
Integument/Wounds	680-709 (all inclusive), 782 (all inclusive), 870-897 (all inclusive), 900-904 (all inclusive), 906 (all inclusive), 910-929 (all inclusive), 940-949 (all inclusive)	NA
Genitourinary	580-629 (all inclusive), 788 (all inclusive), 791 (all inclusive), 793.5, 794.4	NA
Otolaryngology	380-389 (all inclusive), 784.1, 784.2, 784.4, 784.7-784.9 (all inclusive)	NA
Endocrine	240-259 (all inclusive), 783.0-783.6 (all inclusive), 794.5-794.7, 794.9	NA
Psychiatry	290-316 (all inclusive), 780.1, 780.5	NA
Blood/Lymphatic	280-289 (all inclusive), 780.7	NA
Eye	360-379 (all inclusive), 781.8, 802.6, 802.7, 802.8, 870-871 (all inclusive)	NA

*ICD-9 and ICD-10 codes classified based on major diagnostic code classification (ICD-9) and relationship of General Signs and Symptoms category to potential body system of relevance. General Signs and Symptoms diagnosis codes could be assigned to more than one body system categories based on potential relationship.

** Only 3 named physicians had data that extended beyond 10-1-2015 when ICD-10 was implemented. Relevant visit ICD-10 codes related to malpractice claim body system were included based on presence within the data set for named physicians during this time period.

Appendix Table 7 Difference-in-differences Adjusted for State Pre-Notification Laws

Panel A. Named versus Control Physicians Event Dates Adjusted for State Pre-Notification Laws (HI excluded and adjustments for CA and WV): All Included Visits

	Named Physicians (N=63)		***Control Physicians (N=135)***		*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*		
RVUs/Visit	3.69 (0.81)	3.74 (0.85)	3.64 (0.78)	3.73 (0.79)	0.02 (-0.11, 0.15)	-0.033 (-0.104,0.038) SE = 0.036
RVUs/Hour	9.99 (3.59)	10.13 (5.20)	9.56 (2.65)	9.52 (2.79)	0.54 (-0.19, 1.27)	0.061 (-0.422,0.543) SE = 0.241
Visit Length (hrs)	2.68 (1.20)	2.75 (1.18)	2.62 (1.10)	2.70 (1.12)	0.06 (-0.12, 0.23)	-0.036 (-0.136,0.064) SE = 0.050
Proportion of Visits Resulting in Hospital Admission	0.191 (0.110)	0.184 (0.117)	0.188 (0.094)	0.192 (0.097)	-0.004 (-0.018, 0.009)	-0.010 (-0.0200,0.0001) SE = 0.005
Physician Press Ganey Percentile Rank	51.5 (36.9)	57.0 (36.3)	55.2 (36.9)	54.5 (37.6)	0.23 (-3.69, 4.15)	6.747 (0.512,12.983) SE = 3.103
Physician months	706	1,381	685	1,306		
Visits	196,897	358,862	405,516	680,353		

Panel B. Named versus Control Physicians Event Dates Adjusted for State Pre-Notification Laws (HI excluded and adjustments for CA and WV): Failure to Diagnose Claims

	Named Physicians (N=41)		***Control Physicians (N=93)***			
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
RVUs/Visit	3.80 (0.82)	3.82 (0.89)	3.69 (0.78)	3.74 (0.81)	0.09 (-0.10, 0.27)	-0.017 (-0.089,0.056) SE = 0.036
RVUs/Hour	10.10 (4.09)	10.01 (5.35)	9.69 (2.65)	9.83 (2.66)	0.26 (-0.80, 1.32)	-0.399 (-0.912,-0.115) SE = 0.255
Visit Length (hrs)	2.80 (1.35)	2.88 (1.32)	2.69 (1.12)	2.79 (1.143)	0.10 (-0.12, 0.31)	-0.041 (-0.149,0.068) SE = 0.054
Proportion of Visits Resulting in Hospital Admission	0.200 (0.115)	0.198 (0.123)	0.194 (0.092)	0.198 (0.092)	0.002 (-0.017, 0.021)	-0.006 (-0.017,0.004) SE = 0.005
Physician Press Ganey Percentile Rank	45.6 (37.1)	51.4 (36.6)	51.9 (37.1)	52.0 (38.0)	-2.80 (8.21, 2.60)	6.566 (-1.787,14.920) SE = 4.133
Physician months	458	850	445	784		
Visits	124,228	203,619	268,562	426,702		

Panel C. Named versus Control Physicians Event Dates Adjusted for State Pre-Notification Laws (HI excluded and adjustments for CA and WV): Visits Involving Same Body System or Clinical Issue as Malpractice Claim

	Named Physicians (N=61)#		***Control Physicians(N=130)***		*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*		
RVUs/Visit	3.75 (1.00)	3.85 (1.02)	3.78 (1.03)	3.87 (0.99)	-0.02 (-0.14, 0.10)	0.028 (-0.080,0.137) SE = 0.054
Visit Length (hrs)	3.06 (2.00)	3.11 (1.71)	2.94 (1.52)	3.00 (1.63)	0.11 (-0.07, 0.30)	-0.030 (-0.151,0.091) SE = 0.061
Proportion of Visits Resulting in Hospital Admission	0.188 (0.194)	0.186 (0.190)	0.189 (0.193)	0.199 (0.198)	-0.009 (-0.026, 0.008)	-0.011 (-0.027,0.005) SE = 0.008
Physician months	704	1,353	686	1,265		
Visits	198,536	337,356	403,247	631,129		

#One named physician was excluded from this sub-analysis related to their malpractice claim being focused on a non-body system-based issue, and one named physician was excluded due to a lack of subsequent visits corresponding to the body system or clinical issue of the malpractice claim (Table 1) in addition to the two named physicians excluded from HI. This also leads to fewer control physicians in this analysis.

Panel D. Named versus Control Physicians Event Dates Adjusted for State Pre-Notification Laws (HI excluded and adjustments for CA and WV): Non-Failure to Diagnose Claims

	Named Physicians (N=23)		***Control Physicians (N=59)***			
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
RVUs/Visit	3.49 (0.75)	3.59 (0.78)	3.55 (0.75)	3.71 (0.77)	-0.10 (-0.24, 0.04)	-0.027 (-0.146,0.093) SE = 0.059
RVUs/Hour	9.78 (2.35)	10.32 (4.95)	9.31 (2.64)	9.05 (2.92)	1.02 (0.17, 1.86)	0.727 (-0.110,1.565) SE = 0.414
Discharge Visit Length	2.45 (0.79)	2.54 (0.88)	2.46 (1.07)	2.55 (1.08)	-0.01 (-0.26, 0.24)	0.004 (-0.149,0.158) SE = 0.076
Proportion of Visits Resulting in Hospital Admission	0.172 (0.095)	0.163 (0.102)	0.176 (0.097)	0.184 (0.104)	-0.015 (-0.036, 0.005)	-0.013 (-0.029,0.003) SE = 0.008
Physician Press Ganey Percentile Rank	63.2 (33.9)	66.1 (37.9)	62.4 (35.4)	58.7 (36.4)	5.14 (-2.89, 13.17)	5.859 (-2.180,13.898) SE = 3.931
Physician months	248	531	240	522		
Visits	73,153	157,236	158,519	294,866		

Appendix Table 8 Summary Statistics without Propensity-Score Matching

	All Physicians (985)	*Physicians with Malpractice Claims (72)*	*Physicians without Claims (913)*	*Standardized difference in two groups (95% CI)*
Physician characteristics				
Mean Age at First Included Physician-Month	40.9 (9.8)	44.6 (9.1)	40.6 (9.8)	0.42 (0.18, 0.66)
Mean Years since Residency Completion at First Included Physician-Month	8.3 (9.1)	12.1 (9.1)	8.0 (9.1)	0.44 (0.20, 0.69)
% Male	68.1%	72.2%	67.8%	0.10 (-0.14, 0.34)
% Board Certified in Emergency Medicine	78.0%	94.4%	76.7%	0.52 (0.28, 0.76)
Operational characteristics				
Total ED Visits as Attending Physician of Record	6,009,646	669,819	5,339,827	
Per-physician Mean Monthly ED Visits as Attending Physician of Record	255.0 (161.5)	294.1 (130.6)	252.4 (163.1)	0.28 (0.24, 0.32)
Total Physician-Months	21,160	2,238	18,922	
Mean Number of Physician-Months Per Physician	21.5 (17.0)	31.1 (8.3)	20.7 (17.2)	0.77 (0.52, 1.01)
ED Characteristics (same for named and control physicians, by construction)				
ED Mean Annual Visit Volume	35,900			
Mean Annual ED Admission Rate	16.2% (7.5%)			
Percentage with Trauma Designation (Level 1-4)	27.9%			
Percentage in Academic Hospitals	14.8%			
Percentage with Emergency Medicine Residency Program	13.1%			
Malpractice Claim Characteristics (n=77)				
Primary Body System/Clinical Issue of Malpractice Claim (%) #				
Blood/Lymphatic	1 (1.3%)			
Cardiovascular	9 (11.7%)			
ENT	1 (1.3%)			
Endocrine	2 (2.6%)			
Eye	1 (1.3%)			
Gastrointestinal	11 (14.3%)			
Genitourinary	4 (5.2%)			
Skin/Wound	6 (7.8%)			
Neurologic	18 (23.4%)			
OB-GYN	6 (7.8%)			
Orthopedic	7 (9.1%)			
Psychiatric	3 (3.9%)			
Respiratory	7 (9.1%)			
Non-Organ-System Based Clinical Issue (Medical Battery)	1 (1.3%)			
Malpractice Claim Allegation (%) #				
Failure to Diagnose	50 (64.9%)			
Non-Failure to Diagnose	27 (35.1%)			
Claim Disposition (%)				
Physician Voluntarily Dismissed from Claim	52 (67.5%)			
Out-of-Court Settlement	19 (24.7%)			
Trial - Defense Verdict	4 (5.2%)			
Trial - Plaintiff Verdict	2 (2.6%)			
State (%)				
CA	23 (29.9%)			
CT	5 (6.5%)			
HI	2 (2.6%)			
IL	5 (6.5%)			
NC	8 (10.4%)			
NV	9 (11.7%)			
NY	3 (3.9%)			
OH	16 (20.8%)			
OK	2 (2.6%)			
PA	2 (2.6%)			
WV	2 (2.6%)			

Appendix Table 9 Difference-in-differences Analysis without Propensity-Score Matching: Named versus Control Physicians

Panel A. All Visits

	Named Physicians (N=72)		***Control Physicians (N=913)***		*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*		
Relative Value Units (RVUs)/Visit	3.71 (0.81)	3.73 (0.83)	3.76 (0.82)	3.81 (0.82)	-0.07 (-0.16, 0.03)	-0.0004 (-0.08, 0.08) SE = 0.04
RVUs/Hour	10.08 (3.49)	10.09 (4.98)	9.50 (2.62)	9.55 (3.47)	0.55 (-0.06, 1.17)	-0.24 (-0.757, 0.268) SE = 0.26
Visit Length (hrs.)	2.64 (1.17)	2.77 (1.23)	2.65 (1.22)	2.83 (1.29)	-0.04 (-0.23, 0.14)	0.12 (-0.02, 0.25) SE = 0.07
Proportion of Visits Resulting in Hospital Admission	0.192 (0.107)	0.185 (0.116)	0.210 (0.109)	0.203 (0.112)	-0.017 (-0.029, -0.006)	-0.002 (-0.013, 0.008) SE = 0.005
Monthly Physician Press Ganey Percentile Rank	53.3 (37.1)	56.6 (36.9)	55.9 (38.1)	52.8 (38.8)	1.44 (-3.07, 5.96)	5.01 (-0.32, 10.35) SE = 2.67
Physician months	790	1,429	5,911	11,922		
Visits	238,519	431,300	1,900,439	3,439,388		

Panel B. Failure to Diagnose Claims: All Visits

	Named Physicians (N=45)		***Control Physicians (N=827)***		*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*		
RVUs/Visit	3.66 (0.74)	3.72 (0.79)	3.67 (0.81)	3.73 (0.80)	-0.01 (-0.10, 0.08)	-0.01 (-0.10,0.09) SE = 0.05
RVUs/Hour	9.96 (2.91)	9.57 (2.93)	9.72 (2.78)	9.78 (3.83)	-0.05 (-0.49, 0.40)	-0.56 (-1.01,-0.11) SE = 0.23
Visit Length (hrs.)	2.58 (0.90)	2.79 (1.07)	2.62 (1.08)	2.86 (1.14)	-0.06 (-0.21, 0.09)	0.13 (-0.07,0.33) SE = 0.10
Proportion of Visits Resulting in Hospital Admission	0.183 (0.100)	0.180 (0.108)	0.197 (0.110)	0.189 (0.114)	-0.010 (-0.025, 0.004)	-0.002 (-0.014,0.011) SE = 0.006
Monthly Physician Press Ganey Percentile Rank	49.5 (36.3)	57.7 (36.4)	55.6 (38.5)	52.4 (39.0)	1.19 (-4.11, 6.50)	8.92 (2.38,15.46) SE = 3.26
Physician months	492	836	4,746	8,992		
Visits	146,277	264,389	1,560,149	2,785,431		

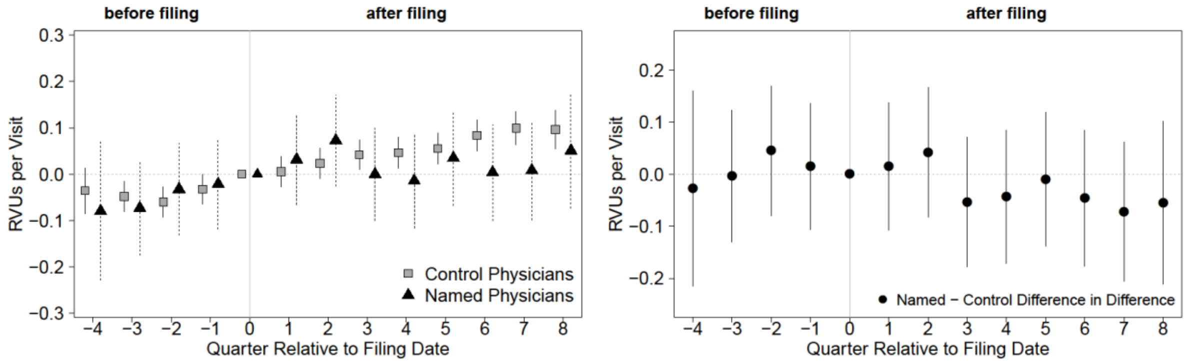
Panel C. Visits Involving Same Body System/Clinical Issue as Malpractice Claim

	Named Physicians (N=70)+		*Control Physicians (N=891)*		*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*		
RVUs/Visit	3.76 (0.98)	3.82 (0.98)	3.83 (1.03)	3.81 (1.00)	-0.02 (-0.11, 0.08)	0.13 (-0.0004, 0.26) SE = 0.07
Visit Length (hrs.)	3.05 (1.94)	3.11 (1.73)	2.87 (1.61)	3.03 (1.68)	0.12 (-0.10, 0.34)	0.12 (-0.06, 0.30) SE = 0.09
Proportion of Visits Resulting in Hospital Admission	0.182 (0.177)	0.177 (0.175)	0.202 (0.189)	0.188 (0.185)	-0.014 (-0.029, 0.001)	0.019 (-0.012, 0.051) SE = 0.016
Physician months	765	1,356	5,812	11,543		
Visits	33,499	58,469	356,430	622,089		

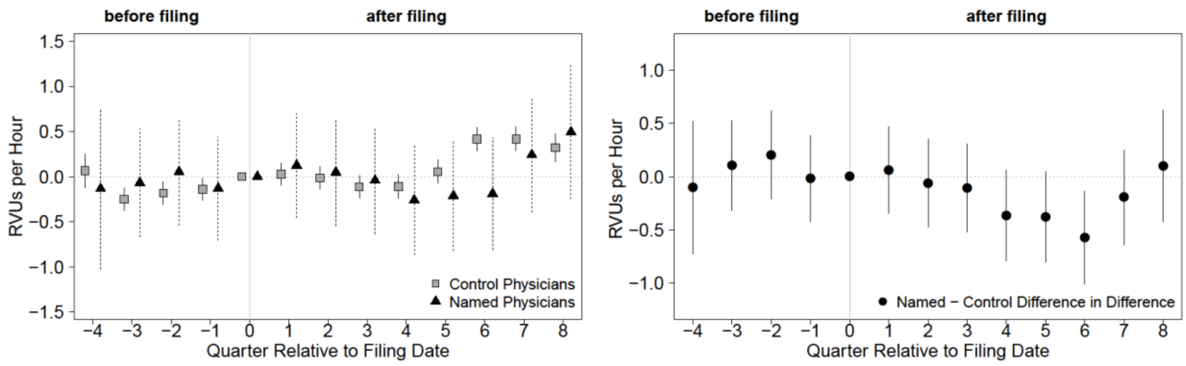
+One named physician was excluded from this sub-analysis because the malpractice claim was for a non-body system-based issue (medical battery), and one named physician was excluded due to a lack of subsequent visits corresponding to the body system or clinical issue of the malpractice claim.

Panel D. Non-Failure to Diagnose Claims: All Visits

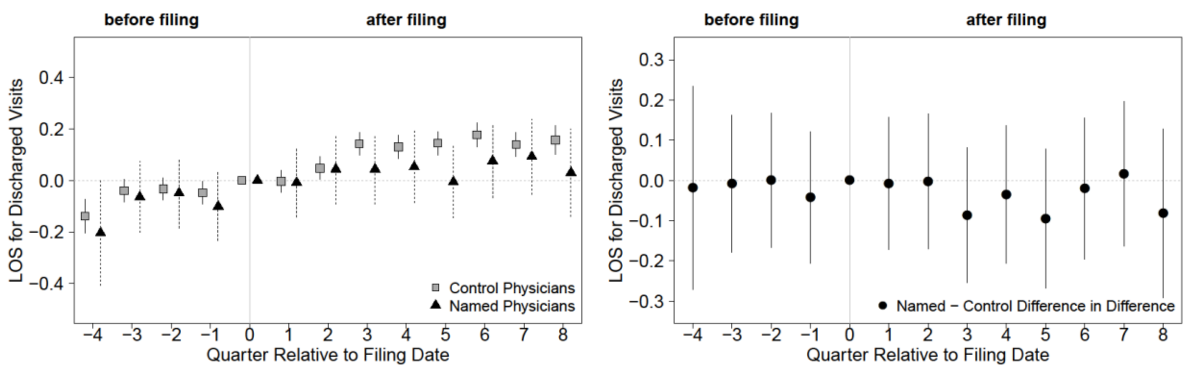
	Named Physicians (N=27)		*Control Physicians (N=397)*		*Pre-claim difference in means (95% CI)*	*DiD Estimate (95% CI), SE*
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*		
RVUs/Visit	3.81 (0.94)	3.75 (0.92)	3.91 (0.83)	3.94 (0.84)	-0.16 (-0.37, 0.05)	-0.004 (-0.12,0.11) SE = 0.06
RVUs/Hour	10.29 (4.38)	11.07 (7.36)	9.13 (2.28)	9.17 (2.77)	1.63 (-0.02, 3.28)	0.29 (-0.67,1.24) SE = 0.46
Visit Length (hrs.)	2.76 (1.55)	2.74 (1.50)	2.71 (1.43)	2.79 (1.51)	-0.02 (-0.31, 0.28)	0.02 (-0.08,0.12) SE = 0.05
Proportion of Visits Resulting in Hospital Admission	0.210 (0.117)	0.195 (0.129)	0.233 (0.102)	0.225 (0.104)	-0.027 (-0.046, -0.008)	-0.0002 (-0.012,0.012) SE = 0.006
Monthly Physician Press Ganey Percentile Rank	59.9 (37.5)	54.5 (37.9)	56.3 (37.7)	53.4 (38.4)	2.00 (-5.82, 9.82)	-2.19 (-8.25,3.87) SE = 2.90
Physician months	298	532	2,565	4,538		
Visits	92,233	166,920	656,949	1,215,936		



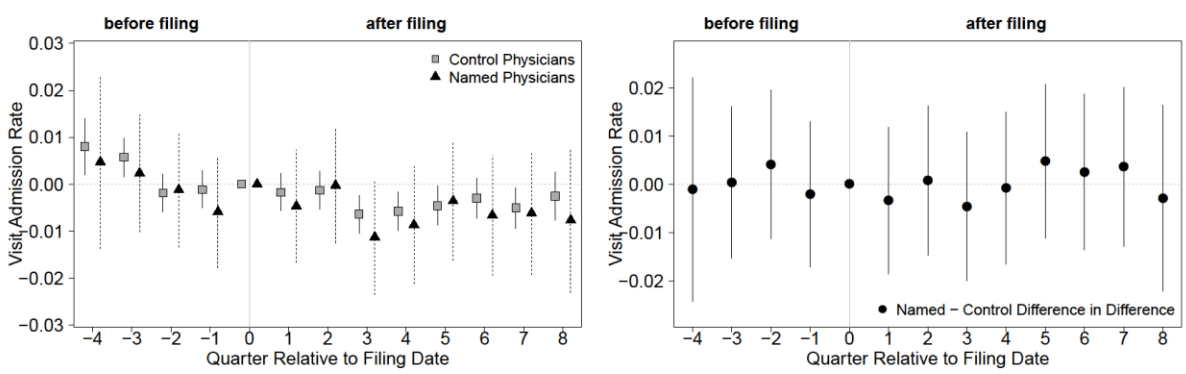
(a) RVUs/visit



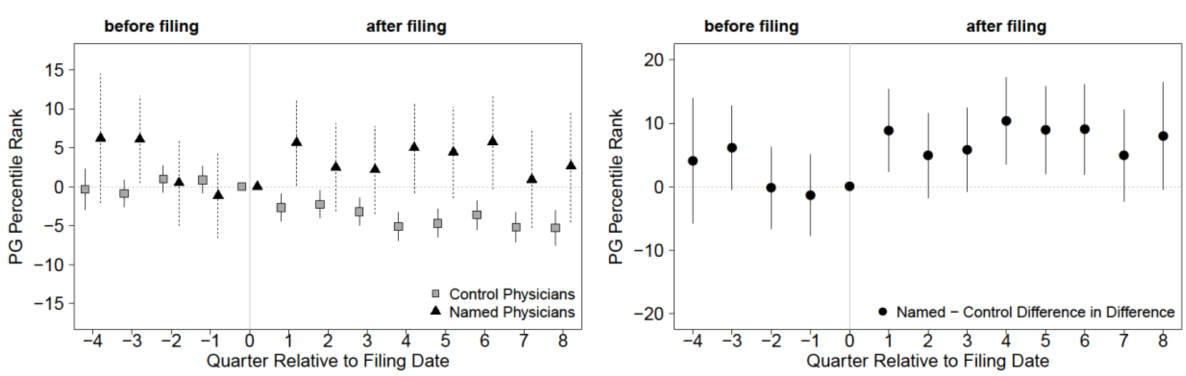
(b) RVUs/hour



(c) Visit Length of Stay (LOS) for Discharged Patients (hours)

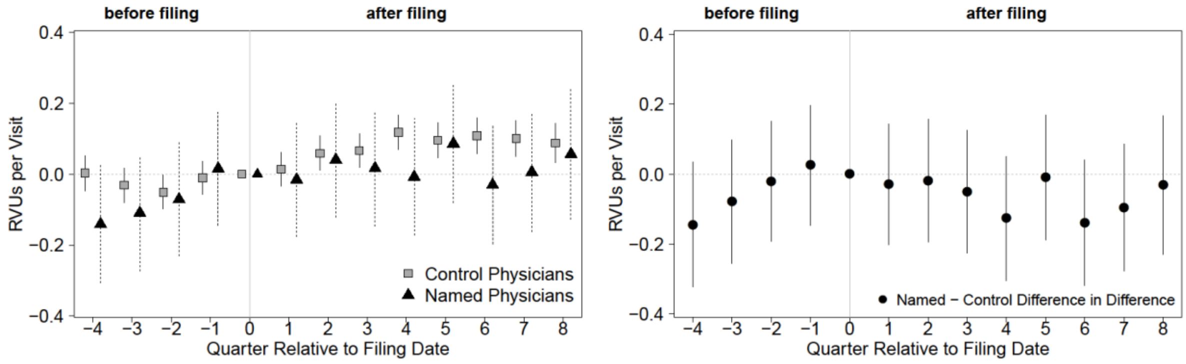


(d) Proportion of Visits Resulting in Hospital Admission

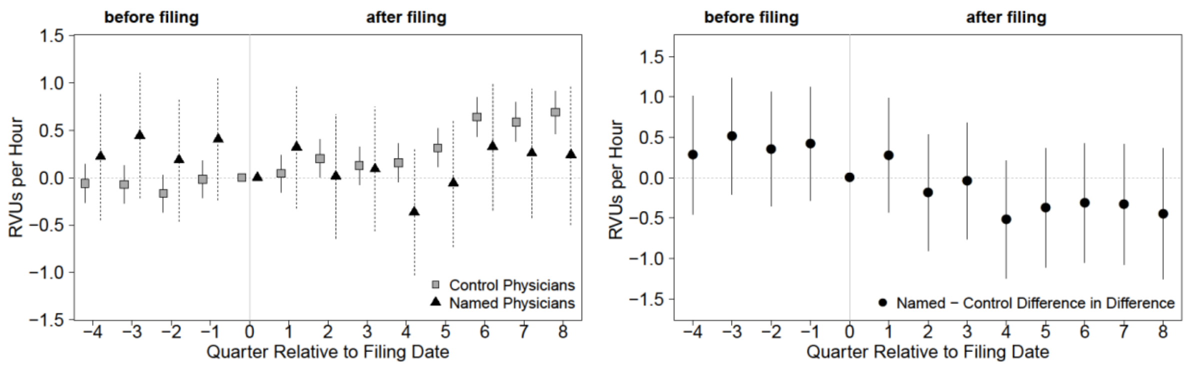


(e) Physician Press Ganey Percentile Rank

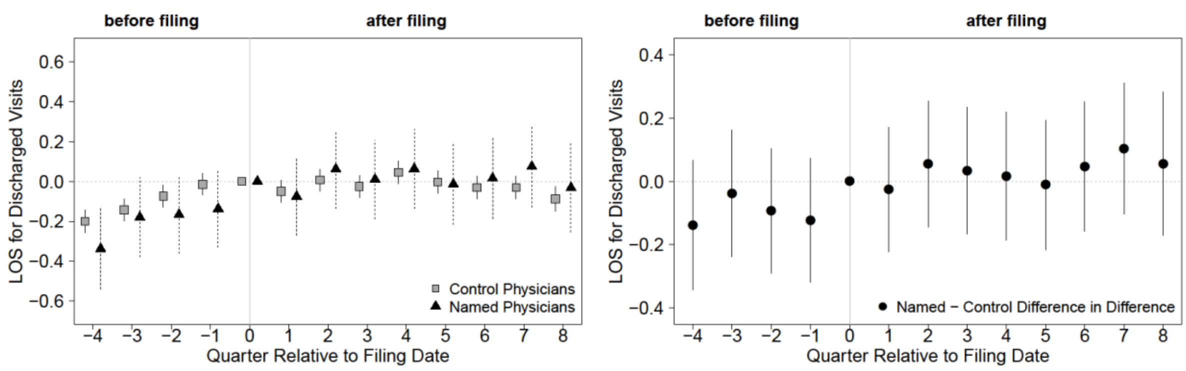
Appendix Figure 3 Leads and Lags for Outcome Measures without Propensity-Score Matching: All Visits



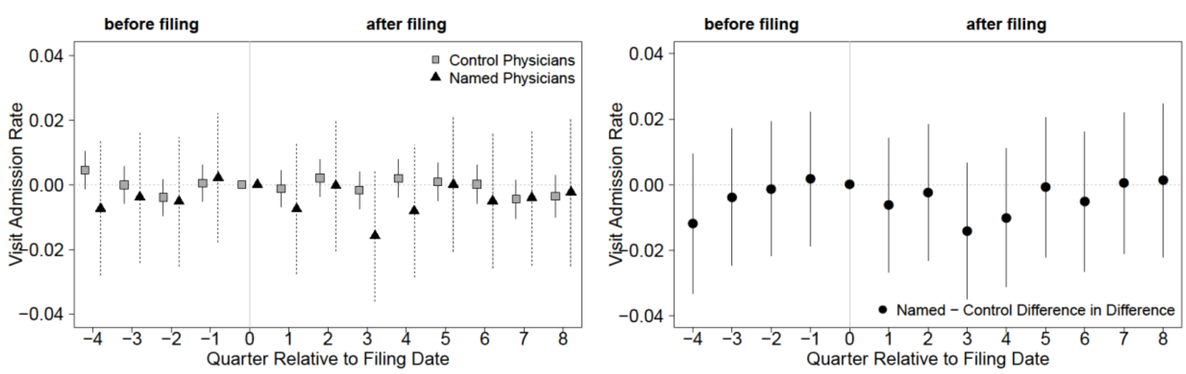
(a) RVUs/visit



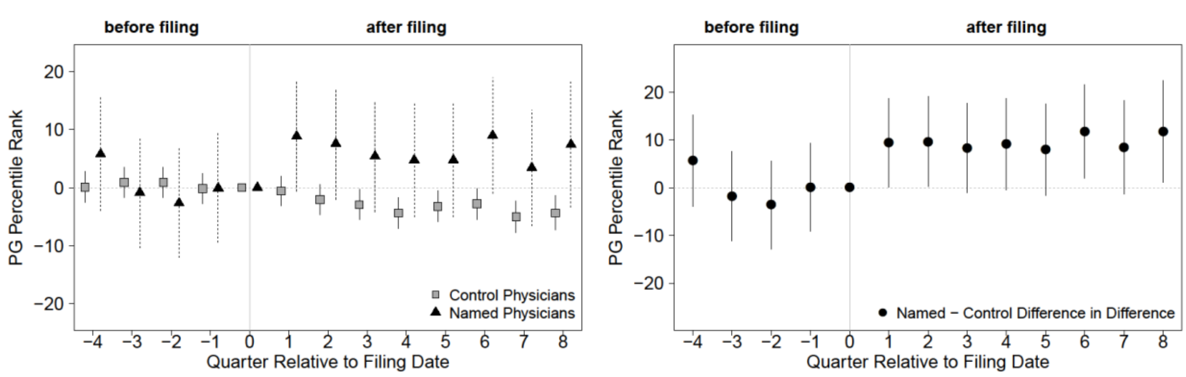
(b) RVUs/hour



(c) Visit Length of Stay (LOS) for Discharged Patients (hours)

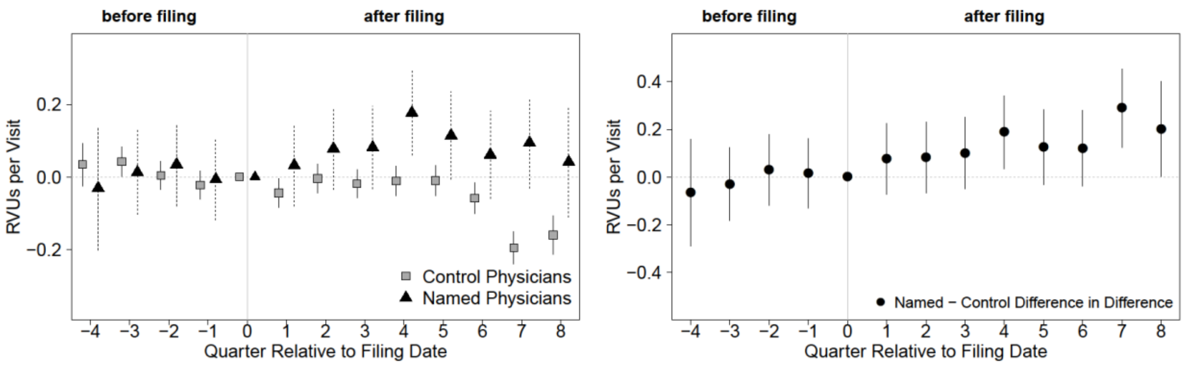


(d) Proportion of Visits Resulting in Hospital Admission

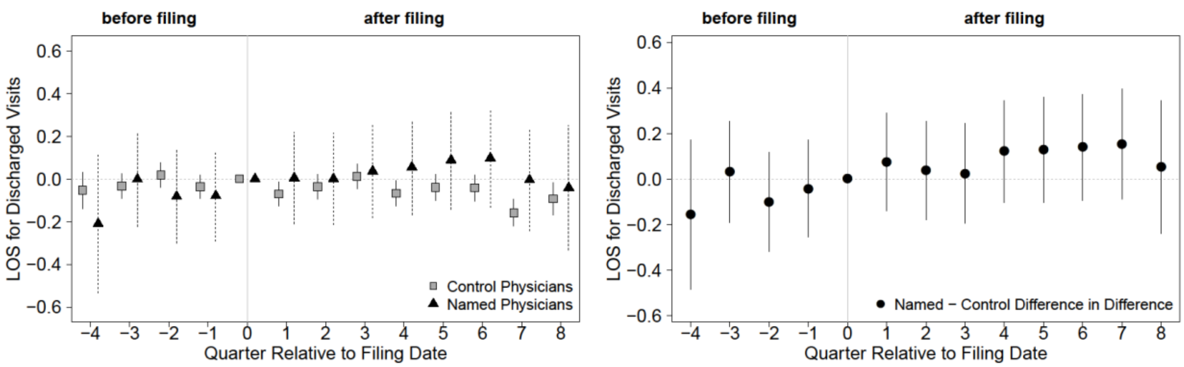


(e) Physician Press Ganey Percentile Rank

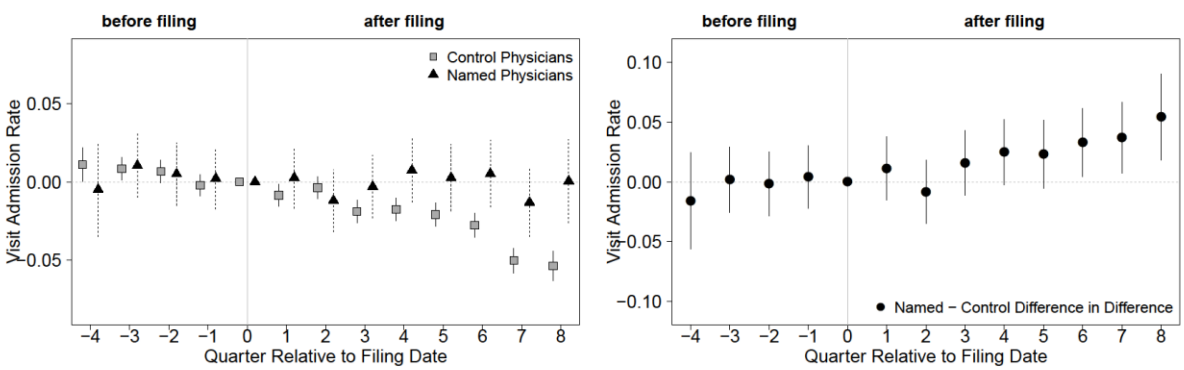
Appendix Figure 4 Leads and Lags Comparing Named Physicians in 50 Failure to Diagnose Claims Versus Their Control Physicians without Propensity-Score Matching for Outcome Measures in All Included ED Visits



(a) RVUs/visit



(b) Visit Length of Stay (LOS) for Discharged Patients (hours)



(c) Proportion of Visits Resulting in Hospital Admission

Appendix Figure 5 Leads and Lags for Outcome Measures without Propensity-Score Matching: Body System/Clinical Issue Specific Visits

Appendix Table 10 Difference-in-Difference-in-Difference Analysis: Named versus Control Physicians by Gender

	Named Physicians (N=65)				***Control Physicians (N=140)***				
	Male Physicians (N=49)		***Female Physicians (N=16)***		***Male Physicians (N=94)***		***Female Physicians (N=46)***		*DDD Estimate (95% CI), SE*
Outcome Measure	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	*Pre-Claim Mean (SD)*	*Post-Claim Mean (SD)*	
Relative Value Units (RVUs)/Visit	3.67 (0.76)	3.74 (0.84)	3.71 (0.90)	3.69 (0.86)	3.68 (0.78)	3.79 (0.77)	3.55 (0.74)	3.57 (0.82)	0.04 (-0.16, 0.24) SE = 0.10
RVUs/Hour	10.24 (4.07)	10.10 (5.44)	9.53 (2.19)	10.01 (4.00)	9.64 (2.75)	9.54 (2.94)	9.33 (2.52)	9.49 (2.26)	0.39 (-0.99, 1.76) SE = 0.69
Visit Length (hrs.)	2.71 (1.28)	2.82 (1.29)	2.56 (0.90)	2.55 (0.91)	2.59 (1.06)	2.69 (1.13)	2.70 (1.12)	2.64 (1.07)	-0.001 (-0.22, 0.22) SE = 0.11
Proportion of Visits Resulting in Hospital Admission*	0.182 (0.104)	0.178 (0.115)	0.202 (0.117)	0.193 (0.118)	0.191 (0.091)	0.199 (0.096)	0.172 (0.095)	0.170 (0.097)	-0.007 (-0.017, 0.031) SE = 0.012
Monthly Physician Press Ganey Percentile Rank	48.77 (36.87)	52.82 (36.82)	58.35 (36.08)	68.19 (32.41)	54.13 (37.62)	53.52 (38.16)	56.97 (35.45)	60.09 (34.87)	4.42 (-9,08, 17.91) SE = 6.72
Physician months	531	974	200	405	1,138	1,926	464	778	
Visits	156,580	256,616	50,922	101,705	300,487	477,669	128,359	202,396	

Bibliography

AANP. (2018). 2017 National Nurse Practitioner Sample Survey Results. https://www.aanp.org/news-feed/2017-national-nurse-practitioner-sample-survey-results#:~:text=Findings%20from%20the%20survey%20suggest,full%2Dtime%20NPs%20was%20%24112%2C923. (accessed date: September 5, 2018).

AAPA. (2018). Women Are Paid Less Than Men in the PA Profession. https://www.aapa.org/wp-content/uploads/2017/11/SR-Realtionship-Between-Salary-and-Gender-for-PAs.pdf. (accessed date: September 5, 2018).

ACEP. (2009). Emergency Department Wait Times, Crowding and Access Fact Sheet. Available at http://newsroom.acep.org/2009-01-04-emergency-department-wait-times-crowding-and-access-fact-sheet. (accessed date: September 5, 2018).

ACEP. (2013). Guidelines Regarding the Role of Physician Assistants and Advanced Practice Registered Nurses in the Emergency Department. Available at https://www.acep.org/patient-care/policy-statements/guidelines-regarding-the-role-of-physician-assistants-and-advanced-practice-registered-nurses-in-the-emergency-department/.(accessed date: September 5, 2018).

ACEP. (2016). *CEDR Frequently Asked Questions*. Retrieved from https://www.acep.org/globalassets/cedr _pdfs/CEDR-FAQs.pdf. Accessed August 27, 2017.

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

Angst, C. M., Devaraj, S., Queenan, C. C., & Greenwood, B. (2011). Performance effects related to the sequence of integration of healthcare technologies. *Production and Operations Management*, *20*(3), 319-333.

ATS. (2019). Trauma Center Levels Explained. Available at https://www.amtrauma.org/page/traumalevels (accessed date: September 30, 2019).

Ba, S., & Nault, B. R. (2017). Emergent themes in the interface between economics of information systems and management of technology. *Production and Operations Management*, *26*(4), 652-666.

Barata, I., Brown, K. M., Fitzmaurice, L., Griffin, E. S., Conners, G. P., American Academy of Pediatrics Committee on Pediatric Emergency Medicine, & Emergency Nurses Association Pediatric Committee. (2015). Best practices for improving flow and care of pediatric patients in the emergency department. *Pediatrics*, *135*(1), 273.

Bhargava, H. K., & Mishra, A. N. (2014). Electronic medical records and physician productivity: evidence from panel data analysis. *Management Science, 60*(10), 2543-2562.

Bishop, T. F., Federman, A. D., & Keyhani, S. (2010). Physicians' views on defensive medicine: a national survey. *Archives of Internal Medicine*, *170*(12), 1081-1083.

Blank, J. L., & Eggink, E. (2014). The impact of policy on hospital productivity: a time series analysis of Dutch hospitals. *Health care management science*, *17*(2), 139-149.

Boudreaux, E. D., & O'Hea, E. L. (2004). Patient satisfaction in the emergency department: a review of the literature and implications for practice. *The Journal of Emergency Medicine*, *26*(1), 13-26.

Boudreaux, E. D., Cruz, B. L., & Baumann, B. M. (2006). The use of performance improvement methods

to enhance emergency department patient satisfaction in the United States: a critical review of the literature and suggestions for future research. *Academic Emergency Medicine*, *13*(7), 795-802.

Brennan, D. F., Silvestri, S., Sun, J. Y., & Papa, L. (2007). Progression of EM resident productivity. *Academic Emergency Medicine*, *14*(9), 790-794.

Brown, T. W., McCarthy, M. L., Kelen, G. D., & Levy, F. (2010). An epidemiologic study of closed emergency department malpractice claims in a national database of physician malpractice insurers. *Academic Emergency Medicine, 17(5), 553-560.*

Brown, D. F., Sullivan, A. F., Espinola, J. A., & Camargo, C. A. (2012). Continued rise in the use of mid-level providers in US emergency departments, 1993–2009. *International journal of emergency medicine*, 5(1), 21.

Büchner, V. A., Hinz, V., & Schreyögg, J. (2016). Health systems: changes in hospital efficiency and profitability. *Health care management science*, 19(2), 130-143.

California Code of Civil Procedure § 364. The Commencement of Actions Based Upon Professional Negligence https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CCP&sectionNum=364. Accessed May 12, 2018.

Carlson, J. N., Foster, K. M., Pines, J. M., Corbit, C. K., Ward, M. J., Hydari, M. Z., & Venkat, A. (2018). Provider and practice factors associated with emergency physicians' being named in a malpractice claim. *Annals of emergency medicine*, *71*(2), 157-164.

Carlson, J. N., Foster, K. M., Black, B. S., Pines, J. M., Corbit, C. K., & Venkat, A. (2020). Emergency physician practice changes after being named in a malpractice claim. *Annals of Emergency Medicine*, *75*(2), 221-235.

Carrier, E. R., Reschovsky, J. D., Mello, M. M., Mayrell, R. C., & Katz, D. (2010). Physicians' fears of malpractice lawsuits are not assuaged by tort reforms. *Health affairs*, *29*(9), 1585-1592.

Carter, A. J., & Chochinov, A. H. (2007). A systematic review of the impact of nurse practitioners on cost, quality of care, satisfaction and wait times in the emergency department. *Canadian Journal of Emergency Medicine*, 9(4), 286-295.

Cawley, J. F., & Hooker, R. S. (2013). Physician assistants in American medicine: the half-century mark. *The American journal of managed care*, *19*(10), e333-41.

Clinkscales, J. D., Fesmire, F. M., Hennings, J. R., Severance, H. W., Seaberg, D. C., & Patil, N. (2016). The Effect of EM Residents on Clinical Efficiency and Staffing Requirements. *Academic Emergency Medicine*, *23*(1), 78-82.

CMS. (2015). Hospital Value-Based Purchasing. Retrieved from https://www.cms.gov/Medicare/Quality -Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html. Accessed September 5, 2016.

Colaco, M., Heavner, M., Sunaryo, P., & Terlecki, R. (2015). Malpractice litigation and testicular torsion: a legal database review. *The Journal of emergency medicine*, *49*(6), 849-854.

Currie, J., & MacLeod, W. B. (2008). First do no harm? Tort reform and birth outcomes. *The Quarterly Journal of Economics*, *123*(2), 795-830.

Cydulka, R. K., Tamayo-Sarver, J., Gage, A., & Bagnoli, D. (2011). Association of patient satisfaction with complaints and risk management among emergency physicians. *The Journal of emergency medicine*, *41*(4), 405-411.

Daniels, A. H., Ruttiman, R., Eltorai, A. E., DePasse, J. M., Brea, B. A., & Palumbo, M. A. (2017). Malpractice litigation following spine surgery. *Journal of Neurosurgery: Spine*, *27*(4), 470-475.

Darcy, C., McCarthy, A., Hill, J., & Grady, G. (2012). Work–life balance: One size fits all? An exploratory analysis of the differential effects of career stage. *European Management Journal*, *30*(2), 111-120.

DePasse, J. M., Ruttiman, R., Eltorai, A. E., Palumbo, M. A., & Daniels, A. H. (2017). Assessment of malpractice claims due to spinal epidural abscess. *Journal of Neurosurgery: Spine*, *27*(4), 476-480.

Dill, M. J., Pankow, S., Erikson, C., & Shipman, S. (2013). Survey shows consumers open to a greater role for physician assistants and nurse practitioners. *Health Affairs*, 32(6), 1135-1142.

Doan, Q., Hooker, R. S., Wong, H., Singer, J., Sheps, S., Kissoon, N., & Johnson, D. (2012). Canadians' willingness to receive care from physician assistants. *Canadian Family Physician*, *58*(8), e459-e464.

Dyrbye, L. N., Varkey, P., Boone, S. L., Satele, D. V., Sloan, J. A., & Shanafelt, T. D. (2013, December). Physician satisfaction and burnout at different career stages. In *Mayo Clinic Proceedings* (Vol. 88, No. 12, pp. 1358-1367). Elsevier.

Eitel, D. R., Rudkin, S. E., Malvehy, M. A., Killeen, J. P., & Pines, J. M. (2010). Improving service quality by understanding emergency department flow: a White Paper and position statement prepared for the American Academy of Emergency Medicine. *The Journal of Emergency Medicine*, *38*(1), 70-79.

Ellaway, R. H., Pusic, M. V., Galbraith, R. M., & Cameron, T. (2014). Developing the role of big data and analytics in health professional education. *Medical Teacher*, *36*(3), 216-222.

Farmer, S. A., Moghtaderi, A., Schilsky, S., Magid, D., Sage, W., Allen, N., ... & Black, B. (2018). Association of medical liability reform with clinician approach to coronary artery disease management. *JAMA cardiology*, *3*(7), 609-618.

Foster, K., Penninti, P., Shang, J., Kekre, S., Hegde, G. G., & Venkat, A. (2018). Leveraging Big Data to Balance New Key Performance Indicators in Emergency Physician Management Networks. *Production and Operations Management*, *27*(10), 1795-1815.

Fraher, E. P., Morgan, P., & Johnson, A. (2016). Specialty distribution of physician assistants and nurse practitioners in North Carolina. *Journal of the American Academy of Physician Assistants*, 29(4), 38-43.

Gandhi, T. K., Kachalia, A., Thomas, E. J., Puopolo, A. L., Yoon, C., Brennan, T. A., & Studdert, D. M. (2006). Missed and delayed diagnoses in the ambulatory setting: a study of closed malpractice claims. *Annals of internal medicine*, *145*(7), 488-496.

Gaudio, F. G., & Borensztein, R. (2018). An Emergency Medicine Residency for Nurse Practitioners. *Advanced emergency nursing journal*, 40(2), 119-126.

Gimm, G. W. (2010). The impact of malpractice liability claims on obstetrical practice patterns. *Health services research*, *45*(1), 195-211.

Grönroos, C., & Ojasalo, K. (2004). Service productivity: Towards a conceptualization of the transformation of inputs into economic results in services. *Journal of Business Research*, *57*(4), 414-423.

Hamden, K., Jeanmonod, D., Gualtieri, D., & Jeanmonod, R. (2014). Comparison of resident and mid-level provider productivity in a high-acuity emergency department setting. *Emerg Med J*, *31*(3), 216-219.

Handel, D. A., French, L. K., Nichol, J., Momberger, J., & Fu, R. (2014). Associations between patient and emergency department operational characteristics and patient satisfaction scores in an adult population. *Annals of Emergency Medicine*, *64*(6), 604-608.

Hawaii Revised Statutes § 671. Medical Claim Conciliation. http://files.hawaii.gov/dcca/oah/hrs/hrs_oah_671.pdf. Accessed May 12, 2018.

Hickson, G. B., Clayton, E. W., Entman, S. S., Miller, C. S., Githens, P. B., Whetten-Goldstein, K., & Sloan, F. A. (1994). Obstetricians' prior malpractice experience and patients' satisfaction with care. *Jama*, *272*(20), 1583-1587.

Hickson, G. B., Federspiel, C. F., Pichert, J. W., Miller, C. S., Gauld-Jaeger, J., & Bost, P. (2002). Patient complaints and malpractice risk. *Jama*, *287*(22), 2951-2957.

Hooker, R. S., Brock, D. M., & Cook, M. L. (2016). Characteristics of nurse practitioners and physician assistants in the United States. *Journal of the American Association of Nurse Practitioners*, 28(1), 39-46.

Horwitz, L. I., Green, J., & Bradley, E. H. (2010). US emergency department performance on wait time and length of visit. *Annals of emergency medicine*, *55*(2), 133-141.

Hwang, C. E., Lipman, G. S., & Kane, M. (2015). Effect of an emergency department fast track on Press-Ganey patient satisfaction scores. *Western Journal of Emergency Medicine*, *16*(1), 34.

Jeanmonod, R., DelCollo, J., Jeanmonod, D., Dombchewsky, O., & Reiter, M. (2013). Comparison of resident and mid-level provider productivity and patient satisfaction in an emergency department fast track. *Emerg Med J*, *30*(1), e12-e12.

Johnson, J. R. (2016). Patient Satisfaction with Care Provided By Physician Assistants in an Orthopedic Clinic. *Ed.D.Dissertations*. 101. https://digitalcommons.olivet.edu/edd_diss/101

Jena, A. B., Seabury, S., Lakdawalla, D., & Chandra, A. (2011). Malpractice risk according to physician specialty. *New England Journal of Medicine*, *365*(7), 629-636.

Johnson, T., Shah, M., Rechner, J., & King, G. (2008). Evaluating the effect of resident involvement on physician productivity in an academic general internal medicine practice. *Academic Medicine*, *83*(7), 670-674.

Karcz, A., Holbrook, J., Auerbach, B. S., Blau, M. L., Bulat, P. I., Davidson III, A., ... & Green, E. D. (1990). Preventability of malpractice claims in emergency medicine: a closed claims study. *Annals of emergency medicine*, *19*(8), 865-873.

Katz, B. (2017). 2017–2018 Compensation report for emergency physicians shows steady salaries. *ACEP Now*.

Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The big-data revolution in US health care: Accelerating value and innovation. *McKinsey & Company*, *2*(8), 1-13.

Kc, D. S. (2013). Does multitasking improve performance? Evidence from the emergency department. *Manufacturing & Service Operations Management*, *16*(2), 168-183.

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

Kraus, C. K., Carlisle, T. E., & Carney, D. M. (2018). Emergency medicine physician assistant (EMPA) postgraduate training programs: program characteristics and training curricula. *Western Journal of Emergency Medicine*, *19*(5), 803.

Krishnaiah, K., & Shahabudeen, P. (2012). *Applied design of experiments and Taguchi methods*. PHI Learning Pvt. Ltd.

Larkin, G. L., & Hooker, R. S. (2010). Patient willingness to be seen by physician assistants, nurse practitioners, and residents in the emergency department: does the presumption of assent have an empirical basis?. *The American Journal of Bioethics*, *10*(8), 1-10.

Lechner, M. (2011). *The estimation of causal effects by difference-in-difference methods*. Now.

Lee, E. K., Atallah, H. Y., Wright, M. D., Post, E. T., Thomas IV, C., Wu, D. T., & Haley Jr, L. L. (2015). Transforming Hospital Emergency Department Workflow and Patient Care. *Interfaces*.

Medical Decision Making And The Marshfield Clinic Scoring Tool FAQ. https://http://www.acep.org/administration/reimbursement/reimbursement-faqs/medical-decision-making-and-the-marshfield-clinic-scoring-tool-faq/. Accessed April 9, 2019.

Mello, M. M., Chandra, A., Gawande, A. A., & Studdert, D. M. (2010). National costs of the medical liability system. *Health affairs*, *29*(9), 1569-1577.

Moghtaderi, A., Farmer, S., & Black, B. (2019). Damage Caps and Defensive Medicine: Reexamination with Patient-Level Data. *Journal of empirical legal studies*, *16*(1), 26-68.

Moskowitz, A., McSparron, J., Stone, D. J., & Celi, L. A. (2015). Preparing a new generation of clinicians for the era of big data. *Harvard medical student review*, *2*(1), 24.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA*, *309*(13), 1351-1352.

Myers, R. A., Parikh, P. J., Ekeh, A. P., Denlinger, E., & McCarthy, M. C. (2014). Scheduling of advanced practice providers at Level 1 trauma centers. *Journal of Trauma and Acute Care Surgery*, *77*(1), 176-181.

Nambiar, R., Bhardwaj, R., Sethi, A., & Vargheese, R. (2013, October). A look at challenges and opportunities of big data analytics in healthcare. In *Big Data, 2013 IEEE International Conference on* (pp. 17-22). IEEE.

Nelson, S. C., & Hooker, R. S. (2016). Physician assistants and nurse practitioners in rural Washington emergency departments. *The Journal of Physician Assistant Education*, *27*(2), 56-62.

Nestler, D. M., Fratzke, A. R., Church, C. J., Scanlan-Hanson, L., Sadosty, A. T., Halasy, M. P., Finley, J.L., Boggust, A., & Hess, E. P. (2012). Effect of a physician assistant as triage liaison provider on patient throughput in an academic emergency department. *Academic Emergency Medicine*, *19*(11), 1235-1241.

Nugus, P., Holdgate, A., Fry, M., Forero, R., McCarthy, S., & Braithwaite, J. (2011). Work pressure and patient flow management in the emergency department: findings from an ethnographic study. *Academic Emergency Medicine*, *18*(10), 1045-1052.

Nyberg, S. M., Keuter, K. R., Berg, G. M., Helton, A. M., & Johnston, A. D. (2010). A national survey: Acceptance of physician assistants and nurse practitioners in trauma centers. *Journal of the American Academy of Physician Assistants*, 23(1), 35-41.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, *375*(13), 1216.

Pavlik, D., Sacchetti, A., Seymour, A., & Blass, B. (2017). Physician assistant management of pediatric patients in a general community emergency department: a real-world analysis. *Pediatric emergency care*, *33*(1), 26-30.

Pines, J. M., Hollander, J. E., Isserman, J. A., Chen, E. H., Dean, A. J., Shofer, F. S., & Mills, A. M. (2009). The association between physician risk tolerance and imaging use in abdominal pain. *The American journal of emergency medicine*, *27*(5), 552-557.

Pines, J. M., Isserman, J. A., Szyld, D., Dean, A. J., McCusker, C. M., & Hollander, J. E. (2010). The effect of physician risk tolerance and the presence of an observation unit on decision making for ED patients with chest pain. *The American journal of emergency medicine*, *28*(7), 771-779.

Pines, J. M., Iyer, S., Disbot, M., Hollander, J. E., Shofer, F. S., & Datner, E. M. (2008). The effect of emergency department crowding on patient satisfaction for admitted patients. *Academic Emergency Medicine*, 15(9), 825-831.

Pines, J. M., Penninti, P., Alfaraj, S., Carlson, J. N., Colfer, O., Corbit, C. K., & Venkat, A. (2018). Measurement under the microscope: high variability and limited construct validity in emergency department patient-experience scores. *Annals of emergency medicine*, *71*(5), 545-554.

Powell, A., Savin, S., & Savva, N. (2012). Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing & Service Operations Management*, *14*(4), 512-528.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, *2*(1), 3.

Reiter, M., Wen, L. S., & Allen, B. W. (2016). The Emergency Medicine Workforce: Profile and Projections. *The Journal of emergency medicine*, *50*(4), 690-693.

Sanbar, S. S., & Firestone, M. H. (2007). Medical malpractice stress syndrome. In *The medical malpractice survival handbook* (pp. 9-15). Mosby.

Sawyer, B. T., & Ginde, A. A. (2014). Scope of practice and autonomy of physician assistants in rural versus urban emergency departments. *Academic Emergency Medicine*, *21*(5), 520-525.

Seabury, S. A., Chandra, A., Lakdawalla, D. N., & Jena, A. B. (2013). On average, physicians spend nearly 11 percent of their 40-year careers with an open, unresolved malpractice claim. *Health Affairs*, *32*(1), 111-119.

Silberman, M., Jeanmonod, D., Hamden, K., Reiter, M., & Jeanmonod, R. (2013). Mid-level Providers Working in a Low-acuity Area are More Productive than in a High-acuity Area. *Western Journal of Emergency Medicine*, *14*(6), 598.

Song, H., Tucker, A. L., & Murrell, K. L. (2015). The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science, 61*(12), 3032-3053.

Sonnati, R. (2015). Improving Healthcare Using Big Data Analytics. *International Journal of Scientific & Technology Research*, *4*(8), 142-146.

Srinivasan, U., & Arunasalam, B. (2013). Leveraging big data analytics to reduce healthcare costs. *IT professional*, *15*(6), 21-28.

Studdert, D. M., Bismark, M. M., Mello, M. M., Singh, H., & Spittal, M. J. (2016). Prevalence and characteristics of physicians prone to malpractice claims. *New England Journal of Medicine*, *374*(4), 354-362.

Sturmann, K. M., Ehrenberg, K., & Salzberg, M. R. (1990). Physician assistants in emergency medicine. *Annals of emergency medicine*, *19*(3), 304-308.

Theokary, C., & Justin Ren, Z. (2011). An empirical study of the relations between hospital volume, teaching status, and service quality. *Production and Operations Management*, *20*(3), 303-318.

Vallejo, S., Medlin, K., Diver, T., & Hartos, J. (2015). Do Patients Prefer Seeing a Physician over a Physician Assistant? A Systematic Review.

van der Linden, C., Reijnen, R., & de Vos, R. (2010). Diagnostic accuracy of emergency nurse practitioners versus physicians related to minor illnesses and injuries. *Journal of Emergency Nursing*, *36*(4), 311-316.

Venkat, A., Kekre, S., Hegde, G. G., Shang, J., & Campbell, T. P. (2015). Strategic Management of Operations in the Emergency Department. *Production and Operations Management*, *24*(11), 1706-1723.

Ward, A. R., & Armony, M. (2013). Blind fair routing in large-scale service systems with heterogeneous customers and servers. *Operations Research*, *61*(1), 228-243.

Ward, M. J., Froehle, C. M., Hart, K. W., Collins, S. P., & Lindsell, C. J. (2014a). Transient and sustained changes in operational performance, patient evaluation, and medication administration during electronic health record implementation in the emergency department. *Annals of Emergency Medicine*, *63*(3), 320-328.

Ward, M. J., Landman, A. B., Case, K., Berthelot, J., Pilgrim, R. L., & Pines, J. M. (2014b). The effect of electronic health record implementation on community emergency department operational measures of performance. *Annals of Emergency Medicine*, *63*(6), 723-730.

Waxman, D. A., Greenberg, M. D., Ridgely, M. S., Kellermann, A. L., & Heaton, P. (2014). The effect of malpractice reform on emergency department care. *New England Journal of Medicine*, *371*(16), 1518-1525.

West Virginia Code § 55-7B. Medical Professional Liability. http://www.wvlegislature.gov/wvcode/ChapterEntire.cfm?chap=55&art=7b. Accessed May 12, 2018.

Wiler, J. L., Welch, S., Pines, J., Schuur, J., Jouriles, N., & Stone-Griffith, S. (2015). Emergency department performance measures updates: proceedings of the 2014 Emergency Department Benchmarking Alliance Consensus Summit. *Academic Emergency Medicine*, *22*(5), 542-553.