Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments

Gautam Gowrisankaran, Keith Joiner, Pierre Thomas Léger

To cite this article:
Gautam Gowrisankaran, Keith Joiner, Pierre Thomas Léger (2023) Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments. Management Science 69(6):3202-3219. https://doi.org/10.1287/mnsc.2022.4544

informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 69, No. 6, June 2023, pp. 3202–3219
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments

Gautam Gowrisankaran,[a,b,c,*] **Keith Joiner,**[d] **Pierre Thomas Léger**[e]

[a] Columbia University, New York 10027; [b] Centre for Economic Policy Research London EC1V 0DK, United Kingdom; [c] National Bureau of Economic Research, Cambridge, Massachusetts 02138; [d] University of Arizona, Tucson, Arizona 85721; [e] University of Illinois Chicago, Chicago, Illinois 60607

*Corresponding author

Contact: gautamg2@gmail.com, https://orcid.org/0000-0001-8549-959X (GG); kjoiner@medadmin.arizona.edu (KJ); ptleger@uic.edu (PTL)

Received: May 31, 2019
Revised: July 7, 2020; May 28, 2021; January 21, 2022; June 27, 2022
Accepted: August 9, 2022
Published Online in Articles in Advance: October 17, 2022

https://doi.org/10.1287/mnsc.2022.4544

Abstract. We study healthcare operations of emergency departments (EDs) by examining the practice styles and skills of ED physicians. Our data include all residents of Montreal, Canada, with an initial ED visit in Montreal during a nine-month period. For each visit, our data record the initial treating hospital, ED physician, ED billed expenditures, and all interactions with the health system within the subsequent 90 days. Physicians in Montreal rotate across shifts between simple and difficult cases, implying a quasi-random assignment of patients to physicians within an ED. We consider three medical conditions that present frequently in the ED and for which mistreatment may have dramatic consequences—angina, appendicitis, and transient ischemic attacks—jointly examining diagnostic and disposition skills. To control for variation in diagnosis, our sample for each condition consists of patients with a broader set of symptoms and signs potentially indicative of the condition. Separately by condition, we regress healthcare usage and cost measures on indicators for physicians to estimate the skill and practice style of each physician. We then evaluate the variation across physicians in their practice style and skills and the correlations between different measures of skill and practice style. We find significant variation across physicians in their practice styles and skills. We also find that physicians with costly practice styles often have worse outcomes in terms of more ED revisits and more hospitalizations. Finally, the practice styles and skills of physicians correlate positively across the three conditions that we consider.

History: Accepted by Jayashankar Swaminathan, operations management.
Funding: Gowrisankaran acknowledges funding from the Center for Management Innovations in Healthcare at the University of Arizona.
Supplemental Material: The online appendix and data are available at https://doi.org/10.1287/mnsc.2022.4544.

Keywords: physician quality • emergency departments • quasi-random variation

1. Introduction

Healthcare systems in many countries are in crisis with system inefficiencies in cost, quality, and access that are important, complex, and have interdependent operational management causes. The interdependencies also imply that it may be possible to simultaneously improve cost and quality. For these reasons, understanding efficient practices and activities in healthcare has become increasingly important to operations management (KC et al. 2019). The purpose of this study is to add evidence on healthcare operations in the context of emergency departments (EDs). EDs are settings that treat heterogeneous and complex medical conditions. Each patient's ED care is coordinated by the ED physician, who must work in concert with many other personnel. For these reasons, EDs are places where the operational issues that typically present in healthcare are particularly important.

This study evaluates the variations in and relations between treatment costs and follow-up visits with the medical system across physicians for similar sets of patients. Specifically, we consider to what extent there is variation in the treatment costs across different ED physicians when presented with patients with similar clinical issues and overall severity of illness; whether physicians who spend more have better outcomes, such as fewer hospitalizations over the medium run; whether physicians who spend more on some health conditions also spend more on others; and whether physicians who perform better in treating some health conditions also perform better in treating others.

We make use of a unique data set on ED patients in the Montreal, Canada, metropolitan area that records follow-up visits over 90 days.[1] Importantly, the initial assignment to ED physician in our setting is quasi-random given the choice of hospital. We exploit the quasi-random assignment of patients to physicians to derive treatment effects of the ED physician in a way that minimizes issues of physician selection and adds to the plausibility of our results.

Understanding variation in healthcare utilization and costs in EDs has broad application from an operations management point of view. EDs are a primary entry point to the healthcare system for many patients, particularly patients who are underserved by primary care physicians. Also, EDs are often seen as a particularly wasteful entry point; that is, researchers focus on the (presumably negative) consequences that insurance can have on increasing ED usage (Taubman et al. 2014). Furthermore, ED care is complex because of the acute, time-sensitive, and wide-ranging nature of the diagnostic and therapeutic processes. Thus, EDs are among the most common sites for misdiagnosis (Pope et al. 2000, Rothrock and Pagane 2000, Pope and Edlow 2012). Finally, EDs are a prominent place of treatment for a number of widely occurring diseases that may be difficult to diagnose and for which appropriate and timely care is particularly important.

To motivate our analysis, we propose a simple model of the role of the ED physician. The ED—or attending—physician is the leader of a team of caregivers that provides care to a patient. We observe the assignment of patients to ED physicians. The ED physician has two central roles: diagnosis and disposition. The ED physician first attempts to diagnose the illness of the patient. The ED physician then chooses patient disposition, which involves both the ED treatments and the recommendations for follow-up care. For both of these steps, the ED physician makes decisions about resource use, for example, laboratory tests, medications, and hospitalization. In the optimal circumstance, the ED physician makes the correct diagnosis, chooses the appropriate disposition, and uses only necessary resources. Overuse or underuse of resources, coupled with and sometimes connected to an incorrect diagnosis or inappropriate disposition, can result in worse outcomes and/or additional resource use.

ED physicians may vary in both their *practice style* and *skills*. A physician's practice style describes the approach that the physician takes to diagnosis and disposition. We measure practice style as the physician's contribution to the patient's billed ED expenditures. A physician's skill refers to the physician's ability to make the correct diagnosis by distinguishing between various alternatives and to ensure that the patient follows an appropriate disposition plan. We measure physician skill as the physician's contribution to desired outcomes, such as fewer ED revisits or medium-term hospitalizations. Practice style and skills may both affect diagnosis and disposition, leading to interdependent system inefficiencies in healthcare operations. An innovation of our study is that we consider the impact of both diagnosis and disposition skills. Further, we examine these skills across a range of illnesses that are commonly seen in the ED and difficult to both diagnose and treat.

Specifically, our paper estimates the extent of variation in practice style and skills and the relation between these measures within and across illnesses. In different medical contexts, some researchers find that a more intensive practice style correlates negatively with skills (e.g., Doyle et al. 2010, Chan et al. 2022), whereas others find a positive correlation (e.g., Currie et al. 2016). Ultimately, understanding this relation in our context requires empirical analysis.

We focus on three illnesses—angina, appendicitis, and transient ischemic attacks (TIAs)—which, together, provide comprehensive evidence on ED operations. Angina refers to heart pain and can be a precursor to a myocardial infarction (heart attack). Appendicitis is the painful inflammation of the appendix. TIA, sometimes called a ministroke, is a self-limited blockage of blood to some portion of the brain and can be a precursor to a full-blown stroke. These illnesses are all difficult to diagnose and treat appropriately. These decisions are also important because appropriate disposition can minimize the probability of potentially catastrophic and expensive future events.

To understand our empirical approach further, consider a single condition: TIA. Suppose first that TIA patients are randomly assigned to physicians in the ED and physicians always diagnose TIAs accurately. In that case, if we compare practice style or skill across physicians conditional on the (proper) diagnosis of TIA, this would identify the treatment effects of each ED physician. We could then correlate treatment effects of the fixed effects for spending and outcomes to understand whether physicians with an expensive practice style disproportionately have high skills. In this case, variation across physicians in future healthcare utilization of their patients would relate exclusively to their decisions regarding patient disposition.

However, a key component of ED physician skill is the ability to accurately diagnose conditions, which is particularly true for the illnesses that we examine. Now, suppose that some physicians are good at recognizing TIAs, others sometimes misdiagnose them as migraines, and still others misdiagnose migraines as TIAs. Then, even if patients are randomly assigned to physicians, the set of patients diagnosed with a TIA by a particular physician will reflect a selected sample. A physician who frequently codes migraines as TIAs may appear to have fewer TIA complications, but this could be due to inaccurate diagnosis rather than better disposition. In order to compare accurately physician practice style and skill for TIA, we can then no longer consider patients coded with TIAs, but we must instead consider patients who *could* have TIAs even if the physician misdiagnosed them as having a different illness.

Thus, our estimation is based on the universe of patients who could have one of the three illnesses noted herein. We refer to the patients who could have angina, appendicitis, and TIAs as angina+, appendicitis+, and TIA+ patients, respectively. For instance, the TIA+ sample (the "plus" sample) includes patients whose primary diagnoses have symptoms and signs that overlap with TIA—such as migraines or epilepsy—whereas TIA (the "base" sample) only includes patients with a primary diagnosis of TIA. We then compare spending and outcomes across physicians conditional on a diagnosis in the plus sample.

Formally, our identifying assumption is that the coding of patients within these plus samples is consistent across ED physicians within an ED or, equivalently, hospitals. We believe that this is reasonable because our plus samples include the most common *differential diagnoses*[2] of the base diagnosis and because accurate diagnosis is relatively easier in the plus sample.

For each of the three conditions that we consider, our empirical analysis proceeds by estimating linear regressions of a number of measures of current and future healthcare utilization on fixed effects for each initial ED treatment physician at each hospital. Our measures of healthcare utilization include costs in the initial ED visit—which allow us to uncover physician practice style—and future ED visits, office visits, and hospital admissions and costs—which allow us to uncover physician skills. We perform regressions on both the base and plus samples. The regressions on the plus samples inform us about the diagnosis and disposition of patients together, whereas the regressions on the base samples inform us about the disposition of patients conditional on a diagnosis of angina, appendicitis, or TIA.

Given quasi-random assignment, the physician fixed effect on the initial ED spending indicates the physician's relative practice style as it indicates the spending level for a similar set of patients. The fixed effect on other outcomes may positively or negatively relate to physician skills, depending on the measure. Following the literature, we interpret three fixed effects of future outcomes as unambiguous (negative) measures of skill given quasi-random assignment: ED revisits within zero to five days and numbers and total costs of hospitalizations within 6–90 days.[3] We view other fixed effects as more ambiguous in their interpretation and, hence, do not assign a normative value for them.

We use the estimated physician fixed effects to examine the variation in outcomes across physicians and the correlations between different indicators for a given physician. Because the quasi-random assignment of patients occurs only within an ED, we compare all physician fixed effects to the mean effect at their ED.

1.1. Overview of Results

For all three conditions, we find that there is large variation within EDs in both physician practice style and skills. Measures of future healthcare usage also vary significantly between the 90th and 10th percentiles. Physician practice style and skills are often negatively correlated, consistent with physicians using extra resources during the initial ED visit when they are uncertain about appropriate diagnosis and disposition. Physician practice style and skills are positively correlated across illness categories. Those who spend more during the initial ED visit or prevent more ED revisits for one of the three illness categories typically do so across the other two though the correlation is far from perfect. Finally, we find positive correlations between the base and plus samples for our three illness, suggesting that both diagnosis and disposition are important.

1.2. Related Literature

Our paper contributes primarily to three recent literatures. One literature leverages the quasi-random assignment of patients to hospitals, physicians, or triage nurses in order to evaluate physician practice style and resource use across different settings (see, e.g., Epstein and Nicholson 2009; Doyle et al. 2010, 2015; Currie et al. 2016; Currie and MacLeod 2017, 2020; Chan et al. 2022; Mullainathan and Obermeyer 2022). Chan et al. (2022) investigate the relation between physician skill and practice style for radiologists diagnosing pneumonia. Currie and MacLeod (2017) separate physician inputs into diagnosis and surgical skills. Currie et al. (2016) consider treatment intensity and heterogeneity in physician treatments across patient characteristics for heart attack patients. Dahlstrand (2022) studies physician skills for primary care services provided online. Relative to this literature, we evaluate both diagnosis and disposition skills, and we examine the relationship of practice styles and skills across conditions, which we believe is critical to understanding the operational effectiveness of ED care.

A second literature considers physician inputs in EDs. Chan (2016) examines how ED physicians respond to different management systems when making care decisions. Doyle (2011) examines ED spending for people who seek treatment far from home in different areas. Silver (2021) examines variations in ED physician productivity by estimating within-physician productivity differences based on shifts. Van Parys (2016) examines ED physician practice styles for minor injuries. Kindermann et al. (2014) and Pines et al. (2009) examine the determinants of the intensity of ED imaging requests. Finally, other studies evaluate ED misdiagnosis (see, e.g., Kachalia et al. 2007, Hastings et al. 2009, Abaluck et al. 2016), ED physician preferences over the set of patients to treat (Chang and Obermeyer 2020), and the geographic determinants of hospitalizations from EDs (Shoff et al. 2019). We contribute to this literature by addressing the relation between physician skills and practice styles. Our unique data—which record the physician of assignment in the ED and track healthcare utilization and expenditures over the medium run—allow us to add evidence here.

A third literature considers the operational effectiveness of ED production. KC (2014) finds that multitasking by ED physicians increases productivity up to a threshold. Song et al. (2015) find that dedicated queueing systems lead to shortened waiting times, whereas Batt and Terwiesch (2017) find that earlier ordering of tests leads to shortened treatment time. Finally, Song et al. (2018) find that the public disclosure of relative performance feedback can improve performance by facilitating best practice dissemination and adoption. Our paper complements this literature with our quasi-experimental study of the determinants of ED skills and practice style that considers the universe of ED admissions for a large city and examines the operational effectiveness of ED physicians in treating three different conditions.

2. Data and Institutional Framework

2.1. Overview

Our study uses administrative data from *la Régie de l'assurance maladie du Québec* (RAMQ). RAMQ pays for all publicly funded healthcare expenditures from the Canadian province of Quebec. Its database tracks enrollees through time and across covered care settings. Almost all residents of Quebec are covered by RAMQ,[4] which provides first-dollar coverage to its enrollees.

We study residents of the Island of Montreal with an initial ED visit during the period April 1, 2006, to December 31, 2006. The Island of Montreal—which we refer to henceforth just as Montreal—had a population of about 1.9 million residents and includes the city of Montreal and some suburban municipalities. We observe information on all ED visits in Montreal for residents of Montreal during this period and on all future healthcare consumption in Quebec covered by RAMQ during the following 90 days. We received access to the data from Montreal's public health department (*le Département de santé publique de Montréal*).

In Quebec, during the study period, RAMQ paid ED physicians on a fee-for-service basis. Our data include the billed physician costs (i.e., the total fee-for-service payments) for the initial ED visit as well as measures of the number of future inpatient hospitalizations and office, hospital-based external clinic,[5] and ED visits. We also observe a proxy for the total costs of each future inpatient stay, called *Niveau d'intensité relative des ressources utilisées* (relative intensity level of resources used, NIRRU). The NIRRU is quite similar to the diagnosis-related groups as defined by Medicare in the United States (Strumpf et al. 2017) and is widely used by academic studies that use Quebec health data (see, e.g., Rivest et al. 2001, Rinfret et al. 2010, Strumpf et al. 2017). RAMQ reports this proxy instead of costs because all nonphysician-related expenditures are covered by a global hospital budget and, thus, not directly associated with the patient.

Each ED visit constitutes one observation in our data set. For each observation, our data include the ED's unique identifier, the physician's unique identifier, month and weekend identifiers, four-digit International Classification of Diseases (ICD)-9 diagnosis codes, procedure codes,[6] fees paid to physicians for services provided, patient gender, age in five-year bins,[7] three-digit postal code, and two measures of socioeconomic status (SES). The data also link the initial ED visit with all future visits covered by RAMQ for 90 days (thus, ending on March 31, 2007). Specifically, Montreal's public health department constructed variables indicating the 0–5, 6–30, and 31–90 day number of ED revisits, office visits, hospital-based external clinic visits, and hospitalizations as well as hospitalization costs as measured by NIRRU.

A limitation of our setting is that RAMQ does not generate billed hospitalization costs. The fact that hospitals are reimbursed with a global budget also leads to a related limitation. Procedures ordered by the ED but provided by the hospital outside the ED—notably, blood tests, CT scans, and MRIs—are also not directly associated with the patient. Finally, we do not have access to precise dates of services. As a result, we cannot determine the sequence of events within time intervals. For instance, we do not know if a patient with a hospitalization within zero to five days of an ED visit was admitted directly from an ED visit or returned home and hospitalized a few days later.

2.2. Assignment of Patients to Physicians

The assignment process of patients to physicians is central to our identification strategy. EDs in Montreal are staffed with one or more physicians. When two or more ED physicians are present, physicians are assigned to either the heavy cases (most often patients who arrive by ambulance) or the light cases (most often patients who enter through the front door) for the duration of their shift. When two or more physicians are assigned to the same shift type, the allocation of patients to physicians within a shift type is random in the sense that it is based uniquely on the triage order (which indicates how much time the patient can wait before seeing a physician or, equivalently, who should be seen next). The triage order is done by a triage nurse. Shift allocations (i.e., the time, day of week, and shift type) are done several months in advance. As ED physicians are paid fee-for-service, and payments are invariant to any physician characteristics, such as experience or tenure, most ED physicians are expected to work all shift types in similar proportions. Thus, shift types are typically done in an equitable manner. Exceptions may nonetheless exist, especially among older physicians who may only work part time or may be given only light cases if they lack advanced emergency medicine training. ED physicians in Montreal told us that there is at least one exception to this assignment rule during the study period: in one ED, the allocation of patients was based uniquely on the triage order rather than having separate streams of light and heavy cases.[8]

We observe patients seen at the 19 general acute care hospitals with EDs in Montreal.[9] From this universe of patients, we exclude initial visits at two EDs that serve fewer than 1,000 patients per year. We also exclude all ED physicians who are not present in all months studied as well as those who treated fewer than 200 cases during the nine-month period as these physicians are less likely to have an equitable mix of cases because they may include older and part-time physicians. Our base sample includes 280 ED physicians who practice in 17 EDs. This sample contains 321,256 ED visits, each of which constitutes one observation. These visits are made by 199,442 distinct patients. In this sample, 26% of visits occur on the weekend.

Although physicians' shift types affect the allocation of patients to physicians, the quasi-random allocation of ED physicians to shifts over time should lead all ED physicians to see very similar pools of patients over the long run. In order to test and deal with the possibility of nonrandom assignment of patients to ED physicians, we examine whether physicians within an ED treat patients with observably different characteristics. We primarily use two patient characteristics that are observable in our data and also easily observable to ED staff: age bin and gender. Although patient pools could still differ in unobservable manners, we believe that this test is informative in identifying the nonrandom assignment of patients.[10] We proceed by regressing the patient characteristic of interest (gender or age bin) on physician dummies with indicators for month and weekend. We perform the regressions separately for each ED. We then test whether the physician fixed effects are statistically different from each other.

Table 1, panel A, presents the p values from F tests for gender and age bins. Of the 17 EDs in our sample, we find evidence that nine do not assign their patients to physicians in a purely random matter with respect to their patients' gender at the 1% level (EDs 3, 5, 6, 7, 8, 10, 11, 13, and 16), whereas we find evidence against random assignment with respect to the patients' age bin for most hospitals (all but EDs 4, 7, and 9). The lack of purely random assignment is likely to be due to differences in shift types and older and part-time physicians. Illness severity may, in turn, be correlated with a patient's gender and/or age bin. Whereas we do not directly observe physician shift type, physician age, or the time or date of patient visits, we further evaluate this point by examining the differences between the mean gender and mean age bin for an individual physician and the means at the ED where the physician works. These distributions (omitted for compactness) suggest that a few atypical physicians may drive the F test results.

Given this, we eliminate ED physicians whose mean gender or mean age are significantly different than peers within their ED.[11] More specifically, we drop physicians whose mean gender is more than five percentage points from the mean for their ED or whose mean age bin is more than 0.5 from the means for their

Table 1. Tests for Randomness

	Panel A: Full sample			Panel B: Restricted sample		
ED no.	Gender	Age bin	No. of physicians	Gender	Age bin	No. of physicians
1	0.024**	0.000***	18	0.119	0.034**	15
2	0.017**	0.000***	13			
3	0.002***	0.000***	24			
4	0.219	0.942	7	0.191	0.935	6
5	0.008***	0.000***	25	0.031**	0.001***	22
6	0.000***	0.000***	14	0.000***	0.000***	12
7	0.000***	0.016	11	0.011**	0.046**	8
8	0.004***	0.000***	17	0.053*	0.010**	12
9	0.118	0.329	16	0.118	0.329	16
10	0.000***	0.000***	22			
11	0.000***	0.000***	22			
12	0.051*	0.000***	14	0.035**	0.000***	12
13	0.003***	0.000***	19	0.049**	0.000***	14
14	0.037**	0.000***	9	0.039**	0.000***	7
15	0.040**	0.000***	12	0.070	0.053*	11
16	0.000***	0.000***	23	0.004***	0.000***	18
17	0.431	0.003***	26	0.393	0.023**	24

Note. Each entry under "gender" and "age bin" provides the p value for one F test of a regression of gender or age bin on physician fixed effects.
***Significance at the 1% level, **5% level, and *10% level of the physician fixed effects.

ED. By doing so, however, we lose a considerable proportion of physicians in some of our EDs (i.e., EDs 2, 3, 10, and 11), which calls into question the randomness of the remaining sample. Thus, we drop these four EDs altogether. We then rerun the same two *F* tests on the remaining EDs, excluding the atypical physicians. Table 1, panel B, presents results from these new *F* tests on the "restricted" sample of ED physicians. Using the restricted samples, we find no violations of the randomness assumption at the 1% level for either age or gender for seven EDs. Thus, our main estimation sample includes seven EDs that do not violate the randomness assumptions: 1, 4, 7, 8, 9, 15, and 17.[12]

Overall, we believe that our assumption of quasi-random assignment of patients to physicians is plausible for three reasons. First, the institutional environment directly implies a quasi-random assignment. Second, our estimation sample is based on statistical tests that exclude physicians who may not have been randomly assigned patients. Finally, we show robustness of our main results to the inclusion of outlier physicians and controls for patient observable characteristics.[13] Nonetheless, because our study is observational, we cannot rule out that our results are in part a result of triage nurses selecting ED physicians based on health status markers that we do not observe.

2.3. Construction of Samples by Illness

Our samples include information on patients who have symptoms that might include angina, appendicitis, and TIA.

Angina is a precursor to myocardial infarction (heart attack). It indicates partial blockage of one or more coronary arteries, which supply blood to the heart muscle. The classic symptom of angina is chest pressure or pain ("an elephant is sitting on my chest"), but other presentations, such as jaw or shoulder pain, indigestion, or nausea, also occur. Diagnosis is considered more difficult in women, in which up to 50% may present without chest pressure/pain (McSweeney et al. 2003). Anginal symptoms usually resolve quickly with cessation of exertion because the relative limitation of blood flow to the heart is relieved. Confirmation of coronary artery disease depends upon prompt assessment using stress echocardiogram, coronary angiography, and/or nuclear medicine scans. The consequences of failing to consider or appropriately manage angina include myocardial infarction and sudden death.

Appendicitis is inflammation of the appendix caused by a blockage of the hollow portion of the appendage. Though pain in the lower right portion of the abdomen (right lower quadrant) is one well-known presenting symptom, the clinical presentation of appendicitis is highly variable. For instance, a classic sequence of symptoms that indicate appendicitis is loss of appetite, followed by nausea, right lower quadrant pain, and vomiting. The accurate diagnosis of appendicitis is often challenging because many other abdominal conditions can mimic appendicitis. A normal appendix is found at surgery in approximately 12% of cases, reflecting misdiagnosis (Seetahal et al. 2011). Common conditions, such as urinary tract infections, gastroenteritis, and gall bladder inflammation (cholecystitis), can all mimic appendicitis. Management of suspected appendicitis varies from administration of antibiotics for early, uncomplicated cases to immediate surgical intervention to circumvent appendiceal rupture with its attendant and often serious complications. This puts a premium on rapid decision making on the part of the ED physician. Unlike for angina and TIA, appendicitis is not a transient event that resolves on its own. Thus, inaccurate diagnosis or disposition almost certainly results in future healthcare usage within our 90-day time period.

TIAs are precursors to strokes and are sometimes called ministrokes. They result from a transient occlusion of a blood vessel in the brain. Unlike strokes, the symptoms and signs resolve quickly (usually within minutes) because the occlusion partially or fully resolves. The symptoms and signs, which vary enormously depending upon the part of the brain that is affected, include visual or speech changes, weakness, and numbness. Confirmation of a TIA depends upon additional tests, often done after discharge from the ED (carotid ultrasound, MRI or CT, echocardiogram). The consequences of failing to consider or appropriately manage a TIA include a nearly fivefold increase in the incidence of stroke over the subsequent 90 days (Rothwell et al. 2007).

We chose these illnesses for six reasons. First, the presentation of each illness at the ED represents a well-recognized diagnostic dilemma in which the base diagnoses must be distinguished from a series of differential diagnoses that often present in similar fashion. This situation arises when the symptoms reported by the patient, the signs present on physical examination, and results from laboratory and imaging studies are broadly consistent with more than one diagnosis. These are distinguished from unambiguous presentations, such as a broken arm from a fall, a dog bite, or an allergic reaction to a bee sting. Second, each illness presents as a new acute event and not as an acute exacerbation of a chronic condition, such as diabetes, heart failure, or emphysema. The latter often result in management difficulties but are rarely diagnostic dilemmas.[14] Third, presentation in the ED is a precursor to a catastrophic event (e.g., heart attack, ruptured appendix, and stroke, respectively), putting a premium on correct diagnosis and disposition. Fourth, the base diagnosis and differential diagnoses are common enough to be seen with substantial frequency in the ED.[15] Fifth, the base diagnosis is a discrete entity rather than representing a spectrum of disorders, such as pneumonia.[16] Finally, each illness is treatable if diagnosed correctly, implying that the initial ED visit is important.

Though diagnosis can be challenging for all of these illnesses, on average, it is most difficult for TIA. The reason is that presentation takes many forms and can vary from one episode to another. Reported rates of misdiagnosis are as high as 40%–60% (Kessler and Thomas 2009). Angina and appendicitis are easier to diagnose (McSweeney et al. 2003, Bhangu et al. 2015) and, hence, likely have lower rates of misdiagnosis. Within an illness, the likelihood of incorrect diagnosis in the ED is influenced by many factors, including gender, age, underlying health status, and whether the symptoms and signs are present at the time of the ED visit.

Recall that, for each of the three base illnesses here, we also want to consider a broader plus set of diagnoses that could reflect a misdiagnosis of the base illness. Table 2 presents the plus diagnoses for each of our three base diagnoses. We include the most frequent differential diagnoses for these base diagnoses as reported by the medical literature (Nadarajan et al. 2014, Hollander and Chase 2016, Martin 2017).

For the most common diagnoses in the plus categories besides the base diagnosis, there are often specific indications that the plus diagnosis is correct. For example, for angina+, a provider can typically determine if symptoms indicate costochondritis (an inflammation of the cartilage in the rib cage) in a clinical setting by whether the provider can reproduce the chest pain through direct palpation of the chest wall (Proulx and Zryd 2009) and/or by eliciting whether the patient suffered some minor traumatic injury (WebMD 2020). The provider can also typically diagnose gastritis or gastro-esophageal reflux by eliciting whether symptoms are precipitated and/or relieved by eating (Johns Hopkins Medicine 2020, Mayo Clinic 2020). Finally, the provider can diagnose chest pain by whether the patient reports a history and presence of pain and by whether the other diagnoses on our angina+ list are not present (WebMD 2020). Similar clues to identifying the plus diagnoses apply to appendicitis (Bhangu et al. 2015) and TIA (Kessler and Thomas 2009). Thus, we believe that there is less ambiguity diagnosing the plus diagnoses than the base diagnoses.

Table 3 presents summary statistics on the set of EDs, physicians, and patients in our main estimation sample. The first panel shows that the estimation samples for the plus diagnoses include seven EDs (as noted), 90 physician–ED pairs,[17] and a total of 15,098 initial ED visits and 12,855 unique patients. The mean age bins for all three illnesses is approximately 12, which represents an age range from 50 to 54. The percentage of males varies from 36.1% for the appendicitis+ sample to 46.3% for the angina+ sample.

The table reports statistics on the set of dependent variables used in the estimation section. "ED $," which denotes the physician-related billed spending during the initial ED visit, has means that range from $45.94 to $68.91 across the three samples. Because the payments here do not include procedures performed outside the ED—notably tests and imaging services—the

Table 2. ICD-9 Diagnosis Codes for Angina+, Appendicitis+, and TIA+

Angina+		Appendicitis+	
Code	Description	Code	Description
413	Angina	540	Appendicitis
786	Chest pain	614.9	Pelvic inflammatory disease
789.6, 530.8	Gastro-esophageal reflux	599	Urinary tract infection
530.5	Esophageal dysmotility	588.9	Noninfectious gastroenteritis
530.1	Esophagitis	617	Endometriosis
535.5	Gastritis	562.1	Diverticulitis
733.6	Costochondrititis	575.0	Cholecystitis and biliary colic
307.8	Psychosematic/psychogenic	620.2	Ovarian cysts
420	Pericarditis		
422	Myocarditis		
441	Acute aortic syndrome		
415.1	Pulmonary embolism		
486	Pneumonia		
489	Asthma		
TIA+			
Code	Description		
435.9	TIA		
346	Migraine		
345	Epilepsy		
780.2	Syncope		
432.1	Subdural hematoma		
431	Intracerebral hemorrhage		
721.1	Compressive myelopathy		

Table 3. Summary Statistics for Main Estimation Sample

Statistic	Angina+	Appendicitis+	TIA+
Number of EDs	7	7	7
Number of unique ED/MD pairs	90	90	90
Number of observations (initial ED visits)	10,942	1,851	2,305
Number of unique ED patients	9,021	1,706	2,128
Percent male	46.3	36.1	40.4
Mean age bin	12.7 (4.0)	11.8 (4.2)	12.8 (4.2)
Initial ED $	53.30 (42.93)	45.94 (35.76)	68.91 (49.50)
ED 0–5 days	0.28 (0.56)	0.25 (0.50)	0.35 (0.63)
ED 6–90 days	0.72 (1.81)	0.52 (1.32)	0.77 (2.31)
External 0–30 days	0.85 (1.36)	1.02 (1.42)	0.93 (1.49)
Office 0–30 days	0.61 (0.99)	0.63 (0.91)	0.61 (0.94)
Hospital stay 0–5 days	0.17 (0.38)	0.23 (0.42)	0.22 (0.43)
Hospital stay 6–90 days	0.16 (0.48)	0.15 (0.44)	0.18 (0.49)
Hospital $ 0–5 days	2,451 (8,937)	2,069 (5,497)	3,671 (10,666)
Hospital $ 6–90 days	2,070 (9,654)	1,300 (6,231)	1,945 (7,290)
Statistic	Angina	Appendicitis	TIA
Number of EDs	7	6	7
Number of unique ED/MD pairs	75	68	46
Number of observations (initial ED visits)	558	229	117
Number of unique ED patients	542	220	108
Percent male	49.6	46.3	49.6
Mean age bin	14.7 (3.0)	9.0 (3.3)	15.3 (3.0)
Initial ED $	61.69 (43.80)	49.39 (32.79)	60.95 (42.68)
ED 0–5 days	0.39 (0.64)	0.26 (0.51)	0.45 (0.73)
ED 6–90 days	0.82 (2.16)	0.27 (0.87)	0.44 (0.98)
External 0–30 days	1.11 (1.42)	1.11 (1.15)	1.48 (1.40)
Office 0–30 days	0.72 (1.03)	0.69 (0.90)	0.62 (0.92)
Hospital stay 0–5 days	0.33 (0.49)	0.62 (0.50)	0.30 (0.48)
Hospital stay 6–90 days	0.22 (0.60)	0.07 (0.30)	0.21 (0.51)
Hospital $ 0–5 days	4,977 (12,225)	4,428 (4,714)	3,111 (5,677)
Hospital $ 6–90 days	3,147 (11,227)	435 (2,301)	2,913 (11,407)

Notes. Standard deviations provided in parentheses. Patient statistics treat each patient/ED encounter as a unique observation. With respect to the age variable, bins 1 and 2 group patients from 0–1 and 2–4 years, bins 3 through 18 group patients in five-year bins, whereas bin 19 groups all patients 85 years of age or older.

variation here comes exclusively from performing different tasks within the ED. We believe that this variation is meaningful from an operations management point of view because ED physicians who perform more tasks in the ED are likely ones who have a more interventional style in general.

"ED 0–5 days" and "ED 6–90 days" denote the number of revisits to the ED during the first 5 and subsequent 85 days following the ED visit, respectively. The number of ED revisits is high; for example, each initial ED visit is followed by an average of 1.00, 0.77, and 1.12 future ED visits during the 90-day period for the angina+, appendicitis+, and TIA+ samples, respectively. "Office 0–30 days" and "External 0–30 days" denote the number of office visits and hospital-based external clinic visits in the first 30 days following an ED visit, respectively. The "Hospital stay 0–5 days" and "Hospital stay 6–90 days" denote the number of hospitalizations during the first 5 and subsequent 85 days following an ED visit.[18] Hospitalizations are quite frequent across all illness categories.

Finally, the "Hospital $" measures denote hospital spending—as measured by NIRRU—during the time period described. The mean contribution to hospitalization costs are $4,521, $3,369, and $5,616 for the angina+, appendicitis+, and TIA+ samples, respectively.

The second panel of Table 3 considers the three base samples separately. There are relatively few patients in the base samples with a total of 904 initial ED visits and 870 unique patients. A comparison of the base samples to the plus samples shows that the base angina and appendicitis samples have much higher mean hospitalization costs over the short run. The higher costs for the base samples reflect the fact that the differential diagnoses for these conditions generally have lower acuity than the base diagnoses.

Interestingly, for appendicitis+, the medium-run hospitalization costs and ED visits are higher than for appendicitis. Similarly, the mean number of ED visits in the 6–90 day period is 0.77 for TIA+ compared with 0.43 for TIA. These results imply that, in both cases, treatments for the base diagnoses are more likely to keep the patient out of the ED in the medium run than are treatments for the common differential diagnoses and, in the case of appendicitis, out of the hospital, too. They also suggest the possibility that some base sample patients are being misdiagnosed, which then results in greater expenditures over the medium run.

3. Model and Estimation Framework

We develop a simple model of ED patient treatment, health outcomes, and healthcare expenditures that can apply to any case in which appropriate diagnosis and disposition of the patient are difficult. We then use the model to motivate our estimation framework.

3.1. Model

We model the ED physician, who directs a team of caregivers. Besides the attending physician, the team includes registered nurses, ED technicians, medical residents, care coordinators, respiratory therapists, specialty consultants, transport personnel, social workers, and other personnel. The ED physician's most important functions are the appropriate diagnosis and disposition of the patient. The disposition includes both the treatments performed in the ED and the recommendations for follow-up care. For instance, discharge should be associated with instructions to the patient, such as medications, tests, and scheduled appointments for physician follow-up. Scheduled appointments could be for office visits, hospital-based external clinics, or testing but not for ED revisits. Discharge could also be associated with a direct hospitalization. ED revisits within the 90-day time period may be for the same or a different condition.

Appropriate disposition of the patient relies, in part, on the ability to make an accurate diagnosis. On the one hand, an ED physician who does not adequately recognize symptoms and signs characteristic of life-threatening illnesses will send patients home without adequate follow-up and with the risk of further complications. On the other hand, an ED physician who is overly intensive in treatments causes extra resources to be used without delivering any extra health benefit. Because many diagnostic and therapeutic interventions occur only after discharge from the ED, disposition also depends on the ED physician's ability to convince patients to adhere to follow-up/treatment recommendations.

Consider a patient i who experiences an illness θ_i and then presents at the ED. The patient is treated by one of J attending physicians. Denote the treating physician by j; $j \in \{1, \ldots, J\}$. We assume that there are two possible illnesses, θ^D, the base illness, and θ^B, the differential diagnosis illness. Let $P_i(\theta^B)$ ($P_i(\theta^D)$) denote the probability that the patient has illness θ^B (θ^D). Empirically, the differential diagnosis illness θ^D represents the plus diagnoses. Consistent with the evidence in Table 3, we model θ^D as having a lower severity during the initial ED visit than θ^B.

The symptoms overlap between the two illnesses, implying that accurate diagnosis is difficult. On one extreme, the physician may have complete information about θ_i. On the other extreme, the physician may assume that θ_i is identical across all patients. Because the reality likely lies somewhere in between these two extremes, we impose no restrictions on the physician's information.

We further assume that there are two possible treatments, $T_i \in \{T^D, T^B\}$. The appropriate treatment for θ^B is given by T^B, and the appropriate treatment for θ^D is given by T^D. Treatment T^B is more expensive than T^D, for example, involving referrals for greater follow-up care.

Following treatment, patient i obtains a health stock H_i^*, which we assume is additive in the patient's baseline health endowment, illness shock, and treatment. The baseline health endowment $\overline{H}_i$ is the endowment prior to the onset of the illness shock. The illness shock is given by $\overline{v}(\theta_i) + \varepsilon_i$. We let $\overline{v}(\theta^B) < \overline{v}(\theta^D) < 0$ so that θ^B confers a worse mean health shock than θ^D and both are worse than no illness. The ε_i term is the deviation of the illness shock from the mean for the illness.

Denote $v(\theta_i, T_i)$ to be the value of treatment $T_i \in \{T^B, T^D\}$ when the patient has health status $\theta_i \in \{\theta^B, \theta^D\}$. We allow the value of a treatment to depend on the underlying condition. We assume that (i) $v(\theta^D, T^B) = v(\theta^D, T^D)$ so that there is no extra value in receiving the intensive treatment for θ^D, and (ii) $v(\theta^B, T^B) > v(\theta^B, T^D)$ so that there is a positive value in getting the intensive treatment for θ^B. Thus, there are three possible treatment utilities that patient i can receive. Taken together, the posttreatment health stock is given by

$$H_i^* = \overline{H}_i + \overline{v}(\theta_i) + v(\theta_i, T_i) + \varepsilon_i. \tag{1}$$

The econometrician does not observe θ_i but does observe the consumption of medical services. More specifically, let H_i denote an observable healthcare event occurring after the initial illness shock and care at the ED. For example, H_i could be an indicator for hospitalization, which occurs if $H_i^* < 0$. Although H_i is binary, our empirical work principally considers count variables (i.e., the number of ED revisits) and continuous variables (i.e., hospitalization cost proxies).

Corresponding to our health stock, we also define dollar expenditures, which are a function of the illness shock, treatment, and a residual term:

$$D_i^* = e(\theta_i, T_i) + u_i, \tag{2}$$

where (i) u_i denotes a person-specific cost shock; (ii) the e function provides the deterministic part in the relationship between the illness–treatment pair and costs at the ED; and (iii) D_i^* is measured in dollars and, hence, theoretically observable (unlike the health stock). We take no stand as to the costs resulting from the different treatments. Whereas T^B costs more than T^D in the current ED visit, it is possible that T^D costs more in the long run because of adverse outcomes.

We use our framework and data to evaluate each physician's average contribution to ED resource use during the initial ED visit (which we call practice style) as well as to desired outcomes (which we call skills). In our model, the practice style and skills of an ED physician depend on (a) the physician's ability to distinguish θ^B from θ^D and (b) the physician's ability to determine and then ensure that the patient receives the appropriate treatment, T^B or T^D, respectively. For some of these attributes, more is always better, whereas for others, this may not be true. To understand the treatment effect of the physician skill, define $p_{ij}(T^B \mid \theta^B)$ to be the probability that physician j picks treatment T^B when patient i's true health state is θ^B. Let d_{ij} denote an indicator for whether patient i is assigned to physician j.

Patient i's expected (latent) health stock from the ED visit is then equal to

$$E[H_{ij}^*] = \sum_{j=1}^{J} d_{ij} p_{ij}(T^B \mid \theta^B) P_i(\theta^B)\big[v(\theta^B, T^B) - v(\theta^B, T^D)\big] + \overline{H}_i + P_i(\theta^D)(v(\theta^D, T^D) + \overline{v}(\theta^D)) + P_i(\theta^B)(v(\theta^B, T^D) + \overline{v}(\theta^B)) + \varepsilon_i, \quad (3)$$

where the expectation here is over θ_i and T_i. In (3), the second line is the expected health status of patient i with the T^D treatment (which is invariant to the physician allocation), whereas the first line provides the extra health impact of being treated by physician j and sometimes receiving T^B when suffering from θ^B. Note that the skill of physician j here has two impacts: first, in the physician's diagnosis of the correct condition and, second, in the physician's choice of the appropriate treatment conditional on diagnosis. These two skills together determine $p_{ij}(T^B \mid \theta^B)$. Finally, whereas physician practice style affects outcomes for both illnesses, physician skill only matters when the patient has illness θ^B because treatments do not affect health outcomes with θ^D.

Let $S_{ij}^H \equiv p_{ij}(T^B \mid \theta^B) P_i(\theta^B)\big[v(\theta^B, T^B) - v(\theta^B, T^D)\big]$ denote the treatment effect in terms of expected incremental health benefit from being treated by physician j. Note that S_{ij}^H is a measure of the impact of the skill of physician j on i's health. Using this definition, we can rewrite (3) as

$$E[H_i^*] = \sum_{j=1}^{J} d_{ij} S_{ij}^H + \overline{H}_i + P_i(\theta^D)(v(\theta^D, T^D) + \overline{v}(\theta^D)) + P_i(\theta^B)(v(\theta^B, T^D) + \overline{v}(\theta^B)) + \varepsilon_i. \quad (4)$$

The expectation of D_i^* as

$$E[D_i^*] = \sum_{j=1}^{J} d_{ij} S_{ij}^D + P_i(\theta^D) e(\theta^D, T^D) + P_i(\theta^B) e(\theta^B, T^D) + u_i, \quad (5)$$

where S_{ij}^D is the treatment effect of physician j in terms of extra expenditures to patient i relative to the expenditures associated with always receiving T^D.

3.2. Estimation Framework

Using our model as a basis, we estimate physician treatment effects across different dimensions, notably costs and health outcomes. Focusing on expected health status from (4), we parameterize the observable part of expected health status that is not a function of the patient's assignment to a particular ED physician as

$$x_i \beta^H \equiv \overline{H}_i + P_i(\theta^D)(v(\theta^D, T^D) + \overline{v}(\theta^D)) + P_i(\theta^B)(v(\theta^B, T^D) + \overline{v}(\theta^B)), \quad (6)$$

where x_i are controls for age, gender, weekend, and SES, and β^H is a vector of parameters to estimate.

We assume that S_{ij}^H is the same across patients for a given physician within a particular plus illness. Thus, we replace S_{ij}^H with a physician fixed effect, FE_j^H. With these substitutions, we can replace (4) with

$$E\big[H_i^*\big] = \sum_{j=1}^{J} d_{ij} FE_j^H + x_i \beta^H + \varepsilon_i. \quad (7)$$

Equation (7) leads directly to our empirical analysis: we regress observable markers of health status, H_i, or dollar costs, D_i^*, on physician fixed effects and the x_i controls for patient characteristics.[19] The estimated physician fixed effects indicate the causal impact of being treated by a given physician relative to other physicians at the same ED. Because ED physicians lead a team of caregivers for each patient, we interpret a physician's fixed effect as the physician's relative average treatment effect across the teams that the physician leads.

To examine the relation between different measures of practice style and skills across physicians, we correlate different estimated fixed effects relative to the ED mean for a given physician, also reporting statistical significance for the correlation measures. Our inference uses Monte Carlo methods to account for the fact that our physician fixed effects are estimated. Online Appendix B provides details on these methods. We view these correlations as observational. They represent the population value of the relation between different practice styles among the population of physicians and not a causal relationship. For example, we do not assert that an intensive practice style in one dimension causes a similar practice style in another dimension even if the two correlate positively.

Our interpretation of physician fixed effects is as practice style, skill, or a mix, depending on the particular measure. Spending in the initial ED visit is our measure of practice style: physicians who spend more for identical patients are ones who have a practice style that uses more resources. Following the literature (Anwar and Fang 2012, Chan et al. 2022), we view revisits to an ED within zero to five days and hospitalizations within 6–90 days as unambiguous indicators of worse care. Other outcomes are more nuanced. A positive coefficient on office visits or hospital-based external clinic visits may be an indicator of better care as such a visit might constitute a useful follow-up (i.e., further testing) that prevents future ED visits or hospitalizations but may also reflect unnecessary caution. A positive coefficient on immediate hospitalization is also ambiguous in its effect as these hospitalizations might forestall future and more expensive treatments, prevent death, or otherwise improve patient welfare but may also be unnecessary. Because of the inherent ambiguity of some of these measures, we do not group measures together, but instead present correlations across seven measures for a physician that separately capture the initial spending and future healthcare usage at EDs, offices, external clinics, and hospitalizations.

For the illnesses we study, the physician fixed effects reflect some function of both diagnosis and disposition skills. Because TIA is the hardest to diagnose, we expect that practice style and skill differences reflect diagnosis skills more than disposition skills the most for this illness.

From (7), identification of the FE_j^H parameters here rests on the assignment function of patients to ED physicians, d_{ij}, being mean independent from the residual ε_i. In other words, identification relies on patients being assigned randomly to physicians within an ED, conditional on x_i. This would then result in expected case severity being balanced across physicians. As we discuss in Section 2.2, the assignment of patients to physicians is quasi-random within an ED for hospitals and physicians in our sample.[20]

4. Results

We first report statistics on the physician effects for the three samples. We then consider the correlations between physician effects within samples, across conditions, and across the base and plus samples for a condition. Finally, we examine the robustness of our main results to different outcomes, to different physician inclusion criteria, and to the inclusion/exclusion of patient observable characteristics.

4.1. Variation in Physician Effects

Table 4 reports the difference between the 90th and 10th percentiles at each ED of each estimated fixed effect of physician skill and practice style separately for the angina+, appendicitis+, and TIA+ samples. The table also reports statistical significances for these differences that account for their joint distribution and that we calculate using Monte Carlo methods.

We find large within-ED variation in spending on physician services during the initial visit across all EDs and all illness categories. Take, for example, the 90th to 10th percentile difference in mean spending during the initial ED visit for the angina+ sample of patients. The within-ED differences range from \$7.82 to \$45.79 on an overall mean in-sample spending level of \$53.30 (Table 3). These differences are even larger for the appendicitis+ and TIA+ samples. These results suggest significant variation in practice style across physicians.

With respect to skills, for angina+ and appendicitis+, we find significant differences in ED revisits within five days for most EDs. For TIA+, we find more significant variation in ED revisits in the 6–90 day period than in the 0–5 day period. This difference likely occurs because inappropriate diagnosis or disposition of TIA is more likely to result in complications in the medium run rather than the short run.

We find even more striking variations for inpatient hospital stays. This is true across all EDs and samples. For example, the 90th to 10th percentile difference in mean number of hospital stays in the first five days post-ED visit are significant for six out of seven EDs for the angina+ and appendicitis+ samples and five out of seven for the TIA+ samples. Among the differences that are significant, the 90th to 10th percentile difference ranges from 0.08 to 0.15 stays for the angina+ sample, from 0.16 to 0.72 stays for the appendicitis+ sample, and from 0.19 to 0.31 stays for TIA+ samples. The 90th to 10th percentile differences in mean hospital stays in the following 85 days are similarly large across EDs and illness categories. There is also significant variation in inpatient hospital spending following the initial ED visit across all illness categories.

The 90th to 10th percentile differences in hospital stays and hospital spending are particularly large for the appendicitis+ sample. These large differences likely reflect complications associated with ruptured appendices. Abscesses in the abdomen are notoriously difficult to manage and identify with certainty. Long courses of antibiotics, interventions to drain the abscesses, fistulas (tracks created by infection and leading to other organs), and other complications can lead to long (and, hence, expensive) hospitalizations.[21]

4.2. Relationships Between Practice Styles and Skills Within Samples

Table 5 reports the correlations between the physician fixed effects—relative to the mean effect of physicians at their ED—for different measures of practice styles and skills separately for our three plus samples. We also report the statistical significance for each reported correlation.

Our most striking result is that physicians with higher mean spending in the initial ED visit—which we interpret as an intensive practice style—have worse skills in keeping patients out of the ED for two of the three samples. Specifically, the correlation between the fixed effect for initial ED spending and the fixed effect for ED revisits in zero to five days is 0.22 for angina+ and 0.48 for appendicitis+; both correlations are statistically

Table 4. Estimated 90–10 Percentile in Physician Fixed Effects

Dependent variable	ED 4	ED 7	ED 8	ED 15	ED 1	ED 9	ED 17
Angina+ sample							
Initial ED $	7.82***	16.07***	17.83***	18.09***	45.79***	18.27***	12.43*
ED 0–5 days	0.26***	0.14**	0.18***	0.18***	0.18***	0.11	0.15*
ED 6–90 days	0.18	0.58*	0.30**	0.21	0.72***	0.33	0.42
External 0–30 days	0.23	0.65***	0.31**	0.22*	0.37**	0.38**	0.44**
Office 0–30 days	0.24**	0.26*	0.21**	0.18*	0.19	0.21	0.40***
Hospital stay 0–5 days	0.15***	0.10**	0.07	0.08**	0.09**	0.12**	0.11**
Hospital stay 6–90 days	0.14**	0.18*	0.09**	0.08	0.17***	0.11	0.11
Hospital $ 0–5 days	2,003*	1,892	1,486*	1,434	3,003***	2,774**	1,817
Hospital $ 6–90 days	1,917	3,471**	2,266	1,734	2,272	1,567	1,724
Appendicitis+ sample							
Initial ED $	38.31	35.03*	13.10**	16.64**	58.91***	20.91**	12.72
ED 0–5 days	0.73	0.52**	0.25***	0.23**	0.66***	0.36**	0.34
ED 6–90 days	1.17	2.84***	0.67***	0.28	0.65	0.92*	0.63
External 0–30 days	1.19	0.89	0.63**	0.67***	0.85	1.24***	1.16**
Office 0–30 days	4.06***	0.78*	0.26	0.45	0.76**	0.58**	0.63**
Hospital stay 0–5 days	0.50	0.72***	0.16*	0.27***	0.35**	0.25**	0.39***
Hospital stay 6–90 days	0.54	0.77***	0.18	0.17	0.31*	0.22	0.29
Hospital $ 0–5 days	5,513	11,581***	1,218	3,169**	5,138**	2,361	3,470
Hospital $ 6–90 days	14,169***	12,532***	2,826**	1,508	3,135	1,348	3,233
TIA+ sample							
Dependent variable	ED 4	ED 7	ED 8	ED 15	ED 1	ED 9	ED 17
Initial ED $	17.99	19.64	18.08**	23.78*	46.34***	28.29***	23.15
ED 0–5 days	0.31	0.31*	0.17	0.23	0.44**	0.28	0.34
ED 6–90 days	0.66	1.79***	0.58	0.78	1.08	1.12*	0.86
External 0–30 days	0.20	0.55	0.39	0.44	0.82**	0.98*	1.56***
Office 0–30 days	0.42	0.20	0.35*	0.56	0.37	0.57	0.52
Hospital stay 0–5 days	0.18	0.19	0.24***	0.22**	0.19	0.24**	0.31*
Hospital stay 6–90 days	0.18	0.41***	0.11	0.16	0.24	0.39**	0.37**
Hospital $ 0–5 days	3,420	6,369*	6,144***	4,893*	4,549*	5,337	6,123*
Hospital $ 6–90 days	2,949	3,910	1,886	2,760	3,958*	3,053	4,133

Note. We order EDs in the table by the number of ED physicians per ED.
***Significance at the 1% level, **5% level, and *10% level.

significant. The relationship is also large and significant when considering revisits to EDs during the 6–90 day period for appendicitis+ (with a correlation of 0.37). The positive relationships may reflect uncertainty about appropriate diagnosis or disposition. That is, physicians who face greater uncertainty (potentially because of lower skills) may also spend more on average during the initial ED visit. They may also be less likely to accurately diagnose appendicitis and also more likely to encourage their patients to return to the ED even when not necessary.

Our results on the relation between the initial ED spending and hospital stays are more mixed. On the one hand, physicians with higher mean spending in the initial ED visit have fewer hospital stays during the 6–90 day period for the angina+ sample with a correlation of –0.18. On the other hand, the correlations between initial ED spending and hospital stays during the first 5 and subsequent 85 days are positive and significant (0.26 and 0.21, respectively) for the appendicitis+ sample.[22] The difference between the correlations for angina+ and appendicitis+ may reflect the fundamental difference in the pathophysiology of the two conditions. Appendicitis is an acute event that typically occurs in an otherwise normal patient and can be completely resolved with appropriate management, often over hours to multiple days (at most). More ED spending may reflect less certainty as to the diagnosis, a less clear disposition plan, and more subsequent visits to clarify the underlying condition. Physicians who spend more for angina may request more tests that help with an appropriate disposition, such as a 64-slice CT coronary angiography, myocardial perfusion SPECT, stress echocardiogram, and coronary arteriogram, which may prevent future hospitalizations (as well as cardiac cripples and deaths).[23] The extra ED visits that correlate with the extra spending for angina+ may also be more likely to distinguish severe from less-severe conditions because angina symptoms are more likely to remit and relapse, thereby lowering the need for hospitalizations.

In addition, physicians who have higher mean spending in the initial ED visit have more subsequent hospital-based external clinic visits for the appendicitis+ (correlation of 0.26) and angina+ (correlation of 0.15) samples and more subsequent physician office visits (correlation of 0.28) for the appendicitis+ sample. These

Table 5. Correlation in Physician Fixed Effects Within Samples

Sample/fixed effect	Initial ED $	ED 0–5 days	ED 6–90 days	External 0–30 days	Office 0–30 days	Hospital 0–5 days	Hospital 6–90 days
Angina+ sample							
Initial ED $	1						
ED 0–5 days	0.22***	1					
ED 6–90 days	−0.11	0.08	1				
External 0–30 days	0.15*	0.10	−0.06	1			
Office 0–30 days	0.04	−0.19**	0.21**	−0.33***	1		
Hospital stay 0–5 days	0.11	−0.01	−0.07	0.40***	−0.17*	1	
Hospital stay 6–90 days	−0.18**	0.00	0.35***	−0.05	−0.10	0.22**	1
Appendicitis+ sample							
Initial ED $	1						
ED 0–5 days	0.48***	1					
ED 6–90 days	0.37***	0.26**	1				
External 0–30 days	0.26***	0.29***	0.19*	1			
Office 0–30 days	0.28*	0.26	0.21	0.12	1		
Hospital stay 0–5 days	0.26**	0.07	0.08	0.14	−0.21	1	
Hospital stay 6–90 days	0.21*	0.06	0.53***	0.09	−0.16	0.34***	1
TIA+ sample							
Initial ED $	1						
ED 0–5 days	0.14	1					
ED 6–90 days	0.12	−0.06	1				
External 0–30 days	0.14	0.20*	−0.01	1			
Office 0–30 days	0.01	−0.18*	0.05	0.08	1		
Hospital stay 0–5 days	0.02	0.07	0.15	0.28***	−0.09	1	
Hospital stay 6–90	0.06	−0.05	0.27***	0.22**	−0.09	0.15	1

***Significance at the 1% level, **5% level, and *10% level.

results may be due to office and hospital-based clinic visits complementing spending in the initial ED visits. Overall, our findings suggest that treatment by physicians who spend more is good for angina+ patients in the medium run but bad for appendicitis+ patients with no significant effect for TIA+.

We also consider the correlation of physician fixed effects measured across different outcomes in order to understand whether there is substitutability or complementarity across different healthcare venues. For both appendicitis+ and TIA+, physicians who have more ED revisits during the first five days after the initial visit also have more hospital-based external clinic visits in the first 30 days, possibly because ED revisits in the first five days are complements to hospital-based external visits for these samples. In contrast, physicians who have more ED revisits in the first five days also have fewer office visits during the first 30 days for the angina+ (–0.19) and TIA+ (–0.18) samples. This suggests that, as expected, office visits substitute for ED revisits. For both angina+ and TIA+, we find a positive correlation between hospitalizations during the first five days and external clinic visits in the first 30 days. This complementarity between the two services is probably because of patients who are hospitalized being referred to posthospitalization care.

Finally, for all three samples, physicians with a greater number of ED revisits in the 6–90 day period also have a greater number of hospital stays. These results suggest that ED revisits within the 6–90 days are complements to hospitalizations. A likely cause of these effects is that the threshold for hospitalization may be lower in association with an ED visit or a hospital-based external clinic visit than with an office visit because providers in the ED or external clinic have a more direct and logistically simpler connection with in-hospital care processes, in part because they have immediate access to laboratory and other ancillary data.

4.3. Relationships Between Practice Styles and Skills Across Conditions

Table 6 reports correlations in practice styles and skills across the angina+ and appendicitis+ samples, across the angina+ and TIA+ samples, and across the appendicitis+ and the TIA+ samples. We find a strong and statistically significant correlation of practice styles across the three illness pairs, ranging from 0.46 to 0.62 across the three pairs. In other words, physicians who spend more on average for one illness also spend more on average when treating the other illnesses.

The diagonals in Table 6 also reveal positive correlations across illness categories for several skills. Specifically, we find positive and statistically significant correlations for ED revisits and hospitalizations during the first five days from the initial ED visit and hospitalizations in the 6–90 day period after the initial ED visit.

Table 6. Correlation in Physician Fixed Effects Across Conditions

Sample/fixed effect	Initial ED $	ED 0–5 days	ED 6–90 days	External 0–30 days	Office 0–30 days	Hospital 0–5 days	Hospital 6–90 days
	Angina+						
Appendicitis+							
Initial ED $	0.59***	0.24***	−0.10	0.01	0.15*	−0.08	−0.22**
ED 0–5 days	0.13*	0.36***	0.10	−0.15	−0.02	−0.28***	−0.11
ED 6–90 days	0.04	−0.05	0.01	−0.09	0.14	−0.09	−0.18*
External 0–30 days	0.07	0.02	0.20**	−0.08	0.18*	−0.06	−0.04
Office 0–30 days	0.13**	0.32***	−0.04	0.12	−0.15*	−0.02	−0.15*
Hospital stay 0–5 days	0.18**	−0.02	0.03	−0.07	0.10	0.09	0.01
Hospital stay 6–90 days	0.11	−0.14	−0.03	−0.10	0.09	−0.03	0.00
	TIA+						
Angina+							
Initial ED $	0.62***	0.35***	−0.00	0.08	−0.04	0.07	0.13
ED 0–5 days	−0.01	0.23**	−0.21**	0.05	−0.16	−0.12	0.04
ED 6–90 days	−0.07	−0.24***	−0.05	0.07	−0.10	0.03	0.07
External 0–30 days	−0.09	0.12	0.14	−0.01	−0.19*	0.07	0.10
Office 0–30 days	0.00	0.08	0.07	0.07	0.09	0.04	−0.00
Hospital stay 0–5 days	−0.17**	−0.04	0.15	−0.05	−0.03	0.01	0.23**
Hospital stay 6–90 days	−0.10	0.00	0.21**	−0.04	0.09	−0.08	0.26***
	Appendicitis+						
TIA+							
Initial ED $	0.46***	0.15*	0.13	0.10	0.16**	0.23**	0.10
ED 0–5 days	−0.01	0.18*	0.11	−0.03	0.17*	−0.10	−0.15
ED 6–90 days	0.04	0.06	0.09	−0.09	−0.06	0.01	0.15*
External 0–30 days	−0.11	0.18**	0.04	0.01	−0.12	0.12	−0.08
Office 0–30 days	0.07	0.06	−0.01	−0.08	0.02	0.09	−0.02
Hospital stay 0–5 days	−0.15*	−0.03	−0.07	−0.08	−0.06	0.16*	−0.18*
Hospital stay 6–90 days	0.01	−0.08	0.09	0.00	−0.18**	0.09	0.21**

***Significance at the 1% level, **5% level, and *10% level.

4.4. Relationships Between Practice Styles and Skills Across Base and Plus Samples

Table 7 provides correlations between the base and plus samples for our three illnesses. The plus sample fixed effects reflect practice styles and skills for diagnosis and disposition, whereas the base sample fixed effects reflect practice styles and skills for disposition conditional on diagnosis. We find a strong positive correlation between the practice styles—measured as spending in the initial ED visit—across the three paired samples, ranging from 0.37 to 0.44. As shown in the diagonal elements, we also find several positive correlations for fixed effects for future healthcare consumption across the plus and base samples.

Physicians who are associated with more hospital stays within the first five days when treating the angina+, appendicitis+, and TIA+ samples are also associated with more hospitalizations during the same period when treating patients specifically diagnosed with angina, appendicitis, and TIA (with positive correlation coefficients of 0.25, 0.38, and 0.28, respectively). These positive correlations suggest that, as expected, both diagnosis and disposition skills contribute to healthcare outcomes. The illness with the fewest statistically significant positive correlations is TIA. Because it is the hardest of the three illnesses to diagnose, it is not surprising that these physician fixed effects measured on the base sample are the most different than on the plus sample.

4.5. Robustness of Results to Sample of Physicians and Patient Covariates

Online Appendix A shows the robustness of our main correlation results, Tables 5–7, to the sample of physicians and patient covariates. First, Online Tables A6–A8 present results in which we drop from each illness physicians who are outliers in terms of the fraction of patients that they treat from that illness.[24] These restrictions leave us with 78 unique physician–ED pairs compared with 90 for our base analysis. For this more restrictive sample of physicians, we find very similar results to those from the initial sample of physicians. Second, we expand our sample to include all ED physicians in our estimation sample of seven EDs, leaving us with 103 to 106 physician–ED pairs. Online Tables A9–A11 show that these results are virtually indistinguishable from the base results.

Third, we examine our correlation results using the original sample but excluding our controls for age, gender, and SES. Online Tables A12–A14 present these results. Finally, we examine our correlation results using the original sample—which includes our controls for age,

Table 7. Correlation in Physician Fixed Effects Across Base and Plus Samples

Sample/fixed effect	Initial ED $	ED 0–5 days	ED 6–90 days	External 0–30 days	Office 0–30 days	Hospital 0–5 days	Hospital 6–90 days
	Angina						
Angina+							
Initial ED $	0.37***	−0.03	−0.06	0.06	0.20***	−0.05	−0.19***
ED 0–5 days	0.01	0.02	−0.05	−0.02	0.13	−0.19**	−0.12
ED 6–90 days	0.05	−0.05	0.31***	0.10	0.09	−0.22**	0.13
External 0–30 days	−0.05	−0.28***	0.04	0.28***	0.13	−0.18**	0.03
Office 0–30 days	0.04	0.17**	0.01	−0.15	0.24**	0.06	−0.10
Hospital stay 0–5 days	−0.02	−0.19**	0.14	0.17*	−0.03	0.25***	0.26***
Hospital stay 6–90 days	−0.12	0.03	−0.02	−0.16	−0.37***	−0.07	0.16*
	Appendicitis						
Appendicitis+							
Initial ED $	0.41***	0.21***	0.09	0.01	0.08	0.00	0.01
ED 0–5 days	0.16**	0.43***	−0.14	0.08	0.23***	−0.21***	0.06
ED 6–90 days	0.02	−0.07	0.32***	0.08	0.21***	−0.13	0.14
External 0–30 days	−0.04	0.17***	0.09	0.36***	0.26***	−0.17**	0.27***
Office 0–30 days	0.02	0.08	0.38***	−0.01	0.31***	−0.07	−0.12
Hospital stay 0–5 days	0.00	−0.12	−0.11	−0.01	−0.08	0.38***	−0.03
Hospital stay 6–90 days	−0.18**	−0.25***	0.02	−0.02	0.13	0.06	0.18
	TIA						
TIA+							
Initial ED $	0.44***	−0.18*	−0.12	−0.16	0.06	−0.01	0.11
ED 0–5 days	0.01	0.26**	−0.07	0.04	−0.10	−0.23**	−0.15
ED 6–90 days	−0.20	−0.12	0.13	0.02	0.01	0.30**	−0.07
External 0–30 days	−0.11	0.10	0.20**	0.29***	−0.14	−0.04	0.02
Office 0–30 days	−0.19	−0.17	0.16	0.14	0.19	−0.10	−0.09
Hospital stay 0–5 days	−0.11	−0.23**	0.14	0.00	−0.09	0.28***	0.11
Hospital stay 6–90 days	0.12	−0.18*	0.24**	0.09	0.11	−0.08	0.06

***Significance at the 1% level, **5% level, and *10% level.

gender, weekend, and SES—and adding controls for the number of secondary diagnoses. Online Tables A15–A17 presents these results. For both cases, we find that the correlations across physician fixed effects are very similar to our base correlations. In the spirit of Altonji et al. (2005), the similarity of these correlations to our base results provides evidence that selection based on unobservables is likely not very important.[25]

5. Conclusions

This paper uses detailed data from emergency departments in Montreal, Canada, to understand how ED physicians contribute to healthcare operations. A central advantage of our data is that we observe the assignment of patients to ED physicians and also observe future interactions with the medical system within 90 days of the initial ED visit. We provide a number of pieces of evidence that show that the assignment of ED physicians in Montreal, conditional on choice of ED, is close to random. We develop a theoretical model under which physicians, who serve as leaders of a team of personnel that cares for each patient, have potentially different skills and practice styles. These differences together affect the diagnosis and disposition of patients. Our outcome measures quantify appropriate diagnosis and disposition.

Using our data, we estimate the effect of being treated by different ED physicians on these outcomes. We perform this estimation for three samples of patients: those who are diagnosed with conditions that might indicate angina, appendicitis, and a transient ischemic attack. Because the differential diagnoses are easier to diagnose, using the set of patients who *might* have one of these conditions lowers the possibility that inaccurate physician diagnosis could affect our results.

We find that physicians vary in both their average contribution to ED costs during the initial ED visit—that is, in practice style—and their future outcomes—that is, in skills. Importantly, for angina+ and appendicitis+, we find that physicians with more intensive practice styles have worse skills in terms of being associated with more ED revisits. We also find that physicians who spend more for one condition are likely to spend more for the other conditions that we consider. Moreover, skills at keeping patients out of the ED also correlate positively across the three conditions that we consider though are far from perfect. Taken together, these results suggest that practice styles and skills correlate across illnesses. Considering the correlations between physician outcomes for the plus and base conditions, our results suggest that both disposition and diagnosis play a large part in contributing to future outcomes.

Our study design includes several limitations noted in Section 2.1. Nonetheless, we offer three main implications for operations management from our results. First, the positive correlation between low costs and high quality suggests that reducing healthcare costs and improving quality can be complementary goals and that more spending on patients does not guarantee better outcomes. It also suggests that physicians with intensive practice styles and low skills could benefit from informational transfer and training.

Second, one cannot simply investigate physician practice style or skills for one condition to perfectly identify overall physician characteristics. This reason for this is that, even though physicians who perform well on one condition on average perform well on other conditions, there is still substantial heterogeneity in practice style and skills across conditions.

Finally, the current system, which treats most ED physicians as substitutes, may not yield the best outcomes for patients. In an ideal world, one would have several ED physicians practicing at the same time and have an allocation mechanism by which each patient is allocated to the ED physician with the best skills for treating the patient's particular disease.

Acknowledgments

The authors thank seminar participants at McGill University, Weill-Cornell Medical School, the University of Chicago, the University of Chile, the University of Connecticut, the University of Illinois Chicago, the University of Illinois Urbana-Champaign, and the University of Washington as well as conference participants at the American Society for Health Care Engineering Conference (Los Angeles, CA), the Highland Health Economics Symposium (Lochailort, Scotland), SIOE (Montreal, Canada), and the Workshop on Healthcare and Industrial Organization (Santiago, Chile). The authors thank Marcelo Olivares and Hummy Song for helpful comments; Michelle Houde and José Pérez at the Département de santé publique de Montréal for data assistance and extraction; Andreas Krull, M.D., and Marie-Josée Ouimet, M.D., for providing institutional details about Montreal emergency departments; and Sebastian Fleitas, Jianjing Lin, Kelli Marquardt, and Yujia Peng for excellent research assistance. Authors are listed alphabetically as study is an equal collaboration between all authors.

Endnotes

[1] Our focus on one metropolitan area is supported by Tsugawa et al. (2017), who find that spending variation across physicians within a hospital is larger than across hospitals, and by Epstein and Nicholson (2009), who find similar results when considering within-market to across-market variation. We focus on Canada, which has a single-payer health system. The single-payer system implies that we do not need to consider the role of insurance coverage in affecting healthcare demand and supply decisions. Physicians in Canada have similar decision-making autonomy to U.S. physicians.

[2] A differential diagnosis is the list of diagnoses that are consistent with the presenting symptoms, signs, and ancillary data. Symptoms are the history reported by the patient, signs are the findings on the physical exams performed by the provider, and ancillary data include laboratory tests and imaging.

[3] Anwar and Fang (2012) interpret ED "bounce back" rates within 72 hours as a negative patient outcome, noting that 18%–30% return because of a "possible medical error" (p. 4). Chan et al. (2022) use a diagnosis of pneumonia within 10 days (which would typically result in hospitalization) without a current diagnosis of pneumonia as a negative measure of physician skill. In some robustness measures, they restrict the negative outcome to a diagnosis of pneumonia for patients who would be likely to be hospitalized for pneumonia.

[4] Although a privately funded sector exists in Quebec, it is very small and generally deals with noncovered services, such as cosmetic surgery. This market remains insignificant as physicians who bill for any services privately must completely opt out of the public sector. An exception to the opt out rule is for imaging facilities.

[5] Hospital-based external clinics are specialty departments that are housed within the hospital from radiology to oncology to psychiatry. Similar specialty departments also exist in office settings.

[6] The data include unique MD identifiers for each diagnosis provided and procedure administered to the patient.

[7] The data report two bins for patients less than five years old and a single bin for patients 85 years or older.

[8] There is no way for us to identify which ED in our data uses the purely random order as we were provided numerical, randomized ED identifiers because of privacy issues.

[9] This excludes patients seen at children's hospitals with EDs.

[10] Other patient characteristics, such as the probability of a certain diagnosis, may be coded in different ways by different ED physicians and are, thus, inappropriate for the purposes of testing random assignment of patients.

[11] EDs 4 and 9 already satisfy this condition with all their physicians included, so we do not remove physicians from these samples.

[12] Online Table A1 provides further tests, analogous to Table 1, panel B, but using our two SES measures. From the 14 EDs in our restricted sample, three fail the randomization test for each measure of SES. Unlike age and gender, it is difficult for the triage nurse to assign patients to a heavy or light stream based on SES. Thus, these failures of randomization are likely caused by small variations in shift times.

[13] Section 4.5 provides robustness results.

[14] We do not consider obstetrical emergencies because in almost all cases, an ob-gyn physician is engaged.

[15] There are many base diagnoses that fit the other criteria but are quite uncommon, such as meningococcal meningitis, herpes encephalitis, or streptococcal toxic shock.

[16] With pneumonia, the extent of lung involvement, the potential for a catastrophic outcome, and the need for treatment and other factors vary so widely that identifying the correct diagnosis and disposition is far less precise.

[17] Not reported in the table, these represent 89 physicians, one of whom treated patients at two EDs.

[18] We combine the 6–30 and 31–90 day categories for ED revisits and hospital stays to separate the visits that directly relate to the initial visit from other ones.

[19] An alternative would be to use an empirical Bayes shrinkage estimator that would reduce mean-squared error by attenuating physician fixed effects toward their mean (Morris 1983). We attempted to implement an estimator based on Morris (1983), but it could not be implemented for our sample because of numerical issues.

[20] We do not assume that the distribution of ε_i is the same across EDs. This is important because patients who are more severely ill

based on unobservable factors may travel further to visit an ED that is higher quality or may live nearer certain hospitals.

[21] Online Table A2 presents analogous results to Table 4 but comparing the 75th and 25th percentiles of physicians within an ED. The results are similar though with fewer significant differences.

[22] Online Tables A3–A5 add hospital spending in addition to hospital stays as outcome variables to our main results in Sections 4.2–4.4. We find fewer significant results with spending than with stays.

[23] Consistent with our finding, Currie et al. (2016) find that cardiologists who use invasive procedures for heart attacks more frequently have better outcomes.

[24] This specification drops ED 4 altogether as it appears to serve a very different patient mix.

[25] Intuitively, Altonji et al. (2005), condition 3, shows that, if selection based on unobservables is no greater than selection based on observables, then the similarity in results across our specifications is informative. We believe that our three specifications include the most likely factors that are observable to triage nurses—comorbidities, age, gender, and SES—and, hence, that the Altonji et al. (2005) condition is likely to hold.

References

Abaluck J, Agha L, Kabrhel C, Raja A, Venkatesh A (2016) The determinants of productivity in medical testing: Intensity and allocation of care. *Amer. Econom. Rev.* 106(2):3730–3764.

Altonji JG, Elder TE, Taber CR (2005) Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *J. Political Econom.* 113(1):151–184.

Anwar S, Fang H (2012) Testing for the role of prejudice in emergency departments using bounceback rates. *B.E. J. Econom. Anal. Policy* 13(3):1–49.

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

Bhangu A, Søreide K, Di Saverio S, Assarsson J, Drake F (2015) Acute appendicitis: Modern understanding of pathogenesis, diagnosis, and management. *Lancet* 386(10000):1278–1287.

Chan DC (2016) Teamwork and moral hazard: Evidence from the emergency department. *J. Political Econom.* 124(3):734–770.

Chan DC, Gentzkow M, Yu C (2022) Selection with skills: Evidence from radiologists. *Quart. J. Econom.* 137(2):729–783.

Chang CY, Obermeyer Z (2020) Association of clinical characteristics with variation in emergency physician preferences for patients. *JAMA Network Open* 3(1):e1919607.

Currie J, MacLeod WB (2017) Diagnosing expertise: Human capital, decision making, and performance among physicians. *J. Labor Econom.* 35(1):1–43.

Currie J, MacLeod WB (2020) Understanding doctor decision making: The case of depression treatment. *Econometrica* 88(3):847–878.

Currie J, MacLeod WB, Van Parys J (2016) Provider practice style and patient health outcomes: The case of heart attacks. *J. Health Econom.* 47:64–80.

Dahlstrand A (2022) Defying distance? The provision of services in the digital age. Working paper, London School of Economics, UK.

Doyle J (2011) Returns to local-area healthcare spending: Evidence from health shocks to patients far from home. *Amer. Econom. J. Appl. Econom.* 3(3):221–243.

Doyle J, Wagner T, Ewer S (2010) Returns to physician human capital: Analyzing patients randomized to physician teams. *J. Health Econom.* 29:866–882.

Doyle J, Graves J, Gruber J, Kleiner S (2015) Measuring returns to hospital care: Evidence from ambulance referral patterns. *J. Political Econom.* 123(1):170–214.

Epstein AJ, Nicholson S (2009) The formation and evolution of physician treatment styles: An application to cesarean sections. *J. Health Econom.* 28(6):1126–1140.

Hastings S, Whitson H, Purser J, Sloane R, Johnson K (2009) Emergency department discharge diagnoses and adverse health outcomes among older adults. *JAMA* 57(10):1856–1861.

Hollander JE, Chase M (2016) Evaluation of the adult with chest pain in the emergency department. Accessed September 29, 2022, https://www.uptodate.com/contents/evaluation-of-the-adult-with-chest-pain-in-the-emergency-department.

Johns Hopkins Medicine (2020) GERD diet: Foods that help with acid reflux (heartburn). Accessed June 3, 2020, https://www.hopkinsmedicine.org/health/wellness-and-prevention/gerd-diet-foods-that-help-with-acid-reflux-heartburn.

Kachalia A, Gandhi T, Puopolo A, Yoon C, Thomas E, Griffey R, Brennan T, Studdert D (2007) Missed and delayed diagnoses in the emergency department: A study of closed malpractice claims from 4 liability insurers. *Ann. Emergency Medicine* 49(2):196–205.

KC DS (2014) Does multitasking improve performance? Evidence from the emergency department. *Manufacturing Service Oper. Management* 16(2):168–183.

KC DS, Scholtes S, Terwiesch C (2019) Empirical research in healthcare operations: Past research, present understanding, and future opportunities. *Manufacturing Service Oper. Management* 22(1):73–83.

Kessler C, Thomas K (2009) An examination of economic outcomes associated with misdiagnosis or undertreatment of TIA. *Amer. J. Managed Care* 15(6):S170–S176.

Kindermann D, McCarthy M, Ding R, Frohna W, Hansen J, Maloy K, Milzman D, Pines J (2014) Emergency department variation in utilization and diagnostic yield of advanced radiography in diagnosis of pulmonary embolus. *J. Emergency Medicine* 46(6):791–799.

Martin RF (2017) Acute appendicitis in adults: Clinical manifestations and differential diagnosis. https://www.uptodate.com/contents/acute-appendicitis-in-adults-clinical-manifestations-and-differential-diagnosis.

Mayo Clinic (2020) Gastritis—Diagnosis & treatment. Accessed June 3, 2020, https://www.mayoclinic.org/diseases-conditions/gastritis/diagnosis-treatment/drc-20355813.

McSweeney J, Cody M, O'Sullivan P, Elberson K, Moser D, Garvin B (2003) Women's early warning symptoms of acute myocardial infarction. *Circulation* 108(21):2619–2623.

Morris CN (1983) Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* 78(381):47–55.

Mullainathan S, Obermeyer Z (2022) Diagnosing physician error: A machine learning approach to low-value healthcare. *Quart. J. Econom.* 137(2):679–727.

Nadarajan V, Perry R, Johnson J, Werring D (2014) Transient ischaemic attacks: Mimics and chameleons. *Practical Neurology* 14(1):23–31.

Pines J, Hollander J, Isserman J, Chen E, Dean A, Shofer F, Mills A (2009) The association between physician risk tolerance and imaging use in abdominal pain. *Amer. J. Emergency Medicine* 27(5):552–577.

Pope J, Edlow J (2012) Avoiding misdiagnosis in patients with neurological emergencies. *Emergency Medicine Internat.* 2012: 949275.

Pope J, Aufderheide TP, Ruthazer R, Woolard RH, Feldman JA, Beshansky JR, Griffith JL, Selker HP (2000) Missed diagnoses of acute cardiac ischemia in the emergency department. *New England J. Medicine* 342(16):1163–1170.

Proulx AM, Zryd TW (2009) Costochondritis: Diagnosis and treatment. *Amer. Family Physician* 80(6):617–620.

Rinfret S, Kennedy WA, Lachaine J, Lemay A, Rodés-Cabau J, Cohen DJ, Costerousse O, Bertrand OF (2010) Economic impact of same-day home discharge after uncomplicated transradial percutaneous coronary intervention and bolus-only abciximab regimen. *JACC Cardiovascular Interventions* 3(10):1011–1019.

Rivest P, Bédard L, Valiquette L, Mills E, Lebel MH, Lavoie G, Carsley J (2001) Severe complications associated with varicella: Province of Quebec, April 1994 to March 1996. *Canadian J. Infectious Diseases* 12(1):21–26.

Rothrock S, Pagane J (2000) Acute appendicitis in children: Emergency department diagnosis and management. *Ann. Emergency Medicine* 36(1):39–51.

Rothwell P, Giles M, Chandratheva A, Marquardt L, Geraghty O, Redgrave J, Lovelock C, et al. (2007) Effect of urgent treatment of transient ischaemic attack and minor stroke on early recurrent stroke (EXPRESS study): A prospective population-based sequential comparison. *Lancet* 370(9596): 1432–1442.

Seetahal SA, Bolorunduro OB, Sookdeo TC, Oyetunji TA, Greene WR, Frederick W, Cornwell EE, Chang DC, Siram SM (2011) Negative appendectomy: A 10-year review of a nationally representative sample. *Amer. J. Surgery* 201(4):433–437.

Shoff C, Caines K, Pines JM (2019) Geographic variation in predictors of ED admission rates in US Medicare fee-for-service beneficiaries. *Amer. J. Emergency Medicine* 37(6):1078–1084.

Silver D (2021) Haste or waste? Peer pressure and productivity in the emergency department. *Rev. Econom. Stud.* 88(3): 1385–1417.

Song H, Tucker AL, Murrell KL (2015) The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Sci.* 61(12):3032–3053.

Song H, Tucker AL, Murrell KL, Vinson DR (2018) Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices. *Management Sci.* 64(6):2628–2649.

Strumpf E, Ammi M, Diop M, Fiset-Laniel J, Tousignant P (2017) The impact of team-based primary care on healthcare services utilization and costs: Quebec's family medicine groups. *J. Health Econom.* 55:76–94.

Taubman SL, Allen HL, Wright BJ, Baicker K, Finkelstein AN (2014) Medicaid increases emergency-department use: Evidence from Oregon's health insurance experiment. *Sci.* 17(17): 263–268.

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

Van Parys J (2016) Variation in physician practice styles within and across emergency departments. *PLoS One* 11(8):e0159882.

WebMD (2020) What's causing my chest pain? Accessed September 29, 2022, https://www.webmd.com/pain-management/guide/whats-causing-my-chest-pain#1.