Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments *

Gautam Gowrisankaran, Ph.D.† Keith A. Joiner, M.D. MPH‡

Pierre Thomas Léger, Ph.D.§

July 8, 2020

Abstract

We study healthcare operations of emergency departments (EDs) by examining the practice style and skills of ED physician-led teams. Our data include all residents of Montreal, Canada with an initial ED visit in Montreal during a 9 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 physician-led teams to estimate the skill and practice style of each team. We then evaluate the variation across teams in their practice style and skills and the correlations between different measures of skill and practice style. We find significant variation across physician-led teams in their practice styles and skills. We also find that physician-led teams with costly practice styles often have worse outcomes in terms of more ED revisits and more hospitalizations. Finally, the practice styles and skills of physician-led teams correlate positively across the three conditions that we consider.

Keywords: physician quality, emergency departments, quasi-random variation.

JEL Classification: I12

*We thank seminar participants at the University of Chicago, the University of Illinois at Urbana-Champaign, the University of Connecticut, the University of Illinois at Chicago, the University of Washington, McGill University, Weill-Cornell Medical School, the University of Chile, conference participants at ASHE (Los Angeles, CA), HHES (Lochailort, Scotland), WHIO (Santiago, Chile) and SIOE (Montreal, Canada), and Marcelo Olivares 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. Gowrisankaran acknowledges funding from the Center for Management Innovations in Healthcare at the University of Arizona.

†University of Arizona, HEC Montréal and NBER, gowrisankaran@eller.arizona.edu

‡University of Arizona, kjoiner@medadmin.arizona.edu

§University of Illinois at Chicago, ptleger@uic.edu

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). Due to the complexity and heterogeneity of the medical conditions of the patients they treat, EDs are places where the operational issues that typically present in healthcare are particularly important. Moreover, like much of healthcare, ED production for a patient depends on a team of individuals, which includes the ED physician and many other personnel. For these reasons, understanding ED production has broad application to the operations management of healthcare and of team production.

This study evaluates the variations in and relations between the costs of treatments and follow-up visits with the medical system across physician-led teams for similar sets of patients. Specifically, we consider: To what extent is there variation in the treatment costs across different teams, when presented with similar clinical issues? Do teams which spend more have better outcomes, such as fewer hospitalizations over the medium-run? Do teams that perform better in treating some health conditions also perform better in treating other conditions?

We make use of a unique dataset on ED patients in the Montreal, Canada metropolitan area for a nine month period in 2006.[1] The starting point for any record in our data is a visit by a patient residing in the Montreal region to an ED in the region. Our data record the initial assignment to the ED physician, who effectively is the leader of the team of medical providers. They also record every encounter with the provincial health system in the 90 days after the initial ED visit. We use the follow-up visits to identify the impact of the ED physician on healthcare utilization over the medium run.

An important attribute of our study is that the initial assignment to ED physician in our setting is quasi-random, given the choice of ED, or equivalently, hospital. This is because at EDs in Montreal (and in most North American hospitals), cases in an ED are assigned on a quasi-random basis. We

[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. Physicians in Canada have similar decision-making autonomy to U.S. physicians.

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 is important 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, i.e., researchers have focused 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 (Rothrock and Pagane, 2000; Pope and Edlow, 2012; Pope et al., 2000). Finally, EDs are a prominent place of treatment for a number of widely-occurring diseases which may be difficult to diagnose and for which appropriate and timely care is particularly important.

To motivate our empirical work, we propose a simple model of the role of the ED physician team. The ED—or attending—physician is the leader of a team of caregivers that provides care to a patient. 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. These other team members perform tasks such as inserting intravenous lines, drawing blood for laboratory analysis, conducting medical imaging, and providing respiratory support. Our analysis considers the production of the team as a whole. In what follows, for brevity, we refer to this team by its leader, the ED physician.

The ED physician has two central roles: diagnosis and disposition. She first attempts to diagnose the illness of the patient. She then chooses patient disposition, which involves both the ED treatments and the recommendations for follow-up care. For both of these steps, she makes decisions about resource use, e.g., laboratory tests, imaging studies, medications, specialty consultations, and hospitalization. In the optimal circumstance, she 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 both in their *practice style*—measured by the physician effect on the resources used in the initial ED visit—and in their *skills*—measured by the physician effect on future healthcare utilization and, by inference, on future healthcare outcomes. Physician skill depends on the ability to make the correct diagnosis and then to outline, and get the patient to follow, an appropriate disposition plan. Practice style and skills may both affect diagnosis and disposition, leading to interdependent system inefficiencies in healthcare operations. While many previous clinical studies consider diagnosis or disposition separately, a strength and innovation of our study is to consider physician variation in both diagnosis and disposition for the same patients. Given the interdependencies, this in turn allows us to obtain a more complete picture of variation in physician practice style and skills.

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 have found that a more intensive practice style correlates negatively with skills (e.g. Doyle et al., 2010; Chan et al., 2020) while others have found 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), that together provide comprehensive evidence on ED operations. Angina refers to heart pain and can be a precursor to a myocardial infarction (heart attack). Appendicitis refers to the painful inflammation of the appendix. TIA, sometimes called a mini-stroke, is a self-limited blockage of blood to some portion of the brain, and can be a precursor to a full-blown stroke. We chose these illnesses for two primary reasons. First, they are all difficult to diagnose and treat appropriately. Second, these decisions are 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 that physicians always diagnose TIA accurately. In that case, if we compare practice style or skill across physicians conditional on the (proper) diagnosis of TIA, this will identify the treatment effects of each ED physician. We can 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 diseases, 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 accurately compare 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 miscoded 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 above. 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—while 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, hospital. 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—which allow us to uncover physician skills. We perform regressions both on the base and plus samples. The regressions on the plus samples inform us about the diagnosis and disposition of patients together, while the regressions on the base samples inform us about the disposition of patients conditional on diagnosis. The physician fixed effect on the initial ED spending indicates her relative practice style. The fixed effect on other outcomes may positively or negatively relate to physician skills, depending on the measure. For instance, a high

[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.

effect on future ED visits is a likely negative indicator of skill.

We then examine the variation in estimated physician indicators for different outcome measures 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. In all cases, our statistical tests of significance use Monte Carlo methods to account for the variance of our estimated physician indicators.

Overview of results. For all three conditions, we find that there is large variation within EDs in both physician practice style and skills. For instance, the differences between physician fixed effects on costs in the initial ED visit from the 90th to the 10th percentile vary, from one ED to another, from \$7.82 to \$45.79 (on a mean \$53.30) for the angina+ sample, with similarly large differences for the appendicitis+ and TIA+ samples.[3]

Furthermore, measures of future healthcare usage also vary significantly between the 90th and 10th percentile. For example, the differences in the physician fixed effects between the 90th and 10th percentile to the number of 0-5-day post-ED-visit hospitalizations vary, from one ED to another, from 0.16 to 0.72 (on a mean of 0.22) for the appendicitis+ sample (again, with similarly large differences in the angina+ and TIA+ samples). Similarly, the 90th to 10th percentile differences in 5-day post-ED-visit hospital costs for the angina+ sample vary, from one ED to another, from \$1,434 (but insignificant) to \$3,003 (across EDs), on a mean of \$2,451. The magnitudes of the differences are even larger for the appendicitis+ and TIA+ samples.

Physician practice style and skills are often negatively correlated. For example, physician-led teams that spend relatively more during the initial ED visit while treating angina+ are also associated with more ED revisits during first five days (correlation of 0.22). Similarly, physician-led teams that spend relatively more during the initial ED visits while treating appendicitis+ are not only associated with more ED revisits during the first 5 days (0.48) and subsequent 6-90 days (0.37) but also more hospitalizations during the first 5 (0.26) and subsequent 6-90 (0.21) days as well. These results are consistent with physicians using extra resources during the initial ED visit when they are uncertain about appropriate diagnosis and disposition.

We also find that physician practice style is positively correlated across illness categories: those who spend more during the initial ED visit for one of the three illness categories typically do so

[3]We list all monetary terms in the paper in 2006 Canadian dollars.

across the other two. More specifically, the correlation coefficient across physicians' resource use for different diseases (measured as MD-related costs during the initial ED visit) is 0.59 between the angina+ and appendicitis+ samples, 0.62 between the TIA+ and angina+ samples, and 0.46 between the appendicitis+ and TIA+ samples, all of which are statistically significant. Physician skills also mostly correlate positively across illnesses. For instance, there are significantly positive correlations between physicians' contribution to ED revisits in the first 5 days across all three illness category pairs (with a correlation coefficient of 0.36 between angina+ and appendicitis+ samples, 0.23 between the angina+, and TIA+ samples, and 0.18 between the appendicitis+ and TIA+ samples). Similarly, there are significant positive correlations between physicians' contribution to hospitalizations during the 0-5 day period across the appendicitis+ and TIA+ samples (0.16) as well as during the 6-90 day period across angina+ and TIA+ (0.26) and across appendicitis+ and TIA+ (0.21), all of which are measures of physician skill. Thus, the physician-led teams that perform well in one context also on average perform well in other contexts. Finally, we find positive correlations between the base and plus samples for our three illness, suggesting that both diagnosis and disposition are important.

Related literature. Our paper contributes primarily to three recent literatures. One literature has leveraged the quasi-random assignment of patients to hospitals and physician groups in order to evaluate physician practice style and resource use across different healthcare settings (see, e.g., Epstein and Nicholson, 2009; Doyle et al., 2010, 2015; Currie et al., 2016; Currie and MacLeod, 2017; Chan et al., 2020). Our study uses similar identifying variation to this literature and has overlapping goals. It builds on this literature through its identification, data, and research question. Closest to our study is Chan et al., who investigate the relation between physician skill and practice style for radiologists diagnosing pneumonia. Relative to our setting, they add external data on the population number of patients with pneumonia, which allows them to estimate the number of false positive tests by radiologists, and ultimately structural utility parameters that underlie practice style decisions. In contrast to Chan et al., who consider only diagnosis, we consider both diagnosis and disposition and in so doing, link physician practice style to future resource use and healthcare utilization over the medium run.

A second literature has considered physician inputs in EDs. Chan (2016) examines how physicians in emergency department settings respond to different management systems when making care decisions. Doyle (2011) finds that visitors to Florida who seek ED treatment in high medical spending areas have lower mortality than visitors who seek ED treatment in lower spending areas. Silver (2016) examines variations in physician productivity by estimating within-physician productivity differences based on shifts. Van Parys (2016) finds that practice styles for minor injuries vary substantially across physicians within an ED. Kindermann et al. (2014) and Pines et al. (2009) examine the determinants of the intensity of ED imaging requests. Finally, other studies have evaluated misdiagnosis in the ED (see, e.g., Hastings et al., 2009; Kachalia et al., 2007; Abaluck et al., 2016). We contribute to this literature by addressing the relation between physician skills and practice styles across diseases. Our unique data—which record the physician of assignment in the ED and which track healthcare utilization and expenditures over the medium run—allow us to add evidence here.

A third literature has estimated the operational effectiveness of team production. Many papers in this literature find that there are benefits from keeping teams intact, in settings ranging from software development to surgery (Reagans et al., 2005; Boh et al., 2007; Huckman et al., 2009; Huckman and Staats, 2013). Huckman and Pisano (2006) find that team production in cardiac surgery is more important than the leader's (surgeon's) skills. Akşin et al. (2015) find that team familiarity is important but so is learning from new team members. We add to this literature through our quasi-experimental identification and by considering the relation between team production for the the same team across different settings, i.e. across different diseases.

The remainder of the paper is organized as follows. Section 2 presents the data and institutional framework. Section 3 presents the theoretical and estimation framework. Section 4 discusses our results and presents a series of robustness checks. Finally, Section 5 concludes.

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. Their database tracks enrollees through time and across four types of care environments: EDs, private offices, hospital-based external clinics, and inpatient hospitalizations.[4] Almost all residents of Quebec are covered by RAMQ,[5] 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 includes the city of Montreal and some suburban municipalities. Henceforth, we refer to this area simply as Montreal. Our study area includes about 1.9 million residents. Our sample contains information on all patients residing in Montreal treated at an ED in Montreal during the sample period, and hence includes information on the universe of ED physicians working in Montreal. We received access to the data from Montreal's public health department (*le Département de santé publique de Montréal*). Each observation in our data starts with an ED visit for a resident of Montreal to an ED in Montreal. It then indicates virtually all future healthcare consumption for the patient in the following 90 days. Specifically, we observe measures of each patient's subsequent consumption of care for all care received within the provincial boundaries and covered by RAMQ.

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 office, hospital-based external clinic, and ED visits. RAMQ records inpatient hospitalization costs differently from in the U.S. Rather than using a DRG- or chargemaster-based system, RAMQ reports a proxy for the total costs of each inpatient stay, called *Niveau d'intensité relative des resources utilisées* (relative intensity level of resources used, NIRRU). RAMQ reports a proxy because all non-physician related expenditures are covered by a global hospital budget and thus not directly associated to the patient. We observe the NIRRU and use it as our measure of inpatient hospitalization costs. A limitation of our data is that NIRRU does not reflect actual billed costs.

The fact that hospitals are reimbursed with a global budget also leads to a related limitation of our data. Procedures ordered by the ED but provided by the hospital outside the ED are also not directly associated to the patient. These include complex imaging services, such as CT scans and MRIs.[6] Finally, our data do not include filled prescriptions.

[4]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 non-hospital settings (i.e., private office settings).

[5]Although a privately funded sector exists in Quebec, it is very small and generally deals with non-covered 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.

[6]Additionally, imaging services are one area for which there exists a robust private market in Quebec, whose services are not included in our data.

For each patient/ED observation, our data include the patient's gender, age group, which is mostly in 5 year bins,[7] three-digit postal code, and two measures of socio-economic status (SES), as constructed by the public health department. The first of these measures, known as the *Indice de défavorisation matérielle* (material deprivation index), is a score from 1 to 5 which seeks to reflect the individual's material (i.e., economic) wellbeing. It reflects mostly variations in education, employment, and income. The second, known as the *Indice de défavorisation sociale* (social deprivation index), which is also a score from 1 to 5, seeks to reflect the individual's social and family support and wellbeing. It reflects mostly variations in family structure and marital status. Both of these are constructed using a variety of sources and are based in large part on geographical (i.e., postal code) location.[8]

Each ED visit—to a Montreal ED by a Montreal resident from April 1, 2006 through December 31, 2006—constitutes one observation in our dataset. For each observation, our data include the ED's unique identifier, the physician's unique identifier, month and weekend identifiers, 4-digit ICD-9 diagnosis codes, procedure codes, and the total fees paid to physicians for services provided.[9] The data also link the initial ED visit with all future visits covered by RAMQ for 90 days. Thus our data terminate on March 31, 2007.

Montreal's public health department used the future revisits to construct the variables indicating the 0-5, 6-30, and 31-90 day number of revisits to EDs, office visits, hospital-based external clinic visits, and hospitalizations, as well as hospitalization costs, as measured by NIRRU. Although precise dates of services are included in the RAMQ data, we do not have direct access to them. As a result, we cannot determine the sequence of events within the time intervals. For instance, we do not know if a patient with a hospitalization within 0-5 days of an ED visit was admitted directly from her ED visit or returned home and was hospitalized a few days later.

2.2 Assignment of patients to physicians

We now discuss the process by which patients are assigned to physicians in the ED and how this affects our estimation sample. 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

[7] There are two bins for patients less than 5 years old and a single bin for patients 85 years or older.

[8] RAMQ refers to these as the Pampalon indices (Pampalon et al. (2009)).

[9] An ED visit can include consultation with several MDs as well as result in several diagnoses and acts/services. The data include unique MD identifiers for each diagnosis provided and procedure administered to the patient.

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, where 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 one type of patients (generally light cases). 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.[10]

We observe patients seen at the 19 general acute care hospitals with EDs in Montreal.[11] From this universe of patients, we exclude initial visits at 2 EDs that serve fewer than 1000 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 9-month period, as these physicians are less likely to have an equitable mix of cases, since they may include the 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 non-random 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 group and gender. Although patient pools could still differ in unobservable manners, we believe that this test is informative in identifying

[10]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, due to privacy issues.

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

the non-random assignment of patients.[12] We proceed by regressing the patient characteristic of interest (gender or age group) 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: Tests for randomness

	Panel A: full sample			Panel B: restricted sample		
ED number	Gender	Age group	Number of physicians	Gender	Age group	Number 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	7
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 group" provides the p value for one F test of a regression of gender or age group on physician fixed effects. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level of the physician fixed effects.

Table 1 panel A presents the p values from F tests for gender and age groups. Of the 17 EDs in our sample, we find evidence that 9 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, 16) while we find evidence against random assignment with respect to the patients' age group 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. For instance, older physicians without advanced emergency-medicine training may be assigned, on average, less-severely ill patients. Illness

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

severity may, in turn, be correlated with a patient's gender and/or age group. While 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 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.[13] More specifically, we drop physicians whose mean gender is more than 5 percentage points from the mean for their ED or whose mean age group is more than 0.5 from the means for their 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 all together. We then rerun the same two F tests on the remaining EDs while 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 and gender for 7 EDs. Thus, our main estimation sample includes 7 EDs that do not violate the randomness assumptions: 1, 4, 7, 8, 9, 15, 17.[14]

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 and we estimate robustness of our main results to a number of different physician exclusion criteria. Finally, we show robustness of our main results to the inclusion of patient observable characteristics.[15]

2.3 Construction of samples by illness

We next discuss how we construct our samples, for each of the three illnesses that we consider, angina, appendicitis, and transient ischemic attacks (TIA).

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

[14] Table A1 in On-line Appendix A provides further tests, analogous to Table 1 panel B, but using our two SES measures. From the 14 EDs in our restricted sample, 3 fail the randomization test for each measure of SES. Unlike age and gender, it would be 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.

[15] Section 4.5 provides robustness results.

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 initial presentation of angina is typically associated with exertion, which puts an increased load on 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 would indicate appendicitis is loss of appetite, followed by nausea, right lower quadrant pain, and vomiting. This sequence occurs in one half to two-thirds of appendicitis patients. The accurate diagnosis of appendicitis is often challenging, since 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). Diagnostic accuracy is substantially lower in adult woman than in adult men, since infection in pelvic organs (pelvic inflammatory disease), urinary tract infections, complications of pregnancy, ovarian cysts, and endometriosis can all mimic appendicitis. Other common conditions, such as gastroenteritis and gall bladder inflammation (cholecystitis), also confound accurate diagnosis. 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, and accordingly limits both the time and extent of diagnostic tests (such as CT scans). Unlike for angina and TIA, appendicitis is not a transient event which resolves on its own. Thus, inaccurate diagnosis or disposition will almost certainly result in future healthcare usage within our 90-day time period.

TIAs are precursors to strokes and are sometimes called mini-strokes. 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 5-fold 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. Specifically, the base diagnoses must be distinguished from a series of alternatives that often present in similar fashion (i.e., the differential diagnoses). 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 must present 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.[16] 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.[17] Fifth, the base diagnosis is a discrete entity, rather than representing a spectrum of disorders, such as pneumonia.[18] Finally, the 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 will 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

[16] We do not consider obstetrical emergencies because in almost all cases, an Ob-Gyn physician would be engaged.

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

[18] 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.

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 which 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).

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	Non-infectious 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	Transient ischemic attack (TIA)
346	Migraine
345	Epilepsy
780.2	Syncope
432.1	Subdural hematoma
431	Intra-cerebral hemorrhage
721.1	Compressive myelopathy

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 she 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 Medical Reference, 2020). The provider can also typically diagnose gastritis (inflammation of the stomach lining) or gastro-esophageal reflux (a digestive disorder) 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 Medical Reference, 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 7 EDs (as noted above), 90 physician-ED pairs,[19] and a total of 15,098 initial ED visits and 12,855 unique patients. The mean age groups for all three illnesses is approximately 12, which represents an age range from 50 to 54. The percentage of males vary 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 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—such as imaging—the 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 will 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 days and the subsequent 85 days following the ED visit, respectively. The number of ED revisits is high, e.g., 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.[20] Hospitalizations are quite frequent across all illness categories.

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

[20]Although our data include the categories of 0-5, 6-30, and 31-90 days, we split the ED visits and hospital stays into these two categories to separate the visits that directly relate to the initial visit—and which would then likely occur within five days—from other ones.

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 or the angina+, appendicitis+ and TIA+ samples, respectively. Although we do not have cost measures for other follow-up healthcare visits, we believe that the bulk of the healthcare resources used by individuals over the 90-day follow-up period are for hospitalizations.

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. Several of the physicians treated too few patients to be able to identify fixed effect for these samples. Thus, the base samples include between 46 and 75 physicians, depending on the illness. For appendicitis, only 6 EDs have physicians with identified fixed effects, while the other two illnesses include physicians from all 7 EDs in our estimation samples for the plus diagnoses.

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 (though in contrast the short-run costs for TIA are similar but somewhat lower than for TIA+). 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 though, 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 to 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

In this section, we develop a simple model of ED patient treatment, health outcomes, and healthcare expenditures that can apply to any case where appropriate diagnosis and disposition of the patient are difficult. We then use the model to develop an estimation framework.

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 group	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 group	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 5 year bins, while bin 19 groups all patients 85 years of age or older.

3.1 Model

The basic premise of our model is that the most important functions of the ED physician 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 different condition although, given the composition of our conditions, they are most likely for a recurrence.

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 her treatments will cause extra resources to be used without delivering any extra health benefit. Since many diagnostic and therapeutic interventions occur only after discharge from the ED, disposition also depends on the ED physician's ability to convince her 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 ED has J attending physician-led teams and the patient will be treated by one of them. Denote the treating physician-led team by j; $j \in \{1, \ldots, J\}$. We assume that there are two possible illnesses, $\theta_i \in \{\theta^D, \theta^B\}$. If $\theta_i = \theta^B$, then the patient has experienced the base illness, while if $\theta_i = \theta^D$, then she has experienced 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 other than the base diagnosis. The symptoms overlap between the two illnesses, implying that accurate diagnosis is difficult. Consistent with the evidence in Table 3, we model θ^D as having a lower severity during the initial ED visit than θ^B.

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. T^B is more expensive than T^D, e.g. involving referrals for greater follow-up care. We assume that the physician chooses T_i using information about the illness shock θ_i. On one extreme, it is possible that the physician has complete information regarding θ_i. On the other extreme, it is possible that she uses a θ_i that is identical across all patients. Because the reality is likely somewhere in between these two extremes, our model does not impose a particular structure on the team's information.[21]

We now exposit the patient's health production function. The post-treatment health stock is a latent value 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 assume that $\overline{v}(\theta^B) < \overline{v}(\theta^D) < 0$ so that θ^B confers a worse health shock on average than θ^D and both are worse than no illness. The ε_i term is the deviation of the illness shock from the mean level 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. Taken together, the post-treatment health stock is given by:

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

The data do not record the information set of the ED physician regarding θ_i at the time of treatment. Thus, we allow the possibility that θ_i is fully or partially observable to the ED physician. Although the health stock is unobservable to the econometrician, the consumption of medical services is observable. More specifically, let H_i denote as an observable healthcare event occurring after the initial illness shock and care at the ED. For example, H_i could be a hospitalization. In such an example, the observable hospitalization occurs (i.e., $H_i = 1$) if latent health falls below some given threshold (i.e., if $H_i^* < 0$). Although the outcome variable presented in this example is binary in nature (i.e., hospitalization or not), our empirical work will principally consider count variables (i.e., the number of ED revisits) and continuous variables (i.e., the costs associated with hospitalizations).

Corresponding to our health production function, we also define a dollar expenditure function.

[21] Also, we do not model patient beliefs, since the physician makes the treatment decision in our model.

Expenditures for patient i 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. While T^B will cost more than T^D in the current ED visit, it is possible that T^D will cost more in the long-run, due to adverse outcomes.

Considering the implications of our model, there are three possible health outcomes that can occur to patient i. She might have θ^D, in which case the treatment does not affect her health status. Or, she might have θ^B, in which case the T^B treatment will result in a higher expected final health status than would an inappropriately low treatment T^D. From an expenditure point of view, her medical expenditures during the initial ED visit will be higher if she receives T^B regardless of her underlying health status.

We use our framework and data to evaluate each physician-led team'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 will depend on (a) her ability to distinguish θ^B from θ^D and (b) her 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, while for others, this may not be true. To understand the treatment effect of the physician skill, define $p_{ij}(T^B|\theta^B)$ to be the probability that physician j picks treatment T^B when the patient i's true health health state is θ^B. Let d_{ij} denote an indicator for whether patient i was assigned to physician j.

Patient i's expected (latent) health stock from her ED visit will then be equal to:

$$\begin{aligned} E[H_{ij}^*] = \sum_{j=1}^{J} d_{ij} p_{ij} \left(T^B|\theta^B\right) P_i(\theta^B) \left[v(\theta^B, T^B) - v(\theta^B, T^D)\right] + \\ \overline{H}_i + P_i(\theta^D)(v(\theta^D, T^D) + \overline{v}(\theta^D)) + P_i(\theta^B)(v(\theta^B, T^D) + \overline{v}(\theta^B)) + \varepsilon_i, \end{aligned} \tag{3}$$

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) while 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 her diagnosis of the correct condition and second, in her choice of the appropriate treatment conditional on diagnosis. These two skills together determine $p_{ij}(T^B|\theta^B)$. Also note that, by assumption, physician skill here only matters to the extent that the patient has θ^B, as treatments do not affect health outcomes for a patient who has θ^D.

Let $S_{ij}^H \equiv p_{ij}\left(T^B|\theta^B\right) P_i(\theta^B)\left[v(\theta^B, T^B) - v(\theta^B, T^D)\right]$ 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:

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

We can also exposit the expectation of D_i^*, in a similar fashion to $E[H_i^*]$ 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.[22]

3.2 Estimation framework

Using our model as a basis, we estimate the component of different outcome measures that is due to the ED physician. Our idea is to uncover the physician treatment effects across different dimensions, notably costs and health outcomes. We focus our discussion on health status and hence equation (4), although the logic for expenditures is identical.

We parametrize the observable part of expected health status that is not a function of the

[22]Note that the equation underlying S_{ij}^D in (5) is different from S_{ij}^H in (4) because expenditures can be affected by the treatment regardless of the underlying health state.

patient's assignment to a particular ED physician as:

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

where x_i are observables and β^H is a vector of parameters to estimate. Our data contain information on the patient condition (angina+, appendicitis+, or TIA+), choice of ED, and the age, gender, and SES of the patient, in addition to an indictor for the treating ED physician. We include controls for age, gender, and SES in x_i and we stratify our sample by condition and choice of ED, as we detail below.

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:

$$E\left[H_i^*\right] = \sum_{j=1}^{J} d_{ij} FE_j^H + x_i\beta^H + \varepsilon_i. \tag{7}$$

Equation (7) leads directly to our estimating equation: we estimate least squares specifications where we regress observable markers of health status, H_i, on physician fixed effects and controls for patient characteristics, separately by ED and by each plus illness.[23] Our regressions for dollar measures $E[D_i^*]$ take an identical form. The estimated physician fixed effects indicate the causal impact of being treated by a given physician relative to other physicians at the same ED or, alternately put, the physician components of resources used.

We then correlate different estimated fixed effects for a given physician in order to examine the relation between different measures of practice style and skills across physicians, also reporting statistical significance of our correlation measures. 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. E.g., 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 some mix of practice style and skill, depending on the particular measure. Spending in the initial ED visit is a measure of practice style: physicians

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

who spend more for identical patients are ones who have a practice style that uses more resources. Revisits to an ED are an unambiguous indicator of worse care. For office visits or hospital-based external clinic visits, a positive coefficient is likely an indicator of better care, as such a visit might constitute a useful follow-up (i.e., further testing) that would prevent future ED visits or hospitalizations. Coefficients on hospitalizations are more ambiguous, as a hospitalization in the medium run might forestall future, and more expensive, treatments, prevent death, or otherwise improve patient welfare. Because of the inherent ambiguity of some of these measures, we do not group utilization across venues 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 will reflect some function of both diagnosis and disposition skills. Since TIA is the hardest to diagnose, we expect that practice style and skill differences will reflect diagnosis skills more than disposition skills for this illness, relative to the others.

We now discuss identification of our empirical specification. 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 EDs. This would then result in expected case severity being balanced across physicians. As we discussed in Section 2.2, the assignment of patients to physicians is quasi-random within an ED for hospitals and physicians in our sample.[24]

We now discuss specifics of how we implement our empirical model. We use linear regressions to uncover all physician fixed effects. The dependent variables are the physician-related costs during the initial ED visit and healthcare costs and visits over the 90 day follow-up period. All specifications include fixed effects for each ED physician and the same set of x_i controls noted above. We first examine the within-ED distribution of fixed effects—specifically, the 90^{th} to 10^{th} percentile difference. We then calculate correlations between fixed effects for a physician—relative to the mean at her ED—across different measures. Our inference uses Monte Carlo methods to account for the fact that our physician fixed effects are estimated. On-line Appendix B provides details on these methods.

[24]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.

4 Results

We now consider the impact of treatment by different ED physicians using the methodology developed above, where we regress each outcome on physician fixed effects and then evaluate the correlation across outcomes. 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 percentile 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. These 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 show that there is significant variation in practice style across physician-led teams.

For angina+ and appendicitis+, we also find significant differences in ED revisits within 5 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 complications from inappropriate diagnosis or disposition of TIA are more likely to result in complications in the medium-run rather than the short-run. These results show that there is also significant variation in skills across physician-led teams.

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 5 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 90-10 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 to identify with certainty. Long courses of antibiotics, interventions to drain the abscesses, fistulas (tracks created by infection and leading to other organs) and more can lead to long (and hence, expensive) hospitalizations.

Table A2 in On-line Appendix A presents analogous results to Table 4 but comparing the 75th and 25th percentiles of physicians within an ED. The results are similar, though most of the significant differences for these measures are for the initial resources used, hospital stays, and hospital spending.

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 hospital—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 physician-led teams that have 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 correlations between the fixed effect for initial ED spending and the fixed effect for ED revisits in 0-5 days are positive and significant for two of the three conditions that we consider, at 0.22 for angina+ and 0.48 for appendicitis+. 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, physician-led teams 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.

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. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Our results on the relation between the initial ED spending and hospital stays are more mixed. On one hand, physician-led teams 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.[25] 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. Physician-led teams that 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).[26]

Overall, our findings suggest that treatment by physicians-led teams that spend more is good for angina+ patients in the medium run, but bad for appendicitis+ patients, with no significant effect for TIA+. Taken together, these results suggest that the correlation between an intensive practice style and skills will vary across conditions.

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+, physician-led teams who have more ED revisits during the first 5 days after the initial visit also have more hospital-based external clinic visits in the first 30 days, suggesting that ED revisits in the first 5 days are complements to hospital-based external visits for these samples. In contrast, physician-led teams that have more ED revisits in the first 5 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. In addition, physician-led teams that have higher mean spending in the initial ED visit have more subsequent external-based hospital clinics for the appendicitis+ (correlation of 0.26) and angina+

[25] Tables A3-A5 in On-line Appendix A adds 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.

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

(correlation of 0.15) samples and more subsequent physician office visits (correlation of 0.28) for the appendicitis+ sample. These results suggest that office and hospital based clinic visits complement spending in the initial ED visits.

Finally, we find that, for all three samples, physician-led teams with a greater number of ED revisits in the 6 to 90 day period also have a greater number of hospital stays. These results suggest that ED revisits within the 6 to 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 an external hospital 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.

Table 5: Correlation in physician fixed effects within samples

	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

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

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. The most striking finding here is the strong correlation of practice styles across the three illness pairs. That is, physician-led teams that spend more on average in the initial ED visit for the angina+ sample also spend more on average when treating the appendicitis+ sample as well as when treating the TIA+ sample. More specifically, the coefficient of correlation between physician average contribution (i.e., practice style) to physician-related costs during the initial ED visit across (i) the angina+ and appendicitis+ samples is 0.59, (ii) the TIA+ and angina+ samples is 0.62, and (iii) the appendicitis+ and TIA+ samples is 0.46. Physician-led teams with costly practice styles for one condition, on average, costly practice styles across illness categories.

The diagonals in Table 6 also reveal positive correlations across illness categories for several skills. More specifically, we find that physicians associated with more revisits to EDs during the first 5 days in one illness category are also associated with more ED revisits during the same period for the other two illness categories. The correlation coefficient for ED 0-5 days across the angina+ and appendicitis+ samples is 0.36, across the TIA+ and angina+ samples is 0.23, and across the appendicitis+ and TIA+ samples is 0.18. Furthermore, we find that physician-led teams associated with more hospitalizations in the 0-5 days post ED-visit when treating TIA+ patients are also associated with more hospitalizations during the same period when treating the sample of appendicitis+ patients (with a correlation coefficient of 0.16). Similarly, we find a positive correlation between hospitalizations in the 6-90 day period after the initial ED visit for the angina+ and TIA+ samples (0.26), and for the appendicitis+ and TIA+ samples (0.21).

4.4 Relationships between practice styles and skills across base and plus samples

Finally, Table 7 provides correlations between the base and plus samples for our three illnesses. Recall that the plus sample fixed effects reflect practice styles and skills for both diagnosis and disposition, while the base sample fixed effects reflect practice styles and skills for disposition only, conditional diagnosis.

We find a strong positive correlation between the practice styles—measured as spending in the initial ED visit—across the angina and angina+ samples (0.37), across the appendicitis and appen-

Table 6: Correlation in physician fixed effects across conditions

	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**

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 7: Correlation in physician fixed effects across base and plus samples

	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

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

dicitis+ samples (0.41) and across the TIA and TIA+ samples (0.44). As shown in the diagonal elements, we find several positive correlations for fixed effects for future healthcare consumption across the plus and base samples. The fixed effects associated with the plus samples reflect diagnosis and disposition skills while the fixed effects associated with the base samples reflect disposition skills conditional on diagnosis.

Physician-led teams that are associated with more hospital stays within the first 5 days when treating the angina+, appendicitis+, and TIA+ samples are also associated with more hospitalizations during the same period when treating patient 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. Since 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

On-line Appendix A replicates our main correlation results, Tables 5-7, but with different criteria for the inclusion of physicians and of patient covariates. First, to mitigate any concern that some physician-led teams may specialize in some illnesses, we reestimated our model, dropping from each illness category teams that are outliers in terms of the fraction of patients that they treat from that category. For this specification, we drop ED 4 altogether, as we found that (i) the fraction of patients with angina+ is bimodal across physician-led teams for this ED and, (ii) this ED appears to serve a very different patient mix, e.g., with many more patients in the appendicitis+ sample. Altogether, these restrictions leave us with 78 unique physician-ED pairs, compared to 90 for our base analysis. Tables A6-A8 present results for this more restrictive sample of physicians. We find very similar results to those from the initial sample of physicians, though Table A6 shows that initial ED spending correlates significantly with 0-5 day ED revisits for TIA+, while the correlation was positive but not significant in Table 5.

Second, we expand our sample to include all ED physicians in our estimation sample of 7 EDs. As a result, our sample goes from 90 physicians-ED pairs to 103 to 106 physicians-ED pairs (depending on the condition studied). Tables A9-A11 presents these results. For all three tables, we find results that are virtually indistinguishable from the base results. Together, these results suggest that correlations in practice styles and skills are robust to outliers.

Finally, we examine our correlation results using the original (main) sample but excluding our controls for age, gender, and SES. We again 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. This further justifies our assumption of quasi-random assignment of patients to physicians.

5 Conclusions

In this paper, we use detailed data from emergency departments in Montreal, Canada to understand how teams led by ED physicians contribute to healthcare costs and outcomes. 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. Nonetheless, we cannot completely rule out non-random selection to physicians. We develop a theoretical model under which physician-led teams have potentially different skills and practice styles, which together affect the diagnosis and disposition of patients. Our outcome measures provide evidence on appropriate diagnosis and disposition.

Using our data, we estimate the effect of being treated by different ED physician-led teams on these outcomes. We perform this estimation for three samples of patients: (i) those who are diagnosed with conditions that might indicate angina (angina+), (ii) those who are diagnosed with conditions that might indicate appendicitis (appendicitis+), and (iii) those who are diagnosed with conditions that might indicate a transient ischemic attack (TIA+). 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, though it does not entirely eliminate this possibility.

We find that physicians-led teams vary in both their average contribution to ED costs during the initial ED visit and in their future outcomes. Importantly, for two of the three conditions that we consider, we find that physician-led teams that use more ED resources achieve worse outcomes. More specifically, physicians-led teams with a high-cost practice style have more ED revisits for the angina+ and appendicitis+ samples (with a positive but insignificant effect for TIA+) and more hospitalizations for appendicitis+. The results for angina+ are more ambiguous: teams that spend more have more ED revisits in the first five days after the initial visit but also have fewer hospital stays in the 6-90 day follow-up period. The differences in our findings between angina+ and appendicitis+ may be due to the fact that appendicitis is an acute event in an otherwise healthy patient which can be resolved completely, while the underlying cause of angina is unlikely to be reversed over a short time frame.

We also find that teams that 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. 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.

We offer three main implications regarding improvements in operations management that stem from our results. First, there is substantial room for learning within organizations to improve ED outcomes and lower costs. Second, teams which perform well on one condition on average perform well on other conditions, implying that team quality is a relatively broad concept and that we can learn about operational effectiveness from high-performing teams. Third, the positive correlation between low costs and high quality suggests that reducing healthcare costs and improving quality are complementary goals, and not goals that are at odds with each other.

References

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

Akşin, Z., Deo, S., Jónasson, J. O., and Ramdas, K. (2015). Learning from many: Partner exposure and team familiarity in fluid teams. Working Paper, Available at SSRN 2672133.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1):151–184.

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

Boh, W. F., Slaughter, S. A., and Espinosa, J. A. (2007). Learning from experience in software development: A multilevel analysis. *Management Science*, 53(8):1315–1331.

Chan, D. C. (2016). Teamwork and moral hazard: Evidence from the emergency department. *Journal of Political Economy*, 124(3):734–770.

Chan, D. C., Gentzkow, M., and Yu, C. (2020). Selection with variation in diagnostic skill: Evidence from radiologists. *Working Paper*.

Currie, J. and MacLeod, W. B. (2017). Diagnosing expertise: Human capital, decision making, and performance among physicians. *Journal of Labor Economics*, 35(1):1–43.

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

Doyle, J. (2011). Returns to local-area healthcare spending: Evidence from health shocks to patients far from home. *American Economic Journal: Applied Economics*, 3(3):221–243.

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

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

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

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

Hollander, J. E. and Chase, M. (2016). Evaluation of the adult with chest pain in the emergency department. `https://www.uptodate.com/contents/evaluation-of-the-adult-with-chest-pain-in-the-emergency-department`.

Huckman, R. and Staats, B. (2013). The hidden benefits of keeping teams intact. *Harvard Business Review*, 91(12):27–29.

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

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

Johns Hopkins Medicine (2020). GERD diet: Foods that help with acid reflux (heartburn). `https://www.hopkinsmedicine.org/health/wellness-and-prevention/gerd-diet-foods-that-help-with-acid-reflux-heartburn`. Accessed: 2020-06-03.

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

KC, D. S., Scholtes, S., and Terwiesch, C. (2019). Empirical research in healthcare operations: Past research, present understanding, and future opportunities. *Manufacturing & Service Operations Management, (forthcoming)*.

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

Kindermann, D., McCarthy, M., Ding, R., Frohna, W., Hansen, J., Maloy, K., Milzman, D., and Pines, J. (2014). Emergency department variation in utilization and diagnostic yield of advanced radiography in diagnosis of pulmonary embolus. *The Journal of Emergency Medicine*, 4.6.6.:791–799.

Martin, R. F. (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. `https://www.mayoclinic.org/diseases-conditions/gastritis/diagnosis-treatment/drc-20355813`. Accessed: 2020-06-03.

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

Morris, C. N. (1983). Parametric empirical bayes inference: theory and applications. *Journal of the American Statistical Association*, 78(381):47–55.

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

Pampalon, R., Hamel, D., Gamache, P., and Raymond, G. (2009). Un indice de défavorisation pour la planification de la santé au canada. *Maladies chroniques au Canada*, 29(4):199–213.

Pines, J., Hollander, J., Isserman, J., Chen, E., Dean, A., Shofer, F., and Mills, A. (2009). The association between physician risk tolerance and imaging use in abdominal pain. *American Journal of Emergency Medicine*, 27.5:552–577.

Pope, J., Aufderheide, T. P., Ruthazer, R., Woolard, R. H., Feldman, J. A., Beshansky, J. R., Griffith, J. L., and Selker, H. P. (2000). Missed diagnoses of acute cardiac ischemia in the emergency department. *New England Journal of Medicine*, 342:1163–1170.

Pope, J. and Edlow, J. (2012). Avoiding misdiagnosis in patients with neurological emergencies. *Emergency Medicine International*, 2012.

Proulx, A. M. and Zryd, T. W. (2009). Costochondritis: Diagnosis and treatment. *American Family Physician*, 80(6):617–620.

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

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

Rothwell, P., Giles, M., Chandratheva, A., Marquardt, L., Geraghty, O., Redgrave, J., Lovelock, C., Binney, L., Bull, L., Cuthbertson, F., Welch, S., Bosch, S., Carasco-Alexander, F., Silver, L., Gutnikov, S., and Mehta, Z. (2007). Effect of urgent treatment of transient ischaemic attack and minor stroke on early recurrent stroke (EXPRESS study): A prospective population-based sequential comparison. *The Lancet*, 370:1432–1442.

Seetahal, S. A., Bolorunduro, O. B., Sookdeo, T. C., Oyetunji, T. A., Greene, W. R., Frederick, W., Cornwell, E. E., Chang, D. C., and Siram, S. M. (2011). Negative appendectomy: A 10-year review of a nationally representative sample. *The American Journal of Surgery*, 201(4):433–437.

Silver, D. (2016). Haste or waste? Peer pressure and the distribution of marginal returns to health care. Working Paper, Princeton University.

Taubman, S. L., Allen, H. L., Wright, B. J., Baicker, K., and Finkelstein, A. N. (2014). Medicaid increases emergency-department use: Evidence from Oregon's health insurance experiment. *Science*, 17(17):263–268.

Tsugawa, Y., Jha, A. K., Newhouse, J. P., Zaslavsky, A. M., and Jena, A. B. (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 Medical Reference (2020). What's causing my chest pain? `https://www.webmd.com/pain-management/guide/whats-causing-my-chest-pain#1`. Accessed: 2020-06-12.

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

On-line Appendix A: robustness tables

Table A1: Additional tests for randomness, using SES measures

ED	Restricted sample SES: material	SES: social	Number of physicians
1	0.026**	0.147	15
4	0.009***	0.839	7
5	0.257	0.016**	22
6	0.086*	0.282	12
7	0.086*	0.008***	8
8	0.000***	0.130	12
9	0.050**	0.003***	16
12	0.088*	0.221	12
13	0.156	0.013**	14
14	0.001***	0.090*	7
15	0.105	0.000***	11
16	0.017**	0.100*	18
17	0.123	0.081*	24

Note: each entry under "SES: material" and "SES: social" provides the p value for one F test of a regression of one SES measure on physician fixed effects. Section 2.1 further describes these SES measures. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level of the physician fixed effects.

Table A2: Estimated 75-25 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 $	1.48	10.60***	10.58***	13.03***	22.35***	10.89**	6.72
ED 0-5 days	0.06	0.14	0.06	0.09*	0.14*	0.06	0.06
ED 6-90 days	0.07	0.37	0.24	0.17	0.40**	0.15	0.16
External 0-30 days	0.12	0.26	0.27*	0.14	0.26	0.22	0.19
Office 0-30 days	0.18*	0.04	0.13	0.09	0.12	0.10	0.20
Hospital stay 0-5 days	0.02	0.06	0.04	0.04	0.04	0.06	0.06
Hospital stay 6-90 days	0.05	0.07	0.04	0.05	0.11**	0.06	0.06
Hospital $ 0-5 days	486	1,287	997	867	1,191	583	1,512
Hospital $ 6-90 days	537	1,402	1,414	1,408	953	834	1,029
Appendicitis+ sample							
Initial ED $	6.12	9.80	9.74	5.72	22.15***	7.32	7.90
ED 0-5 days	0.47	0.49***	0.19*	0.17*	0.30	0.23	0.23*
ED 6-90 days	0.87	0.58	0.41*	0.24	0.54	0.39	0.40
External 0-30 days	0.29	0.36	0.41	0.50	0.65	0.97***	0.68
Office 0-30 days	0.57	0.31	0.24	0.31	0.27	0.27	0.33
Hospital stay 0-5 days	0.23	0.13	0.13	0.23***	0.20	0.19	0.20
Hospital stay 6-90 days	0.30	0.33	0.16**	0.13	0.15	0.17*	0.14
Hospital $ 0-5 days	3,239	540	1,038	2,536	1,881	1,057	2,194
Hospital $ 6-90 days	4,834	4,880	2,380**	1,129	2,120	1,036	1,469
TIA+ sample							
Initial ED $	0.96	14.73	15.77*	18.88*	22.25	21.92**	17.16
ED 0-5 days	0.19	0.18	0.17	0.16	0.25*	0.15	0.16
ED 6-90 days	0.42	0.21	0.49	0.48	0.42	0.69	0.43
External 0-30 days	0.11	0.19	0.24	0.14	0.47	0.60	0.48
Office 0-30 days	0.08	0.11	0.33	0.37	0.15	0.27	0.29
Hospital stay 0-5 days	0.03	0.09	0.20***	0.16*	0.16*	0.19**	0.12
Hospital stay 6-90 days	0.05	0.20*	0.09	0.11	0.16	0.11	0.15
Hospital $ 0-5 days	1,488	2,950	4,268*	3,028	3,739	2,400	4,138
Hospital $ 6-90 days	1,491	1,349	1,452	2,450	2,429	1,692	1,615

Note: we order EDs in the table by the number of ED physicians per ED. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A3: Correlation in physician fixed effects within samples (extra outcomes)

	Initial ED $	ED 0-5 days	ED 6-90 days	External 0-30 days	Office 0-30 days	Hosp stay 0-5 days	Hosp stay 6-90 days	Hosp $ 0-5 days	Hosp $ 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	−0.18**	0.00	0.35***	−0.05	−0.10	0.22**	1		
Hospital $ 0-5 days	0.09	−0.10	0.06	0.03	−0.08	0.40***	0.21**	1	
Hospital $ 6-90 days	0.07	0.07	0.16	0.03	−0.14	0.16	0.64***	0.15	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	0.21*	0.06	0.53***	0.09	−0.16	0.34***	1		
Hospital $ 0-5 days	0.01	0.09	−0.06	0.02	−0.18	0.73***	0.24**	1	
Hospital $ 6-90 days	0.11	0.14	0.53***	0.04	−0.17	0.23*	0.79***	0.21*	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		
Hospital $ 0-5 days	0.16	0.15	−0.07	0.35***	−0.13	0.73***	0.11	1	
Hospital $ 6-90 days	0.04	−0.03	0.31***	0.11	−0.21*	0.12	0.74***	0.15	1

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A4: Correlation in physician fixed effects across conditions (extra outcomes)

	Initial ED $	ED 0-5 days	ED 6-90 days	External 0-30 days	Office 0-30 days	Hosp stay 0-5 days	Hosp stay 6-90 days	Hosp $ 0-5 days	Hosp $ 6-90 days
	Angina+								
Appendicitis+									
Initial ED $	0.59***	0.24***	−0.10	0.01	0.15*	−0.08	−0.22**	−0.05	−0.07
ED 0-5 days	0.13*	0.36***	0.10	−0.15*	−0.02	−0.28***	−0.11	−0.15	−0.16*
ED 6-90 days	0.04	−0.05	0.01	−0.09	0.14	−0.09	−0.18*	−0.04	−0.12
External 0-30 days	0.07	0.02	0.20**	−0.08	0.18*	−0.06	−0.04	0.09	−0.09
Office 0-30 days	0.13**	0.32***	−0.04	0.12	−0.15*	−0.02	−0.15*	−0.02	−0.12
Hospital stay 0-5 days	0.18**	−0.02	0.03	−0.07	0.10	0.09	0.01	−0.08	0.05
Hospital stay 6-90 days	0.11	−0.14	−0.03	−0.10	0.09	−0.03	0.00	−0.05	0.01
Hospital $ 0-5 days	0.02	−0.06	−0.05	−0.08	−0.06	0.08	0.18**	−0.06	0.16*
Hospital $ 6-90 days	0.04	−0.15*	0.08	−0.08	0.08	−0.04	0.09	−0.04	0.03
	TIA+								
Angina+									
Initial ED $	0.62***	0.35***	−0.00	0.08	−0.04	0.07	0.13	0.07	0.18*
ED 0-5 days	−0.01	0.23**	−0.21**	0.05	−0.16	−0.12	0.04	−0.05	−0.07
ED 6-90 days	−0.07	−0.24**	−0.05	0.07	−0.10	0.03	0.07	−0.17*	0.08
External 0-30 days	−0.09	0.12	0.14	−0.01	−0.19*	0.07	0.10	0.07	−0.06
Office 0-30 days	0.00	0.08	0.07	0.07	0.09	0.04	−0.00	−0.14	0.14
Hospital stay 0-5 days	−0.17**	−0.04	0.15	−0.05	−0.03	0.01	0.23**	−0.06	0.11
Hospital stay 6-90 days	−0.10	0.00	0.21**	−0.04	0.09	−0.08	0.26***	−0.24**	0.07
Hospital $ 0-5 days	−0.16**	0.11	0.19*	0.01	0.01	0.05	0.15	−0.11	0.03
Hospital $ 6-90 days	−0.25***	−0.01	0.11	−0.13	0.05	−0.11	0.24**	−0.18*	0.19*
	Appendicitis+								
TIA+									
Initial ED $	0.46***	0.15*	0.13	0.10	0.16*	0.23***	0.10	0.14	0.05
ED 0-5 days	−0.01	0.18*	0.11	−0.03	0.17*	−0.10	−0.15	−0.12	−0.02
ED 6-90 days	0.04	0.06	0.09	−0.09	−0.06	0.01	0.15*	0.03	0.10
External 0-30 days	−0.11	0.18**	0.04	0.01	−0.12	0.12	−0.08	0.14	−0.06
Office 0-30 days	0.07	0.06	−0.01	−0.08	0.02	0.09	−0.02	0.02	−0.04
Hospital stay 0-5 days	−0.15*	−0.03	−0.07	−0.08	−0.06	0.16*	−0.18**	0.19**	−0.14*
Hospital stay 6-90 days	0.01	−0.08	0.09	0.00	−0.18**	0.09	0.21**	0.10	0.19**
Hospital $ 0-5 days	−0.09	−0.02	0.04	−0.09	−0.05	0.19**	−0.15*	0.11	−0.02
Hospital $ 6-90 days	−0.14	−0.13	0.09	0.03	−0.26***	−0.00	0.19**	0.00	0.15

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A5: Correlation in physicians effects across base and plus samples (extra outcomes)

	Initial ED $	ED 0-5 days	ED 6-90 days	External 0-30 days	Office 0-30 days	Hosp stay 0-5 days	Hosp stay 6-90 days	Hosp $ 0-5 days	Hosp $ 6-90 days
	Angina								
Angina+									
Initial ED $	0.37***	−0.03	−0.06	0.06	0.20***	−0.05	−0.19***	0.08	−0.17***
ED 0-5 days	0.01	0.02	−0.05	−0.02	0.13	−0.19**	−0.12	−0.15	−0.02
ED 6-90 days	0.05	−0.05	0.31***	0.10	0.09	−0.22**	0.13	−0.18*	−0.07
External 0-30 days	−0.05	−0.28***	0.04	0.28***	0.13	−0.18**	0.03	−0.20*	0.02
Office 0-30 days	0.04	0.17**	0.01	−0.15	0.24**	0.06	−0.10	0.14	−0.08
Hospital stay 0-5 days	−0.02	−0.19**	0.14	0.17*	−0.03	0.25***	0.26***	0.13	0.17**
Hospital stay 6-90 days	−0.12	0.03	−0.02	−0.16	−0.37***	−0.07	0.16*	−0.16	0.12
Hospital $ 0-5 days	−0.03	−0.15*	0.02	−0.02	−0.07	0.17**	0.20**	0.20**	0.11
Hospital $ 6-90 days	−0.06	−0.10	−0.06	−0.09	−0.38***	−0.14	0.28***	−0.15	0.35***
	Appendicitis								
Appendicitis+									
Initial ED $	0.41***	0.21***	0.09	0.01	0.08	0.00	0.01	0.04	0.03
ED 0-5 days	0.16**	0.43***	−0.14*	0.08	0.23***	−0.21***	0.06	−0.07	−0.02
ED 6-90 days	0.02	−0.07	0.32***	0.08	0.21**	−0.13	0.14	−0.11	0.11
External 0-30 days	−0.04	0.17***	0.09	0.36***	0.26***	−0.17**	0.27***	−0.15**	0.06
Office 0-30 days	0.02	0.08	0.38***	−0.01	0.31***	−0.07	−0.12	−0.02	0.01
Hospital stay 0-5 days	0.00	−0.12	−0.11	−0.01	−0.08	0.38***	−0.03	0.16**	0.05
Hospital stay 6-90 days	−0.18**	−0.25***	0.02	−0.02	0.13	0.06	0.18	−0.03	0.38***
Hospital $ 0-5 days	0.14*	−0.02	−0.08	0.06	−0.02	0.34***	−0.06	0.29***	0.01
Hospital $ 6-90 days	−0.10	−0.18**	−0.15	−0.18	0.10	−0.03	−0.07	−0.03	0.03
	TIA								
TIA+									
Initial ED $	0.44***	−0.18*	−0.12	−0.16	0.06	−0.01	0.11	−0.06	0.09
ED 0-5 days	0.01	0.26**	−0.07	0.04	−0.10	−0.23**	−0.15	−0.18*	0.01
ED 6-90 days	−0.20	−0.12	0.13	0.02	0.01	0.30**	−0.07	0.16	0.30***
External 0-30 days	−0.11	0.10	0.20**	0.29***	−0.14	−0.04	0.02	0.03	0.00
Office 0-30 days	−0.19	−0.17	0.16	0.14	0.19	−0.10	−0.09	−0.20*	−0.16*
Hospital stay 0-5 days	−0.11	−0.23**	0.14	0.00	−0.09	0.28***	0.11	0.16*	−0.15*
Hospital stay 6-90 days	0.12	−0.18*	0.24**	0.09	0.11	−0.08	0.06	−0.10	0.26***
Hospital $ 0-5 days	−0.23**	−0.23**	−0.05	0.10	−0.13	0.18*	0.03	0.22**	−0.06
Hospital $ 6-90 days	−0.05	−0.17	0.02	−0.04	−0.04	−0.16	−0.02	−0.17	0.29***

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A6: Correlation in physician fixed effects within samples (restricted set of physicians)

	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	0.29***	1					
ED 6-90 days	−0.13	0.05	1				
External 0-30 days	0.14*	0.07	−0.05	1			
Office 0-30 days	0.01	−0.12	0.24**	−0.34***	1		
Hospital stay 0-5 days	0.09	−0.05	−0.09	0.43***	−0.12	1	
Hospital stay 6-90 days	−0.23**	0.02	0.33***	0.01	−0.07	0.19*	1
Appendicitis+ sample							
Initial ED $	1						
ED 0-5 days	0.51***	1					
ED 6-90 days	0.33***	0.17	1				
External 0-30 days	0.26***	0.29***	0.18	1			
Office 0-30 days	0.14	−0.04	0.04	−0.14	1		
Hospital stay 0-5 days	0.34***	0.12	0.13	0.22**	−0.16	1	
Hospital stay 6-90 days	0.29***	0.06	0.60***	0.15	0.01	0.33***	1
TIA+ sample							
Initial ED $	1						
ED 0-5 days	0.19*	1					
ED 6-90 days	0.14	−0.11	1				
External 0-30 days	0.13	0.22**	0.01	1			
Office 0-30 days	0.05	−0.17	0.08	0.08	1		
Hospital stay 0-5 days	0.00	0.09	0.19*	0.30***	−0.09	1	
Hospital stay 6-90 days	0.10	−0.05	0.33***	0.23**	−0.08	0.12	1

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A7: Correlation in physician fixed effects across conditions (restricted set of physicians)

	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 $	0.60***	0.23**	−0.05	0.01	0.17*	−0.08	−0.23**
ED 0-5 days	0.17**	0.33***	0.15	−0.19**	0.03	−0.33***	−0.07
ED 6-90 days	0.02	−0.12	0.02	−0.10	0.16*	−0.08	−0.20**
External 0-30 days	0.06	−0.04	0.19*	−0.11	0.25**	−0.05	−0.05
Office 0-30 days	0.27***	0.14	−0.12	0.12	−0.04	−0.01	−0.20*
Hospital stay 0-5 days	0.22***	0.02	−0.04	−0.06	0.10	0.09	−0.04
Hospital stay 6-90 days	0.09	−0.10	−0.08	−0.05	0.10	−0.08	−0.15
	TIA+						
Angina+							
Initial ED $	0.57***	0.39***	−0.03	0.08	−0.03	0.03	0.08
ED 0-5 days	0.05	0.19*	−0.24**	0.02	−0.10	−0.14	0.06
ED 6-90 days	−0.05	−0.21**	0.01	0.08	−0.09	0.03	0.04
External 0-30 days	−0.11	0.13	0.17	−0.01	−0.21*	0.06	0.11
Office 0-30 days	0.00	0.14	0.13	0.07	0.03	0.12	0.09
Hospital stay 0-5 days	−0.17**	−0.06	0.11	−0.04	−0.02	0.05	0.23**
Hospital stay 6-90 days	−0.08	0.00	0.20*	−0.02	0.09	−0.08	0.26**
	Appendicitis+						
TIA+							
Initial ED $	0.44***	0.17*	0.09	0.06	0.13	0.29***	0.07
ED 0-5 days	0.00	0.10	0.10	−0.08	0.05	−0.12	−0.16
ED 6-90 days	0.06	0.09	0.12	−0.08	−0.01	−0.03	0.14
External 0-30 days	−0.13	0.21**	0.04	0.03	−0.20*	0.14	−0.10
Office 0-30 days	0.05	0.04	−0.04	−0.07	0.10	0.13	0.00
Hospital stay 0-5 days	−0.15*	0.02	−0.06	−0.12	−0.17*	0.21**	−0.20**
Hospital stay 6-90 days	0.07	−0.05	0.12	0.01	−0.32***	0.16	0.25**

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A8: Correlation in physician effects across base and plus samples (restricted set of physicians)

	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 $	0.41***	0.11**	−0.03	0.06	0.26***	−0.03	−0.17***
ED 0-5 days	−0.03	−0.02	−0.19**	−0.03	0.11	−0.20**	−0.18**
ED 6-90 days	−0.02	−0.01	0.24***	0.07	0.07	−0.20**	0.09
External 0-30 days	−0.04	−0.21**	−0.02	0.29***	0.07	−0.18*	0.05
Office 0-30 days	−0.00	0.25***	−0.14*	−0.16	0.20**	0.08	−0.07
Hospital stay 0-5 days	0.02	−0.12	0.25***	0.19*	0.02	0.25***	0.26***
Hospital stay 6-90 days	−0.14	−0.05	0.14*	−0.22**	−0.30***	−0.05	0.09
	Appendicitis						
Appendicitis+							
Initial ED $	0.39***	0.21***	0.12	0.00	0.08	−0.03	0.02
ED 0-5 days	0.13*	0.45***	−0.11	0.08	0.24***	−0.27***	0.07
ED 6-90 days	0.01	−0.06	0.34***	0.08	0.20**	−0.16*	0.13
External 0-30 days	−0.05	0.17***	0.09	0.35***	0.27***	−0.15**	0.30***
Office 0-30 days	0.02	0.05	0.38***	−0.02	0.33***	−0.10	−0.13
Hospital stay 0-5 days	−0.00	−0.12	−0.10	−0.02	−0.06	0.35***	0.01
Hospital stay 6-90 days	−0.18**	−0.27***	0.02	−0.04	0.14	0.07	0.19
	TIA						
TIA+							
Initial ED $	0.31***	−0.21**	−0.16	−0.19*	0.01	−0.07	−0.05
ED 0-5 days	0.13	0.37***	−0.07	0.11	−0.06	−0.14	−0.23*
ED 6-90 days	0.27**	0.08	0.11	0.15	−0.01	0.34**	−0.14
External 0-30 days	−0.21**	0.09	0.16	0.28***	−0.11	−0.04	−0.02
Office 0-30 days	0.15	−0.09	0.11	0.07	0.02	−0.11	−0.18
Hospital stay 0-5 days	−0.16	−0.26***	0.09	0.08	−0.18	0.22**	−0.03
Hospital stay 6-90 days	−0.00	−0.26**	0.28**	−0.01	0.08	−0.07	0.07

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A9: Correlation in physician fixed effects within samples (all physicians at EDs)

	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	0.20***	1					
ED 6-90 days	−0.09	0.00	1				
External 0-30 days	0.24***	0.14*	−0.08	1			
Office 0-30 days	0.03	−0.00	0.19**	−0.29***	1		
Hospital stay 0-5 days	0.32***	0.08	−0.10	0.42***	−0.21**	1	
Hospital stay 6-90 days	−0.12	0.03	0.31***	0.00	−0.22**	0.24***	1
Appendicitis+ sample							
Initial ED $	1						
ED 0-5 days	0.22*	1					
ED 6-90 days	0.24*	0.30**	1				
External 0-30 days	0.18	0.30***	0.17	1			
Office 0-30 days	0.36***	0.09	0.12	0.11	1		
Hospital stay 0-5 days	0.42***	−0.09	0.02	0.09	0.06	1	
Hospital stay 6-90 days	0.06	0.02	0.52***	0.03	−0.16	0.18	1
TIA+ sample							
Initial ED $	1						
ED 0-5 days	0.19*	1					
ED 6-90 days	0.03	0.02	1				
External 0-30 days	0.10	0.24**	0.05	1			
Office 0-30 days	−0.05	−0.30***	−0.02	0.01	1		
Hospital stay 0-5 days	0.06	0.12	0.07	0.23**	−0.13	1	
Hospital stay 6-90 days	0.12	0.05	0.30***	0.19**	−0.17	0.16	1

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A10: Correlation in physician fixed effects across conditions (all physicians at EDs)

	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***	0.23***	−0.10	0.01	0.15*	−0.08	−0.22**
ED 0-5 days	0.14*	0.35***	0.09	−0.15	−0.02	−0.27***	−0.12
ED 6-90 days	0.05	−0.05	0.02	−0.10	0.15	−0.08	−0.18*
External 0-30 days	0.07	0.02	0.19**	−0.08	0.18*	−0.07	−0.05
Office 0-30 days	0.13**	0.32***	−0.04	0.12	−0.14*	−0.02	−0.15*
Hospital stay 0-5 days	0.17**	−0.03	0.03	−0.06	0.09	0.10	0.01
Hospital stay 6-90 days	0.12	−0.13	−0.03	−0.10	0.09	−0.03	−0.00
	TIA+						
Angina+							
Initial ED $	0.62***	0.34***	0.00	0.08	−0.04	0.06	0.12
ED 0-5 days	−0.00	0.22**	−0.20**	0.04	−0.15	−0.11	0.03
ED 6-90 days	−0.06	−0.23**	−0.05	0.05	−0.09	0.01	0.05
External 0-30 days	−0.10	0.12	0.13	−0.01	−0.19*	0.07	0.10
Office 0-30 days	0.00	0.08	0.08	0.06	0.09	0.04	−0.00
Hospital stay 0-5 days	−0.17**	−0.04	0.15	−0.05	−0.04	0.01	0.23**
Hospital stay 6-90 days	−0.11	0.01	0.20**	−0.04	0.08	−0.09	0.25**
	Appendicitis+						
TIA+							
Initial ED $	0.46***	0.16*	0.12	0.10	0.16*	0.23***	0.11
ED 0-5 days	−0.00	0.18*	0.11	−0.02	0.16*	−0.10	−0.13
ED 6-90 days	0.05	0.06	0.08	−0.10	−0.05	0.00	0.16*
External 0-30 days	−0.11	0.18**	0.04	0.01	−0.11	0.11	−0.08
Office 0-30 days	0.06	0.06	−0.01	−0.08	0.02	0.08	−0.02
Hospital stay 0-5 days	−0.14*	−0.03	−0.07	−0.08	−0.06	0.17*	−0.18**
Hospital stay 6-90 days	0.01	−0.07	0.09	0.00	−0.17*	0.09	0.21**

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A11: Correlation in physician effects across base and plus samples (all physicians at EDs)

	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 $	0.35***	−0.03	0.09**	0.05	0.22***	−0.04	−0.13***
ED 0-5 days	0.09	0.14**	−0.14**	0.02	−0.07	−0.02	−0.06
ED 6-90 days	0.12	0.01	0.26***	0.09	0.13	−0.21**	0.06
External 0-30 days	−0.03	−0.22***	0.11	0.26***	0.12	−0.14*	0.04
Office 0-30 days	0.10	0.14*	−0.14*	−0.12	0.09	0.07	−0.13*
Hospital stay 0-5 days	−0.03	−0.19**	0.05	0.16*	−0.02	0.25***	0.28***
Hospital stay 6-90 days	−0.07	0.08	0.09	−0.13	−0.24***	0.02	0.17**
	Appendicitis						
Appendicitis+							
Initial ED $	0.51***	−0.03	−0.22*	0.04	0.12**	0.06	−0.18*
ED 0-5 days	0.02	0.47***	0.09	0.05	0.15**	−0.17**	0.13
ED 6-90 days	−0.03	0.00	0.34*	0.07	0.18**	−0.10	0.17
External 0-30 days	−0.03	0.15	0.08	0.35***	0.24***	−0.14**	0.23*
Office 0-30 days	0.31***	−0.20**	−0.24*	0.05	0.33***	0.04	−0.38***
Hospital stay 0-5 days	0.22**	−0.31***	−0.43***	0.02	−0.00	0.35***	−0.25**
Hospital stay 6-90 days	−0.22*	−0.16	0.13	−0.04	0.11	0.03	0.24
	TIA						
TIA+							
Initial ED $	0.47***	−0.06	−0.03	−0.21**	−0.03	0.17*	0.16*
ED 0-5 days	0.08	0.36***	−0.05	−0.01	−0.13	−0.23**	−0.09
ED 6-90 days	−0.26***	−0.15	−0.11	−0.09	−0.03	−0.00	−0.01
External 0-30 days	−0.15*	0.12	0.16*	0.25***	−0.05	−0.11	0.08
Office 0-30 days	−0.04	−0.11	0.15	0.11	0.28***	−0.06	−0.08
Hospital stay 0-5 days	−0.04	−0.16*	0.15	−0.04	−0.08	0.35***	0.10
Hospital stay 6-90 days	0.09	−0.14	0.22**	0.03	0.15	−0.10	0.08

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A12: Correlation in physician fixed effects within samples (no controls for age, gender, or SES)

	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	0.24***	1					
ED 6-90 days	−0.10	0.12	1				
External 0-30 days	0.21***	0.14	−0.08	1			
Office 0-30 days	0.04	−0.19**	0.19*	−0.31***	1		
Hospital stay 0-5 days	0.19**	0.09	−0.07	0.51***	−0.13	1	
Hospital stay 6-90 days	−0.12	0.06	0.35***	0.05	−0.07	0.34***	1
Appendicitis+ sample							
Initial ED $	1						
ED 0-5 days	0.51***	1					
ED 6-90 days	0.41***	0.28**	1				
External 0-30 days	0.29***	0.33***	0.22**	1			
Office 0-30 days	0.32**	0.28*	0.17	0.16	1		
Hospital stay 0-5 days	0.29***	0.12	0.12	0.18	−0.15	1	
Hospital stay 6-90 days	0.26**	0.14	0.55***	0.12	−0.13	0.39***	1
TIA+ sample							
Initial ED $	1						
ED 0-5 days	0.20**	1					
ED 6-90 days	0.13	−0.02	1				
External 0-30 days	0.19*	0.22**	0.01	1			
Office 0-30 days	−0.00	−0.21*	0.07	0.07	1		
Hospital stay 0-5 days	0.12	0.11	0.11	0.28***	−0.10	1	
Hospital stay 6-90 days	0.11	−0.02	0.25**	0.23**	−0.10	0.27**	1

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A13: Correlation in physician fixed effects across conditions (no controls for age, gender, or SES)

	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 $	0.56***	0.26***	−0.06	0.10	0.12	−0.00	−0.18**
ED 0-5 days	0.10	0.31***	0.11	−0.17*	−0.07	−0.28***	−0.16*
ED 6-90 days	0.06	−0.01	0.06	−0.10	0.15	−0.08	−0.16*
External 0-30 days	0.04	0.01	0.17*	−0.05	0.13	−0.10	−0.07
Office 0-30 days	0.11*	0.33***	−0.01	0.11	−0.14*	0.01	−0.12
Hospital stay 0-5 days	0.16**	0.01	0.04	0.01	0.10	0.07	0.00
Hospital stay 6-90 days	0.12	−0.04	0.03	−0.04	0.07	−0.02	0.03
	TIA+						
Angina+							
Initial ED $	0.60***	0.29***	0.02	0.06	−0.02	0.06	0.09
ED 0-5 days	0.00	0.21**	−0.17*	0.07	−0.12	−0.05	0.00
ED 6-90 days	−0.09	−0.23**	−0.10	0.05	−0.09	0.03	0.00
External 0-30 days	−0.09	0.07	0.15	−0.05	−0.23**	0.03	0.11
Office 0-30 days	0.02	0.09	0.08	0.09	0.12	0.13	0.06
Hospital stay 0-5 days	−0.14*	−0.07	0.17*	−0.04	0.01	0.02	0.22**
Hospital stay 6-90 days	−0.11	−0.03	0.21**	−0.03	0.09	−0.08	0.21**
	Appendicitis+						
TIA+							
Initial ED $	0.38***	0.12	0.10	0.06	0.06	0.24***	0.13
ED 0-5 days	−0.04	0.14	0.08	0.01	0.13	−0.07	−0.12
ED 6-90 days	0.02	0.04	0.08	−0.07	−0.03	−0.01	0.10
External 0-30 days	−0.10	0.15*	0.06	0.04	−0.14*	0.14	−0.06
Office 0-30 days	0.07	0.07	−0.03	−0.04	0.00	0.07	−0.05
Hospital stay 0-5 days	−0.14*	−0.07	−0.08	−0.09	−0.15*	0.20**	−0.09
Hospital stay 6-90 days	−0.00	−0.08	0.08	0.02	−0.22***	0.14	0.24**

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table A14: Correlation in physician effects across base and plus samples (no controls for age, gender, or SES)

	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 $	0.35***	−0.01	0.01	0.05	0.16***	−0.07	−0.14***
ED 0-5 days	−0.01	0.04	0.03	−0.03	0.11	−0.19**	−0.01
ED 6-90 days	0.05	0.01	0.37***	0.09	0.08	−0.18**	0.16**
External 0-30 days	−0.07	−0.24***	0.07	0.22**	0.00	−0.18**	0.11
Office 0-30 days	0.08	0.15*	−0.01	−0.09	0.27***	0.07	−0.17**
Hospital stay 0-5 days	−0.04	−0.11	0.16*	0.11	−0.12	0.21***	0.30***
Hospital stay 6-90 days	−0.13*	0.07	0.08	−0.12	−0.40***	0.00	0.27***
	Appendicitis						
Appendicitis+							
Initial ED $	0.41***	0.26***	0.19**	0.08	0.03	0.03	0.03
ED 0-5 days	0.05	0.45***	−0.08	−0.02	0.12*	−0.30***	0.01
ED 6-90 days	0.05	−0.05	0.40***	0.16	0.21**	−0.20**	0.19
External 0-30 days	−0.06	0.15**	0.15*	0.37***	0.07	−0.18**	0.15
Office 0-30 days	0.00	0.03	0.33***	−0.08	0.34***	−0.03	−0.15
Hospital stay 0-5 days	−0.03	−0.11	−0.09	0.10	−0.10	0.42***	0.12
Hospital stay 6-90 days	−0.17**	−0.24***	0.02	0.30***	0.18**	0.15	0.43***
	TIA						
TIA+							
Initial ED $	0.37***	−0.07	−0.11	−0.12	−0.02	−0.05	0.07
ED 0-5 days	0.05	0.30***	−0.06	0.01	−0.16	−0.08	−0.15
ED 6-90 days	−0.17	0.02	0.19	0.12	0.13	0.15	0.06
External 0-30 days	−0.07	0.03	0.20**	0.32***	−0.10	−0.03	0.08
Office 0-30 days	−0.14	−0.18	0.03	0.03	0.33***	−0.15	−0.07
Hospital stay 0-5 days	−0.05	−0.18**	0.13	0.09	0.19**	0.36***	−0.05
Hospital stay 6-90 days	0.08	0.03	0.27**	0.14	0.08	0.05	0.23**

Note: *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

On-line Appendix B: Monte Carlo inference methods

In this appendix, we discuss the calculation of standard errors for Tables 4 through 7, and corresponding tables in the appendix. To estimate the equations jointly, we use a Seemingly Unrelated Regressions (SUR) model (Zellner, 1962). More specifically, in order to get an estimate of the 90th to 10th percentile difference (and its corresponding standard error) in the physician-led team contribution to a given outcome in Table 4, we first estimate the linear fixed effect model (for a given hospital and given outcome variable) and recover the parameter estimates (fixed effects) and estimated matrix of variance/covariance. We then demean these estimated fixed effect and calculate the difference between the 90th and 10th percentile fixed effects using the point estimates. Next we draw fixed effects (one for each physician) from a multivariate normal distribution. We then take the difference in the 90th and 10th percentile fixed effect. We repeat this last step for a total of 2000 draws (i.e., 2000 differences) and calculate the variance/standard error of these 2000 differences.

In order to estimate the correlations across fixed effects and their corresponding standard errors in Tables 5 through 7 (and corresponding tables in the appendix), we first estimate the set of linear fixed effect regressions (one for each outcome variable) using Seemingly Unrelated Regressions (SUR). We then take 2000 draws of the fixed effects from a multivariate normal distribution with mean equal to the vector of point estimates and with the variance-covariance matrix of those estimates. We then demean these point estimates (by the hospital-specific average) and create vectors of the point estimates that are demeaned by hospital for each hospital and dependent variable. We then calculate the correlation coefficients by taking the correlations between the vectors of demeaned fixed-effects. We then estimate the standard errors of the correlations by computing the standard deviation of the resulting vector of correlation coefficients for each pair of dependent variables over the 2000 draws.