PHYSICIAN PRACTICE STYLE AND HEALTHCARE COSTS:
EVIDENCE FROM EMERGENCY DEPARTMENTS

Gautam Gowrisankaran
Keith A. Joiner
Pierre-Thomas Léger

ABSTRACT

We examine the variation across emergency department (ED) physicians in their resource use and health outcomes, and the relationship between ED resource use and future healthcare costs and outcomes. Our data record the initial treating hospital, ED physician, ED billed expenditures, and all interactions with the provincial health system within the subsequent 90 days for EDs in Montreal, Canada. Physicians in Montreal rotate across shifts between simple and difficult cases, implying a quasi-random assignment of patients to physicians conditional on the choice of ED. We consider three medical conditions that present frequently in the ED and for which mistreatment can result in dramatic consequences: angina, appendicitis, and transient ischemic attacks. To control for variation across physicians in their diagnostic acumen, for each condition, our sample consists of patients with a broader set of symptoms and signs that could be indicative of the condition. We regress measures of healthcare costs on indicators for the hospital and ED physician separately by condition. We then evaluate the correlations between different measures of skill and resource use. We find strong positive correlations of physician resource use and skills across the three conditions. However, physicians with costly practice styles are often associated with worse outcomes, in terms of more ED revisits and more hospitalizations. One exception is that for patients in the angina sample, ED physicians with more spending have fewer hospitalizations. Comparisons of physician effects for the base and broader sets of conditions show that both diagnosis and disposition skills are important.

Gautam Gowrisankaran
Department of Economics
University of Arizona
P.O. Box 210108
Tucson, AZ 85721-0108
and HEC Montreal
and also NBER
gowrisankaran@eller.arizona.edu

Keith A. Joiner
College of Medicine and Department of Economics
University of Arizona
kjoiner@email.arizona.edu

Pierre-Thomas Léger
Department of Health Policy
and Administration School of Public Health
University of Illinois, Chicago
ptleger@uic.edu

1 Introduction

An influential literature has found that there is substantial variation across physicians in their spending levels and that this variation is not easily explained by differences in patient health status, prices, or geography (Newhouse et al., 2013). This literature has also examined the relationship between spending and health outcomes, with a meta-analysis concluding that there is no consistent relation between the two (Hussey et al., 2013).

The goals of our study are to further understand the variation across physicians in resource use and health outcomes and in particular, the relation between spending and outcomes, by using a setting where the initial assignment to physician is quasi-random and where we can associate the physician of initial treatment with healthcare usage over the medium run. To what extent is there variation in physician resource use and treatments for similar patients? Do physicians who spend more have better outcomes? If so, do the better outcomes occur because spending facilitates more accurate diagnoses, more appropriate treatments, or both?

We evaluate the impact of physician resource use in the context of emergency department (ED) care in Canada. We make use of a unique dataset that records all ED visits for patients in the Montreal metropolitan area for a nine month period in 2006.[1] Importantly, our data include every encounter with the provincial health system in the 90 days after the initial ED visit, which we use to identify the impact of the ED physician on healthcare outcomes over the medium run. We also observe the initial assignment of ED physician, which allows us to evaluate the extent to which different physicians within the same hospital have different spending patterns and outcomes. Importantly, because of institutional features of EDs in Montreal (and in most North American hospitals), cases in an ED are assigned on a quasi-random basis. We exploit the quasi-random assignment of patients to physicians to derive treatment effects of the ED physician in a way that minimizes issues of physician selection. Our setting also allows us to identify variations in physician behavior that are unrelated to financial incentives.[2]

We focus on EDs because of the quasi-random assignment of patients but also because a number of specifics about EDs imply that an evaluation of the impact of ED care is of significant policy

---

[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.

[2] ED physicians in Montreal are paid on a fee-for-service basis that is same across patients. Thus, financial incentives will be identical across patients within an ED.

interest. In particular, 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.

Our paper estimates the extent of variation in measures of physician skill and practice style and the relation between these measures. We focus on three illnesses: angina, appendicitis, and transient ischemic attacks (TIAs). 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. If not diagnosed and managed, appendicitis can lead to a rupture of the appendix, which has serious consequences. Finally, a TIA is a temporary and partial occlusion of blood supply to the brain and is sometimes called a mini-stroke. It is a precursor to a complete occlusion of the blood supply, which would lead to a full-blown stroke. We consider these conditions for three principal reasons. First, they all require a relatively high level of physician acumen for diagnosis and disposition and as such are likely to capture differential physician ability and/or preferences. Second, inappropriate treatment of these conditions can lead to dramatic consequences, e.g., misdiagnosing a TIA as a less emergent condition such as a migraine leads to the possibility of a stroke that could have been avoided with accurate diagnosis and disposition. Finally, the symptoms, signs, and ancillary data underlying these conditions will be seen with reasonable frequency in an ED.

To motivate our empirical work, we propose a simple model of the roles of the ED physician. In our model, the ED physician has two central roles: diagnosis and disposition. She first attempts to diagnose the illness of the patient and then chooses patient disposition. At both of these steps, she makes decisions about resource utilization, 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.

3

Physicians may vary both in the expected outcomes of their patients—which we call "physician skill"—and in their spending for the care that they provide—which we call "physician resource use" or "physician practice style". It is possible that physician skill and resource use are positively correlated, but the correlation between these two measures could just as well be negative; ultimately understanding this relation requires empirical analysis. Thus, our estimation seeks to uncover different measures of physician skill and resource use. We consider these attributes separately for the three different conditions discussed above and then analyze the correlation between the attributes of a given physician across different conditions.

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. If we compare resource use or outcomes across physicians conditional on a diagnosis of TIA, this will identify the treatment effects of each ED physician on these measures. We can then correlate treatment effects of the fixed effects for spending and outcomes to understand whether physicians with an expensive practice style disproportionately reduce future healthcare consumption such as ED revisits. Variation across physicians in their patient outcomes would relate exclusively to their decisions regarding patient disposition.

However, this model is limited because 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 both patient diagnosis and disposition are challenging and vary across physicians. For instance, suppose 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 coded with a TIA for 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 skills and outcomes 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

4

with TIA—such as migraines or epilepsy—while TIA (the "base" sample) only includes patients with a primary diagnosis of TIA. Given that the three base illnesses that we consider are more specific than the alternatives in their plus samples, we assume that some unknown fraction of patients in the plus sample has the base condition. 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 a hospital. We believe that this is reasonable because our plus samples include the most common *differential diagnoses*[3] upon presentation with the symptoms and signs of the base diagnosis and because the list of differential diagnoses is broad enough that there would be relatively few patients from outside the plus group of diagnoses whose records would indicate these diagnoses. We also examine this proposition by verifying that characteristics of the plus samples are similar across physicians.

For each of the three conditions that we consider, our empirical analysis proceeds by estimating linear regressions of a number of measures of 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 and measures of subsequent encounters with the healthcare system, including future ED visits, office visits, and hospital admissions. 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 resource use. The fixed effect on later measures will be more or less indicative of physician skill, depending on the measure. For instance, a positive effect on future ED visits would generally be seen as a 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 demean all physician fixed effects relative to the mean at their ED. In all cases, we calculate variances for our estimated physician indicators and then account for the variance of these indicators in testing our hypotheses regarding the relationship between physician skill and resource use.

---

[3]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.

Overview of results. For all three conditions, we find that there is large variation in both physician resource use during the initial ED visit and physician skills (measured in terms of outcomes) within EDs. In particular, we find that the differences between physicians' contribution to 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. Furthermore, measures of future healthcare usage also vary significantly between the 10th and 90th percentile. For example, we find that the differences in physician's contribution 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, we find that 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 as well as for the 6-90 day post-ED-visit hospital costs.

Interestingly, physician resource use (during the initial ED visit) and physician skills (measured as the contribution to desired outcomes, or equivalently, the contribution to avoiding bad outcomes) are often negatively correlated. For example, physicians who 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, physicians who 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. One significant exception is that physicians who spend more during the initial ED when treating angina+ are associated with fewer hospitalizations (correlation of -0.18) during the 6-90 day period.[4] Overall, these results are consistent with physicians using extra resources during the initial ED visit when they are uncertain about appropriate diagnosis and disposition.[5]

We find that physician resource use 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 across the

---

[4]Similarly, Currie et al. (2016) find that cardiologists who use invasive procedures for heart attacks more frequently have better outcomes.

[5]This finding is also consistent with Doyle et al. (2010), who find that patients randomly assigned to the physician team from the higher-ranked residency program incurred lower costs and indistinguishable outcomes from those assigned to the team from the lower-ranked program.

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). Overall, these results suggest that physician resource use and skills are highly correlated across illness categories.

We next examine the correlations in practice styles and skills across the plus and base categories. We find positive and statistically significant correlations in the physician practice style, as measured by initial ED spending, across the angina+/angina (0.37), appendicitis+/appendicitis (0.41) and TIA+/TIA samples (0.44). That is, physicians associated with more costly practice styles when treating all patients in the plus sample are also associated with more costly practice styles when treating patients diagnosed with the base sample. We also find a positive correlation between many of the same skills across the "plus" and "base" samples. For example, physicians who are associated with more ED revisits during the first 5 days for the appendicitis+ sample are also associated with more ED revisits during the same time period for the appendicitis sample, with a correlation coefficient of 0.43. Since the treatment effects on the plus samples reflect both diagnosis and disposition skills while the treatment effects on the base samples reflect disposition skill only (though, conditional on a selected sample), our results are consistent with proper disposition, and not just appropriate diagnosis, being important. We also find that later outcomes are more strongly correlated with the initial ED spending for the plus samples than for the base samples. Together with the fact that diagnosis plays a key role in outcomes only for the plus samples, our results are consistent with diagnosis skills also being important.

Related literature. Our paper contributes primarily to two recent literatures. One literature

has leveraged the quasi-random assignment of patients to hospitals and physician groups in order to evaluate different aspects of physician practice style and resource use across different healthcare settings (see, e.g., Epstein and Nicholson, 2009; Doyle et al., 2010, 2014; Currie et al., 2016; Currie and MacLeod, 2017). 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.

Another literature has considered physician inputs in the ED setting. Chan (2016) examines how physicians in emergency department settings respond to different management systems when making care decisions, using as an outcome the amount of time an ED physician spends with her patient. 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 higher mortality areas. Silver (2016) examines variations in physician productivity by estimating within-physician productivity differences based on shifts. Using variation in patient assignment similar to our study, Van Parys (2016) finds that practice styles for minor injuries vary substantially across physicians within an ED. Her data source does not allow her to correlate practice style with later outcomes. Other studies have specifically evaluated misdiagnosis in the ED (see, e.g., Hastings et al., 2009; Kachalia et al., 2007; Abaluck and Agha, 2014). These studies have typically restricted the sample to those subsequently identified to have an adverse event such as complication, death, or hospitalization, sometimes associated with a medical liability suit. We contribute to this literature by addressing a different question, regarding the relation between physician skills and practice styles across diseases. Our unique data—which records the physician of assignment in the ED and tracks healthcare utilization and expenditures over the medium run—allow us to add evidence here.

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. We present a series of robustness checks in Section 5. Finally, Section 6 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.[6] Almost all residents of Quebec are covered by RAMQ,[7] 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*). Although we use ED visits for residents of Montreal as the initial point of interest, we track virtually all future healthcare consumption for patients with an ED visit. Specifically, we observe each patient's subsequent consumption of care (across EDs, offices, hospital-based external clinics and hospitalizations) as long as it is within the provincial boundaries and covered by RAMQ.

In Quebec, during the study period, ED physicians were paid 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 does so 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 is only a proxy for costs, and 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.[8] Finally, our data do not include filled prescriptions.

---

[6]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).

[7]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.

[8]Additionally, imaging services are one area for which there exists a robust private market in Quebec, whose

For each patient/ED observation, our data include the patient's gender, age group, which is mostly in 5 year bins,[9] three-digit postal code, and two measures of socio-economic status, 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.[10]

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.[11] 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.

Future revisits were used by Montreal's public health department 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.

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 patients who arrive by ambulance) or the light cases (most often patients who enter through the

---

services are not included in our data.

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

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

[11]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.

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 provides 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.[12]

We observe patients seen at the 19 general acute care hospitals with EDs in Montreal.[13] 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. 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 the allocation of patients to physicians is affected by physicians' shift types (which we do not observe), 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 use two patient characteristics that will be observable to us 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 the non-random assignment of patients.[14] We proceed by regressing

---

[12]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.

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

[14]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.

11

the patient characteristic of interest (gender or age group) on physician dummies, with indicators for month and weekend visit. We perform the regressions separately for each ED. We then test whether the physician fixed effects are statistically different from each other.

Table 1: P values for randomness

| ED number | Full sample | | | Restricted sample | | |
	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, columns 1-3, 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 (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). In order to get a sense of whether differences uncovered above are meaningful, we examine 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 the F test results may be driven in part by a few atypical physicians. The presence of apparent outliers is consistent with the fact that certain older physicians—especially

those without advanced emergency-medicine training—may be assigned, on average, less-severely ill patients. Illness severity may, in turn, be correlated with a patient's gender and/or age group, which are the regressors of interest in Table 1.

In light of this possibility, we eliminate ED physicians whose mean gender or mean age are significantly different than peers within their ED.[15] More specifically, we drop physicians whose mean gender is more than 5 percentage points from the mean for their EDs 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 call into question the randomness of patient-physician pairs within them. As a result, we drop these four EDs all together. We then rerun the same two F tests on the remaining EDs while excluding the atypical physicians. We present results from these new F tests—on the "restricted" sample of ED physicians—in the last three columns of Table 1. 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 estimation sample includes 7 EDs that do not violate the randomness assumptions: 1, 4, 7, 8, 9, 15, 17.

As further robustness checks, On-line Appendix A provides results including the full set of physicians at these 7 EDs as well as removing more physicians based on the proportion of patients that they treat who might have one of our selected conditions. The appendix tables show that our results are not sensitive to the inclusion criteria of physicians.

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). Recall that we consider these three illnesses because they may be difficult to diagnose, may pose challenges for proper disposition, and are frequently seen in the ED.

Angina is a precursor to myocardial infarction (heart attack). It reflects 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

---

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

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. While pain in the lower right portion of the abdomen (right lower quadrant) is a 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).

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. Our goal is to choose a plus set of illnesses that is sufficiently broad that ED physicians would almost certainly diagnose the base illness inside the plus set but also sufficiently narrow that physician practice styles are similar within the plus category.

Table 2 presents the plus diagnoses for each of our three base diagnoses. We define a patient as being part of the plus sample if she was diagnosed in her initial ED visit with any of the diagnoses listed here, which include the base diagnosis and the diagnoses that are the most likely confounders for each base diagnosis. We include the most frequent differential diagnoses for these base diagnoses, as reported by the medical literature (Hollander and Chase, 2016; Martin, 2017; Nadarajan et al., 2014).

Table 3 presents summary statistics on the set of EDs, physicians, and patients in our estimation sample. The estimation samples for the plus diagnoses includes 7 EDs (as noted above), 90 physician-ED pairs,[16] and a total of 15,098 initial ED visits and 12,855 unique patients included. 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.

These tables report statistics on the set of dependent variables used in the estimation section. More specifically, "ED \$" denotes the billed spending during the initial ED visit, which ranges from a mean of \$45.94 to \$68.91 across the three samples. "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 leads to 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. Although we have data on all these dependent variables separated into the categories of 0-5 days, 6-30 days, and

---

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

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

31-90 days, we split the ED visits and hospital stays into these categories because we wanted to understand the differences between ED revisits and hospitalizations that are directly related to the initial visit—and which would then likely occur within five days—and other ones.

The table shows that hospitalizations are quite frequent across all all illness categories. Finally, the "Hospital $" measures denote hospital spending—as measured by NIRRU—during the time period described. The mean contribution to hospitalization costs are \$4,521, \$3,369 and \$5,616 or the angina+, appendicitis+ and TIA+ samples, respectively. Although we do not have cost measures across different healthcare venues, we believe that the bulk of the healthcare resource used by individuals over the 90-day follow up period are for hospitalizations.

The second panel of Table 3 considers the three base samples. 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 a fixed effect. 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 full estimation sample.

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.

Table 3: Summary statistics for 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 Model and estimation framework

In this section, we develop a simple model of ED patient treatment, health outcomes, and healthcare expenditures. We then use the model to develop an estimation framework.

3.1 Model

The basic premise of our theoretical model is that the most important function of the ED physician is the appropriate disposition of the patient. However, appropriate disposition of the patient relies, in part, on the ability to make an accurate diagnosis. On 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 aggressive 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. Our model is meant to be fairly general, and can apply to any case where appropriate diagnosis and disposition of the patient are difficult.

Consider a patient i who experiences an illness θ_i and then presents at the ED. The ED has J attending physicians and the patient will be treated by one of them. Denote the treating physician 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 the patient 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 it is potentially difficult for the ED physician to make definitely a correct diagnosis.[17] 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

---

[17]For example, a patient who may suffered from a transient ischemic attack (TIA) may present herself at the ED having experienced visual or speech changes, weakness, or numbness, which are consistent with the differential diagnoses.

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 physician's information.[18]

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 mean 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:

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

We do not know 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.

[18]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 seek to use our framework and data to evaluate each ED physician's average contribution to ED resource use during the initial ED visit (which we call the physician's *practice style*) as well as to desired outcomes (which we call the physician's *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. Note that these attributes are partly vertical and partly horizontal. In particular, some physicians may over-treat patients with θ^D, resulting in too many procedures and high expenditures, but good health outcomes; while other physicians may under-treat patients with θ^B, resulting in low expenditures in the ED but potentially adverse health outcomes (and healthcare expenditures) down the road.

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:

$$E[H_{ij}^*] = \sum_{j=1}^{J} d_{ij} p_{ij} \left(T^B | \theta^B\right) P_i(\theta^B) \left[v(\theta^B, T^B) - v(\theta^B, T^D)\right] +$$

$$\overline{H}_i + P_i(\theta^D)(v(\theta^D, T^D) + \overline{v}(\theta^D)) + P_i(\theta^B)(v(\theta^B, T^D) + \overline{v}(\theta^B)) + \varepsilon_i,$$

$$(3)$$

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.[19]

---

[19]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.

3.2 Estimation framework

Using our model as a basis, we seek to 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 now discuss our basic regression analyses. 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 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 socio-economic status, age, and gender of the patient, in addition to an indictor for the treating ED physician. We include controls for age, gender and socio-economic status 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 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. Our regressions for dollar measures $E[D_i^*]$ take an identical form. We then correlate different fixed effects to examine the relation between different measures of practice style and skills for a given physician.[20]

Our interpretation of physician fixed effects is as some mix of practice style and skill, depending on the particular measure. In particular, we interpret spending in the initial ED visit as a measure of practice style; by definition, physicians who spend more for identical patients are ones who have

---

[20]The reported significance on our correlation measures use Monte Carlo methods to account for the fact that the fixed effects are estimated.

a practice style that uses more resources. We consider some other measures— notably revisits to an ED—as 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.

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. Thus, identification relies on the underlying health status of the patient being mean independent from her assignment to ED physician. Focusing on a single ED, identification relies on patients being assigned randomly to EDs. 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.

Nonetheless, it is quite possible that patients choose different EDs based on ε_i. For instance, 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. We avoid endogeneity caused by a correlation between ε_i and the choice of ED by interpreting each physician fixed effect only relative to the mean for physicians at her hospital and not directly comparing the fixed effects across hospitals. We do not attempt to identify the mean FE at any ED, which may reflect some combination of mean physician skill and mean patient severity at that ED.

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 estimate the equations jointly in order to estimate the correlations across residuals of different equations, which is necessary for appropriate inference on the correlations between physician effects across different equations.

To estimate the equations jointly, we use a Seemingly Unrelated Regressions (SUR) model (Zellner, 1962). We estimate each pair of dependent variables with a separate SUR model. We demean each physician fixed effect relative to the mean fixed effects at her hospital to only compare the relative quality of physicians across EDs. We then compute standard errors on each correlation by simulating from the distribution of estimated fixed effects.

Finally, we discuss how we can separate disposition from diagnosis skills. Our base specifications include patients who are diagnosed with one of our plus conditions during the initial ED visit. As we discuss above, physician variation here will capture both diagnosis and disposition skills. However, we also consider specifications where the sample includes patients specifically diagnosed with one of the three base conditions. If diagnosis were identical across physicians for each patient, these regressions would exclusively capture disposition skills. However, we expect variation in diagnosis skills across physicians, implying that these regressions are less plausibly exogenous. Nonetheless, we still view these regressions as informative about disposition. Thus, the plus sample will inform us about diagnosis and disposition skills while the base sample will inform us about disposition skills, though admittedly on a selected sample.

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. Then, we consider the correlations between physician effects within samples. Following that, we examine correlations between physician effects across conditions. We then evaluate correlations between physicians effects across the base and plus samples for a condition. On-line Appendix A considers robustness of our main results to different outcomes and under different physician inclusion criteria.

4.1   Variation in physician effects

Table 4 reports the difference between the 90th and 10th percentile of physician effects within each ED, for each effect and separately for the angina+, appendicitis+, and TIA+ samples. The table also reports statistical significance for the differences in the physician effects between the physician

25

at the 90th percentile and at the 10th percentile.[21]

First, note that there is 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 overall mean in-sample spending of $53.30 (Table 3). These differences are even larger for the appendicitis+ and TIA+ samples.

Even more striking variations can be found when examining inpatient hospital stays. This is true across all EDs and all illness categories. 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 significant for 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. This is particularly true for the appendicitis+ sample, where three of the seven EDs have a significant 90-10 different for spending in 6-90 days and where two of the seven have a significant 90-10 difference in spending in the 0-5 day period.[22]

Interestingly, the 90-10 differences in hospital stays in the 6-90 period are larger for TIA+ than for angina+ and appendicitis+. This is true even though these three conditions have similar mean hospitalization rates (Table 3). We believe that the underlying reason is that identifying the medical causes behind TIA+ symptoms and signs is typically a more uncertain and ambiguous endeavor than identifying the medical causes behind angina+ or appendicitis+ symptoms. This is because angina+ and appendicitis+ typically have presentations directly referable to a specific anatomical location (e.g., heart, esophagus, appendix, and pelvic organs) while TIA+ invokes a variety of symptoms and signs that can involve the entire body.

---

[21]We compute statistical significance with a Monte Carlo process where we simulate from the joint distribution of the estimated fixed effects for the physicians at the 90th and 10th percentiles. In cases where the closest match to the 90th or 10th percentile is a tie, we break the tie randomly. All our simulations use 2,000 simulation draws. We verified that this number of draws is sufficient to perform accurate inference in our setting.

[22]Due to space constraints, we do not report hospital spending results in the remainder of the tables in the main paper, but instead report them in On-line Appendix A.

Table 4: Estimated 90-10 percentile in physician fixed effects

Dependent variable	ED 1	ED 4	ED7	ED 8	ED 9	ED 15	ED 17
Angina+ sample							
Initial ED $	45.79***	7.82***	16.07***	17.83***	18.27***	18.09***	12.43*
ED 0-5 days	0.18***	0.26***	0.14**	0.18***	0.11	0.18***	0.15*
ED 6-90 days	0.72***	0.18	0.58*	0.30*	0.33	0.21	0.42
External 0-30 days	0.37**	0.23	0.65***	0.31**	0.38***	0.22	0.44**
Office 0-30 days	0.19	0.24**	0.26	0.21**	0.21	0.18*	0.40***
Hospital stay 0-5 days	0.09**	0.15***	0.10**	0.07	0.12**	0.08**	0.11**
Hospital stay 6-90 days	0.17***	0.14**	0.18*	0.09**	0.11	0.08	0.11
Hospital $ 0-5 days	3,003***	2,003**	1,892	1,486	2,774*	1,434	1,817
Hospital $ 6-90 days	2,272	1,917	3,471**	2,266	1,567	1,734	1,724
Appendicitis+ sample							
Initial ED $	58.91***	38.31	35.03**	13.10**	20.91**	16.64**	12.72
ED 0-5 days	0.66***	0.73	0.52*	0.25**	0.36**	0.23*	0.34
ED 6-90 days	0.65	1.17	2.84***	0.67**	0.92**	0.28	0.63
External 0-30 days	0.85	1.19	0.89	0.63**	1.24***	0.67**	1.16*
Office 0-30 days	0.76**	4.06***	0.78*	0.26	0.58*	0.45*	0.63**
Hospital stay 0-5 days	0.35*	0.50	0.72***	0.16*	0.25**	0.27***	0.39**
Hospital stay 6-90 days	0.31*	0.54	0.77***	0.18	0.22	0.17	0.29
Hospital $ 0-5 days	5,138**	5,513	11,581***	1,218	2,361	3,169*	3,470
Hospital $ 6-90 days	3,135	14,169***	12,532***	2,826**	1,348	1,508	3,233
TIA+ sample							
Initial ED $	46.34***	17.99	19.64	18.08**	28.29***	23.78*	23.15*
ED 0-5 days	0.44**	0.31*	0.31**	0.17	0.28	0.23	0.34
ED 6-90 days	1.08*	0.66	1.79**	0.58	1.12	0.78	0.86
External 0-30 days	0.82**	0.20	0.55	0.39	0.98**	0.44	1.56***
Office 0-30 days	0.37	0.42	0.20	0.35*	0.57	0.56	0.52
Hospital stay 0-5 days	0.19*	0.18	0.19	0.24***	0.24*	0.22**	0.31**
Hospital stay 6-90 days	0.24	0.18	0.41**	0.11	0.39**	0.16	0.37**
Hospital $ 0-5 days	4,549	3,420	6,369**	6,144***	5,337	4,893*	6,123
Hospital $ 6-90 days	3,958*	2,949	3,910*	1,886	3,053	2,760	4,133

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

4.2 Relationships between practice styles and skills within samples

We now consider the relationships between the different effects for a physician. Table 5 reports the correlations between the physician fixed effects for different measures of practice styles and skills, separately for our three plus samples. All physician fixed effects are demeaned by hospital. The table also reports the statistical significance of each reported correlation.[23]

Our most striking result is that physicians who have higher mean spending in the initial ED visit—which we interpret as an intensive practice style—have worse skill 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 the appendicitis+. The relationship is also large and significant when considering revisits to EDs during the 6-90 day period for appendicitis+ (with correlation of 0.37). That is, patients potentially suffering from angina or appendicitis and assigned to an ED physician with relatively more services for this type of patient are more likely to revisit an ED within the first 90 days.

In addition, physicians who have 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. In contrast, 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. Similarly, for the appendicitis+ sample, physicians who have higher mean spending in the initial ED visits have more subsequent visits to both external-based hospital clinics (correlation of 0.26) and private physician offices (correlation of 0.28).

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. Angina is an acute manifestation of an underlying abnormal coronary vasculature, which is not reversed over a short time frame. Appropriate disposition (leading to tests such as a stress echocardiogram or coronary arteriogram outside of the ED and consequent

---

[23]We calculate statistical significance with a Monte Carlo process where we estimate each pair of outcomes together and then simulate from the joint distribution of the estimated fixed effects.

interventions) may be facilitated by more ED spending, and in return lead to fewer hospitalizations.

Overall, our findings suggest that more spending 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 physician practice style and skills will vary across conditions.

We also consider the correlation of skills measured across different outcomes, in order to understand whether there is substitutability or complementarity across different healthcare venues. For both appendicitis+ and TIA+, physicians who have more ED revisits during the first 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, physicians who 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 suggesting that, as expected, office visits substitute for ED visits. Finally, we find that, for all three samples, physicians 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. Equivalently, physician skills at avoiding ED revisits and hospitalizations correlate positively for all three illness categories. 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 the in-hospital care processes. The immediate availability of both laboratory and other ancillary data then creates a clearer path for the decision to admit.

4.3 Relationships between practice styles and skills across conditions

We next evaluate the correlations between physician practice styles and skills across illness categories. Table 6 reports correlations in physician fixed effects across the angina+ and appendicitis+ samples, across the angina+ and TIA+ samples, and across the appendicitis+ and the TIA+ samples, in three separate panels.

The most striking finding here is the strong correlation of practice styles across the three illness pairs. That is, physicians who 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 con-

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.

tribution (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. Taken together these results suggest that differences in spending during the initial ED visits are driven in part by the assignment to ED physician and that physicians with costly practice styles are generally costly across illness categories.

Not only do we find that practice styles correlate positively across illness categories, but we also find positive correlations for several skills across illness categories, all of which are shown along the diagonals in Table 6. 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. More specifically, 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 the across the appendicitis+ and TIA+ samples is 0.18. Furthermore, we find that physicians 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 physician 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 physician's practice styles—measured as spending in the initial ED visit—across the angina and angina+ samples (0.37), across the appendicitis and appendicitis+ samples (0.41) and across the TIA and TIA+ samples (0.44). We also find several positive correlations for fixed effects for future healthcare consumption especially across the plus and base samples (i.e., presented in the diagonals of the three sub-tables). As one example, for both appendicitis and TIA, physicians who are associated with more ED revisits in the 0-5 day period for the base samples are also associated with more 0-5 day ED revisits for the plus samples. There are similar results for ED revisits in the 6-90 day period for angina and appendicitis.

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.

Given that the fixed-effects associated with the plus samples reflect the physician's diagnosis and disposition skills while the fixed-effects associated with the narrow/severe samples reflect the physician's disposition skills, a strong positive correlation suggests that revisits to the ED during the first 90 days are driven (which holds for all three illness categories), at least in part, by differences in disposition skills rather than just diagnostic skills.

Similarly, physicians who are associated with more hospital stays within the first 5 days when treating the angina+, or appendicitis+ or 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 an 0.28, respectively). Again, these positive correlations suggest that hospitalizations, like ED revisits, are driven in part by differences in disposition skills rather than just diagnostic skills.

To understand the importance of diagnosis skill, we can compare the correlations between outcomes for the plus samples with the practice styles for the base samples (Table 7) and for the plus samples (Table 5), as measured by initial ED spending. For both angina and appendicitis, there are far more significant correlations between practice styles and outcomes for the plus samples than for the base samples. The fact that (i) practice style on the plus sample is more predictive of outcomes and (ii) this practice style includes both diagnosis and disposition suggests that diagnosis skills are also important.

5   Conclusions

In this paper, we use detailed data from emergency departments in Montreal, Canada to understand how 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 use the fact that the assignment of ED physicians in Montreal, conditional on choice of ED, is close to random. We develop this point fully in a theoretical model.

Using our data, we estimate the effect of being treated by different ED physicians on the costs during an initial ED visit and on a number of 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+).

Our results suggest that physicians vary in both their average contribution to ED physician-related costs (and number of services provided) during the initial ED visit as well as in their future outcomes. Thus, who one sees matters on both dimensions.

Importantly, for two of the three conditions that we consider, we find that physicians who systematically use more ED resources do not also systematically achieve better outcomes. More specifically, physicians with a high-cost practice style have higher ED revisits and hospitalizations for appendicitis+ and no significant differences in ED revisits or hospitalizations for TIA+. The results for angina+ are more ambiguous: physicians who 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. Thus, except perhaps for angina, being assigned to a physician who spends more on average at the ED (on physician-related services) does not increase one's chances at lower future care use. 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 physicians who spend more for one condition are also likely to spend more for the other conditions that we consider. Moreover, skills at keeping patients out of the ED also correlate across the three conditions that we consider. 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. More specifically, we find that physicians who are associated with bad outcomes when treating patients in the plus category also are likely to have bad outcomes when treating patients in the base category, where only disposition is likely to be important, but that initial ED spending on the plus samples, where diagnosis is also likely to be important, correlate more with future outcome measures than does initial ED spending on the base samples.

Our results suggest that achieving lower costs without sacrificing patient health may be feasible, perhaps through knowledge transfer and/or financial incentives. They also suggest that if physician practice styles and skills are difficult to influence, then randomizing patients to physicians may be suboptimal. That is, assigning patients to physicians based on the patient's illness category or severity and the physician's practice style and skills may yield lower costs and better outcomes for the population as a whole. Finally, our results are consistent with tools that improve physician's

diagnostic and disposition ability being likely to yield important benefits in both costs and outcomes.

References

Abaluck, J. and Agha, L. (2014). Negative tests and the efficiency of medical care: What determines heterogeneity in imaging behavior? Working Paper, Yale.

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

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 health care 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. (2014). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy*, Forthcoming.

Doyle, J., Wagner, T., and Ewer, S. (2010). Returns to physician human capital: Analzying 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`.

Hussey, P. S., Wertheimer, S., and Mehrotra, A. (2013). The association between health care quality and cost: A systematic review. *Annals of internal medicine*, 158(1):27–34.

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

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`.

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*, (25):2659–2623.

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

Newhouse, J. P., Garber, A. M., Graham, R. P., McCoy, M. A., Mancher, M., Kibria, A., et al. (2013). *Variation in Health Care Spending: Target Decision Making, Not Geography*. National Academies Press.

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.

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.

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 ischemic 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.

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 results

We provide results from a series of robustness checks that consider more outcome variables and narrower and wider sets of physicians. More specifically, we reestimate our models:

1. Adding hospital spending during the 0-5 and 6-90 post-ED visit periods as outcomes, instead of just hospital stays during these periods.

2. Eliminating physicians whose proportion of allocated patients with an angina+, appendicitis+ or TIA+ diagnosis was atypical.

3. Including all physicians who practiced in our final sample of EDs.

We now describe the results from each of these robustness checks.

Results including hospital spending as an outcome variable

Tables A1-A3 replicate Tables 5-7 with the inclusion of two extra outcomes: hospital spending in 0-5 and 6-90 days after the initial ED visit. Recall that our spending measures are the NIRRU proxies, as described in Section 2.

Tables A1 considers the correlations across practice styles and skills within illness category. We find fewer significant results with hospital spending, rather than hospital stays, as the outcome variable. For instance, we find no significant relationship between ED spending and hospital spending for any of the three illness categories and exclude outliers on these three dimensions.

Tables A2 considers the correlations across conditions. The results here show that hospital spending measures are similar to hospital stays, in terms of the correlations across conditions and samples. For instance, the correlation between hospitals stays in 6-90 days for angina+ and TIA+ is 0.26 while it is 0.19 for hospital spending.

Finally, Tables A3 considers the correlations across the base and plus samples. Here, we find stronger evidence of correlations between hospital spending between the base and plus samples than for hospital stays. Thus, this is also consistent with our finding that both diagnosis and disposition are important, for the three conditions that we consider.

Results excluding atypical physicians

As described in Section 2.2, in order to obtain a quasi-random assignment of patients, our estimation sample excluded physicians whose patient allocation failed an F test for random assignment

Table A1: 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 A2: 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 A3: 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.

based on gender and age with P=0.01. Another robustness check of the randomization assumption is to verify the proportion of patients in each physician's patient pool who were in the angina+, appendicitis+ and TIA+ sample.

Table A4 reports the percentage of a physician's patients with diagnoses of angina+, appendicitis+ or TIA+, for physicians at the 10th and 90th percentiles of these distributions at each ED. The table shows the presence of a small number of physicians with disproportionally large or small fractions of patients in these broad diagnosis categories, potentially calling into question the random assignment of patients to physicians. To mitigate this concern, we reestimated our results with a specification that dropped physicians who are outliers for any of the three diagnosis categories.

For this specification, we drop ED 4 altogether, as we found that (i) the fraction of patients with angina+ is bimodal across physicians for this ED and, (ii) it appears to serve a very different patient mix, e.g., with many more patients in the appendicitis+ sample. Furthermore, we omit: (i) 4 physicians from ED 1 because they had disproportionally low or high assignments in the fraction of appendicitis+ patients (leaving us with a range of proportions of angina+ patients from 1.10 to 2.11 percent), (ii) 2 physicians from ED 7 with disproportionally low assignments of appendicitis+ patients (leaving us with a range of proportions from 0.81 to 3.12 percent), and (iii) 1 physician from ED 15 also with a disproportionally low assignment of appendicitis+ patients (leaving us with a range from 1.43 to 3.92 percent). Altogether, these restrictions leave us with 78 unique physician-ED pairs, compared to 90 for our base analysis (as reported in Table 3).

Our results for this more restrictive sample of physicians are presented in Tables A5-A7. Comparing these results to those from the initial sample of physicians, in Tables 5-7, we find very similar

Table A4: Percentage of patients with condition by percentile of physician distribution

ED	Angina+		Appendicitis+		TIA+	
	10th	90th	10th	90th	10th	90th
1	9.32	15.05	0.62	2.11	2.10	3.62
4	11.89	45.34	11.87	59.88	2.13	3.44
7	4.08	12.99	0.32	3.11	1.39	2.93
8	9.49	15.51	2.52	4.21	2.69	4.47
9	7.29	13.19	1.05	3.00	1.10	2.67
15	5.16	12.85	.35	3.91	1.23	3.40
17	9.75	15.02	1.44	3.17	1.73	3.70

Table A5: 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 A6: 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 A7: 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.

results.

Specifically, considering the correlations within conditions presented in Table A5, the correlation coefficients are similar, though generally a bit larger. For example, the correlation coefficient between initial ED spending (physician practice style) and ED revisits during the 5 days (a measure of physician skills) is slightly larger for angina+ (from 0.22 to 0.29), appendicitis+ (from 0.48 to 0.51), and TIA+ (from 0.14 and insignificant to 0.19 and significant). Similarly, with respect to the correlations between physician practice styles and skills measured by hospitalizations, the correlation coefficients are again a bit larger for hospitalizations during the 6-90 post ED visit period for the angina+ sample (from -0.18 to -0.23) and during both the 0-5 day period (from 0.26 to 0.34) and the 6-90 day period (from 0.21 to 0.29) for the appendicitis+ sample.

We next consider the correlations across conditions, in Table A6, and the correlations between the base and plus samples, in Table A7. Again here, the coefficients are very similar to those from the initial sample of physicians.

Taken together these results suggest that variations in physician practice styles and skills as well as correlations across them are not qualitatively be driven by outliers in terms of the proportion of patients in each particular illness categories.

Results adding in all physicians at EDs that we study

In this final robustness check, we expand our sample to include all ED physicians (irrespective of the age, gender and illness profile of their patients) practicing 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).

Our results here are presented in Tables A8-A10. For all three tables, we again find results that are virtually indistinguishable from the base results. We conclude that the our initial exclusion of physicians who did not pass the randomization tests for age and gender did not affect our conclusions in any significant way.

Table A8: 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 A9: 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 A10: 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.