JUST A MINUTE:

THE EFFECT OF EMERGENCY DEPARTMENT WAIT TIME ON THE COST OF CARE

By Lindsey Woodworth and James F. Holmes

Wait times are rising in U.S. emergency departments. Long emergency department wait times are recognized as potentially threatening to patients' health. This study uses the stringency of a patient's emergency department triage nurse as an instrumental variable to identify the impact of wait time on the total cost to care for a patient. Among patients with the most acute conditions, a one hour increase in wait time leads to an approximately 30% increase in costs. The magnitude of this effect dissipates (fading to zero) among patients who arrive with less urgent conditions.

Woodworth: Department of Economics, University of South Carolina, 1014 Greene St, Columbia, SC 29208 (lindsey.woodworth@moore.sc.edu). Holmes: Department of Emergency Medicine, UC Davis School of Medicine, 2315 Stockton Blvd, Sacramento, CA 95817 (jfholmes@ucdavis.edu). We thank Alice Ellyson, Milton Straw, and participants at iHEA World Congress, WEA Conference, SC Applied Micro Day, and University of South Carolina's brownbag series for helpful comments and suggestions. Support for this project was provided by grant number T32HS022236 from the Agency for Healthcare Research and Quality (AHRQ) through the Quality, Safety, and Comparative Effectiveness Research Training (QSCERT) Program, and by grant number UL1000002 from the National Center for Advancing Translational Sciences, a component of the National Institutes of Health (NIH).

"There are a few undeniable observations about emergency departments. Nobody wants to be there; nobody wants to wait; and the longer someone is kept waiting, the more they don't want to be there."

- Diana Mahoney
(Press Ganey Associates)

I. INTRODUCTION

The average wait time across U.S. emergency departments (EDs) has risen 25% since the early 2000s (Hing & Bhuiya, 2012). In 2013, over 17% of ED patients waited more than one hour to see a physician (NCHS, 2017). The medical community has voiced considerable concern that these long wait times are compromising patients' health. Recent work suggests ED patients with pneumonia are more likely to be admitted to the hospital the longer their time between ED arrival and receipt of antibiotic (Fee et al., 2007), and patients who are admitted to the hospital from the ED experience longer inpatient stays and higher rates of inpatient mortality when admitted during periods of high ED crowding (Krochmal & Riley, 1994; Liew, Liew & Kennedy, 2003; Richardson, 2006; Sun et al., 2013). Yet, in spite of the evidence that long ED wait times can be clinically harmful, relatively little attention has been paid to the possibility that long ED wait times might increase the cost of care. The intuition is simple. The longer a patient waits to be seen, the more their condition will worsen. As patients' conditions worsen, additional treatments and medical resources are often required. These "extra" medical resources translate into additional costs.

For economists, the relationship between ED wait time and costs should be significant due to concerns regarding efficiency. Each year, the U.S. spends over $3 trillion in healthcare. As much as 10% of this spending is exhausted in emergency care (Lee, Schuur & Zink, 2013). U.S. policymakers are seeking opportunities to curb the steep upward trajectory in health spending. However, is difficult to reduce healthcare expenditures without jeopardizing individuals' health. Indeed, the competing incentives to save money and improve health outcomes have been dubbed "the health policy conundrum" (Bhattacharya et al., 2014).

This study explores one novel approach to accomplishing the dual aims of promoting health outcomes and minimizing healthcare expenditures. Specifically, we estimate the impact of ED wait time on the total cost to care for a patient. If longer ED wait times *cause* higher costs, then it may be possible for policymakers to target a known predictor of adverse patient outcomes (i.e., untimely care) as a way to *reduce* the amount of money required to care for a patient.

Identifying the effect of ED wait time on costs is difficult because ED wait times are structured to systematically track the acuity of the patients who are currently seeking care. That is, EDs expedite care for the sickest patients. One implication of this design is that the patients with the *shortest* wait times generally require the *most* medical resources. Consequently, naïve estimates for the relationship between ED wait time and costs suggest that costs fall as wait times lengthen. These results provide no useful information other than to demonstrate that EDs are proficient at prioritizing very sick patients.

This study attempts to break the link between ED wait time and patient acuity by exploiting the quasi-random assignment of patients to triage nurses. Importantly, triage nurses may differ in their perception of what constitutes an urgent vs. semi-urgent triage classification. Triage classifications are important because they determine a patient's position in an ED's queue. Most busy hospitals schedule triage nurses with overlapping coverage, so, for any given snapshot of an ED's queue, we will likely see patients who were triaged by different nurses. What does this mean? It means that if a patient happens to be triaged by a “stringent” nurse, then that patient may have a considerably longer wait time than if that same patient had been triaged by a more lenient nurse. In general, triage nurses have no involvement in patients’ care beyond positioning them in the queue. Therefore, if the nurse “behind the door” is effectively random, then so too will be a portion of the patient’s overall wait time.

To conduct this analysis, we collect time stamps and nurse codes from over 180,000 electronic medical records in a large, urban academic ED. We then link these data elements to the corresponding billing record for the patient’s encounter. The billing records we obtain reflect utilization from the moment the patient entered the ED to the moment the patient left the comprehensive health system in which the ED is housed. Following patients throughout the course of their stay in the health system allows us to account for all costs which accrue after an ED patient is admitted to the hospital. With these data, we leverage the plausibly exogenous differences in wait times between patients who were triaged by stringent versus lenient nurses in an instrumental variables framework. This allows us to identify the causal effect of ED wait time on costs.

To the best of our knowledge, the estimates presented in this study are the first to causally identify the impact of door-to-doctor time in the ED on costs. The Centers for Medicare & Medicaid Services now tracks door-to-doctor time in U.S. EDs as a national benchmark. We are aware of four clinical studies that consider the cost of delays in emergency settings. However, these studies focus on different intervals of time during a patient’s ED encounter. Krochmal and Riley (1994) analyze ED patients who are admitted to the hospital. They compare those who are boarded in the ED for more than one day to those who are boarded in the ED for less than one day.[1] They find that ED patients who are not promptly transferred to an inpatient unit have hospital stays that average 10-13% longer. They conclude, “This increased length of stay means an increased cost per patient”. In their hospital, such costs accumulated to $6.8 million over the course of three years. Bayley et al. (2005) similarly examine the impact of ED boarding on hospital costs. They focus on patients with chest pain and consider three hours as a threshold for prolonged ED boarding. Within this narrowly defined patient population, they find no evidence that prolonged ED boarding affects real costs. They do, however, find that opportunity costs are substantially higher when chest pain patients are boarded in the ED for more than three hours.[2] Huang et al. (2010) use yet another threshold to define delayed admission. They find that when an ED patient’s time from triage to hospital admission exceeds 12 hours, the patient’s inpatient costs average 11% higher. Finally, Sun et al. (2013) consider the relationship between ambulance diversion hours and hospital costs.[3] They find that on days when ambulances are diverted away from a hospital’s ED for extended periods of time, the average cost to care for a patient who is admitted to the hospital from that ED is approximately 1% higher.

Each of the above studies does suggest a relationship between timeliness of care in the ED and costs. However, the conclusions that may be drawn from these studies are limited. Each study relies on observational comparisons. These preclude causal inference. We exploit plausibly exogenous variation in wait times. This allows us to ascertain causal effects. Second, each study uses a

---

[1] ED boarding occurs when a hospital does not have enough beds to accommodate the number of ED patients who require admission. When this is the case, ED patients who need to be admitted to the hospital must remain in the ED and wait for an inpatient bed to become available.
[2] In their hospital, opportunity costs amounted to $204 in lost revenue per patient with chest pain patient who was boarded for more than three hours.
[3] Ambulances are diverted away from an ED when the ED does not have the capacity to handle additional patients. When ambulances are being diverted, patients who are being cared for in that ED generally have wait times that are longer than average.

binary indicator to define “delayed care”. As such, it is not clear what the one-to-one correlation is between minutes spent waiting and costs. We measure wait time in one-minute increments. This allows us to quantify the average cost of each minute. Finally, each previous study limits its sample to ED patients who are admitted to the hospital. Only 9% of ED patients are admitted to the hospital, and patients who are admitted to the hospital are characteristically different than those who are not (CDC, 2016). We analyze both patients who are and patients who are not admitted to the hospital and then exploit wait time nudges at multiple thresholds of patient acuity. This positions us as the first study that is capable of capturing effect heterogeneity.

We find that for patients who arrive at the ED with the most acute conditions, a 60-minute increase in wait time increases the hospital’s cost to care for the patient by an average of 30%. For patients who arrive with moderately acute conditions, a 60-minute increase in wait time increases the hospital’s cost to care for the patient by an average of 21%. We detect no statistically significant effect of wait time among relatively healthier patients.

To put these results in perspective, we present back-of-the-envelope calculations for the amount of money a hospital could save if it trimmed 30-minutes from each ED patient’s wait time. When we look at the average U.S. ED (in terms of its distribution of patient acuity) and consider an initial mean wait time of approximately one-half hour for the most acute patients (i.e., those who require immediate life-saving intervention) and one-and-one-half hours across all other patients, then we find that an across-the-board trimming of 30-minutes from each ED patient’s wait time could lower the hospital’s overall cost to care for individuals who present at the ED by a magnitude of 7%. This effect size is large but conceivable. The sickest patients are the most resource-intensive and therefore generate a disproportionately large share of costs. These are also the patients whose costs deflate most rapidly; that is, with the least amount of pressure on wait. These findings suggest large returns to minimizing ED patients' wait times. We interpret this evidence as a considerable opportunity to safely drive down U.S. healthcare expenditures.

II. INSTITUTIONAL BACKGROUND

To understand the identification strategy used in this study, it will be helpful for us to describe particular aspects of how EDs in U.S. hospitals operate. Of particular importance are how patients move through an ED and how triage decisions are made. We address each of these items, in turn, below.

II.A. Where Do the Patients Go?

There are generally two entrances through which a patient might arrive in an ED. Patients who arrive by ambulance will come in through the ED's "internal entrance". This is a private door connected to the ambulance bay. All other patients come in through the ED's "external entrance" – otherwise known as the front door.

In many hospitals, a triage station is adjacent to each ED entrance.[4] Each station is staffed with a separate triage nurse who is responsible for providing an initial evaluation of the patients who enter through that door. Based on the nurse's assessment of the patient's condition, the nurse will here assign each patient a triage level. Triage levels effectively categorize patients into broad groups that loosely correspond to the severity of the patients' conditions and their anticipated resource requirements. In most EDs, including the ED whose data is used in this study, triage levels are based on the Emergency Severity Index. The

[4] This is true of the ED whose data is used in this study.

Emergency Severity Index consists of five levels. Level 1 is reserved for patients who are "actively dying". Level 5 is given to patients with near perfect health.

Once a patient receives his or her triage level, they proceed to the ED's waiting room if they are not immediately placed in a patient care area. All patients – regardless of their triage level or point of entry – are consolidated into a single queue in this step. Patients who are transported by ambulance are not given preferential emergency care. They are triaged in the same manner.

From the waiting room, patients are called to see the physician in descending order of triage level. Level 1 patients are seen before Level 2 patients, all Level 2 patients must be seen before any Level 3 patient is seen, and so on. If the group that is first-up to be seen contains multiple patients with the same triage level, then these patients are routinely prioritized by their time of arrival.

To reiterate what the triaging process means for patients with relatively high triage levels, consider John who has a triage assignment of Level 5. Suppose John has been waiting in the ED all day and is finally first-up to be seen. Immediately before John's name is called, another patient enters the waiting room with a triage assignment of Level 4. This new patient will be seen before John – even though John waited much longer. In short, patients are perpetually pushed down the line as long as other patients enter the waiting room with more critical levels.

There is one caveat to the standard triaging process described above which applies specifically to the ED whose data we use in this study. In this ED, patients with triage Levels 4 & 5 are eligible for "upfront care" during daytime hours. Upfront care is provided as a substitute for (regular) advanced ED care. The provision of upfront care effectively generates a fork in the queue. All patients who arrive at the ED will still "line-up" behind the last individual who arrived with the same triage level. However, rather than patients only exiting the queue from the front of the line, Level 4+ triaged patients also exit the queue from the notch where the Level 4 patients start. This additional outlet means that if a patient is on the margin of triage Levels 3 & 4, then being assigned Level 4 could shorten the patient's wait time relative to being assigned Level 3.[5] Figure I shows the patient-flow process.

II.B. What Triage Level Will the Patient Receive?

Given the prominence of triage level assignments in determining a patient's wait time, it is worth asking: What are the factors that determine which triage level a patient receives?

The criteria for assigning particular triage levels are not exact. However, for facilities utilizing the Emergency Severity Index, there are a series of questions that should guide a triage nurse's decision (Gilboy et al., 2012). These questions should progress in the following order:

Question 1: Does this patient require immediate life-saving intervention?

> If the answer to the first question is "yes", then the triage nurse should assign Level 1 and ask no further questions. If, however, the triage nurse believes that the answer to the first question is "no", then the nurse should proceed to the following question.

Question 2: Is this a patient who should not wait?

---

[5] We account for this caveat in the analysis.

If the triage nurse believes that the answer to the second question is “yes”, then the nurse should assign Level 2 and suspend further questioning. However, if the triage nurse believes that the answer to the second question is “no”, then the nurse should proceed to the following question.

Question 3: How many resources will this patient need?

To answer the third question, the triage nurse will need to make an educated guess regarding the number of resources he or she believes that the physician will need to use in their eventual care for the patient. Medical resources consist of large services and supplies such as laboratory tests, radiographs, and intravenous fluids. They do not consist of relatively trivial items such as physical examinations, crutches, or splints. If the triage nurse believes that the physician will not need to utilize any resources to care for the patient, then the nurse should assign Level 5. If the triage nurse believes that the physician will need to utilize exactly one resource, then the nurse should assign Level 4. If the triage nurse believes the physician will need to utilize more than one resource, then the triage nurse should re-evaluate the patient’s vital signs and proceed to the next and last question.

Question 4: Does this patient have abnormal vital signs?

Vital signs (temperature, heart rate, blood pressure, and respiration rate) consist of continuous variables. There are no discrete thresholds for what constitutes “normal”. Accordingly, abnormal vital signs exist in nebulous ranges. If the triage nurse believes that the patient’s vital signs are abnormal, then the nurse should assign Level 2. If the triage nurse does not believe that the patient’s vital signs are abnormal, then the nurse should assign Level 3.

Figure II shows how the triage process unfolds.

II.C. Additional Items to Note

So far, we have seen that triage assignments can have a nontrivial impact on a patient's time-to-physician and that there is subjectivity involved in the triaging process. We highlight two additional facts of importance and then summarize key details.

First, it is not uncommon for non-acute patients to arrive by ambulance and very acute patients frequently arrive by personal transportation. This results in a full range of Level 1-5 patients passing through both the internal and external triage stations. A peek into the ED's waiting room will likely reveal some Level 3 patients, for example, who were triaged by the internal nurse and some who were triaged by the external nurse. The same is true of other levels. This means that if the two triage nurses who are concurrently working are not triaging identically, then the counterfactual wait time for a patient who is triaged by the *other* nurse (than the nurse who actually performed triage on the patient) may be markedly different than the patient's actual wait.

Second, triage nurses provide no care to patients. Triage nurses only assign triage levels. Triage levels are only important in that they influence time-to-physician. Once a patient reaches the physician, their triage level assignment is no longer relevant. Consequently, there is little chance that a triage nurse will affect any aspect of a patient's ED experience other than the time it takes for the patient to reach the physician.

For brevity and emphasis, we summarize in rapid order the details that are important to keep in mind for identification:

(i) A one unit change to a patient's triage level can have a substantial effect on their wait time.

(ii) A triage nurse can effect a one level change.

(iii) At least two triage nurses are always working concurrently.

(iv) There is overlap in the distributions of patients' acuity between those who are triaged by the internal and external nurse.

(v) Once triaged (by either the internal or external nurse), all patients converge into a single queue where they are called in order of triage assignment.

III. IDENTIFICATION

The logistics of emergency care lend themselves to natural experiments. In this section, we describe the natural experiment we set-up to capture variation in patients' wait times that is haphazard. We then discuss the decisions we make to empirically execute this natural experiment. Finally, we probe the validity of our decisions.

III.A. The Natural Experiment

This study identifies the causal effect of ED wait time on costs by exploiting the quasi-random assignment of patients to triage nurses in an instrumental variables framework. Conceptually, this identification strategy intends to mimic a randomized controlled trial (RCT). Whereas a RCT participant might open an envelope to determine if they are in the treatment (long wait time) or control (short wait time) group, here, triage nurses act as the randomizer. Some patients "open the door" and are greeted by a stringent triage nurse. Other patients "open the door" and are greeted by a lenient triage nurse. Patients who are greeted by a stringent triage nurse are expected to wait longer, on average, than patients who are greeted by a lenient nurse. Hence, the former subset of patients forms a natural treatment group. The latter, a natural control group.

Importantly, patients are not greeted by a stringent or lenient triage nurse based on their clinical characteristics. Patients in either group should be comparable in their average projected costs at arrival. The amount of time that either group spends in the waiting room might systematically differ, but neither groups is expected to begin their wait with an unfair advantage. Therefore, if we observe systematic differences in costs between patients in the natural treatment and natural control group by the end of their stays, then these cost differences might reasonably be attributed to the marginal difference in the groups' wait times.

The identifying assumptions underpinning this design are that (i) triage nurse stringency affects patients' wait times, (ii) whether a patient is triaged by a stringent or lenient triage nurse is effectively random, and (iii) exposure to a stringent triage nurse will move patients' wait times monotonically; that is, being triaged by a stringent nurse would never reduce a patient's wait time relative to being triaged by a more lenient nurse.

III.B. Measuring Triage Nurse Stringency

The execution of this natural experiment relies on the ability to identify which nurses are stringent. To elucidate nurses' tendencies, we rely on clinical expertise which says that sprained ankles are easy for triage nurses to correctly identify and that sprained ankles are relatively homogenous in terms of acuity. This means that across the universe of patients whose final primary diagnosis is a sprained ankle (these we can identify in the data), there should be consistency from nurse to nurse in which triage level they assign. If we observe variation in triage level assignments to patients with a sprained ankle across nurses, then this must stem from idiosyncratic triaging tendencies across nurses.

We quantify nurses' tendencies in a binary fashion. Nurses whose median triage assignment to their patients whose final primary diagnosis is a sprained ankle is Level ≥4 are coded *Stringent*=1. Nurses whose median triage assignment to patients whose final primary diagnosis is a sprained ankle is Level <4 are coded *Stringent*=0. For exposition, we refer to the latter group of nurses as lenient.[6] Importantly, zeroes and ones are imputed across all observations, not just observations that have sprained ankle codes.

III.C. Validity of Triage Nurse Stringency as an Instrument

If *Stringent*, given the way we have constructed it, is a valid randomizing mechanism (i.e., it selects patients independently of their characteristics for inclusion in the natural treatment group), then the patients exposed to *Stringent*=1 nurses should be observably similar to the patients exposed to *Stringent*=0 nurses. We test whether this is the case by looking for balance in exogenous patient characteristics between nurses' caseloads. In Section 4, we will explain in more detail how we arrived at our sample. For now, we just preview the data to verify that the instrument, given the way we have constructed it, appears to be valid.

As shown in Table I, when we look at the mean characteristics of patients in the natural treatment group and compare these to the mean characteristics of patients in the natural control group, values are generally similar. The representation of male patients differs by only 0.2 percentage-point between *Stringent*=1 and *Stringent*=0 nurses' caseloads; their average patient's age differs

[6] To be precise, Level 4 is appropriate for patients with a sprained ankle. However, so that the instrument and endogenous treatment variable are positively correlated, it is convenient to use less strict triagers as the reference.

by just 1.1 years.[7] When we compare the share of patients diagnosed (by the physician, long after triage) with each of the five most common conditions, shares differ by no more than 0.2 percentage-point between *Stringent*=1 and *Stringent*=0 nurses' caseloads.

It is important to note that even in the case of RCTs, randomization can break down. When a finite number of people are randomly partitioned into one of two groups (treatment or control), it is almost never the case that the treatment and control group end up identical, on average. The same holds true in instrumental variable analysis. Here, the instrument does the assignment. Yet, even if the instrument is completely indiscriminate in who it assigns to the natural treatment group, there is a (good) chance that patients sent to the natural treatment group will look slightly different than patients sent to the natural control group, on average. Slight imbalances in mean patient characteristics between groups do not necessarily mean that the natural experiment has failed. It is permissible to adjust for confounding. What instrumental variables analysis instead requires is that groups be effectively identical *after* conditioning. This is the weaker assumption of "conditional independence". In the analysis, we condition on observed patient characteristics. However, for reassurance, we point out now that cost-predictors appear to be fairly balanced pre-conditioning.

IV. DATA AND DESCRIPTIVE STATISTICS

IV.A. Data Source

The starting point for our sample is encounter-level electronic medical records for patients who visited a large, urban, academic ED between July 2008

---

[7] The fact that patients in the natural control group are slightly older implies that, if anything, the 1.1 years age disparity will bias the results away from finding an adverse effect.

and April 2013. We consider patients between the ages of 18-89 years at their time of arrival.[8] The ED we consider a part of a comprehensive health system and is home to a designated Level I Trauma Center.

The variables we extract from patients' electronic medical records include age, sex, encounter date, triage nurse, triage start time, and ED bed placement time (the ED physician begins evaluation shortly after ED bed placement). We use the latter two variables to construct a variable for wait time. This measures the approximate number of minutes that elapse before an ED patient is evaluated by an ED physician. This is our independent variable of interest.

We separately obtain patients' billing records. Billing records are stored by the study site health system in an entirely separate file from medical records. We are provided with no crosswalk to link the records. Therefore, we manually match the medical records to the billing records using unique combinations of patient ID and date of presentation. One limitation of this linkage process is that it is infeasible to match records in which a one patient visits the ED more than once in a day. This is because it is unclear which medical record from that day is associated with which billing record from that day.[9]

From the billing records, we are able to deduce aggregate charges and payments for all services provided to the patient over the course of their time in the health system.[10] That is, from the moment the patient entered the health system at the ED to the moment the patient left the health system from either the ED or an inpatient unit. It is important that we consider billing records that span

[8] This is to comply with IRB protocol.
[9] While we must exclude such observations, we note that such observations would likely be excluded anyway. We later omit observations in which the patient prematurely discontinues care. Presumable, many patients who visit the ED more than once in a day do this because they do not see their treatment to completion the first time.
[10] During the referent encounter.

the duration of patients' encounters so that we are able to capture information on any utilization that occurs after an ED patient transfers out of the ED and is admitted to the hospital for inpatient care. Notably, billing records also clarify patients' final diagnoses. A patient's final diagnosis is recorded by the ED physician if the patient is discharged from the ED or an inpatient physician if the ED patient is admitted to the hospital.

We multiply each patient's aggregate charge by the hospital's cost-to-charge ratio in that year. The resultant product is an approximation for the total cost to the health system to care for each patient who presented at the ED. This is our primary outcome of interest.

IV.B. Sample Used to Construct the Instrument

Given that we determine whether a patient is in the natural treatment or natural control group by observing how his or her triage nurse behaves when confronted with a sprained ankle, it is necessary that we exclude encounters in which the patient's triage nurse is never observed triaging a sprained ankle. These observations will otherwise have missing values for the instrumental variable.

Omitting observations where the patient's triage nurse's tendency is indeterminable reduces the size of the sample by 3%. We use the remaining 97% of observations to construct the instrument, *Stringent*.

IV.C. Sample Used in the Main Analysis

After constructing the instrument, we make three further exclusions before performing the main analysis. First, we omit observations in which the patient prematurely terminates their visit. It is not clear what the cost of caring for a patient who leaves the health system before completing their treatment *would* have been had this patient seen their care to completion. Approximately 5% of observations are omitted in this exclusion. Second, we omit observations in which the patient is triaged out of the ED. Patients who are triaged out of the ED are referred to a clinic; both their wait times and costs are effectively null. This exclusion further reduces the sample size by <0.001%. Finally, we omit observations that have key variables missing from either the medical and/or billing record. Of the remaining observations, <0.01% have key variables missing. This leaves us with a final sample of 187,149 observations.

Descriptive statistics for the final sample are presented in Table II. On average, patients wait 1 hour 23 minutes to see a physician. The average cost per encounter is $6,395. There is an approximately even split between male and female patients. The average patient's age is 45 years. On a typical day, roughly 18% of patients arrive with a condition that is life-threatening (Level 1), 30% are acutely injured or ill (Level 2), 37% are moderately injured or ill (Level 3), 15% are mildly injured or ill (Level 4), and 1% require no medical resources (Level 5).

Figure III shows the density of triage level assignments across a typical week. On top of this, we overlay average wait times. Unsurprisingly, there are diurnal patterns in when patients arrive. Less noticeably, acuity is correlated with time of arrival. During both early morning hours and early evening hours, a greater share of the patients who arrive at the ED come with conditions that are severe.

V. ESTIMATION

For motivation, consider the structural equation below quantifying the association between ED wait time and costs. Subscript i denotes a single encounter:

$$Cost_i = \beta_0 + \beta_1 Wait_i + \beta_2 X' + \varepsilon_i \qquad (1)$$

Wait time is endogenous in this equation. EDs expedite care for patients with the most acute conditions, and acuity is a predictor of costs. It is infeasible to perfectly quantify acuity, and so acuity will be nested within ε_i. This implies that $Wait_i$ and ε_i will be correlated, making the parameter of interest (β_1) biased.

To account for the endogeneity of ED wait time, we instrument for a patient's wait time using two-stage least squares (2SLS). The first-stage is specified as follows:

$$Wait_i = \alpha_0 + \alpha_1 Stringent_i + \mathbf{T}_t + \mathbf{C}_i + \eta_i \qquad (2)$$

The endogenous treatment variable, $Wait_i$, measures time-to-physician.[11] The instrument, $Stringent_i$, is a binary variable equal to one when the patient is triaged by a nurse who is stringent. The vector **T** contains controls for the patient's time of arrival (hour, weekday, month, and fiscal year fixed effects), and the vector **C** contains controls for the patient's clinical characteristics (age, sex, and diagnosis fixed effects).[12]

In the second-stage, predicted values from equation (2) are substituted into the equation below:

---

[11] Measured in one minute increments scaled up to the hour.

[12] Age is measured in years. Diagnosis is measured as the ten integer grouping for the patient's final primary ICD-9 diagnosis (e.g., ICD-9 diagnoses codes 420.XX through 429.XX are treated as one single group).

$$ln(Cost)_i = \delta_0 + \delta_1\widehat{Wait}_i + \mathbf{T}_t + \mathbf{C}_i + \mu_i \tag{3}$$

The coefficient of interest is δ_1. This is interpreted as the average percent change in the hospital's total cost to care for a patient who is made to wait one hour longer. Importantly, δ_1 is identified off of the variation in $Wait_i$ that stems from the variation in $Stringent_i$. This means that if triage nurse assignments are effectively random, then $Stringent_i$ will be independent of acuity, and $\widehat{Wait}_i$ will be uncorrelated with μ_i. As a result, δ_1 might be interpreted causally.

Importantly, instrumental variables analysis identifies a local average treatment effect (LATE). The "local" component of LATE implies that δ_1 will reflect the impact of waiting for patients whose wait time is affected by their triage nurse. If a patient clearly deserves a triage Level 3, for example (any nurse would assign the patient this level), then δ_1 will not reflect the average impact of wait time on this patient's costs. The coefficient δ_1 only speaks to the impact of waiting for patients who are on the margin of two triage levels who would be triaged up (and made to wait longer) if seen by a stringent nurse and triaged down (and allowed to wait less) if seen by a lenient nurse. These patients' wait times *comply* with the stringency of their triage nurse. The "average" component of LATE implies that δ_1 will be a medley of as many treatment effects as exist across compliers in the sample. If the sample contains compliers at the margins of triage Levels 1 & 2, Levels 2 & 3, Levels 3 & 4, and Levels 4 & 5 – and the compliers at each of the different margins are affected differently by waiting – then δ_1 will yield an average across the heterogeneous treatment effects.

It is likely ED wait time affects costs heterogeneity. Patients who come to the ED with life-threatening conditions are probably more susceptible to harm given a one hour increase in their wait time, relative to patients who arrive in better health. To not overlook effect heterogeneity, we stratify the sample by neighboring triage levels and estimate equations (2) and (3) separately over each stratum. This allows us to separately identify the impact of wait time for compliers at the margin of Levels 1 & 2, at the margin of Levels 2 & 3, at the margin of Levels 3 & 4, and at the margin of Levels 4 & 5.[13]

For just the triage Level 3 & 4 subsample, we omit patients who arrive 6:00am to 10:00pm. This is to prevent a violation of the monotonicity assumption. During the night (when upfront care is not being offered), patients who arrive on the margin of Levels 3 & 4 can only be impeded if exposed to a stringent nurse. However, during the day (when upfront care is being offered), patients on the margin of Levels 3 & 4 might experience a reduction in their wait time if exposed to a stringent nurse. This is because a (stringent) Level 4 assignment would potentially send the patient to the front of the line for upfront care, whereas a Level 3 assignment would send the patient to the back of the line for regular (advanced) ED care. We avoid a non-monotonic relationship between $Stringent_i$ and $Wait_i$ if, among patients on the margin of qualifying for upfront care, we omit time periods when upfront care is provided.

We make one final point regarding statistical inference. It is appropriate to cluster standard errors when observations contribute repetitive noise.

[13] It is likely that within the triage Levels 1 & 2 subsample there will be patients who arrived on the margin of Levels 2 & 3 who were exposed to a lenient nurse (and therefore assigned Level 2). The presence of Level 3ish patients in the Levels 1 & 2 subsample will, if anything, lead to understated effect sizes for the impact of wait time on costs at the margin of Levels 1 & 2. This presumes that Level 3ish patients are less (adversely) affected by waiting than compliers on the margin of Levels 1 & 2 and that their smaller treatment effect feeds into the average. On the other hand, within the Levels 4 & 5 subsample, there may be patients who arrived on the margin of Levels 3 & 4 who were triaged by a stringent nurse (and therefore assigned Level 4). The presence of Level 3ish patients in the Levels 4 & 5 subsample could generate overstated effect sizes in terms of the estimated impact of wait time at the margin of Levels 4 & 5; this presumes Level 3ish patients are relatively more affected by waiting and their larger treatment effect feeds into the average. In short, the “spillover” of patients into different subsamples is only a concern in that it may dampen the extent to which we are able to see the extent of the effect heterogeneity.

Practitioners frequently cluster over the assignment variable in 2SLS models. This would here be the patient's triage nurse.

Clustering over triage nurses may be not be appropriate in this particular setting. Conceptually, the dividing lines between clusters should be drawn such that observations within a cluster share correlated error terms, but observations in different clusters have error terms that are uncorrelated (Cameron & Miller, 2015). This pattern is unlikely to hold if we draw lines around individual nurses. A single triage nurse may work a variety of shifts: some on weekdays and some on weekends; some during the day and some overnight; some at the internal entrance and some at the external entrance. Patients who arrive on weekdays are characteristically different than those who arrive on weekends. Patients who arrive during the day are characteristically different than those who arrive at night. Patients who arrive by ambulance are characteristically different than those who arrive on their own. Therefore, within a given triage nurse, error terms are unlikely to be correlated across shifts. Accordingly, we cluster standard errors over a proxy for the distinct shift within each nurse (nurse×date combination). We also present unclustered standard errors for comparison.

VI. RESULTS

To provide a baseline for our instrumental variable results, we begin by presenting OLS estimates. These are given in Table III. Along rows, we add controls for patients' clinical characteristics. These estimates demonstrate the endogeneity of ED wait times.

As shown in Row A, naïve estimates for the impact of ED wait time on costs suggest that a one hour increase in time-to-physician is associated with an 8% reduction in costs. In Row B, the association between ED wait time and costs is approximately halved when we control for patients' age, sex, and diagnosis. In Row C, the estimated impact of wait time falls substantially when we further control for assigned triage level.[14] However, this point-estimate remains statistically significant and negative. This suggests an underlying gradient of acuity within triage level that the OLS model cannot account for.[15]

VI.A. Effect of Triage Nurse Stringency on the Marginal Patient's Wait

To harness variation in patients' wait times that is exogenous, we focus on just the component of each patient's wait time that is determined by their triage nurse's level of stringency. Whether this component is large should be random across patients. Therefore, if we can identify the magnitude of this component within each patient's wait time, then we can work off variation in the magnitudes. Our first step, then, is to determine, for each patient, the amount of leverage their triage nurse exerted on their time-to-physician.

First-stage estimates for the impact of exposure to a stringent triage nurse on patients' wait times are presented in Table IV. For patients who arrive at the ED on the margin of triage Levels 1 & 2, exposure to a stringent triage nurse increases wait time an average of 0.06 hour. For patients who arrive on the margin of triage Levels 2 & 3, exposure to a stringent triage nurse increases wait time an average of 0.15 hour. For patients who arrive (at night) on the margin of triage Levels 3 & 4, exposure to a stringent triage nurse increases wait time an

---

[14] In the Appendix, we present OLS results stratified by triage assignment. These help explain the source of the persistent endogeneity.

[15] For intuition, consider patients who deserve triage Level 2. Within this broad category, some patients will still be sicker than others. Suppose John and Jane both have a condition that deserves triage Level 2, but John is slightly sicker than Jane. If John and Jane both wake up in the middle of the night feeling ill, then John will be more likely to go to the ED immediately rather than wait until morning, relative to Jane. Wait times are shorter at night than in the morning. Hence, for these two patients who deserve the same triage level, acuity will remain linked to time in the waiting room.

average of 0.06 hour. We detect no statistically significant impact of triage nurse stringency for patients at the margin of triage Levels 4 & 5. We note, however, that Level 5 is rarely assigned. Only 4% of patients in our sample who are assigned either Level 4 or Level 5 actually receive a Level 5 assignment. Thus, it may not be surprising that the subtle nudge of a nurse's own triaging tendency is not a strong enough force to push some patients into the Level 5 category.

A general rule-of-thumb is that an instrument is sufficiently strong if the first-stage point estimate is statistically significant and the first-stage F-statistic exceeds 10 (Hall, Rudebusch & Wilcox, 1996; Stock, Wright & Yogo, 2002). This criteria is not satisfied for the triage Levels 4 & 5 stratum. Therefore, we do not proceed with second-stage analysis here; this could lead to misguided conclusions about the relationship between ED wait time and costs for patients in this region of the acuity spectrum (Bound, Jaeger & Baker, 1995). We do, however, move forward with second-state analysis elsewhere. All other triage level stratums are supported by a strong first-stage.

VI.B. Effect of Emergency Department Wait Time on Costs

Instrumental variable results for the causal effect of ED wait time on costs are presented in Table V. For patients who arrive at the ED on the margin of triage Levels 1 & 2, a one hour increase in wait time is estimated to increase the hospital's total cost to care for the patient by an average of 30%. The magnitude of this effect dissipates as we look at relatively healthier patients.

For patients who arrive at the ED on the margin of triage Levels 2 & 3, a one hour increase in wait time is estimated to increase costs by an average of 21%. For patients who arrive on the margin of triage Levels 3 & 4, a one hour increase in wait time is estimated to have no statistically significant impact on costs. This coefficient (in the triage Levels 3 & 4 stratum) is positive and comparatively smaller, but the standard error suggests it is driven by noise. Thus, we cannot rule out a null effect. We can make no claims regarding the impact of wait time for patients on the margin of triage Levels 4 & 5.

VI.C. Robustness Checks

VI.C.1. Sensitivity to Alternative Specifications

To test the robustness of our findings, we re-estimate the base model but add and remove controls. If all predictors of costs are perfectly balanced between stringent and lenient nurses' caseloads, then 2SLS estimates should be not be sensitive to covariate adjustments. We also re-estimate the base model but replace $0 costs with $0.01. This is to assess whether our results are sensitive to an alternative method of handling zeroes. In the base model, observations with $0 costs are omitted from the analysis as a result of the log-transformation.

Table VI presents instrumental variable estimates across a range of specifications. The sign and statistical significance of the point-estimates generally conform with previous findings. While the point-estimates' magnitudes vary slightly from the base model, these estimates tell a similar story.

Wait time is estimated to have a positive and statistically significant effect on costs for patients at the margins of both triage Levels 1 & 2 and triage Levels 2 & 3. The range of precisely estimated effect sizes at the triage Levels 1 & 2 margin is 28-40% (compared to 30% in the base model). The range of precisely estimated effect sizes at the triage Levels 2 & 3 margin is 18-39% (compared to 21% in the base model). Similar to before, wait time is found to have no statistically significant impact on costs for patients at the margin of triage Levels

3 & 4. These findings support the conclusion that wait times exacerbate costs and there is effect intensification as acuity heightens.

VI.C.2. Sensitivity to Alternative Measure of Costs

To further test the robustness of our findings, we consider an alternative measure of costs. Here, we focus on the cost to the consumer. We define the cost to the consumer as the total amount of money that the hospital is paid for the encounter.

Payments to the hospital often differ substantially from the hospital's charge. In our data, there is a $33,420 disparity between the average payment ($6,334) and average charge ($39,754). Hence, measuring the impact of wait time on payments is not redundant.[16]

As shown in Table VII, first-stage point-estimates are far less precise in the payment analysis. This is likely due to the additional observations with *ln*(Payment=$0) which are excluded in the first-stage to ensure sample size consistency with the second-stage. Triage nurse stringency does affect wait time for patients on the margins of triage Levels 1 & 2 and Levels 2 & 3. However, the first-stage is weak at all other margins. Accordingly, we restrict second-stage analysis to just the triage Levels 1 & 2 and Levels 2 & 3 stratums.

Instrumental variable estimates for the impact of wait time on payments are generally imprecise. However, statistical significance aside, the magnitudes of the point-estimates are similar to those in the main cost analysis. For patients

[16] There is significant heterogeneity across payers in how and how much hospitals are paid. Therefore, in the payment analysis we control for insurance category. This is the only change we make to the base specification relative to the main cost analysis. The payer variable we obtain is measured at a very granular level and does not include meaningful labels. There are over 300 distinct (non-descriptive) payer codes, so we simply adjust for these codes using fixed effects.

who arrive on the margin of triage Levels 1 & 2, a one hour increase in wait time is associated with a 30% higher cost to the consumer; this is the same effect size as the estimated increase in the cost to the hospital. For patients who arrive on the margin of triage Levels 2 & 3, a one hour increase in wait time is associated with a 14% higher cost to the consumers; this is not far off from the estimated 21% higher cost to the hospital.

VII. DISCUSSION

VII.A. Explaining the Pattern of the Results

To provide intuition for our results, we present examples of two realistic scenarios. First, we focus on a patient who is typical of one who arrives at the ED on the margin of triage Levels 1 & 2. We then turn our attention to a typical patient on the margin of triage Levels 2 & 3.

Patients who come to the ED experiencing myocardial infarction deserve a Level 1 triage assignment. However, the symptoms of myocardial infarction (e.g., chest pain) are frequently mistaken for less serious conditions. These include myocardial ischemia, acute pneumonia, anxiety, etc. As a result, patients with myocardial infarction sometimes receive a Level 2+ triage assignment. Other times they are assigned Level 1.

Suppose a patient who is experiencing myocardial infarction receives a Level 2 triage assignment. This will prolong the patient's wait time relative to being assigned Level 1. The mean difference in wait time between a Level 1 and Level 2 patient in our sample is 53 minutes. Spending an additional 53 minutes waiting will greatly increase a myocardial infarction patient's probability of complications from their myocardial infarction including the development of heart failure. Treatment for heart failure often involves placement of an intra-aortic balloon pump; intra-aortic balloon pumps are expensive and could easily double the cost of care.

For rough validation purposes only, let use assume that no patient with myocardial infarction who is assigned Level 1 will require an intra-aortic balloon pump because their infarction is treated immediately, but all patients with myocardial infarction who are assigned Level 2+ will require an intra-aortic balloon pump due to their treatment delay. Let us further assume that provision of an intra-aortic balloon pump exactly doubles the cost of care[17] and that the marginal increase in wait time from Level 1 to Level 2+ is exactly one hour (i.e., close to 53 minutes). In our sample, 65% of the patients whose final primary diagnosis is myocardial infarction are assigned Level 1 and 35% are assigned Level 2+. Given these parameters, a one hour increase in wait time would be consistent with a 35% increase in costs. We hold this up against the 30% 2SLS point-estimate we obtain in the triage Levels 1 & 2 stratum.

Pneumonia is another condition commonly seen in the ED. Patients who come to the ED with pneumonia generally deserve either a Level 2 or Level 3 triage assignment. As pneumonia progresses, patients experience increasing need for oxygen and/or intravenous fluids. When a pneumonia patient's condition has worsened to the point of hypoxia, the standard of care is to provide supplemental oxygen. This treatment is not required for pneumonia patients whose condition has not worsened to the point of hypoxia. Albin et al. (1992) report that the provision of oxygen costs approximately $6.20/day; this is roughly $9.73 in 2010 dollars.

[17] Assuming an otherwise comparable baseline cost.

In our data, the average cost to care for a pneumonia patient is $7,400; an additional day in the hospital would cost approximately $1,680. Hence, if the development of hypoxia results in one extra day in the hospital with supplemental oxygen provided each day, then the cost to care for a pneumonia patient would increase by approximately 23%. The magnitude of this approximation is similar to the 21% 2SLS point-estimate we obtain in the triage Levels 2 & 3 stratum.

VII.B. Formula for Cost-Savings Potential

It is important to recognize that the effect sizes we estimate in this study are not a derivative of our sample's particular distribution of patient acuity. Each point-estimate represents a treatment effect at a particular position along the spectrum of acuity. This makes our results generalizable to other EDs with different distributions of patient acuity.

Any ED utilizing the Emergency Severity Index need only to weight our point-estimates to obtain an approximation for the amount of money it could save if it reduced its own patients' wait times, regardless of their patients' distribution of acuity. To ease with the calculation, we provide a simple formula for EDs to use. This formula approximates the percent reduction in total costs that would be realized if one-half hour were trimmed from each ED patient's time-to-physician. We emphasize that this approximation: (i) reflects a reduction relative to the original total cost to care for just those patients who present at the ED, and (ii) assumes an initial mean wait time of 26 minutes for Level 1 patients and 95 minutes across Level 2-5 patients. The base formula is given below:

$$\begin{array}{c}\text{Potential savings from}\\ \text{reducing all ED patients'}\\ \text{wait times 30 mins}\end{array} = \frac{-\left(\frac{0.30}{2}\right)\overline{C_1}\,Share_1 - \left(\frac{0.21}{2}\right)\overline{C_2}\,Share_2}{\overline{C_1}\,Share_1 + \overline{C_2}\,Share_2 + \overline{C_3}\,Share_3 + \overline{C_4}\,Share_4 + \overline{C_5}\,Share_5}$$

$\overline{C_L}$ reflects the average cost to care for a patient with triage Level L, and $Share_L$ reflects the share of ED patients who are triaged with Level L.

If we substitute the $\overline{C_L}$ values from our data into the equation above, then we obtain a more informative (albeit less generalizable) formula. The magnitudes of the $\overline{C_L}$ parameters are not as important as their relative sizes. We believe that the *relative* treatment costs for a Level 1 vs. 2 vs. 3 vs. 4 vs. 5 patient in our sample should be about the same as in others hospitals. Hence, our $\overline{C_L}$ imputations are not believed to greatly compromise external validity. Imputing mean costs yields the following:

$$= \frac{-\left(\frac{0.30}{2}\right)\ \$\mathbf{15,225}\ Share_1 - \left(\frac{0.21}{2}\right)\ \$\mathbf{8,239}\ Share_2}{\$\mathbf{15,225}\ Share_1 + \$\mathbf{8,239}\ Share_2 + \$\mathbf{2,883}\ Share_3 + \$\mathbf{838}\ Share_4 + \$\mathbf{403}\ Share_5}$$

In the average U.S. ED, 1-3% of patients are assigned Level 1, 20-30% are assigned Level 2, 30-40% are assigned Level 3, and 20-35% are assigned Level 4 or 5 (Gilboy et al., 2012). This means that the average U.S. ED could save approximately 7% if it trimmed one-half hour from each ED patient's wait time.[18] In our relatively high-acuity hospital, cost-savings might approach 10%.[19] The larger an ED's share of high acuity patients, the greater the potential for cost-savings.

VIII. CONCLUSION

Wait times in U.S. emergency departments are on the rise. This study examines the effect of ED wait time on the total cost to care for a patient. We

[18] The ~7% approximation assumes $Share_1$=0.02, $Share_2$=0.25, $Share_3$=0.40, $Share_4$=0.30 and $Share_5$=0.03

[19] Approximation based on $Share_1$=0.178, $Share_2$=0.304, $Share_3$=0.365, $Share_4$=0.146 and $Share_5$=0.007.

address the endogeneity of ED wait time by exploiting the quasi-random assignment of patients to triage nurses. Triage nurses frequently differ in their triaging stringency so patients who are triaged by a stringent nurse experience wait times that are systematically longer.

Our results suggest that prolonging a patient's ED wait time one hour will increase the cost of care for the most acute patients by an average of 30% and will increase the cost of care for moderately acute patients by an average of 21%. We detect no statistically significant impact of waiting among relatively healthier patients. There are two points of action which might come from these findings.

First, it would be beneficial to both patients (in terms of their outcomes) and hospitals (in terms cost-minimization) to drive down ED wait times in large, busy hospitals. Accomplishing this feat will likely require action from healthcare administrators and policymakers. Of course, some initiatives aimed at reducing ED wait times will, themselves, cost the health system money – for example, encouraging more hospitals to open an ED. However, not all initiative need be expensive. For instance, it is estimated that to treat a non-urgent condition in a primary care setting costs one-half to one-fifth the amount that it does to treat that same condition in an ED (NEHI, 2010). Encouraging some ED patients to seek care outside of the ED (in a primary care clinic, for example) could plausibly reduce congestion and improve patient flow within the ED. This, in turn, might lead to shorter wait times for the remaining patients who continue to frequent EDs.

Second, our finding of effect heterogeneity reiterates the importance of triaging. Costs increase most rapidly among the patients with the worst conditions. Hence, if sicker patients are not seen relatively faster than others, then this will independently exacerbate costs. This suggests a relatively easy first step that hospitals can internally take to realize cost-minimization. With presumably little investments, hospitals could focus efforts on triaging ED patients with greater precision. This alone could save the health system money.

REFERENCES

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444-455.

Bayley, M. D., Schwartz, J. S., Shofer, F. S., Weiner, M., Sites, F. D., Traber, K. B., & Hollander, J. E. (2005). The financial burden of emergency department congestion and hospital crowding for chest pain patients awaiting admission. *Annals of Emergency Medicine*, *45*(2), 110-117.

Bernstein, S. L., Aronsky, D., Duseja, R., Epstein, S., Handel, D., Hwang, U., ... & Schafermeyer, R. (2009). The effect of emergency department crowding on clinically oriented outcomes. *Academic Emergency Medicine*, *16*(1), 1-10.

Bhattacharya, J., Hyde, T., & Tu, P. (2014). The health policy conundrum. In *Health Economics* (pp. 306-327). Palgrave Macmillan.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, *90*(430), 443-450.

Centers for Disease Control and Prevention. (2016). National Hospital Ambulatory Medical Care Survey: 2013 Emergency Department Summary Tables. *Centers for Disease Control and Prevention, Atlanta, Georgia*. Retrieved from https://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2013_ed_web_tables.pdf

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317-372.

Fee, C., Weber, E. J., Maak, C. A., & Bacchetti, P. (2007). Effect of emergency department crowding on time to antibiotics in patients admitted with community-acquired pneumonia. *Annals of Emergency Medicine*, *50*(5), 501-509.

Gilboy, N., Tanabe, T., Travers, D., & Rosenau, A. M. (2012). Emergency Severity Index (ESI): A Triage Tool for Emergency Department Care, Implementation Handbook. *Agency for Healthcare Research and Quality, Rockville, MD, 4th ed. AHRQ Publication*, (12-0014). Retrieved from http://www.ahrq.gov/sites/default/files/wysiwyg/professionals/systems/hospital/esi/esihandbk.pdf

Hall, A. R., Rudebusch, G. D., & Wilcox, D. W. (1996). Judging instrument relevance in instrumental variables estimation. *International Economic Review*, 283-298.

Hing, E., & Bhuiya, F. A. (2012). *Wait Time for Treatment in Hospital Emergency Departments, 2009*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/data/databriefs/db102.pdf

Huang, Q., Thind, A., Dreyer, J. F., & Zaric, G. S. (2010). The impact of delays to admission from the emergency department on inpatient outcomes. *BMC Emergency Medicine*, *10*(1), 16.

Imbens, G., & Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467-475.

Krochmal, P., & Riley, T. A. (1994). Increased health care costs associated with ED overcrowding. *The American Journal of Emergency Medicine*, *12*(3), 265-266.

Lee, M. H., Schuur, J. D., & Zink, B. J. (2013). Owning the Cost of Emergency Medicine: Beyond 2%. *Annals of Emergency Medicine*, *62*(5), 498-505.

Liew, D., Liew, D., & Kennedy, M. P. (2003). Emergency department length of stay independently predicts excess inpatient length of stay. *Medical Journal of Australia*, *179*(10), 524-527.

Mahoney, D. (2015, October 19). The Word on ED Wait Times [Blog post]. Retrieved from http://www.pressganey.com/resources/blog/the-word-on-ed-wait-times

National Center for Health Statistics. (2017). National Hospital Ambulatory Medical Care Survey: 2013 Emergency Department Summary Tables. *Centers for Disease Control and Prevention*. Retrieved from https://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2013_ed_web_tables.pdf

New England Healthcare Institute (NEHI). (2010). A Matter of Urgency: Reducing Emergency Department Overuse, A NEHI Research Brief – March 2010. Retrieved from http://www.nehi.net/writable/publication_files/file/nehi_ed_overuse_issue_brief_032610finaledits.pdf

Richardson, D. B. (2006). Increase in patient mortality at 10 days associated with emergency department overcrowding. *Medical Journal of Australia, 184*(5), 213.

Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, *20*(4), 518-529.

Sun, B. C., Hsia, R. Y., Weiss, R. E., Zingmond, D., Liang, L. J., Han, W., ... & Asch, S. M. (2013). Effect of emergency department crowding on outcomes of admitted patients. *Annals of Emergency Medicine*, *61*(6), 605-611.

Table I. Balance in Patient Characteristics

	Natural Treatment Group (*Stringent*=1)	Natural Control Group (*Stringent*=0)	P-Value
Share Male	50.5%	50.3%	0.621
Mean Age	45.0 years	46.1 years	0.000
Final Diagnosis:			
Chest pain, unspecified (ICD-9: 786.50)	3.4%	3.3%	0.187
Headache (ICD-9: 784.0)	1.7%	1.7%	0.473
Abdominal pain, other specified site (ICD-9: 789.09)	1.6%	1.4%	0.003
Urinary tract infection, site not specified (ICD-9: 599.0)	1.5%	1.6%	0.563
Unspecified septicemia (ICD-9: 038.9)	1.3%	1.4%	0.006
	N=134,884	*N=52,265*	

The five most frequent final diagnoses (given by physician, not the triage nurse) are presented above.

Table II. Descriptive Statistics

	Mean
Wait Time	1hr 23min
Total Charge	$39,754
Total Payment	$6,334
Total Cost‡	$6,395
Male	50.4%
Age	45.3 years
Triage Assignment:	
Level 1	17.8%
Level 2	30.4%
Level 3	36.5%
Level 4	14.6%
Level 5	0.7%
	N=187,149

‡ Computed by multiplying patient's charge by the cost-to-charge ratio in the corresponding year.

Table III. OLS Estimates

	ln(Cost)
A. No Controls for Patient Characteristics	
Wait Time (hours)	-0.0750 ***
	(0.0013)
	N=186,465
B. Add Controls for Patients' Age, Sex, Diagnosis	
Wait Time (hours)	-0.0389 ***
	(0.0011)
	N=186,465
C. Add Triage Level Fixed Effects	
Wait Time (hours)	-0.0032 ***
	(0.0010)
	N=186,465

All regressions control for hour of arrival, weekday, month, and fiscal year. Observations with $0 costs are omitted. Values in parentheses are heteroskedastic-robust standard errors with * if p≤0.10, ** if p≤0.05, and *** if p≤0.01.

Table IV. First-Stage Estimates

	Wait Time (hours)
A. Level 1 & 2 Patients	
Stringent Nurse	0.0612 ***
	(0.0128)
	F-Stat: 55.77
	N=89,789
	Mean Wait: 59min
B. Level 2 & 3 Patients	
Stringent Nurse	0.1450 ***
	(0.0150)
	F-Stat: 80.59
	N=124,700
	Mean Wait: 1hr 38min
C. Level 3 & 4 Patients[†]	
Stringent Nurse	0.0642 *
	(0.0376)
	F-Stat: 17.17
	N= 17,433
	Mean Wait: 1hr 54min
D. Level 4 & 5 Patients	
Stringent Nurse	0.0287
	(0.0299)
	F-Stat: 28.49
	N=28,520
	Mean Wait: 1hr 21min

All regressions control for hour of arrival, weekday, month, fiscal year, age, sex, and diagnosis. For consistency, first-stage estimates do not include observations with $0 costs. Values in parentheses are standard errors with * if p≤0.10, ** if p≤0.05, and *** if p≤0.01.

† Omits 6am-9:59pm arrivals (so Level 4 patients never susceptible to upfront care).

Table V. Main 2SLS Estimates

	ln(Cost)
A. Level 1 & 2 Patients	
Wait Time (hours)	0.3045
	(0.1444) **
	[0.1795] *
	N=89,789
	Mean Cost: $10,822
B. Level 2 & 3 Patients	
Wait Time (hours)	0.2135
	(0.0527) ***
	[0.0714] ***
	N=124,700
	Mean Cost: $5,315
C. Level 3 & 4 Patients[†]	
Wait Time (hours)	0.1446
	(0.2884)
	[0.3107]
	N=17,433
	Mean Cost: $2,560
D. Level 4 & 5 Patients	
Wait Time (hours)	X
	N=28,520
	Mean Cost: $819

All regressions control for hour of arrival, weekday, month, fiscal year, age, sex, and diagnosis. Values in parentheses are unclustered heteroskedastic-robust standard errors and values in brackets are heteroskedastic-robust standard errors clustered by nurse×date (i.e., a proxy for the nurse's shift) with * if p≤0.10, ** if p≤0.05, and *** if p≤0.01.

† Omits 6am-9:59pm arrivals (so Level 4 patients never susceptible to upfront care).

Table VI. Sensitivity of 2SLS Estimates to Alternative Specifications

	Main Specification	ln(Cost)				
A. Level 1 & 2 Patients						
Wait Time (hours)		0.3078	0.1856	0.2806	0.2347	0.4040
		(0.1681) *	(0.1680)	(0.1513) *	(0.1593)	(0.1575) ***
		[0.2000]	[0.2279]	[0.2086]	[0.1950]	[0.1972] **
		N=90,221	*N=89,789*	*N=89,789*	*N=89,789*	*N=89,789*
				Mean Cost: $10,822		
B. Level 2 & 3 Patients						
Wait Time (hours)		0.1925	0.3224	0.3860	0.1804	0.2404
		(0.0579) ***	(0.0573) ***	(0.0640) ***	(0.0475) ***	(0.0516) ***
		[0.0740] ***	[0.0879] ***	[0.0983] ***	[0.0642] ***	[0.0691] ***
		N=125,236	*N=124,700*	*N=124,700*	*N=124,700*	*N=124,700*
				Mean Cost: $5,315		
C. Level 3 & 4 Patients†						
Wait Time (hours)		0.1884	0.0539	0.3364	0.0027	0.0824
		(0.3306)	(0.2149)	(0.4426)	(0.1608)	(0.2596)
		[0.3560]	[0.2366]	[0.5040]	[0.1702]	[0.2741]
		N=17,474	*N=17,433*	*N=17,433*	*N=17,433*	*N=17,433*
				Mean Cost: $2,560		
Controls:						
Time of Presentation	✓	✓		✓		✓
Clinical Characteristics	✓	✓			✓	✓
Payer						✓
Replace Cost=$0 with $0.01	No	Yes	No	No	No	No

All regressions control for fiscal year. Time of presentation controls include hour of arrival, weekday, and month fixed effects. Clinical characteristic controls include age, sex, and diagnosis fixed effects. Values in parentheses are unclustered heteroskedastic-robust standard errors and values in brackets are heteroskedastic-robust standard errors clustered by nurse×date (i.e., a proxy for the nurse's shift) with * if p≤0.10, ** if p≤0.05, and *** if p≤0.01.

† Omits 6am-9:59pm arrivals (so Level 4 patients never susceptible to upfront care).

Table VII. Sensitivity of 2SLS Estimates to Alternative Measure of Cost

	First-Stage	2SLS
	Stringent → Wait	Wait → ln(Payment)
A. Level 1 & 2 Patients	0.0492 ***	0.3020
	(0.0137)	(0.2420)
	F-Stat: 19.13	[0.2728]
	N=71,516	*N=71,516*
	Mean Wait: 59min	*Mean Payment: $11,367*
B. Level 2 & 3 Patients	0.1304 ***	0.1407
	(0.0170)	(0.0773) *
	F-Stat: 24.72	[0.0884]
	N=91,496	*N=91,496*
	Mean Wait: 1hr 38min	*Mean Payment: $4,722*
C. Level 3 & 4 Patients†	0.0366	X
	(0.0452)	
	F-Stat: 6.01	
	N=11,440	*N=11,440*
	Mean Wait: 1hr 54min	*Mean Payment: $1,847*
D. Level 4 & 5 Patients	0.0085	X
	(0.0384)	
	F-Stat: 8.13	
	N=16,598	*N=16,598*
	Mean Wait: 1hr 21min	*Mean Payment: $484*

All regressions control for hour of arrival, weekday, month, fiscal year, age, sex, diagnosis, and payer category. Values in parentheses are unclustered standard errors (heteroscedastic-robust in the 2SLS column) and values in brackets are heteroskedastic-robust standard errors clustered by nurse×date (i.e., a proxy for the nurse's shift) with * if p≤0.10, ** if p≤0.05, and *** if p≤0.01.
† Omits 6am-9:59pm arrivals (so Level 4 patients never susceptible to upfront care).

Figure I. Patient Flow Chart

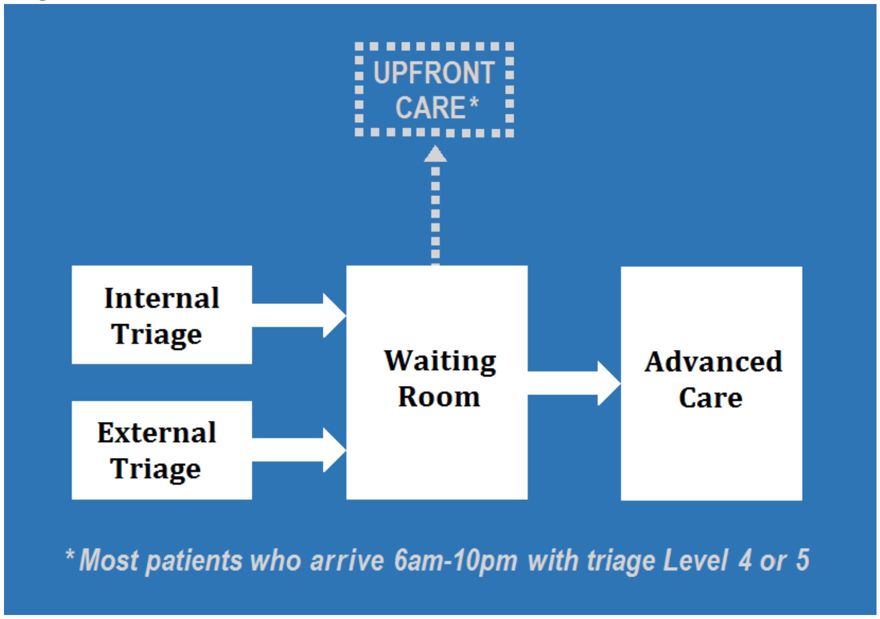
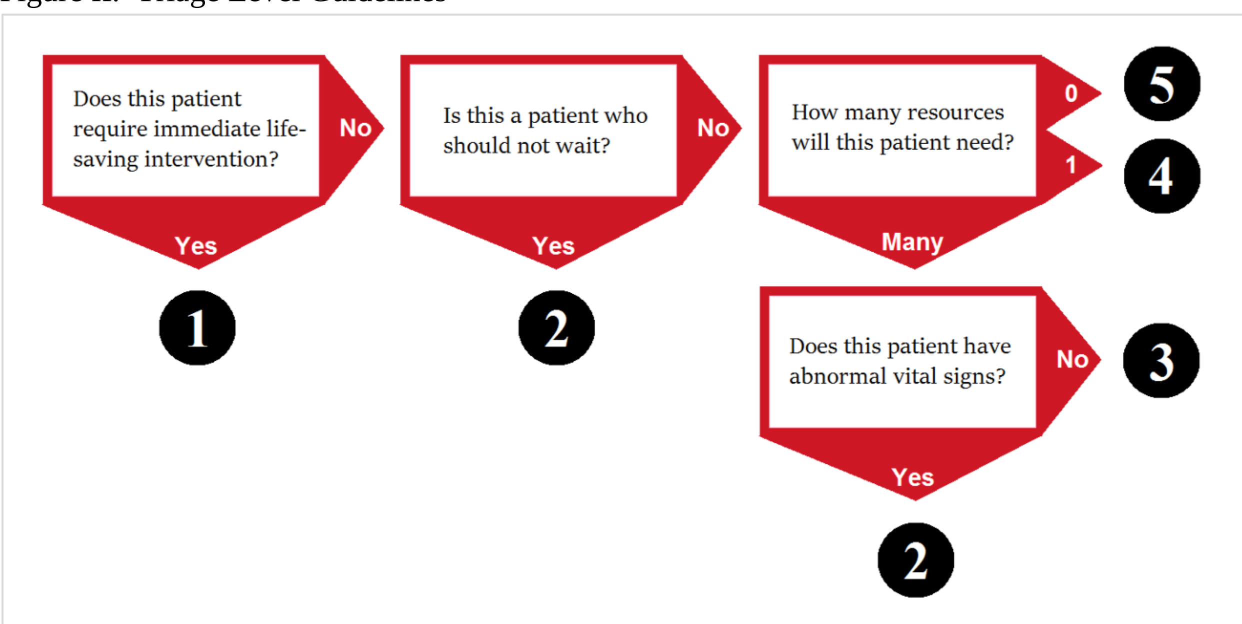
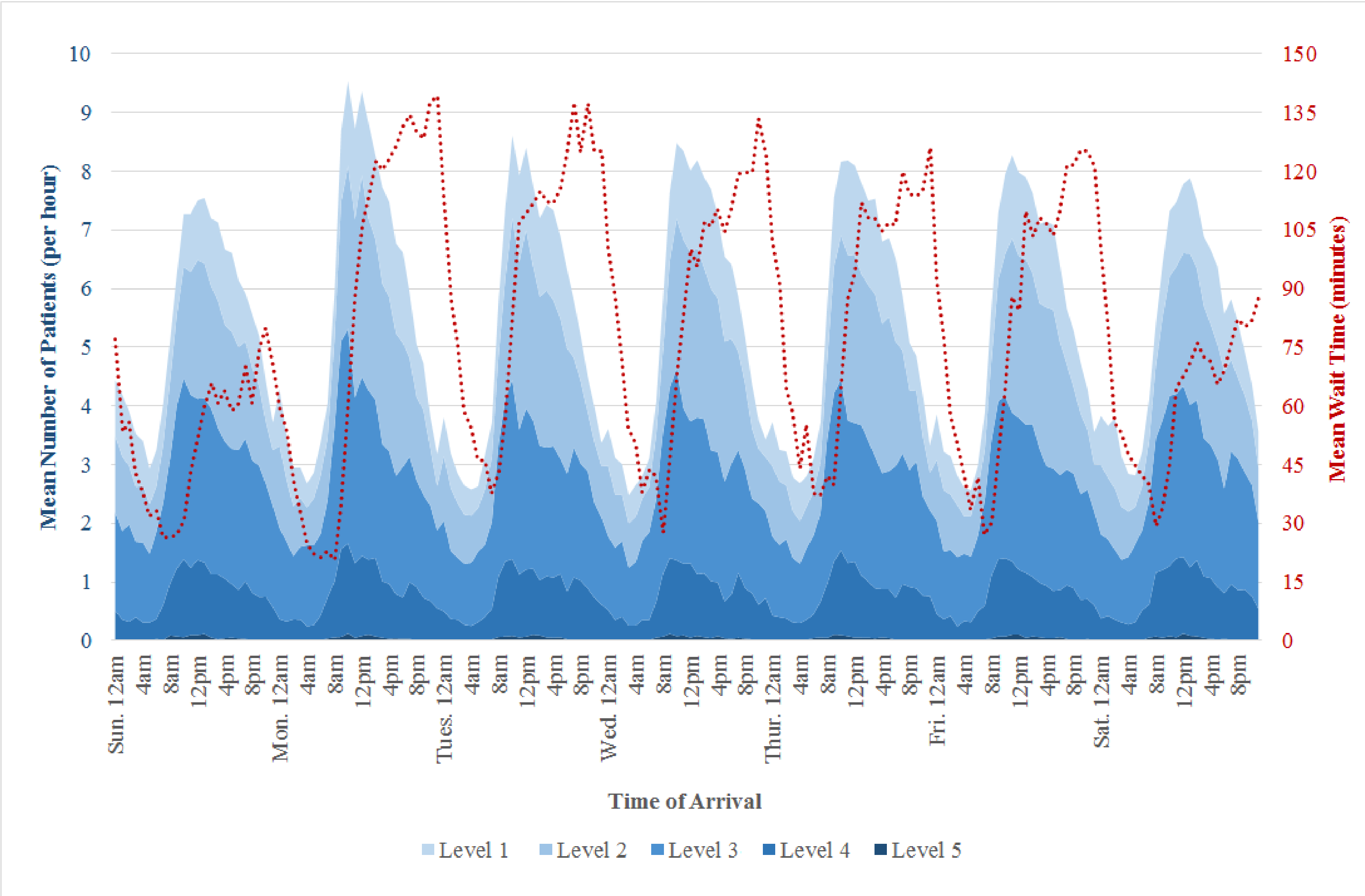


Figure II. Triage Level Guidelines



Adapted from Gilboy et al., 2012.

Figure III. Patient Volumes and Mean Wait Times, by Time of Arrival



Patient volumes reflect the frequencies in the final sample (excludes <18 and >89 year olds, walk-outs, etc.).

APPENDIX

To probe the source of the persistent endogeneity in the OLS model, we present OLS estimates stratified by triage level assignment.

Table A.1. Stratified OLS Estimates

	ln(Cost)
Wait Time (hours) | Level 1	-0.1146 ***
	(0.0056)
	N=33,245
Wait Time (hours) | Level 2	-0.0555 ***
	(0.0019)
	N=56,544
Wait Time (hours) | Level 3	0.0111 ***
	(0.0013)
	N=68,156
Wait Time (hours) | Level 4	0.0465 ***
	(0.0027)
	N=27,283
Wait Time (hours) | Level 5	0.0691 ***
	(0.0165)
	N=1,237

All regressions control for hour of arrival, weekday, month, fiscal year, age, sex, and diagnosis. Observations with $0 costs omitted. Values in parentheses are heteroskedastic-robust standard errors, with * if $p \leq 0.10$, ** if $p \leq 0.05$, and *** if $p \leq 0.01$.

These indicate a strong negative correlation between wait time and costs for Level 1 and Level 2 patients, but a positive correlation for Level 3, 4, and 5 patients. Moreover, the magnitude of the positive correlation increases with the “health” of the patient. This is a surprising result because we would expect wait times to have little effect on healthy patients’ costs. A straightforward explanation for this curious finding is self-selection in walk-outs. Willingness to wait will be relatively low for Level 5 patients; willingness to wait will be lowest among the healthiest Level 5 patients. This means that, conditional on triage assignment, “sicker” Level 5 patients will demonstrate more patience (and thereby have longer wait times) than the relatively healthier Level 5 patients who are quick to leave (because they have less need to be there). The same will hold true for Level 3 and Level 4 patients. Walking out is a viable option for these relatively non-acute patients. However, sorting will be most pronounced for patients with higher triage levels; these patients really don’t need to be there. This heterogeneous tendency to leave early is consistent with the larger positive coefficients we observe among patients with higher triage levels. In short, endogeneity remains a concern in OLS models even when coefficients are positive.