Essays on Decision-Making

Dissertation Advisors:
Professor Lawrence Katz
Professor David Cutler

Author:
Priya V. Shanmugam

Essays on Decision-Making

Abstract

I investigate how high-stakes decision-makers simplify complex choices, respond to choice fatigue, and learn from their mistakes. In chapter 1, I find that, when constrained, emergency room doctors engage in "satisficing" by shifting their attention away from the financial implications of their treatment choices and focusing solely on clinical considerations. In doing so, quality of care improves and the gap in care between insured and uninsured patients narrows significantly. In chapter 2, I find that when parole judges are fatigued by repeated choice, they resort to simple rules of thumb that increase the rate of parole to the least-risky incarcerated individuals but decrease the rate of parole for individuals serving life sentences. Accounting for the costs of future recidivism and incarceration, I find that the decisions made when judges are fatigued are optimal and cost-effective. Finally, in chapter 3, I show that emergency room physicians overreact to recent adverse events. Female physicians, however, react to, extrapolate from, and remember adverse patient events for longer. Gender differences in how physicians learn from mistakes shed light on one channel through which gender gaps in performance emerge and persist.

Contents

List of Tables

List of Figures

Acknowledgments

I would first like to thank Larry Katz, David Cutler, and Isaiah Andrews for advising me along this journey. They allowed me to pursue the ideas I was most genuinely curious about, and finish this program in a way that I feel proud of: assured of my own capacity to be curious, and to work hard. I also owe many thanks to Bruce Sacerdote, Doug Staiger, Jon Skinner, Chris Snyder, Sandip Sukhtankar, Annamaria Lusardi, and the entire faculty at the Dartmouth College economics department for fostering my interest in this field. I am delighted to reflect on the origins of Chapter 2 in Econ80 with Professor Staiger.

I am deeply grateful to Lisa Cook, Josh Goodman, Jamin Speer, Peter Blair, Gary Chamberlain, Claudia Goldin, and Ed Glaeser for providing helpful comments and encouragement throughout. I feel lucky to have been able to learn from all of you.

I am grateful to Brenda Piquet, Peter Brown, and Paul Millett for their help with the many logistical nightmares on which this work was built. Colleen Fiato at NY SPARCS, and Joan Schwartz, John Eckert, Kathleen Charron-Folsom and Andrea Goldberg at the New Hampshire Adult Parole Board were exceptionally generous with their time. I gratefully acknowledge funding support from the National Bureau of Economic Research Predoctoral Fellowship in Health and Aging, the Becker-Friedman Institute Predoctoral Fellowship in Health Economics, the Institute for Quantitative Social Sciences, and the Lab for Economic Applications and Policy.

I would not have been able to finish graduate school without the support of the many amazing people and communities who made my Cambridge life special. I'm grateful to the whimsical, funny, and deeply caring performers at ImprovBoston. It has been a privilege and a delight to perform with you all, be it at the PYPO Showcase, the Econ Department Holiday Party, Somerville Night Live, Club Drosselmeyer, Poseidon, The Last WASP, This Changes Everything!, Cinnamon Challenge, Harold Night (the Women Of shows, Houdini, Skateboard Graveyard, Gloria!, Mothership, and Babylon), All Access, The Democracy Center, Flotilla, DCM, The Field or 50 Vassar. In particular, I'm grateful to have shared this experience with Rachel Andelman, Kaitlin Buckley, Marc Campasano, Nick Garner, Max

Kreisky, Devin Quinlan, Jess Schmidt, and Kristina Stapelfeld. IB is a moveable feast, and I'm grateful to have had a seat at the table.

I am lucky to have lived in several incredibly supportive communities. I'm grateful to Ayda Ramadan, Maha Malik, and Zoe Mercer-Golden, who made 44 Columbia St. a beautiful first home. I cherish the memories of sleeping on bare floors while waiting for our IKEA mattresses, battling our first mice, becoming AMC regulars and throwing some great parties as we took on the real world.

I thank the students and fellow tutors who made my three years as a resident tutor at Winthrop House a delight. In particular, I am grateful to Winthrop's dining hall manager, Michelle Gray. Michelle was always the first person to notice if I missed a meal. Her warm presence made Winthrop feel like home.

I thank Jared Donohue, Caroline Jaffe, Aviva Musicus, Adam Reeve, Sam Spaulding, and of course, one Zephyr Spaulding-Jaffe for welcoming me into the curiosity, care, and revelry of Ellsworth House. The last year of my PhD was the most challenging yet, and I couldn't have asked for a more supportive group of people to go through it with.

I am doubly blessed to have had the opportunity to live the life of the mind outside of the economics department. In particular, I am grateful to have serving as a Teaching Fellow for Steven Pinker's course on Rationality, and to have taken three years of Tamil for heritage speakers with the inimitable Jonathan Ripley.

Can you thank a river? I thank the Charles River.

My dear, dear friends have made the tough parts of this journey easier to bear, and the fun parts even sweeter. I thank Lisa Abraham for many productive and not-so-productive days and nights at the Smith Campus Center, the grad student lounge, and the Woodberry Poetry Room. I thank Cristina Herren for our runs around the river, yoga classes, and many Tatte coffee chats. I am grateful to Soeren Henn for the years of friendship and support that he provided me. I thank Ayda Ramadan for her endless love and support, whether by leaving signs on our front door before my exams, or by cheerily accompanying me to the New Hampshire State Prison to help me photocopy the data that became the basis of

chapter 2. I thank Medha Raj, Joyce Cho, Lauren Magnuson, and Angela Dunnham for being my companions and sounding boards as we took on the challenges and delights of grad school, life, and growing up.

Raja Mama, Ayah, and Iyya, I'm grateful that grad school allowed me to be so close by. I will always cherish our meals in Chinatown and the Square, discussions about research, and evenings spent laughing.

Appatha, thank you for always being so loving. I consider the fact that I was able to learn Tamil here, to speak to you better, one of the greatest blessings of the PhD.

Appa, thank you for being so calm, funny, thoughtful, and supportive of me. I often think of a phone call I had with you while waiting for a poké bowl at Wholesome Fresh. I was bemoaning the research process, and you told me the tale of R.G. Gallager and the parity-checking bits. You always know exactly what to say.

Amma, thank you for always being so positive, energetic, and supportive of me throughout this entire experience. Your own journey in science is an inspiration to us all. You always having a metaphor to inspire me with at a moment's notice - be it about a frog, a sunflower, or a trampoline - and I am so lucky that you are never more than a phone call away.

Divya and Nithya, thank you both for being amazing people in your own rights, and amazing sisters to me. I often lay claim to teaching you everything you know about style and life, but truthfully, you teach me a lot, too. Divya, thank you for always being there to code, scrape, model or have dinner or coffee together. I am lucky to have had you so close by. Nithya, thank you for always encouraging me and listening to me, and for sending me off to the AEAs in true parent-style. Sister Summer with you both was truly one of the best parts of the PhD. And Titan, thank you for showing us unbeatable love and strength.

These years showed me how truly loved and supported I am, and I hope to pay these blessings forward every day.

To M&M, my two favorite Ms, and T.

Introduction

In this dissertation, I empirically investigate the mechanisms behind, and consequences of, theoretical concepts from the behavioral economics literature using administrative datasets from high-stakes decision-making settings.

In chapter 1, I investigate how high-stakes decision-makers take complicated choices and transform them into simple choice. Complex, high-stakes decisions are often made solely by human experts. However, many of these decisions are made under significant cognitive constraints. I estimate the causal impact of an increase in cognitive constraints on the quality and equity of Emergency Department care using the universe of ED visits across New York from 2005-2015. I define cognitive constraints as a function of variation in the number and complexity of other patients a doctor sees at the same time. Patients arriving when the ED is busy versus empty are of similar ex-ante health, but differ in how cognitively constrained their physician is. My empirical analysis focuses on two common complaints: chest pains, where decision-making aids in the form of simple risk-scoring tools are plentiful, and abdominal pains, where no such aids are available. I show that, when constrained, doctors reallocate care away from low-risk, insured patients and towards high-risk, uninsured patients. These reallocations significantly reduce the disparity between insured and uninsured patients in hospital admission, specialty inpatient services, and 1-year patient mortality. When decision-making aids are available (versus absent), treatment reallocations are highly cost-effective; variation in treatment both within and across hospitals is reduced; and doctors' algorithms for evaluating uninsured patients converge to the algorithms of insured patients. I rule out changes in ED staffing, triage, and binding physical capacity constraints as alternative mechanisms. Overall, cognitive constraints can cause both the quality and equity of high-stakes decision-making to improve, and their effects hinge critically on the presence of decision-making aids.

In chapter 2, I investigate how decision fatigue affects important repeated choices. Evidence from previous studies shows that the likelihood of granting parole decreases dramatically as judges get fatigued. I examine parole outcomes in New Hampshire, where parole hearings are ordered alphabetically. I find that as judges become fatigued, they are more likely to grant parole. This suggests that decision fatigue is a psychological phenomenon unrelated to emotional changes. Second, I find that, when fatigued, judges trend toward a simple rule of thumb: they grant parole more to all incarcerated individuals, except those serving life sentences, who become less likely to receive parole as the day wears on. Lastly, I use data on future criminal activity and incarceration costs to analyze the cost-effectiveness of these fatigue-induced changes in decision-making. I find that each slot later in the day a hearing is held translates to a nearly $400 reduction in societal costs of incarceration and future crime.

In chapter 3, I show that men and women perceive and learn differently from their own mistakes and the mistakes of others. Using an event study of every patient death occurring in Emergency Departments (EDs) in New York between 2005 and 2015, I show that female ED physicians adjust their treatment behavior significantly more than male physicians after a patient they are treating dies in the ED. I decompose this effect into three channels. After the patient's death, female physicians - relative to male physicians - 1) become significantly more cautious, admitting more patients with the same concern as the patient who died; 2) over-extrapolate from the adverse event, admitting more patients with medical concerns that differ from that of the index patient; 3) change their behavior significantly more when their peer physicians experience a patient death; 4) remember the adverse event for longer, maintaining increased risk-aversion and over-extrapolation after the adverse event has occurred.

Chapter 1

Decision-Making Under Cognitive Constraints

"No one is busier or needs more bandwidth than a generalist physician."

David Loxterkamp
Physician & Author

Emergency department (ED) physicians perform taxing mental work[1]. They must effectively diagnose and treat a wide variety of health concerns - from suspicious rashes to motor vehicle trauma - while also considering resource constraints and financial incentives. Mistakes can cause waste, costly delays in treatment, and even death (Fordyce *et al.*, 2003). However, ED physicians face another unique constraint that distinguishes ED care from every other part of the healthcare system: they must handle *all* simultaneous patient traffic, regardless of how many patients arrive at once, or how complicated these patients are. I investigate an important, but understudied consequence of ED traffic: its effect on the *cognitive* constraints faced by doctors, and the subsequent decision-making strategies they employ to turn these high-stakes, complex clinical choices into simple, solvable ones.

[1] Staying ahead of getting behind: Reflections on"scarcity". British Medical Journal (Loxterkamp, 2014)

Patient traffic already poses a problem for EDs, as the rate of ED usage is growing at twice the rate of the population (Gonzalez Morganti *et al.*, 2013). The efficient operation of EDs has important financial and equitable implications. EDs are responsible for nearly half of all hospital admissions, which make up the biggest fraction of healthcare spending. EDs also serve as the primary healthcare safety net for uninsured patients (Burke and Paradise, 2015), as EDs are the only part of the healthcare system mandated to serve patients without regard to their ability to pay.[2]. Indeed, in 2011 the Society for Academic Emergency Medicine specifically identified the impact of ED traffic on the equity of ED care as a pressing issue (Hwang *et al.*, 2011)

Due to the complex nature of ED care, patient traffic could affect the quality and equity of care through several channels. Increased wait times delay the provision of treatment and worsen short-term patient mortality (Woodworth, 2019). Distortions in treatment choices caused by physicians working more rapidly could improve or worsen patient mortality. Binding physical resource constraints that wholly prevent patients from receiving timely treatment worsen patient mortality (Johnson and Winkelman, 2011).

Less attention has been paid to how ED traffic affects how doctors allocate their attention and arrive at their decisions. These cognitive constraints could improve or worsen the quality and equity of ED care depending on the decision-making shortcuts that doctors take, and the ways in which they reallocate their attention, when constrained. For example, there is evidence that fatigue and stress can increase racial bias (Ma *et al.*, 2013). However, cognitive constraints caused by increased ED traffic could also induce efficient reallocations of attention or stricter adherence to guidelines such that quality of care actually improves (Pines, 2017). Evidence from the implementation of laws mandating a reduction in ED treatment times in the UK shows that as doctors work faster, patient survival improves (Gruber *et al.*, 2018).

---

[2]The Emergency Medical Treatment and Active Labor Act, passed by Congress in 1986, is a landmark federal mandate guaranteeing uninsured patients nondiscriminatory access to emergency healthcare. However, EMTALA only covers the patient's right to (1) receive a medical screening exam and (2) receive stabilizing care in the event of a medical emergency.

In this paper, I investigate the causal impact of ED traffic on how doctors allocate their attention when making clinical decisions, and further investigate the subsequent impacts on the quality, as measured by patient survival and cost-effectiveness, and equity, as measured by disparities in care between insured and uninsured patients, of ED care. I leverage the random nature of ED patient arrivals to isolate quasi-exogenous variation in ED traffic. I define traffic as a function of the number and complexity of patients arriving at an ED in the past two hours: doctors who must treat more (and more serious) patients at once have less bandwidth available for each patient. Comparing similar patients who arrive when doctors have their "hands full" versus doctors who are less busy isolates the effect of the cognitive constraint. I rule out changes in physical capacity constraints, ED staffing, patient triage, and waiting times as alternative mechanisms.

Decision-making aids such as checklists or scoring tools have emerged as a possible solution to the problem of cognitive constraints, though they reduce the amount of discretion available to decision-makers. My analysis focuses on two common complaints: chest pains, where decision-making aids in the form of simple risk-scoring tools are widely available, and abdominal pains, where such tools are not available. I compare the causal impact of ED traffic on clinical decision-making for these two types of ED visits.

I conduct my empirical analysis using administrative data from every Emergency Department visit occurring across the State of New York from 2005 to 2015. The data, which cover approximately 70 million ED visits, include patient demographics, codes for every procedure performed and every diagnosis code given over the course of the ED visit, and most importantly, timestamps indicating the hour of patient arrival and hospital IDs, which I use to create a detailed measure of hourly fluctuations in facility-level complexity-scaled ED traffic.

Conditional on a small set of observable patient demographic traits and visit characteristics, patients who arrive when their doctor's hands are full versus empty are similar, both in terms of their likelihood of having pre-existing health conditions, the composition of patients by race and by insurance type, and their overall likelihood of adverse outcomes such as one-year mortality or hospital admission. However, they receive markedly different amounts of diagnostic and therapeutic care, and experience different levels of care quality and cost overall.

I show that when doctors become more cognitively constrained, they reallocate hospital admission and therapeutic treatments toward high-risk, uninsured patients, and away from low-risk, insured patients. They reallocate diagnostic testing in the opposite direction. These reallocations significantly reduce the disparity between insured and uninsured patients in hospital admission, specialty inpatient services, and patient mortality.

These reallocations of care induced by changes in ED traffic are highly cost-effective only when decision-making aids are available, as is the case for chest pain patients. For these patients, ED traffic causes doctors' algorithms for evaluating uninsured patients to converge to the algorithms they employ in assessing insured patients. ED traffic also reduces variation in treatment for observably similar patients both within and across hospitals in the presence of decision-making aids. For abdominal pain patients, where no decision-making aids are available, changes in care induced by ED traffic are not cost-effective, and within- and across-hospital variation in treatment increase.

Overall, I show that cognitive constraints can improve both the quality and equity of ED care, but their effects hinge critically on the presence of decision-making aids. My research contributes to four distinct literatures: the effects of ED crowding, the role of physician behavior in improving healthcare delivery, the dynamics of cognitive constraints, and the role of decision-making guidelines and rules in high-stakes choice settings.

My research establishes cognitive constraints as a specific, important channel through which ED traffic affects the quality of care. I add to the literature on ED crowding, which has established the effects of ED traffic on hospitals' financial losses (Foley *et al.*, 2011), costly delays in treatment (Johnson and Winkelman, 2011), decreased patient satisfaction (Zibulewsky, 2001), and short-term patient deaths (Woodworth, 2019).

I show that changes in doctors' choice strategies have significant impacts on patient spending, diagnostic and therapeutic intensity of care, and patient survival. Physician education (Schnell and Currie, 2017), training (Chan, 2016), beliefs (Cutler *et al.*, 2013), race (Alsan *et al.*, 2018), hospital-wide practice styles (Molitor, 2018), procedural skill and ability to effectively identify patients with the highest marginal benefit (Currie and MacLeod, 2017) have all been shown to play a role in treatment choices and disparities therein. I further show that the cognitive constraints channel also contributes to the "unwarranted variation" problem in healthcare: persistent differences in observed treatment choices for similar patients across geographic areas and hospital types (Wennberg, 2002).[3]

I further shed light on the specific mechanisms by which cognitive constraints themselves act. My research has important implications for understanding the effects of cognitive constraints in any high-stakes choice setting: judges deciding who to jail and who to parole, police officers deciding whom to stop and with how much force. I show that, when standardized shortcuts are available, cognitive constraints cause doctors to rely on them, and when such aids are absent, doctors rely on less-effective shortcuts. The impacts of cognitive constraints have been demonstrated in judicial (Danziger *et al.*, 2011; Yang, 2015a) and consumer settings (Iyengar and Lepper, 2000b) as well as in individual labor supply decisions (Thakral and To, 2018). Recent work has shown that doctors who are more behind-schedule are more likely to prescribe opioids (Neprash and Barnett, 2019). Much attention has been devoted to trying to understand the source of these mistakes, whether through a limited ability to process information, incorrect understanding of what information is necessary, inherent errors or information acquisition frictions (Handel and Schwartzstein, 2018).

Lastly, my paper speaks to a more recent literature on decision-making with the use of guidelines or recommendation systems. I show that the optimal allocation of discretion between physicians and decision-making aids is one channel through which the quality of care can be improved, and that doctor discretion benefits ex-ante low risk patients, while potentially harming ex-ante high-risk patients. The role of discretion in high-stakes decisions

[3]Unwarranted variation is a key target for healthcare cost reduction and quality improvement (Sabbatini *et al.*, 2014)

has more recently been explored in hiring (Hoffman *et al.*, 2018), doctors' decisions to pursue diagnostic imaging (Abaluck *et al.*, 2016), health insurance plan choices (Bundorf *et al.*, 2019), usage of chest CT scans in emergency departments (Venkatesh *et al.*, 2018) and vaccination choices (Rao and Nyquist, 2018).

This paper is organized as follows. In Section 2, I briefly describe the emergency department setting. In Section 3, I describe the theoretical concepts relevant to understanding clinical decision-making under cognitive constraints. In Section 4, I describe how I construct measures of cognitive constraints, doctor decisions, and patient outcomes. In Section 5, I describe my empirical methodology. In Section 6, I describe my results in detail, and in Section 7, I discuss and rule out several alternative mechanisms. I conclude in Section 8.

1.1 Emergency Departments

Emergency departments across the state of New York handle seven million visits each year. When a patient seeks care in an ED - as opposed to scheduling a visit with their primary care provider - the visit is typically unscheduled and perceived as somewhat urgent. The patient checks in with a triage nurse, briefly describes the issue, and then waits to be seen. The triage nurse distills the patient's brief, verbal description of their issue into one "chief complaint". These chief complaints generally take the form of non-technical symptom descriptions such as "chest pain", "abdominal pain", "fever" or "head injury". Table 1.1 shows the twenty most common chief complaints and their frequencies across New York in 2009. The triage nurse then assigns the patient to a physician based on the information available: the patient's chief complaint, age and gender. Physician assignments are also partially determined by caseload: as patients arrive, they are assigned to physicians to balance work across all physicians working in the ED.

When the patient sees the physician, a verbal description of the issue is given and the patient's relevant health history and symptoms are reviewed. The physician may perform a verbal or physical examination, order diagnostic tests or therapeutic procedures, make a diagnosis, and prescribe medications or follow-up care. One of every seven ED visits

Table 1.1: Top 20 Chief Complaints

	Number of Visits	% of Visits
Fever	316760	11.52
Chest Pain	232465	8.46
Abdominal Pain	230442	8.38
Cough	219511	7.99
Limb Pain	186986	6.80
Headache	160180	5.83
Abdominal Pain, other	125124	4.55
Leg Injury	116365	4.23
Backache	114270	4.16
Sore Throat	113110	4.12
Shortness of Breath	106658	3.88
Skin Issue	105348	3.83
Head Injury	104720	3.81
Fainting	96385	3.51
Vomiting	95734	3.48
Lower Back Pain	94729	3.45
Dizziness	93608	3.41
Chest Pain, other	90558	3.29
Upper Resp Infection	74042	2.69
Dental Issue	71699	2.61
Total	2748694	100.00

Notes: This table reports the top twenty most common chief complaints and their frequencies across all EDs in the state of New York in 2009. Each ED visit is given a "chief complaint": a broad, non-technical summary of the patient's symptoms at the time of ED arrival. Each visit is given just one chief complaint, which is recorded using ICD-9-CM diagnosis codes.

results in the patient being admitted into the hospital to receive more intensive therapeutic or diagnostic care. These choices are determined based on the physician's assessment of the patient's risk of adverse outcome, but simultaneously constrained by the facility's resources.

Specialized procedures may be performed by a consulting physician who does not work primarily in the ED, though these choices are constrained by whether the specialist is working at the time.[4] ED physicians must also consider whether the patient requires a transfer to another ED or hospital. Psychiatric inpatient wards, for example, are not available or guaranteed to have availability at most hospitals.[5] Patient health histories and vital signs are taken by physician's assistants, nurse practitioners, or registered nurses. These other licensed healthcare professionals often also perform simpler diagnostic or therapeutic procedures, like IV placement. Billing is handled after the encounter, when hospitals negotiate with the patient's insurance plan. The ED visit concludes when the physician decides to either discharge a patient whose needs have been met, or admit into the hospital a patient who requires inpatient care. Patients who are admitted into the hospital usually receive inpatient care for more than one day.

Emergency room doctors thus have a challenging task: they must manage an unpredictable flow of patients - each with a medical situation that could range from a foreign object stuck in the body to a possible stroke - while bearing in mind facility-level capacity constraints, various financial incentives, the reliability of information given by patients, and the relative costs of over- and under-treatment errors. Appendix A.1 gives an example of a typical ED encounter.

---

[4] For example, "cancer doesn't grow on the weekends" is a common refrain referencing the fact that certain specialists, such as radiologists, are usually unavailable on weekends and evenings.

[5] Despite the fact that hospitalizations overall have decreased since 2012, hospitalization of patients with mental illnesses has increased. The number of inpatient beds dedicated to psychiatric care in New York, however, has decreased, primarily due to private hospitals shuttering their inpatient psychiatric services. The bulk of these patients have been subsumed by public hospitals.

1.2 Modeling Doctor Decision-Making

My approach to understanding the effects of "decision-making bandwidth" on doctors' choices builds on two concepts familiar to the behavioral economics literature. The first is the concept of bounded rationality (Simon, 1955), or the idea that the rational economic agent does not have enough "bandwidth" to consider all information that is relevant to a given choice. That is, mental work is constrained by a budget - the amount of cognitive bandwidth available - and for many important decisions, it creates a binding constraint.

Second, the literature on scarcity shows that having more considerations on one's mind leads to decreases in cognitive function and executive control - capacity is fixed and tasks thus compete with each other for limited mental resources (Mani *et al.*, 2013). In the medical setting, I assume that doctors have a limited, fixed amount of bandwidth with which to make medical decisions and they must split this bandwidth across all patients they are dealing with at once. I build on work showing that bandwidth deteriorates over time (Danziger *et al.*, 2011), as a function of choice complexity (Iyengar and Lepper, 2000b), and as a result of having to simultaneously juggle many choices. I examine the ED doctor-patient interaction with a focus on the mental work - the statistical, information-gathering and information-processing choices - a doctor must perform.

1.3 Variable Construction and Summary Statistics

1.3.1 Data Sources

I combine data from four sources to draw conclusions about how physicians respond to bandwidth constraints and the consequences of constrained decision-making. I first describe these four data sources, and then describe my methodology for constructing several measures of physician bandwidth, decisions, patient characteristics and outcomes used throughout the analysis.

My primary data source is the New York Statewide Planning and Research Collaboration (SPARCS) administrative inpatient and outpatient datasets. These anonymized, identifiable data contain every emergency department, ambulatory surgery clinic, urgent care center and hospital visit in the state of New York from 2005-2015 (NY SPARCS, 2014a,b). SPARCS data is ideal for this study for three reasons. First, the data is a census of all inpatient, outpatient, urgent care clinic and ambulatory surgery activity across the state of New York, allowing for my analysis to incorporate the rich variation in hospital types, physician practice styles, and case types.[6] Second, the data includes the hour, date, and physician IDs of each visit, which allow me to estimate physician work schedules and emergency room traffic flows by the hour, as well as track patients over the decade-wide sample. Third, detailed patient demographics, payment information, subsequent mortality measures, the patient's chief complaint, anonymized patient IDs and all diagnosis, procedure and billing codes are included, allowing my analysis to create rich measures of patient health histories and investigate the variety of doctor decisions - from diagnostic testing, to therapeutic treatments, to final medical diagnoses - made during an ED visit.[7]

I augment these visit-level data with medical schooling and licensing information for the doctors in my sample. These data were acquired via a FOIL request to the New York State Education Department. The data contain each practitioner's name and license number, date and expiration status of medical license, city, name of medical school, and date of medical school degree for the approximately 90,000 unique doctors across the SPARCS dataset.

Using the doctor's full name and license number, I supplement the information on medical schooling and licensing with data on graduate and specialty medical training obtained from the New York State Physician Profile website (`www.nydoctorprofile.com`), a website maintained by the New York State Department of Health. These data include the self-reported dates, institutions and fields of specialty for all training - including

[6]The only visits which are entirely redacted from the data are visits in which the patient has HIV/AIDS, or is receiving an abortion. Outside of these two protected categories, every visit to an emergency department, hospital, ambulatory surgery center or urgent care clinic is included in the data.

[7]SPARCS data is constructed from medical billing records, and as such is skewed towards data that appears on claims. Both diagnostic and therapeutic procedures that are provided during a visit are recorded, and medical diagnosis codes that are given to justify the provided procedures are also recorded. Data on the details of physician-patient interactions are sparse: physician notes, test results, or any verbal data collected during the interaction are not included.

residencies and fellowships - obtained after a physician has graduated from medical school.

I further characterize hospital-specific measures of quality, utilization and cost from the American Hospital Directory Hospital Profile. Cost-to-charge ratios, facility-level Total Performance Scores (TPS) quality indicators, number of hospital beds and average number of inpatient days are pulled from the AHD Hospital Profile and are used to deflate reported visit charges. Information on hospital capacities and clinical decision support technology usage comes from the American Hospital Association (AHA) Annual Survey and Healthcare IT Databases.

1.3.2 Approximating Physician Bandwidth

In order to study the effects of changes in ED traffic on physicians' cognitive constraints, I start with the simple assumption that cognitive resources available are inversely proportional to cognitive resources occupied. That is, doctors have a fixed amount of decision-making bandwidth and must allocate it towards all of the patients, of varying levels of complexity, that they are treating at once. If a doctor is seeing three patients, she has less bandwidth available for the fourth patient than a doctor who has not been assigned any patients at all. Likewise, if a doctor is seeing three complicated patients, she has less bandwidth available to see a fourth patient than a doctor who is currently seeing three very simple patients.

An ideal proxy for occupied physician bandwidth would capture variation in how many choices - each scaled by its complexity - a doctor must consider at any one time. To capture how much of a doctor's bandwidth is currently occupied, I construct a measure of ED crowding based on how many patients - scaled by how complicated each patient is - are being seen in the emergency room when the index patient arrives.

Appendix A.2 details the provider IDs that appear on each record. Because SPARCS data does not preserve the IDs of the ED doctors who see patients who are eventually admitted into the hospital, this creates a purely mechanical relationship between patient health characteristics and apparent ED staffing patterns. Appendix A.3 contains a detailed discussion and simulation of this issue. For this reason, I do not explicitly scale ED traffic by the number of doctors working in the ED. Note that while complexity-scaled ED traffic is measured at the ED level, it should be strongly correlated with each doctor's individual level of complexity-scaled traffic: if the ED as a whole is busy, each doctor should be busy.

1.3.3 Measuring ED Traffic

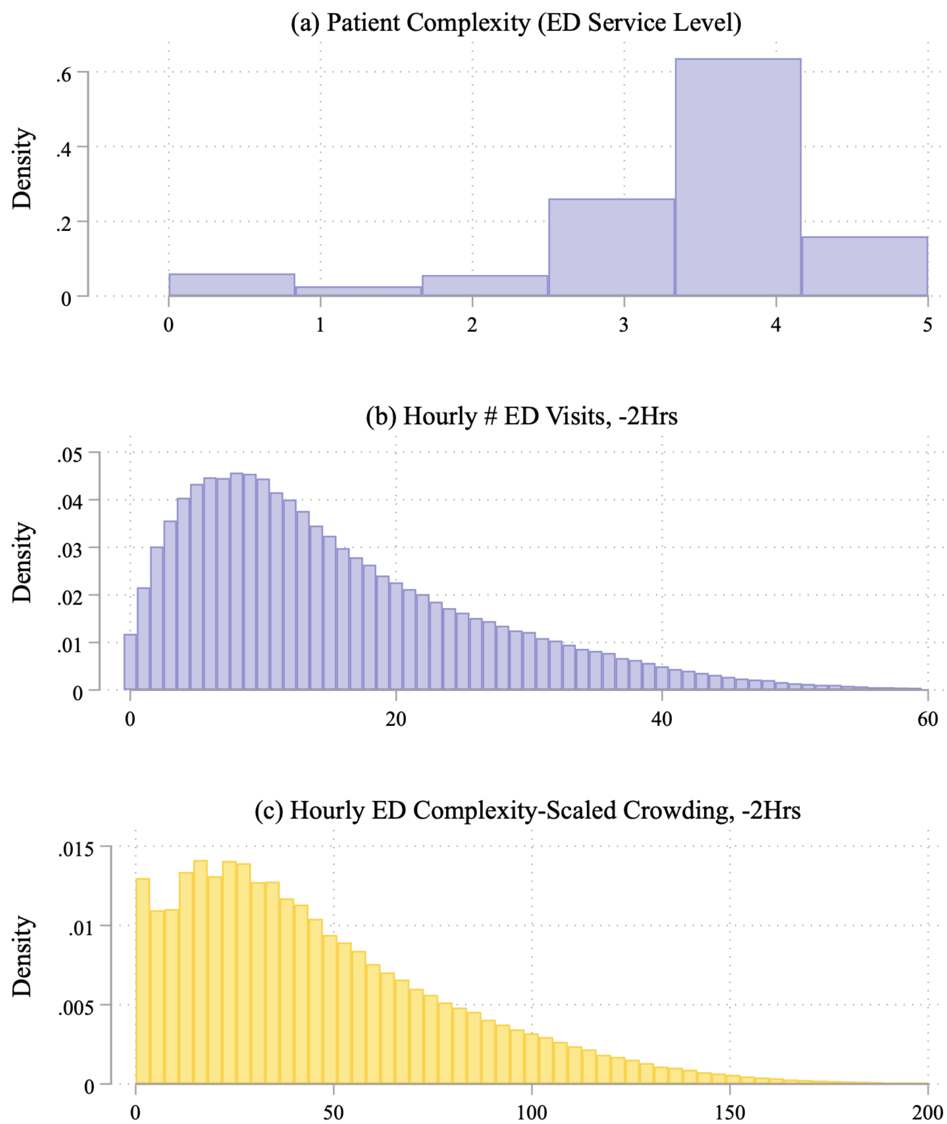
Because ED patients receive treatment for an average of two hours, I define the available bandwidth a doctor d has to treat patient p who arrives at hour h as follows. I take all patients arriving at the index ED in hours $h-1$ and $h-2$. Each ED visit is assigned a billing "level" 1 through 5, based on the required level of detail in diagnosing and treating the patient, as well as the severity of the patient's problem. The billing critera for ED service levels is given in Appendix A.4. The total amount of ED services amassed by all patients seen in the ED in the two hours prior to the index patient's visit yields my proxy for physician cognitive constraints.

For example, consider two scenarios. Patient P arrives at 2pm. The ED received one patient per hour for the past two hours, and each of these patients received ED Level 2 and 3 services respectively. The complexity-scaled traffic in the ED when Patient P is being seen is 2+3=5. Patient Q arrives at 2pm. The ED has not received any patients in the last two hours. The ED traffic measure for patient Q is 0. Figure 1.1 describes the distribution of patient complexity, two-hour patient volume, and the combined complexity-scaled two-hour patient volume measure.

My measure of ED traffic is most similar to the Emergency Department Work Index (EDWIN)[8] and the Boston ED Work Score, both of which take the sum of the number of patients in the ED, scales them by their triage category, and divides this sum by the number of physicians working and beds in the ED.[9] Among four well-known ED crowding scores - the Real-time Emergency Analysis of Demand Indicators (READI), EDWIN, the National

[8] https://www.hindawi.com/journals/emi/2012/838610/tab1/

[9] The Boston ED Work Score then adds other aspects of ED crowding, including patients waiting to be admitted, and patients in the waiting room.



Notes: This figure describes the construction and distribution of the complexity-scaled ED traffic measure. This measure is defined as the sum of the ED Service Levels for each patient arriving in the ED in the past two hours. Panel (a) shows the raw distribution of ED Service Levels. Panel (b) shows the distribution of the number of patients arriving in the ED in the past two hours. Panel (c) shows the distribution of the number of patients, each scaled by their ED Service Level, arriving in the ED in the past two hours. 80% of the variation in complexity-scaled traffic is driven by changes in patient complexity, and 20% is driven by patient volume.

Figure 1.1: Distribution of Complexity-Scaled 2-Hour ED Traffic

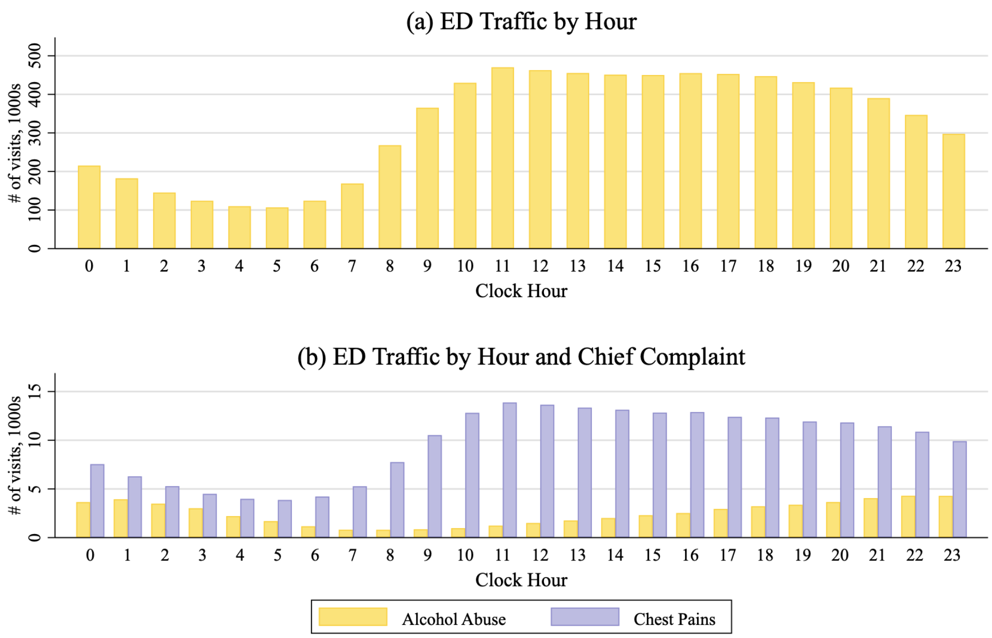
Emergency Department Overcrowding Study (NEDOCS) scale, and the Emergency Department Crowding Scale (EDCS), EDWIN was among those found to have good scalability and predictive power across various levels of actual crowding (Jones *et al.*, 2006).

My measure differs in that I do not scale patient traffic by the number of physicians working or the number of beds in the ED. Because I aim to capture differences in how many choices a doctor must think about at once, physical capacity constraints are unlikely to affect the number of choices a doctor must make for the patients they are seeing. Due to the selective omission of ED doctor IDs for admitted patients, which I discuss in detail in A.3, I leave variation in ED staffing on the table. Should EDs be able to fully compensate for changes in traffic with changes in staffing, these changes would bias my analysis against finding any effect of crowding on cognitive constraints.

1.3.4 ED Arrival Timestamps

Each ED visit includes a timestamp for the date and hour of arrival. This timestamp identifies the time at which the patient arrives in the ED and checks in with the triage nurse. Importantly, this is not the time at which the patient physically sees the physician or the time at which diagnosis and treatment decisions are made. The time of arrival is the only timestamp available in the data, and the only timestamp that is plausibly random. I use this timestamp throughout my analysis.

This timestamp is recorded accurately, as it is needed in case of malpractice. Several descriptive statistics of the arrival hour variable support that it is accurately reported. Figure 1.2 describes the distribution of the arrival hour variable overall, and for two different chief complaints: chest pain and alcohol abuse. Chest pain complaints peak at noon, while alcohol abuse complaints peak in the early hours after midnight. Both distributions show that patient flows do not appear to "bunch" at any hour, further supporting that the arrival hour variable is recorded accurately.



Notes: This figure describes the relationship between the hour of the day and the number of ED visit arrivals. Panel (a) shows this distribution for all complaints. Panel (b) shows the distribution separately for two common chief complaints: chest pains and alcohol abuse. Chest pains peak in the morning, while alcohol abuse cases peak in the early hours after midnight. The distributions of these visits, as well as the distribution of ED visits overall, do not indicate excess masses or bunching at "round" hours like noon or midnight, supporting that the arrival hour is recorded accurately.

Figure 1.2: Distribution of ED Traffic by Hour and Chief Complaint

1.3.5 Characterizing Treatment Decisions

Each diagnostic and therapeutic procedure available to a patient represents a binary decision on the part of the ED physician. I construct binary indicators for each specific testing or treatment decision, and also create aggregate measures of the intensity of overall diagnostic and therapeutic care provided.

The procedures performed during the visit are characterized by a set of procedure codes and billing codes. For ED patients admitted into the hospital, every single procedure provided during the visit is recorded using International Statistical Classification of Diseases Volume 9 (ICD-9) codes. For ED patients who are not admitted into the hospital, these procedures are reported using Current Procedural Terminology (CPT) codes. Both ICD-9 and CPT codes are highly specific (e.g. ICD-9 87.06 "Contrast radiograph of nasopharynx" or CPT 70150, "Complete radiograph of facial bones").

To create broader measures of procedure choice, I rely on billing information. Revenue codes are reported using National Uniform Billing Committee (NUBC) revenue codesets. If a revenue code appears on the record for a visit, a line-item for the corresponding category was billed by the facility for that visit. For example, I construct a binary indicator for whether or not a patient was given prescription medication based on whether any "pharmacy" revenue code (NUBC 025X) appears on the patient's record. Appendix A.5 describes the most common revenue line-items and their frequencies for several chief complaints. Appendix A.6 describes how I categorize NUBC revenue codes in greater detail.

I use the Healthcare Cost and Utilization Project Procedure Classes tool to create aggregate measures of the intensity of diagnostic and therapeutic care. The Procedure Classes tool classifies each CPT and ICD-9 procedure codes as either "diagnostic" or "therapeutic". I create measures for the number of diagnostic procedures, number of therapeutic procedures, and total charges for the visit. These measures together characterize the intensity of care provided during the visit. Appendix A.7 summarizes these variables for common chief complaints. Appendix A.8 describes the HCUP Procedure Classes tool in greater detail.

1.3.6 Characterizing Diagnosis Decisions

Each ED visit in the SPARCS dataset includes up to 25 ICD-9 diagnosis codes: one "primary" diagnosis, representing the physician's conclusion as to the primary cause of the chief complaint, and up to 24 "ancillary" codes representing pre- and co-existing conditions. Hospitals are partially reimbursed based on whether the reported primary and ancillary diagnosis codes medically justify the procedures provided to the patient; hospitals therefore have an incentive to report these codes thoroughly. Appendices A.9 and A.10 describe the most common primary diagnoses for specific chief complaints.

To characterize the physician's diagnostic decisions, I construct binary indicators for whether or not the patient received each of the first through tenth most common primary and ancillary diagnoses for their given chief complaint. I create a binary indicator for whether or not the patient receives a primary diagnosis that is different from their chief complaint, and also construct the number of ancillary diagnoses given. These measures represent the diagnostic intensity of the patient's ED visit. Appendix A.7 reports summary statistics for these measures for selected chief complaints.

1.3.7 Measuring Treatment Quality

To measure the quality of care provided during an ED visit, I focus on two common clinical endpoints: whether patients *die* after their ED visit, and whether patients *return* to the ED after their visit. SPARCS data includes mortality indicators (derived from New York Vital Statistics death records) at 7, 15, 30, 180 and 360-day intervals following the patient's date of discharge from the initial ED visit. Table 1.2 summarizes mortality rates for selected chief complaints.

Hospital revisit rates are regarded as proxies for the quality of care provided based on the simple logic that issues unresolved during the index visit are likely to cause a patient to return for further unscheduled care. (Rising *et al.*, 2014). I identify whether patients subsequently return to any ED or hospital for further treatment within 7 or 30 days of discharge from their initial visit. This measure includes any subsequent visit in which the

Table 1.2: Rate of Patient Mortality n Days Post-Visit, by Chief Complaint

	(1) Chest Pain	(2) Chest Pain, Other	(3) Ab Pain	(4) Ab Pain, Other
7d Mortality	0.444	0.400	0.461	0.233
	(6.646)	(6.312)	(6.772)	(4.817)
15d Mortality	0.556	0.494	0.585	0.304
	(7.434)	(7.010)	(7.628)	(5.506)
30d Mortality	0.749	0.674	0.807	0.436
	(8.619)	(8.183)	(8.945)	(6.591)
180d Mortality	2.116	1.962	2.093	1.307
	(14.39)	(13.87)	(14.32)	(11.36)
360d Mortality	3.269	3.062	2.985	1.965
	(17.78)	(17.23)	(17.02)	(13.88)
Observations	2441502	963370	2416394	1425590

Notes: This table summarizes the mortality rate, in percentage points, of patients at various intervals after their initial ED visit, for patients arriving in the ED with various chief complaints. The n-day mortality rate represents the rate at which a patient dies within the n days since being discharged from their index visit. If a patient visits the ED on day 1 and is admitted into the hospital for a 14-day stay, the 7-day mortality variable indicates whether this patient had died by day 22 ($1 + 14 + 7$). Mortality records are provided in the SPARCS data and are derived from New York Vital Statistics Records.

patient returns to an ED and is discharged (revisit), or returns and is admitted into the hospital (readmission). I further identify cases in which a patient returns with a complaint that is medically related to the chief complaint for their index visit. For example, I count a patient as returning with the same or related complaint if they presented with chest pains in their index visit, and subsequently return to any ED or hospital with any cardiological chief complaint within the specified time window.[10]

1.3.8 Characterizing Patient Health

I construct several variables to describe a patient's health status prior to their ED visit. Race, ethnicity, gender, and age (in months) are provided in the data. I construct a patient's insurance status based on up to six "methods of payment" that appear on the record. These fields may include a health insurance plan, Medicare, Medicaid, or may indicate that the patient paid in cash or did not pay. If the two latter categories are the only forms of payment that appear across the six available insurance fields on the record, I classify the patient as uninsured. Appendix A.11 provides summary statistics for these patient demographics. Appendix A.12 further details how forms of payment are recorded in the data.

I further characterize a patient's health status by utilizing diagnosis codes given during encounters prior to the index ED visit. I create indicator variables for specific prior health events or conditions such as a previous heart attack, hypertension, high cholesterol or diabetes. Importantly, using past diagnosis codes creates a detailed picture of a patient's health prior to their index visit, and allows me to characterize aspects of patient health that are clinically important. For example, of the 15 aspects of patient health that comprise the HEART Score for Major Adverse Cardiac Events (detailed in Appendix A.14), approximately 12 can be detected using the ICD-9-CM diagnosis codes in my data. Appendix A.13 provides summary statistics of these constructed binary variables for selected chief complaints.

[10]Because my data includes every facility across the state of New York, I am able to capture an important fraction of healthcare facility revisit behavior that is missing from studies that are restricted to single facilities. Readmission rates are approximately 30% higher when visits to non-index facilities are included (Duseja *et al.*, 2015).

Importantly, while some prior health conditions may be transient, my prior health variables capture whether a patient has *ever* or *never* been previously given a certain diagnosis.

1.3.9 Protocolized and Unprotocolized Case Types

To study the effects of constraints on decision-making strategies, I draw a distinction between ED visits where decision-making aids are available, and ED visits where physicians receive very little standardized guidance and must rely on their own training, previous clinical experience, and *gestalt*. My analysis sample consists of two separate subsamples: chest pains and abdominal pains, two chief complaints respectively with and without plentiful aids.

Medical protocols generally fall into three categories. The first aims to prevent tail events (such as sepsis or conflicting medications) from occurring by flagging a standard set of warning signs associated with such an event (Larsen *et al.*, 2011). The second aims to universalize the use of low-cost, high-value procedures such as hand-washing for all medical professionals (Allegranzi and Pittet, 2009) or the administration of aspirin within 30 minutes of arrival for patients presenting with a possible heart attack (Centers for Medicaid and Medicare Services, 2010; Saketkhou *et al.*, 1997). The third aims to standardize the processes by which doctors arrive at their final decisions - the inputs doctors attend to, and the relative weights they give to these inputs - and thus decrease interobserver variability. Arguably the most successful protocol of this type is the APGAR score, a tool created by Dr. Virginia Apgar in 1953 to quickly risk-stratify newborns non-invasively, as objectively as possible, and within sixty seconds (Apgar, 1953).

The first and second category of protocol can be thought of as tools that encourage corner solutions: aids that aim to increase best-practice adherence or procedure provision to 100%. Given the low cost ($\sim$\$0.28 for an aspirin; several seconds for hand-washing) and large benefits of these practices, the optimal rate of aspirin provision and physician hand-washing cannot be less than 100%. The optimal rate of EKG provision, however, is strictly less than 100%, and this distinguishes the third category of decision-making aids.[11] Therefore, this

[11]Technically, the optimal rate of EKG provision is 100% for *some subset* of patients and 0% for the remainder,

category of medical protocol aims to standardize the process by which doctors arrive at these strictly non-corner solutions.

I focus on two common ED complaints: chest pains and abdominal pains, which are the second and third most common ED complaints, each accounting for approximately one quarter million yearly ED visits in New York. Because of the prevalence of rare but often fatal Major Adverse Cardiac Events (MACEs, such as heart attacks and pulmonary embolisms), several risk-scoring tools have been developed to help doctors quickly identify patients who may have these conditions. Appendix A.14 details several of these scoring tools. Tintinalli's Emergency Medicine Manual, a popular reference handbook for ED physicians, explicitly recommends that the practitioner use these scoring tools in its entry on chest pain (Tintinalli *et al.*, 2011). By contrast, there is no such risk-scoring tool for abdominal pain patients. While there are fewer major adverse endpoints for abdominal pain patients - the most common acute reasons for abdominal pains are appendicitis or burst ovarian cysts - untreated or under-treated chronic conditions exert a considerable burden on the healthcare sector and the emergency room specifically.

1.4 Methodology

1.4.1 Identification Assumptions

To understand the effect of physician bandwidth constraints on decision-making, I compare the testing and treatment decisions doctors make, and subsequent patient mortality, for visits occurring when the ED is busy versus empty, where "busy" and "empty" are defined by the complexity-scaled measure of two-hour ED traffic described previously. The key identifying assumption underlying this approach is that potential treatment decisions and potential health outcomes - that is, the treatments and outcomes that would have been realized had the patients arrived at the ED at different level of crowding - are not systematically different

---

but to accurately partition patients into these two subsets would require data that physicians, computers, and researchers do not possess.

between patients who arrive when the ED is busy compared to those who arrive when the ED is empty. Under this assumption, differences in treatment and quality outcomes between patients who arrive at different levels of ED busyness can be interpreted as causal effects. I discuss three potential violations of this assumption, and describe my approach to addressing these violations.

First, since my estimation sample contains every ED across the state of New York from 2005-2015, variation across ED facilities and the underlying health of the populations they serve could generate a spurious relationship between physician bandwidth and patient outcomes. For example, hospitals in urban areas tend to be more crowded than those in suburban areas, and the typical patient in an urban area is more likely to be low-income and have poor health. Similarly, economic changes within a geographic area, such as expansions of healthcare insurance coverage, area-wide health shocks, or changes in the number or quality of available health facilities could simultaneously lead to changes in the health of the patient population, and changes in observed ED utilization. I control for facility-by-year fixed effects to remove potential confounds due to variation across ED facilities and within facilities over time.

Second, comparing the outcomes of ED visits during "busy" and "empty" times could introduce confounding due to differences in potential patient outcomes and resource availability during daytime versus nighttime hours. It is plausible that a chest pain patient arriving in the dead of the night, when ED traffic is usually light, might be different from a chest pain patient arriving at noon. Indeed, heart attacks are most common in the morning because blood pressure tends to be highest then. The resources available to ED physicians also vary across the hours of the day: specialists such as radiologists may not be available during nighttime hours, for example. I control for variation in potential patient outcomes and ED resource availability across the hours of the day with clock-hour fixed effects.

Lastly, patients may have different potential treatment and health outcomes during busy versus empty EDs, not due to differences in physicians' choices, but due to differences in how patients are triaged. However, triage nurses can only make these prioritization decisions off of a very small set of ex-ante observable patient characteristics: chief complaint, age, and gender. I thus include these triage variables as controls in the model to alleviate this concern. It is also unlikely that the relative priority assigned to chest pain and abdominal pain patients would vary significantly with changes in ED traffic, as both chief complaints are associated with potentially fatal, acute health concerns, and both accordingly receive high priority in the ED.

1.4.2 Estimating Equation

I estimate the causal impact of physician bandwidth on treatment choices and patient outcomes using the following estimation equation. I estimate this equation separately for chest pain visits and abdominal pain visits. For instance, for all patients arriving in the ED with a chief complaint of chest pain, the treatment choices and outcomes for patient i who arrives at facility f at date-hour h in year y is

$$y_{ifc,h|c(i)=\text{chest pain}} = \beta \mathbf{X}_{i,h} + \delta_{f,y(h)} + \alpha_{\text{clock}(h)} + \text{triage}_{i,h} + \gamma \text{crowding}_{f,h} + \epsilon_{ifc,h}. \tag{1.1}$$

$y_{ifc,h}$ represents each of the outcomes of interest associated with the encounter: binary indicators for the provision of specific tests and treatments, variables indicating overall treatment intensity, and subsequent patient mortality, as described previously. $X_{i,h}$ is a vector of patient i-specific health history measures at hour h, patient race and insurance type indicators. $\delta_{f,y(h)}$ represents facility-by-year fixed effects. $\alpha_{\text{clock}(h)}$ represents clock-hour fixed-effects 0 through 23. $\text{triage}_{i,h}$ represents the set of triage controls: an indicator for patient gender and a quadratic age term. $\text{crowding}_{f,h}$ represents the measure of complexity-scaled 2-hour ED crowding. $\epsilon_{ifc,h}$ represents the error term.

$\hat{\gamma}$ represents the effect of variations in ED crowding on doctor decision-making and patient outcomes. While the most obvious sources of potential confounds in the relationship between ED crowding and patient outcomes are the ones detailed above, it is possible that, after flexibly controlling for these sources of variation, there could still be a relationship between ED crowding and observable and/or unobservable determinants of patient potential outcomes. For example, patients could decide to leave if the ED appears to be crowded, and could do so differentially based on their potential health outcomes. I present four tests that suggest that this is unlikely.

In Figure 1.3, I plot the relationship between ED crowding and patient health conditions prior to their index visit. Panel (a) shows the raw relationship between these variables. Panel (b) shows the relationship after controlling for variation across the hours of the day. Panel (c) further residualizes these variables on the full set of facility-by-year fixed effects. After controlling for both sources of variation, patients appearing when the ED is busy versus empty appear to have similar levels of prevalence of ex-ante health conditions. Appendices A.15 and A.16 repeat the same exercise for the composition of patients by race and by form of insurance, respectively.

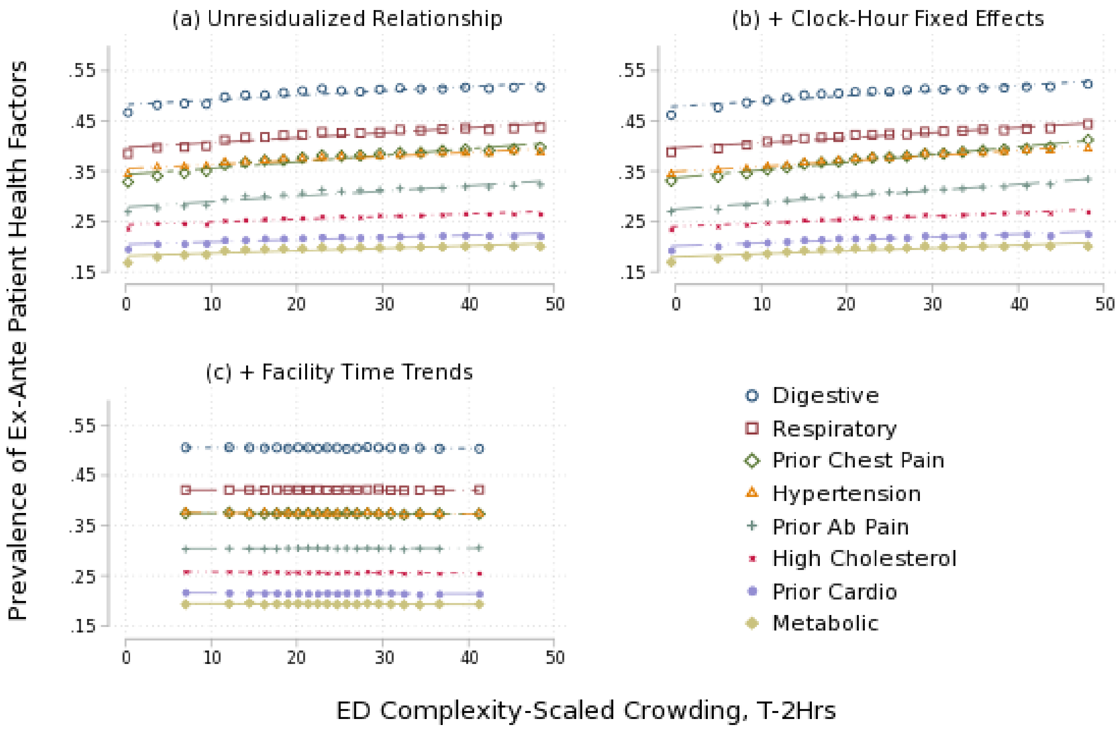
1.4.3 Heterogeneity by Patient Type

I then take the full set of patient prior health conditions, race, and insurance type, and use them to create a set of risk scores. These risk scores capture a patient's likelihood of experiencing an adverse health outcome, or receiving a certain treatment. I estimate the following equation for several outcomes: 1-year mortality, 1-month mortality, 30-day hospital revisit, total visit charges, and a binary indicator of whether or not the patient is admitted into the hospital.

$$y_{ifc,h|c(i)=\text{chest pain}} = \beta \mathbf{X}_{i,h} + \delta_{f,y(h)} + \alpha_{\text{clock}(h)} + \text{triage}_{i,h} + \epsilon_{ifc,h} \tag{1.2}$$

The patient's risk score is $\hat{\beta}X_{ih}$ where X_{ih} represents the full set of patient prior health conditions, race, and insurance type. I then repeat the same exercise as in Figure 1.3, Appendix A.15 and Appendix A.16 using these patient risk scores. Appendix A.18 shows that, once facility-specific time trends and time-of-day variation are controlled for, patients do not appear to differ by their overall risk of experiencing an adverse outcome when the ED is busy versus empty.

I create two measures of patient type, meant to capture two key dimensions of a patient



Notes: This figure shows the relationship between ED complexity-scaled crowding and patient health characteristics prior to the index visit. Panel (a) shows the raw relationship between this measure of crowding and binary indicators for patient health conditions. Panel (b) shows the relationship after the variables have been residualized on fixed-effects for each hour of the day. Panel (c) shows the relationship after further residualizing on facility-by-year fixed-effects.

Figure 1.3: Relationship between Ex-Ante Patient Health and ED Crowding

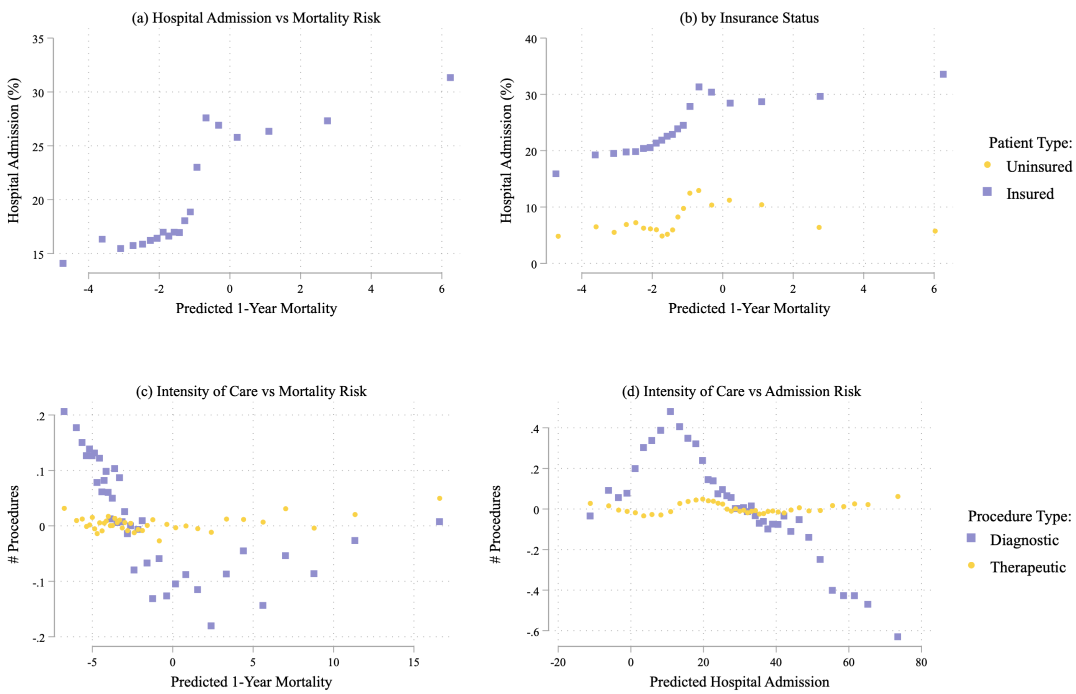
that a doctor must consider when deciding on treatment. The first is the patient's level of health, and the second is the patient's ability to pay. I investigate the effects of cognitive constraints on doctors' treatment choices across these two dimensions of patient type.

Mortality Risk: Since patients of different risk levels require different amounts of testing and treatment, the effect of cognitive constraints on observed treatment choices and patient outcomes may depend on the risk level of the patient. To account for this potential source of heterogeneity in the effect of cognitive constraints, I classify patients into groups based on whether their predicted 1-year mortality risk is "high", "medium" or "low" if their risk score falls in the top 25, middle 50, or bottom 25 percent of the risk distribution for their given chief complaint. The mortality risk score captures the patient's likelihood of dying within a year of their index visit, as predicted by their gender, age, ex-ante health characteristics, race, and insurance status as in Equation 1.2. The mortality risk score aims to capture differences in how "healthy" or "unhealthy" a patient is.

Importantly, I use 1-year mortality - a strictly health-based endpoint - to capture differences in patients' overall health. Using predicted treatment outcomes, such as total healthcare costs, as a proxy for health - instead of using health endpoints explicitly - can create bias in predicted patient "health" if patients receive differing amounts of care due to discrimination by race or insurance status (Mullainathan and Obermeyer, 2019). For example, a patient's admission risk score captures differences both in the patient's health status *and* their ability to pay[12]. For my analysis, I aim to create a measure of patient risk that isolates the patient's overall level of health.

The nonparametric relationship between patient risk and rates of testing and treatment illustrates the difference between a health endpoint-based risk score and a treatment endpoint-based risk score. Panel (a) of Figure 1.4 shows the nonparametric relationship between the patient's risk of mortality and their risk of admission. Sicker patients are more likely to be admitted. Panel (b) shows the same relationship for insured and uninsured

[12]In my sample, chest pain and abdominal pain patients who have any form of insurance are nearly six times more likely to receive hospital admission (25% versus 4% for insured versus uninsured patients, conditional on having the same overall mortality risk.)



Notes: Panel (a) of this figure shows the nonparametric relationship between a patient's risk of mortality and their risk of hospital admission. Panel (b) shows the same relationship separately for insured and uninsured patients. Panel (c) shows the relationship between a patient's risk of 1-year mortality and the average amount of diagnostic (testing) and therapeutic (treatment) care they receive. Panel (d) shows the same relationship based on the patient's risk of hospital admission.

Figure 1.4: Patient Risk and Intensity of Care

patients separately. While the positive relationship between mortality risk and admission risk remains, the disparities in admission likelihood between uninsured and insured patients is large - uninsured patients are 10 to 20 percentage points less likely to be admitted into the hospital than insured patients with the same mortality risk. Panel (c) shows the relationship between the patient's mortality risk and the number of diagnostic tests and therapeutic treatments they receive. Sicker patients - patients with higher mortality risk scores - receive fewer tests and more therapeutic treatments. Panel (d) show the relationship between the patient's risk of admission and the average number of diagnostic tests and therapeutic treatments they receive. Patients with a higher risk of admission also receive more treatments, but the relationship between admission risk and diagnostic testing is inversely U-shaped: the patients least likely to be admitted are also less likely to receive testing.

Ability to Pay: My data reports up to six "forms of payment" used by the patient to pay for their visit. I designate a patient as "insured" if they report any form of insurance other than "Self-Pay". I designate a patient as "uninsured" if the only form of payment appearing in any of the six Forms of Payment fields is “Self-Pay". Thus, insured patients include those on Medicare, Medicaid, Workers' Compensation, Blue Cross, Disability, and other smaller federal and state health insurance programs. Appendix A.12 details the various forms of insurance coded in my data.

I create a categorical variable, $\mathbf{R}_i$, that represents the intersection of the patient's mortality risk group and insurance status. I interact my measure of ED crowding with this six-cell composite patient type variable, yielding the following estimation equation:

$$y_{ifc,h|c(i)=\text{chest pain}} = \beta \mathbf{X}_{i,h} + \delta_{f,y(h)} + \alpha_{\text{clock}(h)} + \text{triage}_{i,h} + \gamma \text{crowding}_{f,h} \times \mathbf{R}_i + \mathbf{R}_i + \epsilon_{ifc,h} \tag{1.3}$$

$\hat{\gamma} \times \mathbf{R}_i$ represents the causal impact of cognitive constraints on the treatment choices and subsequent health outcomes for patients in each of the six groups created by the intersection of the mortality risk and insurance status groups. Table 1.3 describes the distribution of patients across these six subgroups.

Table 1.3: Distribution of Patients by Insurance and Mortality Risk Subgroups

	Death Risk: Low	Mid	High	Total
Uninsured	12.27	10.31	2.72	25.30
Insured	12.73	39.69	22.28	74.70
Total	25.00	50.00	25.00	100.00

Notes: This table describes the distribution of abdominal pain and chest pain patients by both their 1-year mortality risk, and by their insurance status. I categorize patients into one of six subgroups based on the intersection of their mortality risk group and their insurance status. Insured patients are much more likely to be medium- and high-risk, while uninsured patients are overrepresented among the low-risk group, likely due to differences in access to primary care.

1.5 Results

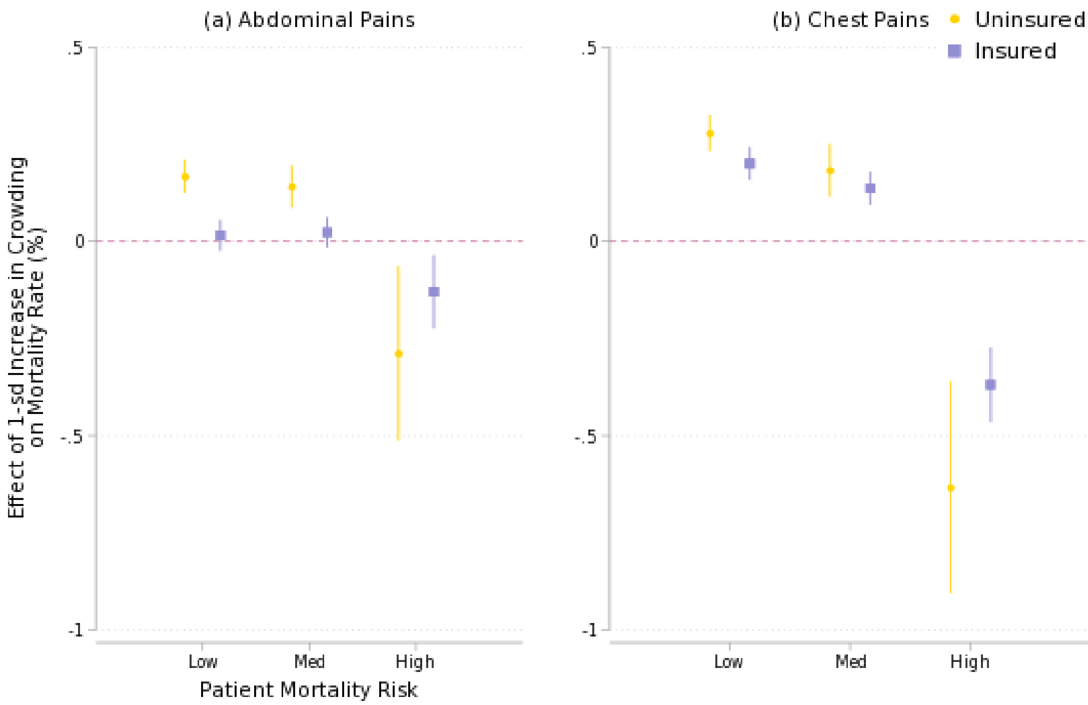
I first discuss the effects of ED traffic on patient mortality. I then turn to the changes in hospital admission, diagnostic and therapeutic care that drive the effects on patient mortality. I discuss the cost-effectiveness of these distortions in care, and then provide evidence supporting that ED traffic causes doctors to reallocate their decision-making bandwidth, and the efficiency of these reallocations hinge critically on the presence of decision-making aids.

1.5.1 Cognitive Constraints Improve Quality of Care for the Sickest Patients

Both in the presence (chest pains) and in the absence (abdominal pains) of decision-making aids, cognitive constraints causes 1-year patient survival to improve among the highest-risk patients, while slightly worsening mortality among low-risk patients.

Figure 1.5 plots the coefficients on the effects of ED traffic on patient mortality across six patient subgroups defined by mortality risk and insurance status, following Specification 1.3. Table 1.4 reports these effects and their corresponding effects on the gap in mortality between insured and uninsured patients.

For abdominal pain patients, where decision-making aids are absent, the gains in survival for sickest patients - and the corresponding reductions in survival for the least-sick



Notes: This figure plots the coefficient estimates of the causal impact of a 1-sd increase in ED complexity-scaled two-hour traffic on the 1-year patient mortality of abdominal pain patients, in Panel (a), and chest pain patients, in Panel (b). An increase in ED traffic causes large improvements in survival for high-risk patients, and small increases in patient mortality for low-risk patients. These effects are more pronounced for uninsured patients.

Figure 1.5: Effect of ED Traffic on Patient Mortality

Table 1.4: Effect of Crowding on Insured-Uninsured Mortality Gap

	Abdominal Pains			Chest Pains		
	(1) Low	(2) Med	(3) High	(4) Low	(5) Med	(6) High
Insured	0.016	0.023	-0.130	0.201	0.137	-0.369
s.e.	0.020	0.020	0.048	0.022	0.022	0.049
Uninsured	0.167	0.142	-0.288	0.279	0.183	-0.634
s.e.	0.022	0.028	0.115	0.024	0.035	0.139
$\overline{Y}_{\text{insured}} - \overline{Y}_{\text{uninsured}}$ gap	0.773	0.403	-0.202	1.131	0.357	-1.465
Effect of 1-*sd* $\uparrow$ in crowding						
on insured-uninsured gap:	-0.152	-0.119	0.158	-0.078	-0.047	0.265
as % of gap:	-19.634	-29.452	-78.351	-6.886	-13.041	-18.068

Notes: This table reports the effect of a 1-sd increase in ED crowding on the rate of patient mortality for each of six patient subgroups defined by whether or not the patient has insurance, and whether they have low, middle or high predicted 1-year mortality. These effects are calculated using specification 1.3. The differential effect on crowding for uninsured relative to insured patients, both on the rate of hospital admission, and on the gap in hospital admission by insurance status, is reported.

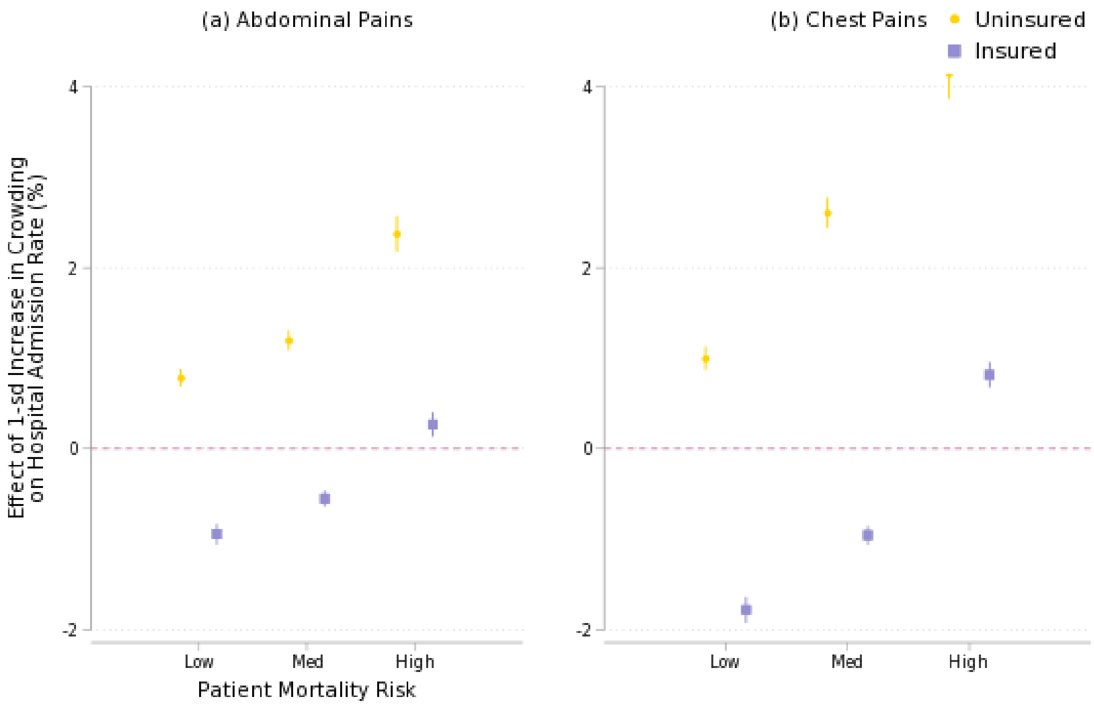
patients - are relatively small in magnitude. A 1-*sd* increase in ED traffic, which corresponds to approximately three additional highly complex patients arriving at the ED in the past two hours, improves 1-year survival among high-risk insured patients by 0.130pp (1.20%), and uninsured patients by 0.288pp (3.22%). These changes shrink the mortality gap between the sickest insured and uninsured abdominal pain patients by 78%. A 1-*sd* increase in ED traffic worsens mortality among the lowest-risk insured patients by 0.016pp, and among uninsured low-risk patients by 0.167pp (56%).

For chest pain patients - where decision-making guidelines are plentiful - the effects of ED traffic on patient survival are significantly larger. A 1-*sd* increase in ED crowding leads to a 0.369pp (3.31%) reduction in the mortality rate among the sickest insured patients, and a 0.634pp (5.81%) reduction among the sickest uninsured patients. These changes correspond to an 18% reduction in the mortality gap between the sickest insured and uninsured patients. Low-risk patients experience worse mortality as a result of increases in ED traffic. A 1-*sd* increase in ED complexity-scaled traffic increases 1-year mortality by 0.201pp (38%)for the least-sick insured patients, and by 0.279pp (62%) for the least-sick uninsured patients.

1.5.2 Hospital Admission is Reallocated Towards High-Risk, Uninsured Patients

I next examine the changes in hospital admission, diagnostic and therapeutic care induced by ED traffic that drive the previously discussed changes in patient mortality. Figure 1.6 plots the effect of a 1-*sd* increase in ED traffic on hospital admission for all six patient subgroups. Table 1.5 reports these effects and calculates the subsequent changes in the hospital admission gap between insured and uninsured patients.

For both abdominal pain and chest pain patients, ED traffic causes hospital admission to be reallocated away from low-risk and towards high-risk insured patients. However, ED traffic unilaterally causes hospital admission to increase for *all* uninsured patients. These effects are much larger for chest pain patients than for abdominal pain patients. A 1-*sd* increase in ED traffic raises the hospital admission rate for high-risk uninsured chest pain patients by 4pp. Given that the overall rate of hospital admission for this patient subgroup



Notes: This figure plots the coefficient estimates of the causal impact of a 1-sd increase in ED complexity-scaled two-hour traffic on the hospital admission rates of abdominal pain patients, in Panel (a), and chest pain patients, in Panel (b). An increase in ED traffic causes large increases in hospital admission for high-risk, uninsured patients. These effects are more pronounced for chest pain patients than for abdominal pain patients.

Figure 1.6: Effect of ED Traffic on Hospital Admission

Table 1.5: Effect of Crowding on Insured-Uninsured Admission Gap

	Abdominal Pains			Chest Pains		
	(1) Low	(2) Med	(3) High	(4) Low	(5) Med	(6) High
Insured	-0.943	-0.551	0.269	-1.781	-0.956	0.817
s.e.	0.058	0.045	0.068	0.073	0.054	0.071
Uninsured	0.784	1.199	2.375	0.999	2.608	4.151
s.e.	0.049	0.056	0.100	0.066	0.085	0.145
$\overline{Y}_{\text{insured}} - \overline{Y}_{\text{uninsured}}$ gap	3.836	5.176	18.475	19.048	25.208	42.190
Effect of 1-*sd* $\uparrow$ in crowding						
on insured-uninsured gap:	-1.728	-1.750	-2.106	-2.780	-3.564	-3.335
as % of gap:	-45.035	-33.813	-11.401	-14.595	-14.140	-7.904

Notes: This table reports the effect of a 1-sd increase in ED crowding on the rate of hospital admission for each of six patient subgroups defined by whether or not the patient has insurance, and whether they have low, middle or high predicted 1-year mortality. These effects are calculated using specification 1.1, using within-hospital, within-hour variation in ED crowding. The differential effect on crowding for uninsured relative to insured patients, both on the rate of hospital admission, and on the gap in hospital admission by insurance status, is reported.

is only 7.36%, a 1-*sd* effectively more than doubles the rate at which the sickest uninsured patients are admitted into the hospital.

The reallocation of hospital admission away from insured and towards uninsured patients reduces hospital admission gaps between these patients significantly, as reported in Table 1.5. A 1-*sd* increase in ED traffic reduces the hospital admission gap between insured and uninsured patients by 8% among high-risk chest pain patients, and by 45% among low-risk abdominal pain patients.

ED traffic causes corresponding changes in the rates of diagnostic and therapeutic care. Increases in hospital admission are accompanied by reductions in the rate of diagnostic testing and increases in the rate of therapeutic treatments. Appendices A.19 and A.20 plot the coefficient estimates for a 1-*sd* increase in ED traffic on the number of diagnostic tests and therapeutic treatments, respectively.

In Appendices A.21, A.22, A.23, and A.24 I decompose the effects of ED traffic on aggregate diagnostic and therapeutic care into its effects on specific diagnostic and therapeutic procedures for abdominal pain and chest pain patients respectively. The effects of ED traffic on the probability of test or treatment provision is highly heterogeneous across procedures.

I highlight two important patterns. First, the changes in diagnostic testing rates are largely driven by EKG usage for chest pain patients, and CT scan usage for abdominal pain patients. Second, changes in therapeutic care are also heterogeneous across procedure type, but an increase ED traffic significantly reallocates specialty inpatient services away from insured patients and towards uninsured patients. A 1-*sd* increase in ED traffic raises the rate of services from the hospital's Coronary Care Unit by 1.263% for high-risk uninsured chest pain patients - doubling the rate at which they receive specialized coronary care. The same increase in ED traffic raises the rate of referral of uninsured abdominal pain patients to Gastrointestinal Care by .373% - also nearly double the rate. These changes reduce the specialty services referral gap between insured and uninsured patients by 28% and 11.6% for chest and abdominal pain patients, respectively.

1.5.3 Are ED Traffic-Induced Reallocations of Care Cost-Effective?

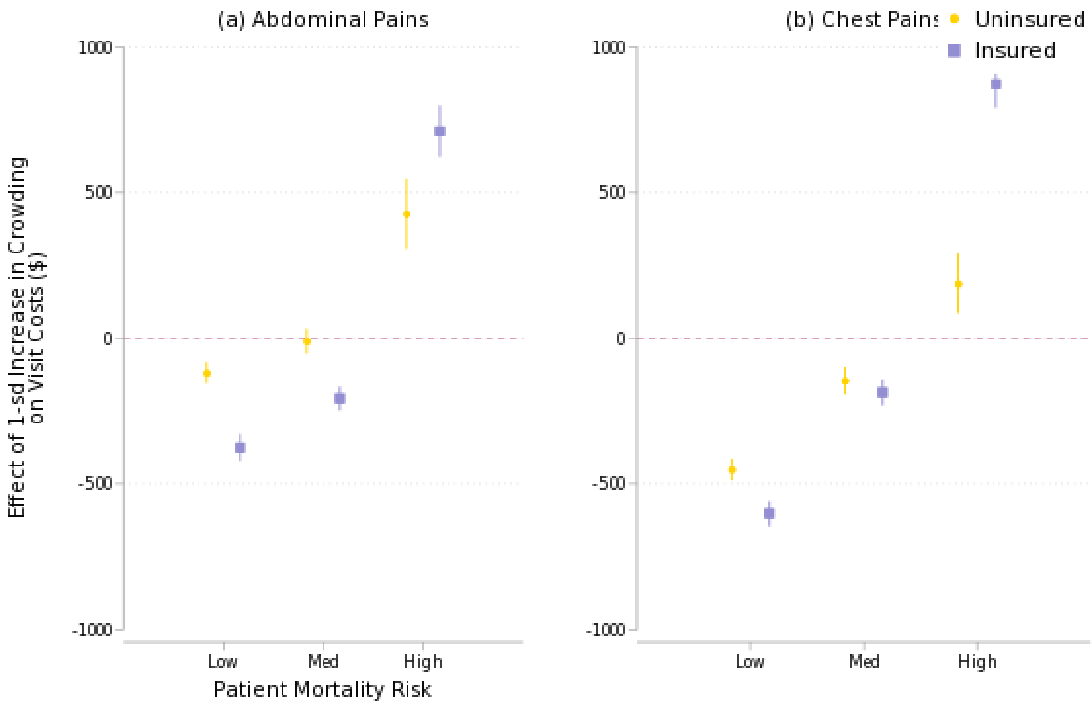
I next turn to the natural question of whether the reallocations of care described above - and the subsequent changes in patient mortality they accompany - are cost-effective. If ED physicians respond to cognitive constraints by indiscriminately providing or removing treatment to all patients, the changes in care described above may not be cost-effective. However, if cognitive constraints actually cause ED doctors to reallocate their attention in ways that improve their ability to identify the patients with the highest expected marginal benefit of treatment, these changes may be efficient.

The effect of a 1-*sd* increase in ED traffic on total visit costs are plotted in Figure 1.7. I calculate the ratio of the change in spending to the change in 1-year patient mortality induced by a 1-*sd* increase in ED traffic. These ratios are reported in Table 1.6 for low- and high-risk insured and uninsured patients, for both chest pain and abdominal pain visits. I benchmark these costs, which can be interpreted as the amount of additional spending required to gain one additional life-year, or the amount of additional healthcare costs saved at the loss of one additional life-year, relative to the $100,000 per Quality-Adjusted Life-Year (QALY) healthcare standard.

I highlight two patterns. For uninsured patients, *both* the increases in spending for high-risk patients and the decreases in spending for low-risk patients are cost-effective for chest pain patients. However, *both* the increases and decreases in spending for high- and low-risk uninsured patients respectively are cost-ineffective for abdominal pain patients. Thus, changes in ED traffic that redirect therapeutic care and hospital admission towards uninsured patients do so in ways that are *cost-effective* when guidelines are available, and *not cost-effective* when guidelines are absent.

1.5.4 ED Traffic Causes Physicians to Reallocate Attention More Effectively when Guidelines are Present

I turn next to the question of *why* changes in ED traffic induce cost-effective reallocations of care when guidelines are present, and cost-ineffective reallocations of care when guidelines



Nots: This figure plots the coefficient estimates of the causal impact of a 1-sd increase in ED complexity-scaled two-hour traffic on the total visit costs for abdominal pain patients, in Panel (a), and chest pain patients, in Panel (b). An increase in ED traffic causes large increases in spending for high-risk patients, and reductions in spending for low-risk patients.

Figure 1.7: Effect of ED Traffic on Total Visit Costs

Table 1.6: Cost-Effectiveness of ED Traffic-Induced Care Distortions

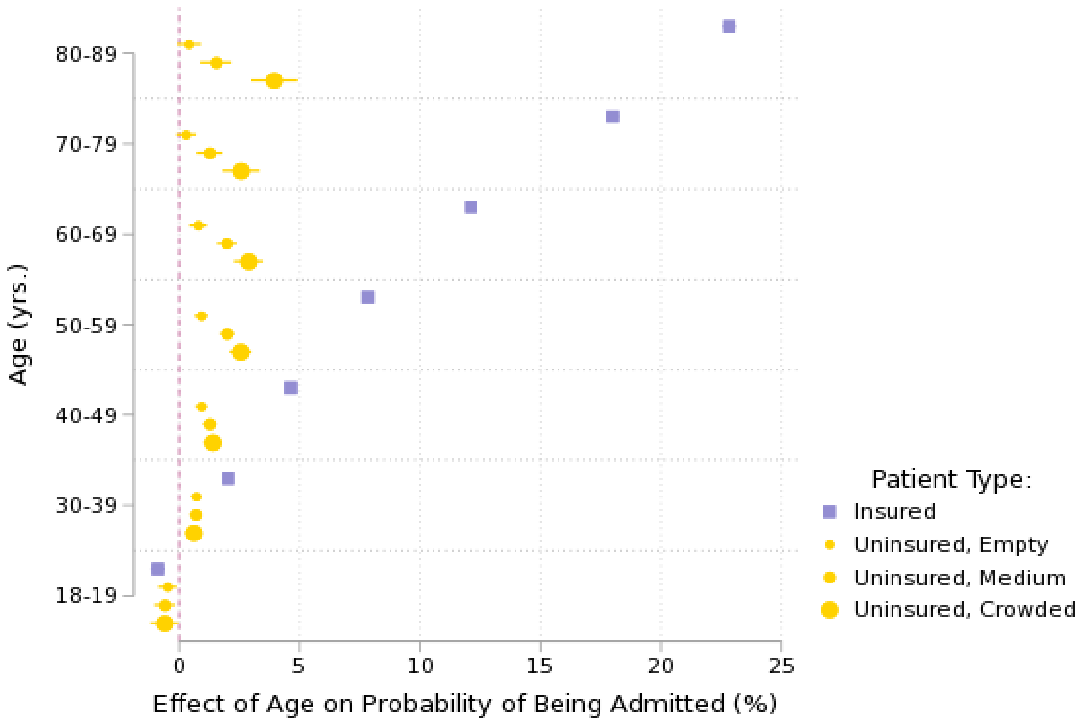
	(1) Abdominal Pains	(2) Chest Pains
Low-Risk, Uninsured:		
$\hat{\beta}_{\$}/\hat{\beta}_{\text{1-Yr Mort.}}$	70,662	161,589
High-Risk, Uninsured:		
$\hat{\beta}_{\$}/\hat{\beta}_{\text{1-Yr Mort.}}$	148,347	29,743
Low-Risk, Insured:		
$\hat{\beta}_{\$}/\hat{\beta}_{\text{1-Yr Mort.}}$	2,383,315	299,700
High-Risk, Insured:		
$\hat{\beta}_{\$}/\hat{\beta}_{\text{1-Yr Mort.}}$	547,702	236,526

Notes: This table reports the ratio of the causal impact of a 1-sd increase in ED traffic on visit costs to the impact on 1-year patient mortality. These ratios are separately for abdominal pain and chest pain patients, across patient subgroups based on mortality risk and insurance status.

are absent. I hypothesize that as ED traffic induces cognitive constraints, doctors rely more heavily on guidelines - when they are present - that reallocate their attention towards patient characteristics that are the most relevant for identifying a patient's expected marginal benefit of treatment. In the absence of such guidelines, doctors respond to cognitive constraints by reallocating their attention in ways that do not improve their prediction of patients' expected marginal benefit of treatment.

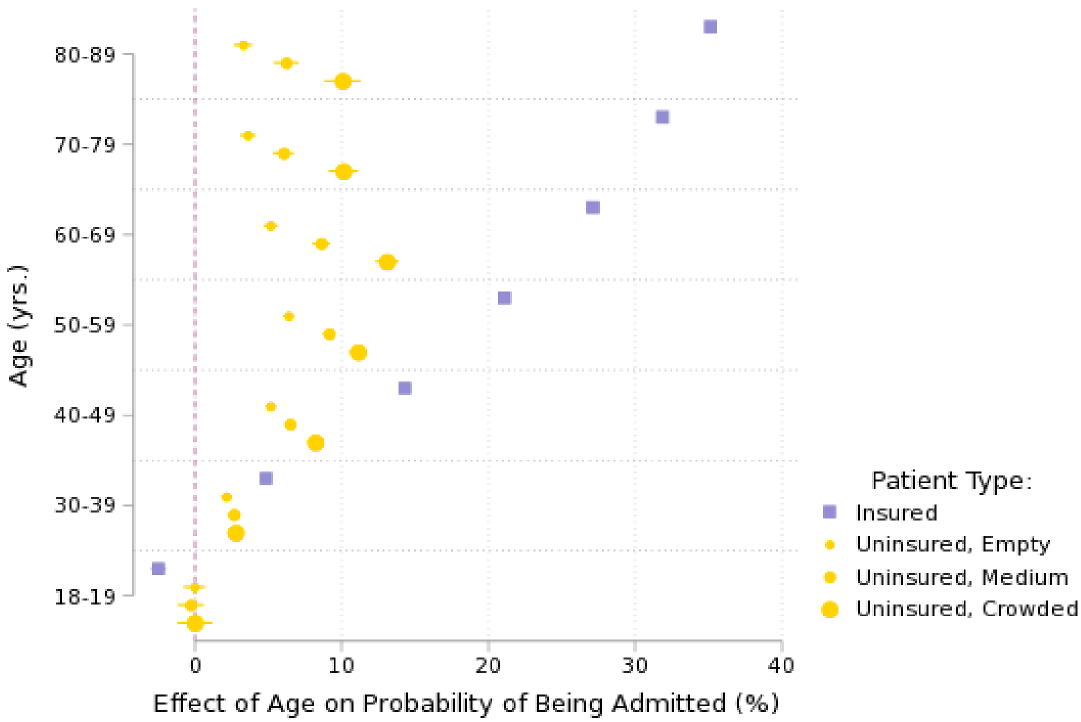
To test this proposed mechanism, I estimate the effect of all health history indicators and patient demographic traits on the patient's likelihood of being admitted into the hospital, following specification 1.2. I estimate this specification separately for insured patients and for uninsured patients conditional on empty, medium and busy levels of ED traffic. I investigate whether the weights on patients' health characteristics and indicators for age categories change for uninsured patients as EDs become more crowded. Figures 1.8 and 1.9 report these coefficient estimates for abdominal pain and chest pain patients respectively.

For uninsured abdominal pain patients, the effect of patient age on the probability of admission increases only marginally as ED traffic increases. In an empty ED, an uninsured



Notes: This figure plots the coefficients of each decadal age bin on an abdominal pain patient's probability of hospital admission. These coefficients are estimated for insured patients overall, and for uninsured patients at low, medium and high levels of ED traffic.

Figure 1.8: Effect of Age on Probability of Admission for Abdominal Pain Patients



Notes: This figure plots the coefficients of each decadal age bin on a chest pain patient's probability of hospital admission. These coefficients are estimated for insured patients overall, and for uninsured patients at low, medium and high levels of ED traffic.

Figure 1.9: Effect of Age on Probability of Admission for Chest Pain Patients

60-69 year old abdominal pain patient is 1% more likely than a 20-29 year old to be admitted into the hospital. When the ED is crowded, this premium rises to just 3%. For uninsured chest pain patients, as the ED becomes more crowded, the effect of patient age on the probability of hospital admission increases significantly. For example, in an empty ED, 60-69 year-old uninsured patients are 5% more likely than 20-29 year olds to be admitted into the hospital. In a crowded ED, these patients are 13% more likely to be admitted. These large shifts in the apparent weights that physicians place on health characteristics like age suggest that, when decision-making guidelines are present, when the ED becomes more crowded, doctors evaluate uninsured patients more similarly to how they evaluate insured patients.

1.5.5 ED Traffic Causes Physicians to Behave More Similarly when Guidelines are Present

I conduct a second test of whether increases in ED traffic cause doctors to rely on guidelines - when they are available - by testing whether or not ED traffic reduces interobserver variability when guidelines are present. A key goal of clinical decision-making guidelines is to reduction variation stemming from similar patients being treated differently by different providers[13]

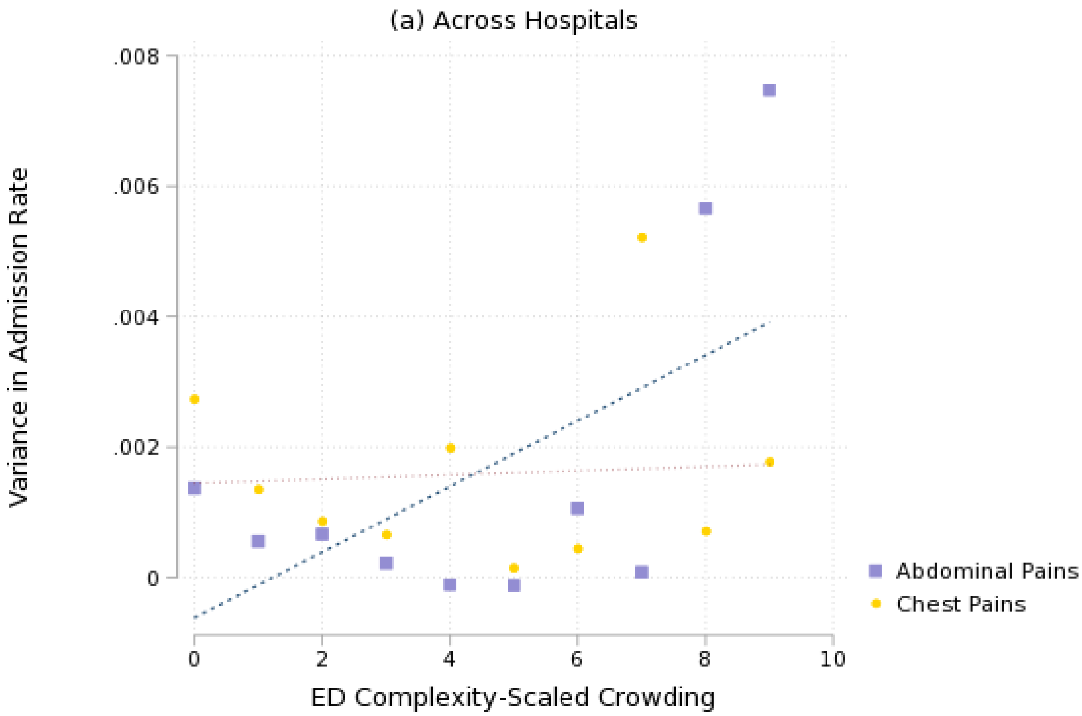
I decompose the variance of hospital admission rates across three dimensions to test this mechanism. I regress a binary indicator variable for whether or not the index patient was admitted into the hospital on facility-by-year fixed effects and hour-of-day fixed effects:

$$y_{ifc,h|c(i)=\text{chest pain}} = \alpha_{\text{clock}(h)} + \delta_{f,y(h)} + \epsilon_{ifc,h} \tag{1.4}$$

I decompose the variance of the residuals from this specification into within- and across-hospital variation, for various levels of ED traffic. Figures 1.10 and 1.11 plot the across-hospital and within-hospital variation in admission rates respectively, for both chest

[13] The first and perhaps most famous clinical decision-making guideline - the APGAR score - aimed to improve and standardize how physicians assessed the health of newborn babies. Originally, the APGAR score was meant to be computed by two physicians and then averaged together. The score was so effective at assisting different physicians in arriving at the same assessment that it is now performed by just one physician (Apgar, 1953).

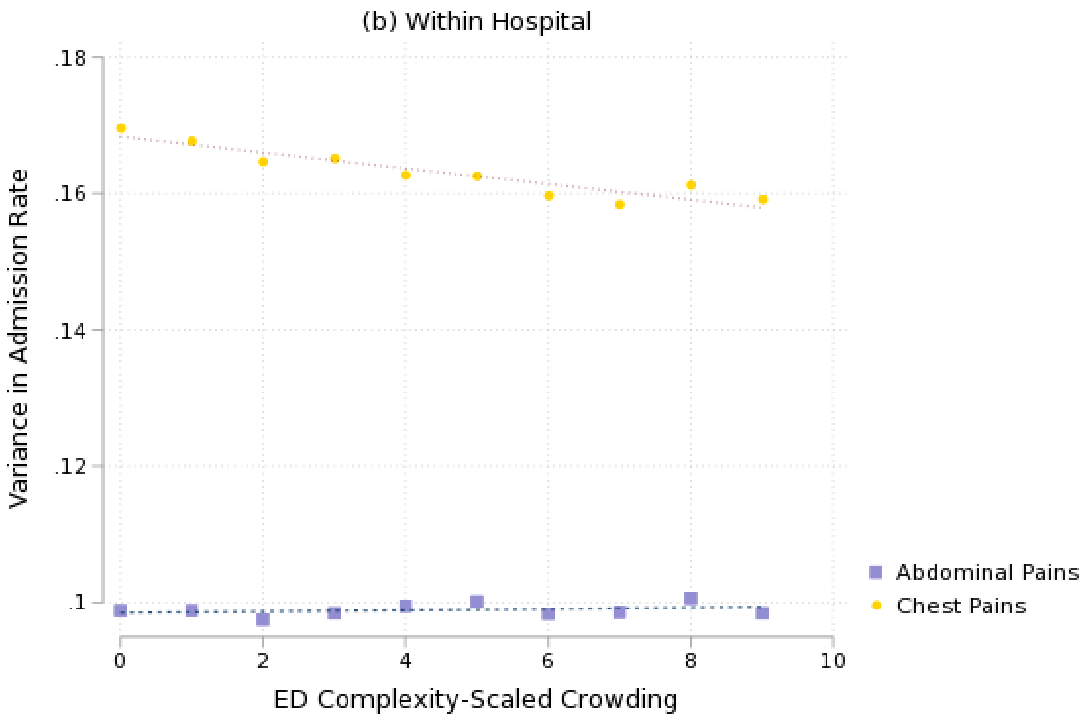
pain and abdominal pain patients. Figure 1.10 shows that across-hospital variation in



Notes: This figure shows the relationship between the level of complexity-scaled, two-hour ED patient traffic, and the amount of variation in admission - conditional on facility-by-year fixed-effects and hour-of-day fixed-effets - driven by differences across hospitals, for abdominal pain and chest pain patients separately.

Figure 1.10: Effect of ED Traffic on Variation in Admission Across Hospitals

hospital admission significantly increases with ED traffic for abdominal pain patients, and stays approximately level for chest pain patients. Figure 1.11 shows that within-hospital variation in hospital admission *decreases* with ED traffic for chest pain patients, but slightly *increases* with ED traffic for abdominal pain patients. Taken together, these changes in variance are consistent with a story in which ED traffic causes physician choices to converge when guidelines are present, and diverge in the absence of a guideline, perhaps because different physicians use their own rules of thumb or heuristics in the absence of a centralized guideline.



Notes: This figure shows the relationship between the level of complexity-scaled, two-hour ED patient traffic, and the amount of variation in admission - conditional on facility-by-year fixed-effects and hour-of-day fixed-effets - driven by differences within hospitals, for abdominal pain and chest pain patients separately.

Figure 1.11: Effect of ED Traffic on Variation in Admission Within Hospitals

1.6 Discussion of Alternative Mechanisms

I consider four alternative explanations that may explain my results. I show empirical evidence from my own analysis, as well as discuss evidence from the relevant literatures, that jointly suggest that my results are not driven by changes in either ED patient triaging and subsequent wait times, changes in ED staffing, decision-making economies of scale, or binding physical capacity constraints rather than cognitive constraints.

1.6.1 Verifying Robustness via Coefficient Stability

In addition to verifying that patients appear to be of similar ex-ante health, race, insurance status, and overall risk across different levels of ED traffic after facility-by-year and hour-of-day variation are controlled for, as shown in Figure 1.3 and Appendices A.15, A.16 and A.18, I verify that my coefficient estimates are relatively stable when additional patient health and demographic controls are added to my regression specification.[14]

Table 1.7 reports the coefficient estimates for the effect of ED traffic on hospital admission for all six patient subgroups. I first report the coefficient estimates from a baseline regression with no controls. I then add hospital-by-year fixed effects and hour-of-day fixed effects. I show that, after the inclusion of these controls, the addition of controls for patient gender, age, and health history do not meaningfully change the coefficient estimates, and - if anything - tend to move coefficient estimates away from zero.

1.6.2 Changes in Triage and Subsequent Wait Times

It is possible that as EDs get more crowded, triage nurses change the ways in which they assess and assign priority arriving patients. I discuss three ways in which this explanation is both unlikely to occur in my setting, and unlikely to explain the patterns of treatment choice and patient survival that I observe.

[14] In accordance with (Oster, 2019), I verify that these additional controls have reasonable predictive power for hospital admission itself: the R^2 increases from 0.241 to 0.303 with the introduction of gender, age and patient health history controls.

Table 1.7: Testing for Coefficient Stability

	Baseline	After Additionally Controlling For:			
	(1) None	(2) Hosp & Hour FEs	(3) Triage	(4) Health History	(5) Clustered SEs
Crowding X: Uninsured, Low Risk	0.374*** (0.0537)	0.874*** (0.0568)	0.816*** (0.0547)	0.870*** (0.0545)	0.870*** (0.0423)
Uninsured, Mid Risk	1.139*** (0.0586)	1.803*** (0.0608)	1.837*** (0.0585)	1.863*** (0.0584)	1.863*** (0.0504)
Uninsured, High Risk	2.524*** (0.114)	3.015*** (0.110)	3.473*** (0.106)	3.269*** (0.106)	3.269*** (0.0879)
Insured, Low Risk	-0.436*** (0.0537)	-1.546*** (0.0546)	-1.521*** (0.0525)	-1.454*** (0.0524)	-1.454*** (0.0478)
Insured, Mid Risk	0.399*** (0.0304)	-0.692*** (0.0362)	-0.809*** (0.0349)	-0.779*** (0.0348)	-0.779*** (0.0360)
Insured, High Risk	1.108*** (0.0406)	0.500*** (0.0439)	0.654*** (0.0423)	0.554*** (0.0422)	0.554*** (0.0503)
R-Squared	0.107	0.241	0.298	0.303	0.303
# Obs	3986052	3986050	3986050	3986050	3986050

Notes: This table reports estimates of the impact of a 1-sd increase in ED traffic on each of the six patient subgroups defined by mortality risk and insurance status. In column (1), I include no controls. In columns (2) through (5), I add controls for facility-specific yearly time trends and hour-of-day variation, followed by triage controls (age, chief complaint, gender), and a host of ex-ante patient health characteristics. After the inclusion of facility-specific yearly time trends and hour-of-day variation, the coefficient estimates are relatively stable to the introduction of additional controls.

ED patients are triaged based on a small set of patient characteristics: chief complaint, age, gender, and vital signs (Gilboy *et al.*, 2011). A typical chest pain or abdominal pain patient would likely receive an Emergency Severity Index triage level of 3 or 4 (out of 5). Racial bias in ESI assignment levels has also been documented (Vigil *et al.*, 2015). My regression specifications account for differences in triage level due to chief complaint, age, gender, and race by controlling for these patient characteristics flexibly and directly. The only aspect of a patient's health that might affect their triage level that I do not observe in my data is the patient's vital signs (e.g. heart rate or blood pressure.) With data on patient ESI levels, chief complaint, gender, age, race and vital signs, it is possible to estimate and bound the effects of missing vital signs on myestimation of the patient's ESI level.

The literature on the effects of ED nurse triage also suggests that changes in triage are unlikely to be driving my results. First, patients receiving high ESI scores will always receive first priority, whether they arrive in a crowded or empty ED. Changes in triage due to crowding should only affect patients with low triage ratings, which is inconsistent with my findings.

Second, assigning patients to higher triage levels leads to lower hospital admission rates, and undertriaging patients leads to higher admission rates (Hinson *et al.*, 2018). In order for these changes to be driving my results, crowding would need to induce nurses to specifically *undertriage* the highest-risk patients and *overtriage* the lowest-risk patients. Empirical evidence on the relationship between ED traffic and triage patterns finds that traffic unliterally causes an increase in triage levels for all patients, and that these changes arise at levels of crowding twice as large as the levels of crowding to which my analysis is limited (Chen *et al.*, 2019). Lastly, triage decisions largely affect short-term patient survival (Grossmann *et al.*, 2012), while the magnitude of the effects of ED traffic on patient survival up to 1-year post-visit which I observe are not explained by the magnitude of changes in short-term patient deaths.

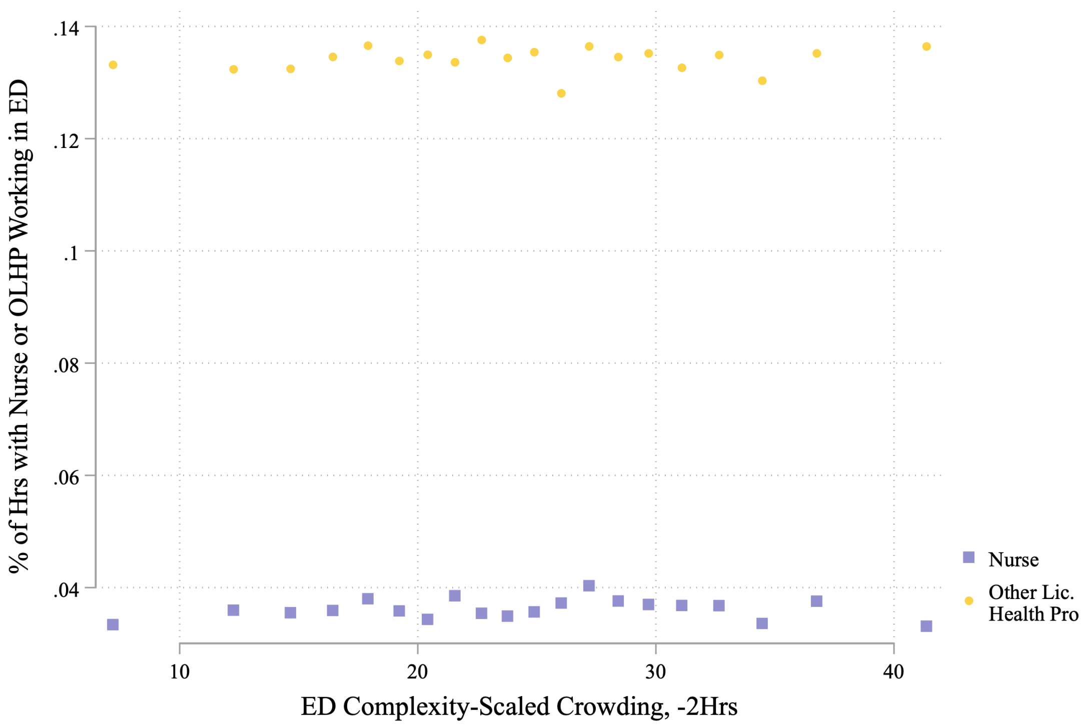
1.6.3 Changes in Staffing

An alternative explanation for the effects I find is that fluctuations in ED patient traffic cause changes in not only the number, but the composition of healthcare professionals serving patients in the ED. For example, as patient traffic increases, hospitals may bring in more nurses or physician's assistants to assist with treating patients. I show that changes in the composition of ED staff are unlikely to be driving my results in two ways.

First, I use provider ID codes to assess the frequency with which two types of non-physician healthcare professionals are present in the ED: nurses, and physician's assistants (coded in the data as "Other Licensed Healthcare Professionals"). Appendix A.2 details how these professional designations are coded. Since nurses and PAs work with the entire team of doctors - and thus all patients seen in the ED in any hour - they are less susceptible to the selectively omitted doctors problem discussed in Appendix A.3 and their observed working patterns are less likely to be mechanically related to random variation in the sickness of patients appearing in the ED. In Figure 1.12, I show that the proportion of hours in which either a nurse or a PA is reported to be working in the ED does not vary with respect to two-hour, complexity-scaled crowding. At all levels of crowding, nurses and PAs are present 14% and 3% of the time, respectively.

Second, I run an alternative regression specification that removes possible heterogeneity in hospitals' staffing patterns across the hours of the day. If some hospitals are better able to anticipate hours of high or low traffic, and adjust their ED staff accordingly, facility fixed effects and hour-of-day fixed effects alone would not capture this variation in staffing patterns. In Appendices A.25 and A.26, I add facility-by-hour-of-day fixed effects and show that these do not meaningfully alter the pattern of results I observe.

Third, I note that my measure of ED traffic is based on two-hour patient flows, conditional on facility, year, and hour-of-day. 80% of the variation in ED traffic is driven by changes in patient complexity, and the remaining 20% is driven by changes in patient volume. Given that most ED workers adhere to shift-like schedules, it is unlikely that a specific ED during an unexpectedly busy hour of the day will be able to change its ED staffing



Notes: This figure shows the relationship between the level of complexity-scaled, two-hour ED patient traffic, and the rate at which either a nurse or a physician's assistant is present in the ED in the index hour, conditional on facility-by-year and time-of-day fixed-effects. Nurses are present and assisting patients in about 13% of all hours, and PAs are present in about 3% of all hours. These rates do not appear to vary with respect to ED traffic.

Figure 1.12: Prevalence of Nurses and PAs by Levels of ED Traffic

meaningfully in response to unexpected fluctuations in complexity-scaled traffic - driven largely by fluctuations in patient complexity - within a two-hour time frame.

1.6.4 Economies of Scale

I consider whether crowding might induce choice-specific economies of scale that distort the allocation of admission, diagnostic and therapeutic care. For example, if several chest pain patients are being seen at once, rather than separately, doctors may make different treatment choices both due to the doctor's ability to compare similar patients to each other, the possibility that doctors "narrow-bracket" their treatment choices[15], and because a single test will be informative not just for the index patient, but for other patients with similar concerns.

I directly test this theory by adding controls for the specific portion of complexity-scaled ED crowding comprised of patients with the same concern - either chest pains or abdominal pains - as the index patient. I interact this "economies of scale" control with my six-cell patient risk- and insurance-type indicator and report the results of this alternate specification in Table 1.8. The introduction of controls for "similar" patient crowding does not appear to change the broad patterns of reallocation of care by risk and by patient insurance status. Curiously, economies of scale appear to reinforce the effects of crowding for chest pain patients, but counteract the effects of crowding for abdominal pain patients. I investigate the possibility of economies or diseconomies of scale for high-stakes ED treatment and testing choices in a separate project (Shanmugam, 2019).

1.6.5 Cognitive Constraints versus Physical Capacity Constraints

Substantial increases in ED traffic trigger facility-level physical resource constraints - such as space in the ED waiting room, hallway space for patients waiting to be admitted, ED

[15]For example, if a chest pain patient and a limb pain patient are seen at the same time, the doctor may decide to admit them both. If two chest pain patients arrive at the same time, a doctor may be less willing to admit two patients with the same concern.

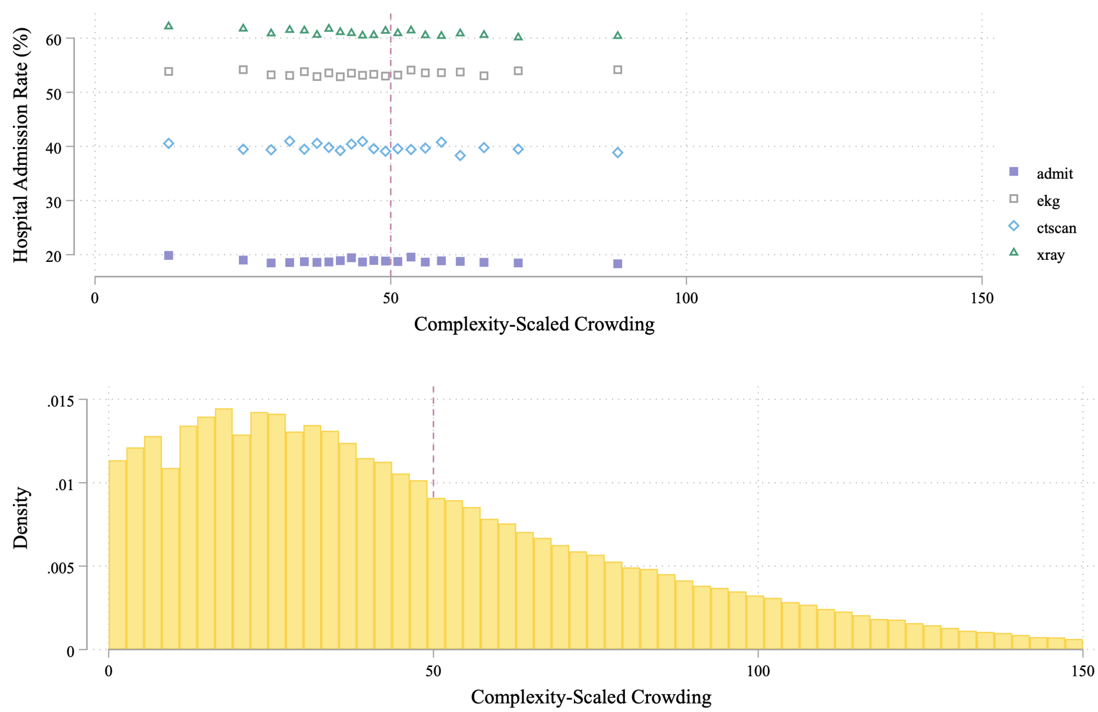
Table 1.8: Testing for Economics of Scale

	Abdominal Pains		Chest Pains	
	(1) Admitted	(2) 360d Mortality	(3) Admitted	(4) 360d Mortality
Core Effects:				
Crowding X Uninsured, Low	0.853*** (0.0511)	0.182*** (0.0228)	0.704*** (0.0693)	0.285*** (0.0242)
Uninsured, Mid Risk	1.268*** (0.0580)	0.150*** (0.0288)	2.223*** (0.0900)	0.178*** (0.0362)
Uninsured, High Risk	2.420*** (0.107)	-0.288* (0.121)	4.001*** (0.155)	-0.648*** (0.147)
Insured, Low Risk	-1.016*** (0.0607)	0.0145 (0.0206)	-1.771*** (0.0765)	0.189*** (0.0220)
Insured, Mid Risk	-0.597*** (0.0462)	0.0211 (0.0204)	-0.838*** (0.0559)	0.137*** (0.0222)
Insured, High Risk	0.233** (0.0717)	-0.108* (0.0505)	0.956*** (0.0749)	-0.388*** (0.0516)
Economies of Scale:				
Uninsured, Low Risk	-0.0891*** (0.0165)	-0.0172** (0.00560)	0.396*** (0.0249)	-0.00884 (0.00603)
Uninsured, Mid Risk	-0.0880*** (0.0199)	-0.00861 (0.00934)	0.495*** (0.0354)	0.00667 (0.0130)
Uninsured, High Risk	-0.0554 (0.0446)	0.00117 (0.0497)	0.210** (0.0719)	0.0180 (0.0648)
Insured, Low Risk	0.0859*** (0.0208)	0.00228 (0.00529)	-0.0138 (0.0247)	0.0131* (0.00574)
Insured, Mid Risk	0.0532*** (0.0140)	0.00330 (0.00499)	-0.136*** (0.0165)	-0.000370 (0.00551)
Insured, High Risk	0.0409 (0.0251)	-0.0256 (0.0182)	-0.159*** (0.0258)	0.0215 (0.0184)
Observations	1995938	1995938	1990105	1990105

Notes: This figure reports the results of an alternative specification test in which I directly control for the amount of complexity-scaled crowding driven by patients with the same chief complaint as the index patient, in addition to my overall measure of complexity-scaled crowding. I show that, even after controlling for these possible "economies of scale", the overall pattern of the effects of ED traffic on doctor decisions and patient outcomes remains unchanged.

beds and observation units, and inpatient beds - to become binding. I discuss two reasons why binding physical capacity constraints are unlikely to be driving my results.

First, when EDs become full, hospital admission rates tend to increase overall. I limit my analysis to levels of crowding at which an overall increase in the hospital admission is not observed. Figure 1.13 describes how the overall rate of hospital admission strongly decreases as ED traffic increases, and flattens once complexity-scaled two-hour patient traffic reaches the level associated with 80 units of ED service, or just over 25 patients arriving in the ED. I limit my analysis to ED crowding of less than 50 units of ED service. The variation



Notes: This figure shows the relationship between complexity-scaled two-hour crowding and the hospital admission rate. When the ED becomes physically full, doctors increase the rate of hospital admission. I limit my analysis to levels of crowding that are too low to trigger an increase in admission, as seen in Panel (a).

Figure 1.13: Overall Admission Rate by ED Traffic

in patient traffic driving my analysis ranges from zero patients arriving in the ED in the past two hours, to approximately 14 patients arriving in the ED over the same timeframe.

These levels of traffic are unlikely to be large enough to create binding physical capacity constraints. Figure 1.13 shows that hospital admission rate increases are not triggered by these lesser levels of ED patient traffic.

Second, binding physical capacity constraints should weakly *decrease* the amount of care received by all patients. However, my results show that ED traffic is just as likely to *increase* the provision of treatment for high-risk and uninsured patient subgroups. Heterogeneity in the effects of crowding by patient subgroup are inconsistent with facility-level physical capacity limits.

1.7 Conclusion

Experts often make high-stakes decisions under significant cognitive constraints. In this paper, I estimate the causal impact of those cognitive constraints on the quality and equity of these important decisions. I leverage random variation in hourly ED traffic flows to estimate the effect of increases in physicians' cognitive load on the amount of diagnostic and therapeutic care they provide, the types of final diagnoses they arrive at, and the subsequent impacts on patient survival. I further investigate heterogeneity in these effects by comparing chest pains and abdominal pains, two common ED complaints which differ significantly in the availability of decision-making aids.

I show that the effect of cognitive constraints hinges critically on the presence of decision-making guidelines. When decision-making aids are both present and absent, increases in ED traffic - and the cognitive constraints they induce - cause doctors to reallocate care towards high-risk and away from low-risk patients, and broadly toward all uninsured patients. These reallocations significantly reduce the disparities between insured and uninsured patients in treatment and survival. When guidelines are present, these reallocations are highly cost-effective, but when guidelines are absent, these reallocations are not cost-effective. Furthermore, ED traffic has differing impacts on both within- and across-hospital variation in treatment depending on the presence of a guideline: for chest pain patients, an increase in ED traffic significantly reduces within- and across-hospital variation, relative to when guidelines are absent. Lastly, I show that when guidelines are present, doctors' evaluations of uninsured patients converge to their evaluations of insured patients, suggesting that cognitive constraints and decision-making aids combined directly reallocate physicians' attention and result in more equitable decision-making.

I show that cognitive constraints, contrary to classical theory on bounded rationality, can improve both the quality and equity of high-stakes decision-making in the ED, and that these effects hinge critically on the presence of decision-making aids. These results suggest that optimal clinical decision-making involves a combination of decision-making aids and doctor discretion, rather than either on its own.

Chapter 2

Decision Fatigue and Not-So-Grumpy Judges

2.1 Introduction

On average, nearly a thousand cases are heard each day across thirteen courtrooms at the Bronx Housing Court.[1] High-stakes decisions are regularly made in staggering sequence by human decision-makers in courtrooms, emergency rooms, and even by law enforcement. In these settings, “choice fatigue" may have serious consequences (Vohs *et al.*, 2008).

A highly publicized 2011 study found that Israeli parole judges quickly became fatigued when deciding cases: the rate of parole dropped from 65% at the beginning of a day of hearings to nearly 0% by the end of the day (Danziger *et al.*, 2011). I revisit the tale of the grumpy parole judges using administrative data on every hearing held at the New Hampshire Parole Board from 2009 to 2011. I exploit the alphabetical ordering of parole hearings to estimate the causal effect of decision fatigue on these important choices.

In striking contrast with earlier findings, I find that having a parole hearing occur one time-slot later in the day *increases* the likelihood that an incarcerated person (IP) receives

[1] Chief Administrator of the Courts (2017).

parole by 2 percentage points.[2] An individual seen at the end of their group is 10 percentage points more likely to receive parole than an individual seen at the beginning of their group.

Second, I find that judges appear to implement a simple, reasonable rule of thumb when fatigued: they grant parole to all prisoners, except those serving life sentences. If an incarcerated individual is not serving a life sentence, they become more likely to receive parole as the day wears on; if they are serving a life sentence, they become significantly *less* likely to receive parole as judges become fatigued.

Lastly, using detailed data on past and future incarceration events, I show that fatigued decision-making is significantly more cost-effective than decision-making at full bandwidth: paroled individuals are unlikely to re-offend, and cost burdens on the criminal justice system are significantly reduced. Accounting for the societal and criminal justice system costs of future criminal activity and the reduction in costs of actively incarcerating IPs, and using conservative estimation methods, I find that for each time-slot later in the day a hearing is held, approximately $400 in net costs are saved over the next 15 months.

My findings contribute to three literatures. First, I isolate the causal impact of decision fatigue in a high-stakes setting, and estimate its social cost. I build on work from psychology and behavioral economics that demonstrates the existence of decision fatigue and cognitive limitations, both theoretically and empirically in consumer settings (Iyengar and Lepper, 2000b,a; Tversky and Kahneman, 1974). Second, I find evidence for the usage of simple heuristics in response to these cognitive limitations in high stakes settings. Similar patterns have been highlighted in survey and lab settings (Dhar, 1997; Besedeš *et al.*, 2012). Lastly, I highlight the importance of decision fatigue in criminal justice decision-making, building on work that investigates the existence and causes of type I and II errors (Clair and Winter, 2016; Gottfredson and Gottfredson) and racial bias (Albonetti *et al.*, 1989; Bynum and Paternoster, 1984; Yang, 2015a,b) in the criminal justice system. In particular, my findings shed light on possible mechanisms driving the relationship between judicial system capacity and

[2] I refer to the subjects of these parole hearings as Incarcerated Persons (IPs) throughout this paper, in keeping with recommendations by The Marshall Project to use terminology that separates the individual from their incarcerated status, and does not assume guilt or criminality (Keller, 2015)

decision-making (Huang, 2011; Yang *et al.*, 2016).

The rest of this chapter is organized as follows. In Section 2, I describe the proceedings of the New Hampshire Adult Parole Board. In Section 3, I describe my data sources and various constructs used throughout the analysis. In Section 4, I describe my empirical methodology, and in Section 5 I describe my results. Section 6 discusses the implications of these results. Section 7 concludes.

2.2 The New Hampshire Adult Parole Board

The New Hampshire Adult Parole Board conducts approximately 1,000 hearings annually. Hearings are held on Tuesdays, beginning at 9:00am and proceeding until all hearings are concluded. Each hearing is conducted in ten minutes, by a panel of three parole board members, who are appointed on a volunteer basis for five-year terms. These members may be former law enforcement officers, academics, or community members. Board members are given information on the slate of hearings they will oversee a week before the hearings are held. On average, thirty hearings are held per day.

Hearings are ordered first by the facility in which the IP is incarcerated, and then alphabetically by the IP's last name.[3] At the hearing, the parole board reviews information on the IP's behavior while incarcerated, program participation, training or treatment programs undertaken while incarcerated, proceedings from any previous hearings, prior criminal record. IPs and their case managers propose employment and housing plans if paroled. Victims are notified if an IP is eligible for parole, and may attend and/or deliver comments to the parole board at the hearing (NHDOC).

New Hampshire is an "indeterminate sentencing" state, which means that each IP is given a minimum ("released no earlier than") and maximum ("incarcerated not later than") sentencing date. Initial parole hearings occur within the two month prior to an IP's

[3]Hearings are ordered first by facility to minimize the costs associated with transporting IPs from each facility to the parole board and back, and to accommodate different levels of security and availabilty for IPs housed at different facilities.

minimum sentence date.

Several non-standard parole hearing categories are included in my sample. A "mandatory" hearing is a hearing held due to the requirements of New Hampshire Senate Bill 500, which was enacted in 2010. SB500 increased the early release of nonviolent IPs by mandating parole after the IP had served 120% of their minimum sentence. A "review" hearing is held if an inmate who was previously granted parole commits an infraction before they can be released, and their parole decision must now be reviewed. A "reconsideration" hearing is held if an IP is initially denied parole, and applies for parole again. A "medical" hearing involves a request for parole due to an IP's medical condition.

If an IP is denied parole, the board must specify reasons for the denial. The board may opt to "continue" the hearing if more information (e.g. the results of a pending drug test) is needed.[4] If an IP is granted parole, the parole board must approve the IP's plan for housing and employment. Twelve standard rules of conduct apply to all parolees, and the board may impose any additional conditions, ranging from geographic restrictions, contact restrictions, extra supervision including random job or home visits, curfews, or mandatory completion community-based treatment programs.

2.3 Data

I combine data from two sources for this analysis. The first is a hand-collected dataset that contains the parole hearing schedules and final decisions for all hearings held at the New Hampshire Adult Parole Board between January 2009 and August 2011. These data were retrieved from paper records stored at the NH Parole Board. The data consists of 3,302 parole hearings conducted over 114 hearing days.[5]

The second data source is administrative data on the incarceration histories of the IPs

[4] In my sample, fewer than 5% of hearings end in a continuation. I exclude these hearings from my analysis.

[5] Recent and upcoming hearing schedules are now posted on the New Hampshire Department of Corrections website. The alphabetical hearing ordering convention can easily be verified by perusing these schedules, available at `https://www.nh.gov/nhdoc/divisions/parole/hearings.html`.

in my sample. These data were obtained from the Office of Research and Planning at the NH Department of Corrections, and they cover every movement an IP experiences within the prison system. These data include movements into, within, and out of incarceration. Each movement includes a date, reason for movement, initial location, and destination. Movements into incarceration are due to criminal activity or parole/probation violations and are reported using Major Offense codes. Movements within incarceration represent the relocation of IPs due to medical needs, security needs, or facility capacity constraints. Movements out of incarceration are due to parole, probation, escape, death, or the completion of an IP's sentence.

2.4 Hearing Characteristics

The parole hearing schedule data indicates whether a hearing is for medical, review, or reconsideration reasons. I create a binary indicator for each of these hearing types. The schedules also indicate if victims or victims' representatives will be present. I create a binary indicator for the presence of any nonzero number of victims or victims' representatives. These hearing characteristics are summarized in Table 2.1.

The schedule lists the date and time of day at which the hearing is held, and the facility in which the IP is incarcerated. Using the combination of these two variables, I create a measure for the ordinal position of each hearing within its day and facility group. The ordinal position variable takes a value of k if the IP's hearing is the kth hearing from his facility that day.

2.5 Incarcerated Person Characteristics

The administrative data include a binary indicator for the IP's gender, a categorical variable for the IP's race, and the IP's age at the time of their hearing, in years. I classify the IP's county of citizenship based on the county court from which their first incarceration event occurs, and use this county variable to control for additional prisoner characteristics that may be correlated with their county of residence prior to their first incarceration.

The facility an IP is housed in is determined by (and therefore captures) the IP's custody level, which is a classification based on the severity of the crime committed, security restrictions, and history of behavior while incarcerated. Appendix B.1 describes how custody levels are determined. There are five custody levels, ranging from minimum to maximum security, as well as a separate custody designation for inmates undergoing medical care.

I characterize incarceration events prior to the IP's hearing as follows. Reasons for incarceration are reported using detailed Major Offense codes. For example, there are twenty Major Offense codes for sexual assault that distinguish whether the victim is a man, woman, or child; whether the victim is elderly or disabled; and whether a weapon, drugs or the threat of force is used. I aggregate these codes into four offense categories: violent, drug, public order, and property offenses. For each IP, I calculate the number of prior offenses in each category, as well as the number of prior parole violations.

I use the same convention to characterize recidivism after an IP is paroled. I calculate the number of parole violations and violent, drug, public order and property offenses an IP is charged with in the 15 months following their index hearing. These IP characteristics are summarized in Table 2.1.

2.6 Empirical Methodology

2.6.1 Estimating Equation

I estimate the causal impact of decision fatigue on the choices made by the parole board using the following specification.

$$y_{if,t} = \beta(\text{ord. pos.}_{if,t}) + \gamma X_{i,t} + \delta H_{if,t} + \eta_f + \epsilon_{if,t} \tag{2.1}$$

$y_{if,t}$ is a binary variable that represents whether or not parole is granted to IP i housed at facility f whose hearing is heard on day t. β represents the effect of IP i's hearing's ordinal position on the hearing outcome. $X_{i,t}$ is a vector of IP characteristics, including

Table 2.1: Summary Statistics

	μ	σ
Hearing Characteristics:		
Mandatory	.0199	(0.140)
Review	.0115	(0.107)
Reconsideration	.0103	(0.101)
Medical	.00374	(0.0610)
Victim Present	.0501	(0.218)
Received Parole	.813	(0.390)
Incarcerated Person Demographics:		
Male	.897	(0.305)
White	.846	(0.361)
Hispanic	.0545	(0.227)
Black	.0629	(0.243)
Asian	.0028	(0.0529)
Other	.0343	(0.182)
Age at Hearing (yrs.)	35.3	(11.13)
Age at 1st Crime (yrs.)	30.6	(12.03)
Incarceration History:		
# Parole Violations	.343	(0.789)
# Violent Offenses	.428	(0.596)
# Property Offenses	.383	(0.632)
# Drug Offenses	.246	(0.506)
# Public Order Offenses	.256	(0.508)
Serving Life Sentence	.0607	(0.239)
Time Served (yrs.)	3.41	(3.640)
# Violent Offenses	.00161	(0.0400)
# Property Offenses	.00161	(0.0400)
# Drug Offenses	0	(0)
# Public Order Offenses	.000401	(0.0200)
Observations	3211	

Notes: This table reports the average characteristics of the hearings, incarcerated individuals, incarceration histories and recidivism for each hearing in the sample. One observation is one parole hearing. Recidivism characteristics are calculated for the 81.3% of parole hearings that resulted in parole being granted.

time-invariant demographics, and time-varying measures of prior offenses. $H_{if,t}$ is a vector of hearing characteristics, including whether a victim is present, and whether the hearing is a review, reconsideration, mandatory or medical hearing. η_f is a vector of prison facility fixed-effects, which capture differences in IP custody levels. $\epsilon_{if,t}$ represents the error term. Errors are unclustered, as the sample constitutes a census of all parole hearings occurring at the NH Adult Parole Board between 2008 and 2011, and ordinal position varies at the hearing level (Abadie *et al.*, 2017).

2.6.2 Identifying Assumption

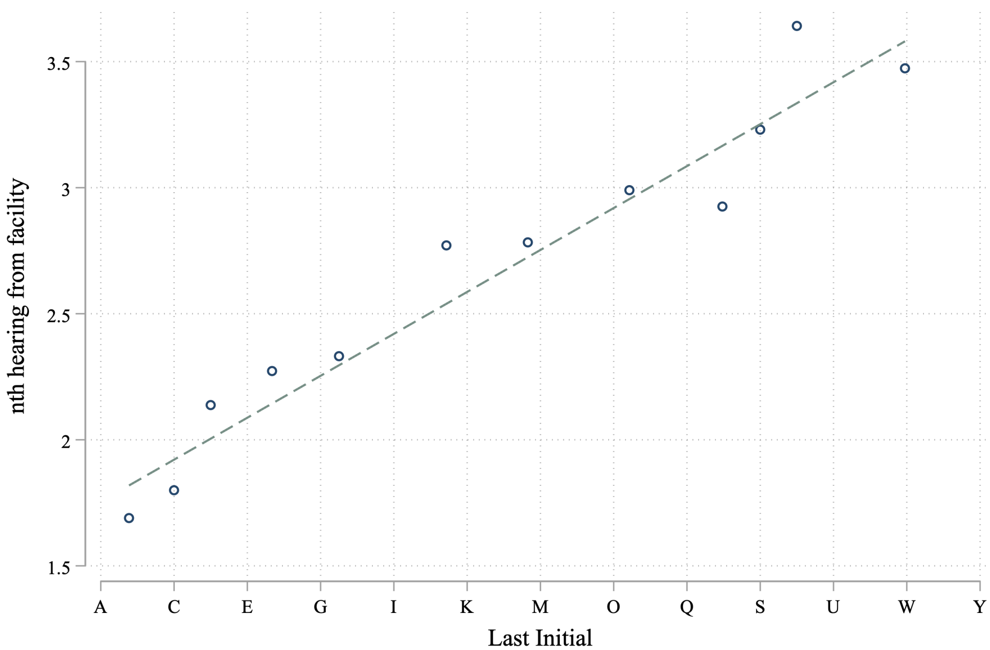
In order to interpret β as the causal effect of a hearing's ordinal position on the hearing outcome, the key assumption is that the potential hearing outcomes of hearings scheduled later and earlier within a facility group are comparable. That is, if not for the direct effect of decision fatigue itself, the IPs whose hearings occur at the end of the day would have the same potential hearing outcomes as those occurring at the beginning of the day. If this assumption is met, I can attribute any differences in observed hearing outcomes to the effect of fatigue itself, and not to observable or unobservable IP characteristics.

Hearings are ordered first by prison facility and then alphabetically by the prisoner's last name. The relationship between the alphabetical ranking of the prisoner's last name and the ordinal position of his hearing within his prison facility is displayed in Figure 2.1.

That hearings are ordered alphabetically within prison facility syggests that the ordering is plausibly unrelated to IP characteristics. I validate this claim by plotting IP characteristics such as race, gender, age, and number of prior offenses against the ordinal position of the IP's hearing in Figure 2.2.

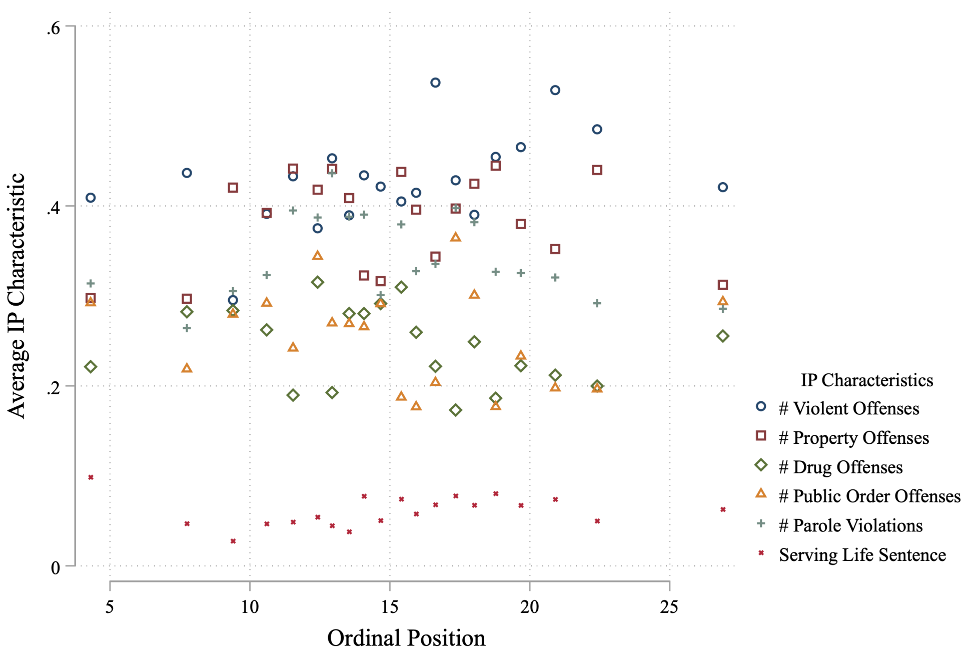
2.7 Results

I find that prisoners are more likely to receive parole if they have a later hearing, when judges are more fatigued. These results are reported in Table 2.2. I begin by regressing the hearing outcome on ordinal position after accounting for variation across facilities. In



Notes: This figure plots the relationship between the last initial of an IP and the ordinal position of their hearing within the group of hearings from their facility on the same day. Having a last name beginning with a letter later in the alphabet results in one's hearing occurring later within their facility group.

Figure 2.1: Alphabetical Ordering of Hearings



Notes: This figure plots the nonparametric relationship between a hearing's ordinal position and the criminal history characteristics of the IP. One observation is one parole hearing. Variables are residualized on a vector of facility indicators.

Figure 2.2: Relationship between Hearing Order and IP Characteristics

Columns 2, 3 and 4, I add controls for the hearing type, IP demographics, and IP criminal histories. A one unit increase in the ordinal position of a hearing increases the likelihood of parole by 0.4%, relative to an average of 81.3% overall.

I next modify the regression specification to allow for the effect of a hearing's ordinal position to interact with the IP's type. I include an interaction term for whether or not the IP is serving a life sentence. Table 2.3 reports these results. IPs not serving life sentences are 0.45% more likely to receive parole for each slot forward in time their hearing occurs, while IPs serving life sentences are approximately the same amount *less* likely to receive parole.

Table 2.2: Effect of Ordinal Position on Parole Likelihood

	(1)	(2)	(3)	(4)
Ordinal Position	0.00307*	0.00314*	0.00455***	0.00405**
	(0.00129)	(0.00127)	(0.00136)	(0.00128)
# Violent Offenses				-0.0208
				(0.0194)
# Property Offenses				0.0221
				(0.0161)
# Drug Offenses				0.0338
				(0.0188)
# Public Order Offenses				0.0373*
				(0.0179)
# Parole Violations				0.00249
				(0.0100)
Observations	3066	3066	2609	2927
Hearing Type	No	Yes	Yes	Yes
Demographics	No	No	Yes	Yes
Criminal History	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: I regress a binary indicator of whether or not the IP receives parole on the ordinal position of their hearing, as well as hearing type indicators (victim present, mandatory, review, reconsideration or medical), prisoner demographics (facility fixed-effects, race fixed-effects, age and gender), and criminal history measures (number of violent, property, drug and public-order offenses, number of previous parole violations, and a binary indicator for whether or not the IP is serving a life sentence

Table 2.3: Heterogeneity of Ordinal Position Effect

	(1) Received Parole
Ordinal Position	0.00456***
	(0.00130)
Serving Life Sentence=1	-0.0556
	(0.0523)
Serving Life Sentence=1 $\times$ Ordinal Position	-0.00812*
	(0.00324)
# Violent Offenses	-0.0193
	(0.0194)
# Property Offenses	0.0234
	(0.0161)
# Drug Offenses	0.0350
	(0.0188)
# Public Order Offenses	0.0378*
	(0.0179)
# Parole Violations	0.00215
	(0.0100)
Observations	2927

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: I regress a binary indicator of whether or not the IP receives parole on the ordinal position of their hearing, a binary indicator for whether or not the IP is serving a life sentence, and an interaction term for these two varibles. Controls include hearing type indicators (victim present, mandatory, review, reconsideration or medical), prisoner demographics (facility fixed-effects, race fixed-effects, age and gender), and criminal history measures (number of violent, property, drug and public-order offenses, number of previous parole violations.

I next examine the effect of a hearing's ordinal position on the subsequent costs incurred by the criminal justice system and society due to recidivism after parole, or due to the costs of continuing to incarcerate individuals who are denied parole. I use estimates from the crime costing literature on the criminal justice system cost and the tangible and intangible victim costs for property, drug, public order and violent offenses (McCollister *et al.*, 2010). To make the cost-benefit analysis as conservative as possible, I use the highest available cost estimates for each category of crime. For example, I assign the cost of any violent offense as

$8,000,000, the cost of murder, even though robberies are also considered violent offenses. Drug offenses and property offenses are assigned values of $28,000 and $26,000 respectively. Public order offenses, which are also known as "victimless" crimes, are given a cost of $6,000. Crime costs are recorded in 2008 dollars.

I use estimates reported by the New Hampshire Department of Corrections to account for the costs of incarcerating and paroling the individuals in my sample. The cost to incarcerate an IP annually is $34,155. The cost of paroling an IP annually is $520. Note that the costs of incarceration are underestimated, as they do not include the social and psychological toll of incarceration.

I estimate the future costs in the 6, 12 and 15 months following a hearing. I then estimate the impact of a hearing's ordinal position on these costs. These results are reported in Table 2.4. The effect of fatigue on decision-making appears to reduce costs overall, and these effects become more apparent the longer IPs spend paroled. In the 15 months after an IP's original parole hearing, nearly $400 in criminal justice and societal costs are reduced for each slot forward in time a hearing is held.

Table 2.4: Welfare Analysis of Parole Decisions

	Future Costs		
	(1)	(2)	(3)
	6 mo	1 yr	15 mo
Ordinal Position	-126.8	-310.8*	-376.9**
	(91.14)	(133.8)	(140.9)
Observations	3061	3061	3061

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: I regress measures of future criminal justice and societal costs at six, twelve, and fifteen months after a parole hearing on the ordinal position of their hearing. Controls include hearing type indicators (victim present, mandatory, review, reconsideration or medical), prisoner demographics (facility fixed-effects, race fixed-effects, age and gender), and criminal history measures (number of violent, property, drug and public-order offenses, number of previous parole violations, and a binary indicator for whether or not the IP is serving a life sentence.

2.8 Discussion

Decision Fatigue Distinct from Deteriorations in Mood

I show that as judges face an increasing number of decisions, the probability of granting parole increases by 2-3 percentage points for every ten minutes by which a hearing is delayed. While a decrease in parole-granting could likely be caused by emotional changes associated with fatigue, an increase in parole-granting behavior is much less likely to be emotionally driven. This indicates that changes in decisions are distinguishable from the changes in mood associated with fatigue, implying that decision fatigue is a psychological phenomenon entirely separate from emotion. The mechanism at work is also not indicative of an increased reliance on maintaining the status quo or default by keeping prisoners imprisoned. Instead, I propose that decision-makers seek to decrease their mental load by implementing a simple heuristic, simplifying the cost of making a decision, and opting for the choice they make most often - paroling IPs - except for IPs serving the longest sentences to begin with.

Marginal Effects are Small

The original study of Israeli parole judges found drastic changes in parole probability, while this study suggests much smaller effects (Danziger *et al.*, 2011). There are three possible explanations. First, in Israel, judges went through fifty hearings per day, as opposed to thirty in New Hampshire. One would expect decision fatigue to take a more drastic toll in Israel because more decisions were being made over the course of the day, causing fatigue to accumulate further.

Second, in New Hampshire, parole board members are presented with relevant hearing information one week beforehand. If a decision-maker has more time to process the relevant information, the mental cost of the decision can be spread over a larger period of time, leading to less fatigue. We would expect that conducting thirty hearings in one day would be less taxing than if no information was provided beforehand. Thus, any decision fatigue found in this setting should naturally be smaller in magnitude than if the decision-maker's time and access to pre-information is restricted.

A final difference is that a panel of three judges makes the final decision in New Hampshire. In the Israeli judges study, one judge makes the final decision. If the three judges split the mental cost of the decision, the magnitude of decision fatigue may be smaller. These factors and their effects on the direction and magnitude of decision fatigue are ripe for further investigation.

2.9 Conclusion

I exploit exogenous variation in the ordering of parole hearings in New Hampshire to investigate the impact of decision fatigue on high-stakes parole decisions. In New Hampshire, where the majority of inmates are granted parole, I find that as fatigue sets in, parole probabilities increase, contrary to previous studies. It is extremely unlikely that an increase in parole granting is caused by the typical changes in emotion associated with fatigue. This indicates that decision fatigue is a psychological phenomenon separate from changes in mood.

I show that judges implement simple heuristics when fatigued, and that this process appears to improve the quality of the parole decisions made. I conclude that decision fatigue is not a one-dimensional phenomenon; rather, I propose that decision-makers employ simple rules of thumb. In New Hampshire, when fatigued, parole judges increasingly grant parole to all prisoners except those serving life sentences. The impact of decision fatigue is multi-directional and predictable.

Decision fatigue has strong implications for any setting in which repeated decisions are made. The impacts on judicial proceedings, college admissions, and even medical procedures are costly. Understanding the mechanisms behind decision fatigue is crucial for improving effective decision-making.

Chapter 3

Perceiving and Learning from Mistakes

3.1 Introduction

> "The practice of medicine is difficult enough without having to bear the yoke of perfection."
>
> *David Hilfiker*
> *Physician & Author*

Women are persistently hired, compensated, and promoted less than men. In this paper, I show that how women and men perceive and learn from their own mistakes, and the mistakes of their peers, is an important channel through which gender disparities in the labor market emerge and persist.

I study how workers perceive and learn from mistakes in the context of medical decision-making. Medical errors are the third most common cause of death in the United States (Makary and Daniel, 2016). Medical mistakes are particularly salient in emergency medicine, wherein generalist physicians must quickly triage and treat patients presenting with a wide range of concerns while managing capacity and time constraints (Goldberg *et al.*, 2002;

Hilfiker, 1984). In a medical setting like the emergency department (ED), it can be especially difficult to identify the cause of an adverse event, given that such an outcome is jointly determined by the patient's health prior to the visit, the provider's decision-making in the visit, team decision-making, and larger system failures (Yee, 2002; Wears and Wu, 2002).

Using administrative data on the universe of Emergency Department (ED) visits occurring across the state of New York between 2005 and 2015, I conduct an event study of the impact of patient deaths occurring in the ED on subsequent physician treatment behavior. The detailed nature of my data allow me to identify physicians who experienced an ED patient death, their peer physicians working in the same ED, and the treatment decisions made by these physicians in the days before and after an ED patient death.

I identify several channels through which these adverse events differentially affect the subsequent treatment behavior of male and female physicians. First, I show that after experiencing a patient death, female physicians - relative to male physicians - become more aggressive in their treatment of patients who have the same medical concern as the patient who died. Second, I show that female physicians also overextrapolate from the mistake more than male physicians by becoming more cautious in how they treat patients with other medical concerns. Third, I show that female physicians not only overreact to their own patient deaths; they overreact to patient deaths experienced by peer physicians in their ED. Fourth, I show that female physicians take approximately twice as long for their treatment behavior to return to prior levels.

I consider, and rule out, several alternative mechanisms that might also explain these results. First, I show that changes in the treatment behavior of providers who experienced a patient death do not appear to be driven by changes in the composition of patients they are assigned. The patients appearing at a given ED after that ED has experienced a patient death do not differ based on demographics or prior health characteristics. Patients also do not appear to sort differently to physicians within a given ED after a patient death.[1] The

[1] In the ED, patients cannot pick their physicians; physicians, too, have a very limited ability to select their patients. They can only express preferences based on the chief complaint, age, and gender of the patients - the three pieces of information available at the time of triage (Chang and Obermeyer, 2020).

types of ED patient deaths experienced by male and female physicians are similar, given the rare nature of the event, and female and male physicians who experienced a death display similar patterns of treatment behavior over time prior to experiencing the patient death.

Second, I consider possible gender differences in malpractice lawsuits following a patient's adverse event, and internal formal or informal consequences taken by hospitals and ED staffing agencies after such an event. I conclude that both channels are unlikely to be driving my effects, especially because the gender disparity in response to an adverse events holds whether or not the adverse event occurs to a provider, or to their peer.

My results contribute to three distinct literatures. First, I show that how men and women respond differently to similar adverse events is one channel through which gender disparities in worker performance appear and persist. The literature has shown that differences in how male and female workers interpret their own qualifications (Abraham and Stein, 2020), trade wages for gender-specific amenities (Bolotnyy and Emanuel, 2019), and assess their own abilities all contribute to the gender wage gap (Blau and Kahn, 2007; Exley and Kessler, 2019). In particular, a large body of work in psychology and sociology shows that women are more avoidant of errors, and externalize positive outcomes while internalizing negative ones (Kanze *et al.*, 2018; Deaux and Emswiller, 1974; Deaux and Farris, 1977; Etaugh and Brown, 1975; Coffman, 2014; Feather, 1968; Coffman and Klinowski, 2020). Gender gaps in pay and promotion are present in both medical practice and academic medicine (Esteves-Sorenson and Snyder, 2012; Lo Sasso *et al.*, 2011; Jagsi *et al.*, 2012), and in physician referral networks (Zeltzer, 2020; Sarsons, 2019).

Second, I contribute to the literature on how agents react to new information, much of which explores updating in financial contexts by shedding light on the specific mechanisms of updating and extrapolating, and exploring the distinction between learning from one's own experiences and peers' experiences. Agents overreact to both positive and negative experiences of their own and their peers in portfolio choice (Choi *et al.*, 2009; Malmendier and Nagel, 2011; Gennaioli *et al.*, 2015; Calvet *et al.*, 2009) and the housing market (Bailey *et al.*, 2018).

Third, I build on the literature that explores the determinants of physicians behavior. I leverage unique features of the generalist ED setting to simultaneously 1) estimate both direct and spillover learning effects and 2) understand how agents use information gained in one task to extrapolate to other tasks. The literature establishes theoretically and empirically that doctors exhibit decidedly non-Bayesian updating (Gong, 2018; Camacho *et al.*, 2011; Rottman, 2017), and explores the subsequent effects on medical technology adoption (Smythe, 2002; Ferreyra and Kosenok, 2011). Related work by Singh (2020) and Fiedler (2013) show that physicians overreact to recent negative experiences in the labor and delivery setting, and Lecate (2014) explores how physicians extrapolate when learning about the effectiveness of various cancer drug regimens.

The rest of this paper is structured as follows. In Section 2, I briefly describe the Emergency Department setting. In Section 3, I describe my data sources, sample and variable construction. Section 4 outlines my empirical methodology and identifying assumptions. Section 5 summarizes my key findings. In Section 6, I consider and rule out alternative mechanisms, and Section 7 concludes.

3.2 Decision-Making in the ED

3.2.1 Context

The ED is a compelling setting in which to study decision-making, both because of the urgent, high-stakes nature of ED visits, and because of the ED's place within the healthcare system. ED usage is currently growing at twice the rate of the population, and EDs are responsible for half of all hospital admissions (Gonzalez Morganti *et al.*, 2013). EDs are also the primary source of equity in the healthcare system, as they are the only setting in which a minimum standard of care is guaranteed without regard for the patient's ability to pay (Burke and Paradise, 2015; Zibulewsky, 2001).[2]

[2]This minimum standard is the "screening and stabilization" of patients presenting in the ED, as mandated by the Emergency Medical Treatment and Labor Act (EMTALA). EDs must perform a medical screening exam for all patients, and stabilize all patients in unstable condition.

Patients arrive at the ED in an unscheduled manner, describe their symptoms to a triage nurse, and are assigned priority based on perceived severity of their concern.[3] The patient's concern is distilled into a single "chief complaint", which usually take the form of a broad, nontechnical symptom description. Table 3.1 lists the twenty most common ED visit chief complaints. Patients cannot express a preference for a physician, but physicians

Table 3.1: Top 20 Chief Complaints

	Number of Visits	% of Visits
Fever	316760	11.52
Chest Pain	232465	8.46
Abdominal Pain	230442	8.38
Cough	219511	7.99
Limb Pain	186986	6.80
Headache	160180	5.83
Abdominal Pain, other	125124	4.55
Leg Injury	116365	4.23
Backache	114270	4.16
Sore Throat	113110	4.12
Shortness of Breath	106658	3.88
Skin Issue	105348	3.83
Head Injury	104720	3.81
Fainting	96385	3.51
Vomiting	95734	3.48
Lower Back Pain	94729	3.45
Dizziness	93608	3.41
Chest Pain, other	90558	3.29
Upper Resp Infection	74042	2.69
Dental Issue	71699	2.61
Total	2748694	100.00

Notes: This table reports the top twenty most common chief complaints and their frequencies across all EDs in the state of New York in 2009. Each ED visit is given a "chief complaint": a broad, non-technical summary of the patient's symptoms at the time of ED arrival. Each visit is given just one chief complaint, which is recorded using ICD-9-CM diagnosis codes.

may select their preferred patients based on the limited information available at the time:

[3]Since 1999, most US EDs have used the Emergency Severity Index to assign priority to patients in the ED. The ESI assigns patients to a five-point severity scale based on perceived urgency, vital signs, and anticipated resource utilization (Gilboy *et al.*, 2011).

chief complaint, age, and gender. The ED physician reviews the patient's health history, pursues further testing, offers a final diagnosis, and decides to either admit into the hospital a patient who needs inpatient care, or discharge a patient in stable condition. Nurse practitioners and physician's assistants may assist with collecting the patient's vital signs, health histories, or administering treatment.[4] While adverse events undoubtedly affect the entire ED team, I focus on the impact of adverse patient events on the physician associated with that patient.[5] Impacts on the ED's nursing staff would likely affect all patients, as nurses work collaboratively with the entire ED team.

3.2.2 Errors

While both under- and over-treatment errors are common in medicine, I focus on the impact of perceived undertreatment errors, for which there are several measures.[6] Whether or not a patient *dies* shortly after an ED visit is often a sign of a missed diagnosis (McCarthy *et al.*, 1993; Obermeyer *et al.*, 2017). Patients returning to the ED or hospital in an unscheduled manner in the days following an initial ED visit also represent preventable errors, and are a key target for quality improvement (Khera *et al.*, 2020; Wadhera *et al.*, 2019).[7] Table 3.2 reports the 3- and 30-day mortality rates and hospital revisit rates for patients with different chief complaints.

I focus on in-ED mortality, for two reasons. Mortality and hospital revisit behavior in the days following ED discharge create two sources of ambiguity: when or whether the provider is notified about the adverse event. Focusing on in-ED deaths allows me to

---

[4]Alexander and Schnell (2019) explore the impact of changes in legislation that give nurse practitioners independent prescriptive authority.

[5]Hogan *et al.* (2016) describes an ED patient's death from the perspective of the ED nursing staff.

[6]The consequences of overtreatment go well beyond wasteful medical spending. Overtesting often leads to incidental finding cascades, hospital-acquired infections or injuries, and fewer limited resources such as inpatient beds for other patients in need (Mirilas and Skandalakis, 2002; De Vries *et al.*, 2008; Ganguli *et al.*, 2019; Jha *et al.*, 2013).

[7]The Hospital Readmissions Reduction Program is a pay-for-performance program that incentivizes hospitals to reduce the number of hospital readmissions within 30 days of a patient's discharge.

Table 3.2: Rate of Adverse Events

	(1) Cardiac Arrest	(2) Syncope	(3) Chest Pain	(4) Fever
7-Day Mortality	0.886	0.0118	0.00455	0.00435
	(0.317)	(0.108)	(0.0673)	(0.0658)
30-Day Mortality	0.889	0.0188	0.00768	0.00747
	(0.314)	(0.136)	(0.0873)	(0.0861)
3d Revisit	0.0106	0.0447	0.0612	0.0696
	(0.102)	(0.207)	(0.240)	(0.254)
30d Revisit	0.0112	0.0488	0.0723	0.0721
	(0.105)	(0.216)	(0.259)	(0.259)
N	5040	48275	123374	85044

Notes: This table reports the average rates of 7- and 30-day mortality, and 3- and 30-day ED revisits, for selected chief complaints. The n day intervals are calculated from the date of discharge. If a person visits the ED on day 1, stays in the hospital for 14 days, and is discharged on day 15, their 3-day revisit rate is calculated based on whether they revisited an ED between days 15 and 18. Revisits to the original ED and to non-index EDs are included in my calculation of the revisit rate, which results in a more accurate picture of revisit behavior, as almost 30% of revisits are to non-index facilities (Duseja et al., 2015).

precisely identify *when* an ED physician experiences the adverse event, and isolate changes in treatment behavior in the days immediately following such an event. Patient deaths occurring in the ED can also clearly be attributed to treatment received by the patient within the index ED visit, and are less likely to be attributed to the patient's medication adherence, follow-up care, or health behaviors occurring after being discharged.

3.2.3 Consequences of Patient Deaths

I focus on the impact of adverse events on providers' beliefs about their own capabilities and confidence in their subsequent treatment choices. Doctors describe these events as causing grief, guilt, burnout, and resolving not to allow similar events to occur again (Yee, 2002).[8] I briefly outline other channels through which adverse patient outcomes may also have

[8] Robertson and Long (2018); Kaldjian *et al.* (2008); Leape (1994); Newman (1996) and Helo and Moulton (2017) discuss the emotional impact that such deaths have on providers.

an effect. In Section 6, I describe how I distinguish the internal effects of adverse patient outcomes from either institutional responses or legal responses.

Institutional Response

How patient deaths are handled varies by hospital. There are surprisingly few norms around how doctors communicate with each other, or with patients and their families, about adverse outcomes (Kroll *et al.*, 2008; Gallagher *et al.*, 2013; Kaldjian *et al.*, 2008; Newman, 1996; Shoenberger *et al.*, 2013). Morbidity and Mortality conferences are standard practice at most academic hospitals; particularly salient recent cases in which adverse events occurred are prepared and discussed with providers on a weekly or monthly basis (Sinitsky *et al.*, 2019).[9] Hospitals most often search for "process" or "system" changes that can be implemented to prevent similar errors from occurring again, de-emphasizing individual culpability.

Legal Response

ED physicians are more likely than other specialties to be the target of a malpractice lawsuit.[10,11] Female physicians, however, are significantly less likely to be sued than male physicians (Guardado, 2017).[12]

In order to separate the effect of legal responses from the physician's belief updating process, I will collect data from the New York State Department of Health on medical malpractice actions occurring to ED physicians in my sample between 2005 and 2015, to control directly for this channel. Importantly, spillover effects of deaths on peer physicians

---

[9] In "When Doctors Make Mistakes", Dr. Atul Gawande recounts his experience having his own adverse patient outcome featured in an M&M conference: `https://www.newyorker.com/magazine/1999/02/01/when-doctors-make-mistakes`.

[10] ED visits represent 12% of all malpractice suits in New York (Farley, 2014).

[11] Frakes and Gruber (2019) study the impact of complete legal immunity on the cost and quality of care.

[12] In addition to differences in work hours and specialty choice, female physicians establish better communication with patients and loved ones when conveying news of an adverse event, which is believed to lower their risk of getting sued (Carroll, 2015). The practice of "defensive medicine" increases healthcare costs, but appears not to meaningfully improve communication or prevent adverse events (Harris, 1987).

would not be driven by malpractice proceedings.

3.3 Data

I combine data from three sources for this analysis. The first is the New York Statewide Planning And Research Cooperative System (NY SPARCS) outpatient administrative dataset, which includes every ED visit occurring across the state of New York from 2005 - 2015. These data cover approximately 70 million ED visits, including those that result in hospital admission. The data include every ICD-9-CM diagnosis code [13] given to the patient as part of the index visit, and every NUBC revenue code[14] and CPT procedure code[15] for services provided during the visit. The records also contain anonymized patient IDs and patient demographics such as age, race, and forms of payment.

The SPARCS dataset includes facility identifiers, physician identifiers, patient mortality indicators within the visit and at subsequent intervals after discharge, and the date of the patient's visit. These variables allow me to identify physicians who experienced the death of a patient in the ED, and identify the patients treated both prior to and after that patient death, both by the index physician, and by the physician's peers at the same ED. Appendix A.2 describes the types of providers identified in my data.

I combine the SPARCS dataset with physician licensing data from the New York State Department of Education. The licensing data include each provider's six-digit license number, full name, date of board exam, medical school name and date of graduation.[16]

---

[13] International Classification of Disease, Volume 9, Clinical Modification. A full list of these diagnosis codes is available at `http://www.icd9data.com/`.

[14] National Uniform Billing Committee codes are used to record line-items that appear in claims data. The NUBC codeset is publicly available at `https://med.noridianmedicare.com/web/jea/topics/claim-submission/revenue-codes`.

[15] Current Procedural Terminology codes are used to record every therapeutic and diagnostic procedure a patient is *charged for* during their visit. Common procedures that patients are not charged for, such as the provision of aspirin within 30 minutes of arrival for patients with suspected heart attack, are not recorded in CPT data.

[16] These data were obtained via Freedom of Information Law request FL-OP-17/127 to the New York State Department of Education. Data on individual licensed physicians is publicly available at `http://www.`

I combine the physician licensing data with a dataset of gender-labeled first names from two sources: US census birth records, and social media accounts from within and outside the US via the Genderize API.[17] Appendix C.1 details how first names are isolated from the text data containing each provider's full name. Appendix C.2 describes how both datasets classify first names as either male or female, and highlights examples in which gender classification is not possible. 90.5% of all ED physicians in my sample are assigned a gender. 35% of these physicians are identified as female.

3.4 Methodology

3.4.1 Characterizing Provider Choices

I focus on the effects of adverse events on an emergency medicine provider's propensity to admit patients into the hospital. The decision to admit is an important one, as EDs are the source of half of all hospital admissions, and many hospitals face high demand for a limited number of inpatient beds.

A challenging feature of SPARCS data is that the only patients who can be associated with an ED provider are the patients who were treated and *discharged* by that provider. If an ED patient is admitted into the hospital, the only providers that appear on the patient's record are the providers who work in the hospital. The providers who treated this patient in the ED do not appear on the patient's record.

I therefore infer changes in an ED provider's propensity to admit patients by studying changes in the composition of patients the provider chooses to discharge before and after experiencing an adverse event. I then rule out that these changes in the composition of discharged patients are driven either by changes in the composition of patients arriving at the provider's ED, or by changes in how patients at the index ED are allocated between the index physician and his peer providers.

`nydoctorprofile.com`.

[17] Available at `http://genderize.io`

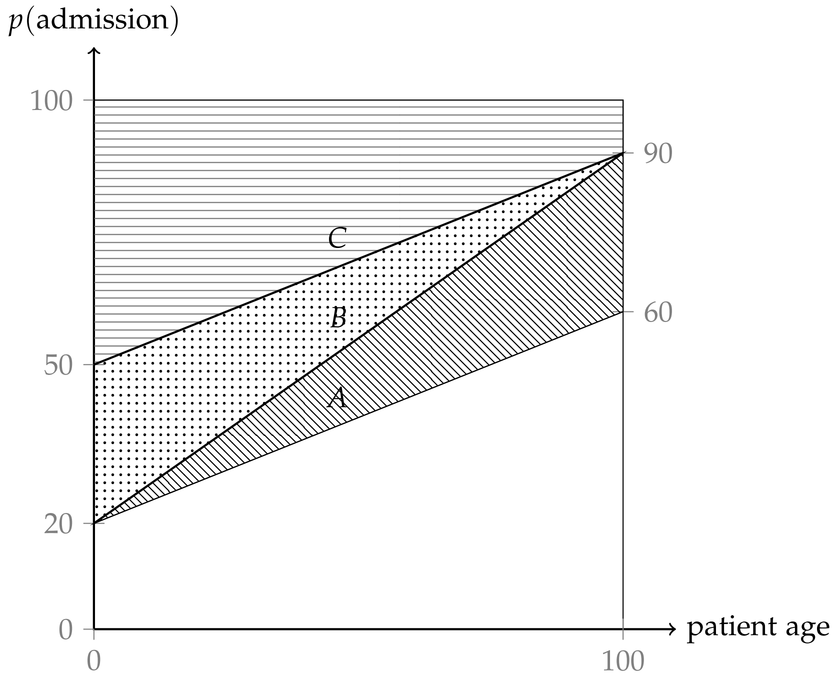


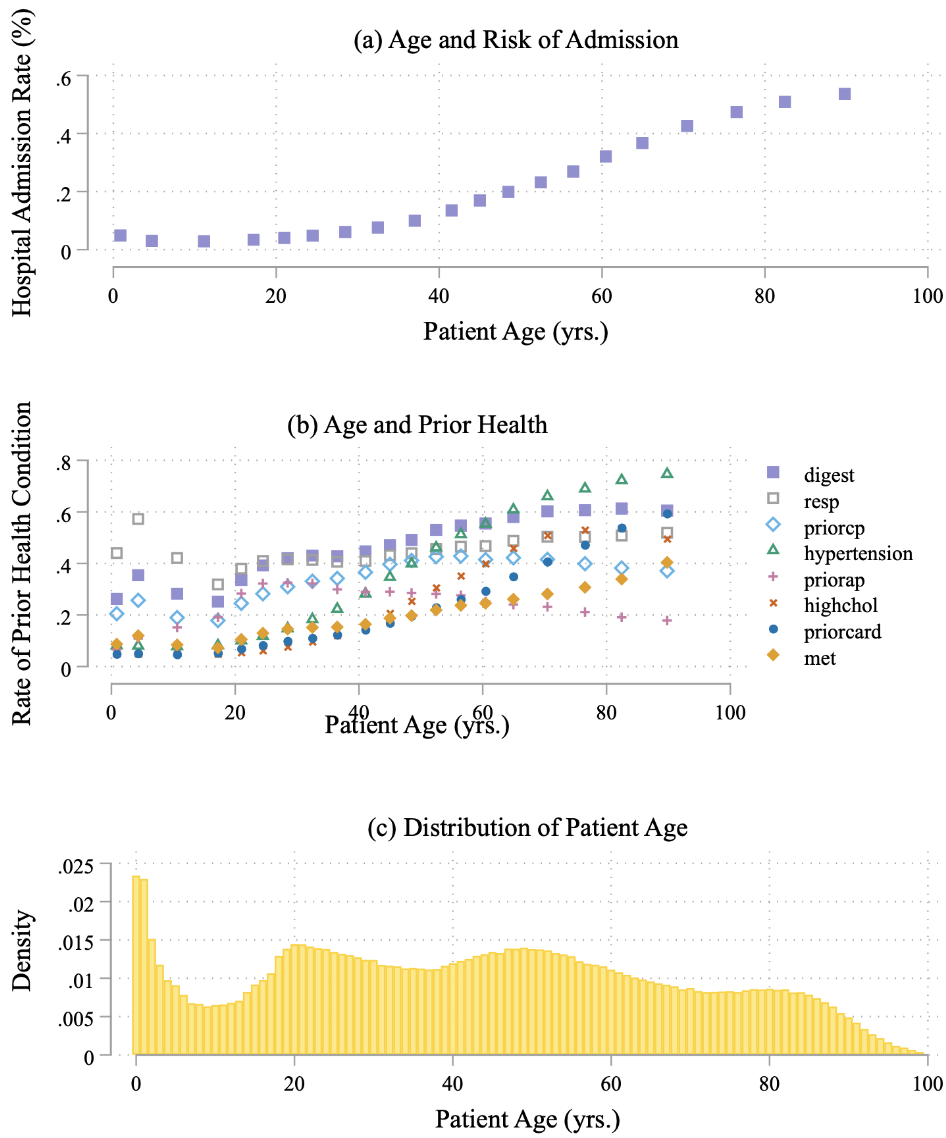
Figure 3.1: Imputing Admission Behavior from Discharged Patient Characteristics

Figure 3.1 illustrates how changes in the average characteristics of discharged patients can be used to infer changes in a provider's propensity to admit patients into the hospital, under certain assumptions. The figure plots the probability of being admitted into the hospital as a function of the patient's age for three providers: A, B, and C. The population of patients discharged by provider C is the shaded region above line C. The population of patients discharged by provider A is the combination of the three shaded regions above line A. The empirical relationship between age, prior health characteristics and probability of admission is shown in Figure 3.2. The average age of the patients a provider discharges is given by

$$\bar{x} \mid \text{discharge} = \int_0^{100} (1 - (\alpha + \beta x))(\lambda_x)(x)$$

where λ_x represents the share of the patient population that is age x, and the provider's propensity to admit a patient is represented by $p(\text{admission}) = \alpha + \beta \times \text{age}$.

I assume a normally distributed patient population: $\lambda_x = \frac{1}{100}$. Then $\bar{x} \mid \text{discharge} =$



Notes: Panel (a) of this figure shows the unconditional empirical relationship between a patient's age and their risk of hospital admission, pooled across chief complaints. Panel (b) shows relationship between age and various prior health conditions. Panel (c) shows the raw distribution of patient age for the same population. The distribution is approximately trimodal, with three modal ages: early 20s, 40s, and 80s.

Figure 3.2: Patient Age and Likelihood of Hospital Admission

$50(1-\alpha) - \frac{\beta(100^2)}{3}$. The average age of discharged patients is a function of both the provider's base rate of admission, and the relationship between their propensity to admit a patient and the patient's age. Examining changes in the composition of discharged patients will therefore capture changes in provider behavior along either of these two dimensions: if, in the wake of an ED patient's death, provider A becomes more like provider B or provider C - either by increasing their baseline propensity to admit (α) or by increasing their propensity to admit with respect to a patient's age or other health characteristic (β), both changes will be reflected in the composition of the provider's discharged patient population.

A key assumption of this exercise is that the distribution of patients assigned to a given provider does not change in the wake of an adverse patient outcome. I break this assumption into two key components: first, that the composition of patients arriving at the index ED does not change as a result of the adverse patient outcome, and second, that the allocation of patients to the index provider and his peers within a given ED does not change. I show that both assumptions hold in Section 6.

3.4.2 Characterizing Patient Health

I characterize the composition of an ED physician's discharged patients by their average health characteristics at the time of their visit: age, demographics, and prior health conditions. I classify patients as having a given health condition if they have ever been given a primary or ancillary diagnosis of that condition in any visit to an ED, hospital, urgent care clinic or ambulatory surgery center prior to their index visit.[18,19]

These measures create a detailed picture of the patient's health at the time of their visit. Table 3.3 gives the rates of common health conditions for patients with select chief complaints. Appendix C.3 discusses how ICD-9 diagnosis codes are aggregated to create measures of health conditions which are both clinically relevant, and amenable to statistical

[18] Visits to EDs and hospitals are included in the entire SPARCS sample, from 2005 to 2015. Visits to ambulatory surgery centers and urgent care clinics are included in SPARCS data beginning in 2011.

[19] Note that while some health conditions are transient, these measures are permanent - once a patient has been given a diagnosis code of high cholesterol, they are classified as having high cholesterol in all future visits.

analysis.

Table 3.3: Characterizing Ex-Ante Patient Health

	(1) Cardiac Arrest	(2) Syncope	(3) Chest Pain	(4) Fever
Cancer	0.134	0.0944	0.0754	0.0687
	(0.340)	(0.292)	(0.264)	(0.253)
Metabolic	0.287	0.203	0.183	0.148
	(0.452)	(0.402)	(0.386)	(0.355)
Hypertension	0.493	0.427	0.418	0.172
	(0.500)	(0.495)	(0.493)	(0.377)
High Cholesterol	0.341	0.295	0.301	0.115
	(0.474)	(0.456)	(0.459)	(0.320)
Thyroid	0.111	0.106	0.0997	0.0479
	(0.314)	(0.308)	(0.300)	(0.214)
Diabetes	0.278	0.193	0.203	0.0816
	(0.448)	(0.395)	(0.402)	(0.274)
Endocrine	0.0455	0.0306	0.0295	0.0193
	(0.208)	(0.172)	(0.169)	(0.138)
Obesity	0.120	0.0841	0.130	0.0523
	(0.325)	(0.278)	(0.336)	(0.223)
Prior Heart Attack	0.0901	0.0388	0.0595	0.0147
	(0.286)	(0.193)	(0.236)	(0.120)
Anemia	0.279	0.197	0.180	0.119

(continued on next page)

(continued from previous page)

	(0.448)	(0.398)	(0.384)	(0.324)
Immune	0.0540	0.0355	0.0391	0.0353
	(0.226)	(0.185)	(0.194)	(0.185)
dementia	0.0343	0.0273	0.00979	0.00901
	(0.182)	(0.163)	(0.0984)	(0.0945)
Mental Health	0.0881	0.0801	0.106	0.0398
	(0.283)	(0.271)	(0.307)	(0.195)
Alcohol	0.0459	0.0358	0.0630	0.0150
Dependence	(0.209)	(0.186)	(0.243)	(0.122)
Drug Dependence	0.0353	0.0291	0.0634	0.0138
	(0.185)	(0.168)	(0.244)	(0.117)
Multiple	0.00419	0.00343	0.00463	0.00291
Sclerosis	(0.0646)	(0.0585)	(0.0679)	(0.0539)
Respiratory	0.436	0.353	0.435	0.522
	(0.496)	(0.478)	(0.496)	(0.500)
Digestive	0.437	0.422	0.480	0.384
	(0.496)	(0.494)	(0.500)	(0.486)
Prior Cardiac	0.396	0.274	0.254	0.107
	(0.489)	(0.446)	(0.436)	(0.309)
Prior Chest Pain	0.282	0.267	0.432	0.281
	(0.450)	(0.443)	(0.495)	(0.449)
Prior Ab Pain	0.134	0.166	0.250	0.145

(continued on next page)

				(continued from previous page)
	(0.340)	(0.372)	(0.433)	(0.352)
Nutritional	0.0843	0.0440	0.0376	0.0280
	(0.278)	(0.205)	(0.190)	(0.165)
Kidney Function	0.232	0.117	0.0905	0.0550
	(0.422)	(0.322)	(0.287)	(0.228)
N	5016	48035	123120	84871

Notes: This table reports the summary statistics of prior health descriptors for a set of chief complaints. One observation represents one ED visit. Note that although some health conditions are transient, the indicators described above are ever or never indicators. That is, a patient is recorded as having a given health condition if they have ever been given a diagnosis code for that condition in any visit in my sample that occurred prior to their index visit.

Using the following specification, I aggregate these patient health characteristics into a predicted 1-year mortality risk score. I use the patient's age and this aggregate risk score as my main measures of the overall health of a provider's discharged patient population.

$$y_{iph,d} = \beta \mathbf{X}_{i,d} + \delta_{h,y(d)} + \text{triage}_{i,d} + \epsilon_{iph,d} \tag{3.1}$$

For patient i treated (and discharged) by provider p at hospital h on day d, I regress an indicator $y_{iph,d}$ of the patient's mortality within 1 year following their index visit on a vector of prior health characteristics $\mathbf{X}_{i,d}$, fixed effects for each hospital by year $\delta_{h,y(d)}$, and fixed effects capturing the patient's chief complaint, age and gender, represented by $\text{triage}_{i,d}$. $\hat{y}_{iph,d}$ represents the patient's aggregate risk of 1-year mortality.

3.4.3 Sample Construction

Patient deaths in the ED are extremely rare, representing fewer than 0.01% of all patient visits. Table 3.4 lists the chief complaints that are most commonly represented among in-ED deaths, their frequencies, as well as the overall probability of in-ED mortality for each chief complaint. I restrict my analysis to an event study of the ED deaths of patients presenting with these chief complaints. Table 3.5 describes the characteristics of patients who expire in the ED relative to all ED patients. To capture the immediate effects of an adverse event, I restrict my analysis to the six months leading up to, and following, an ED patient death. I further restrict my sample to the 90.5% of providers for whom a gender can be classified, and to the 189 facilities which include both an ED and an inpatient treatment center.[20]

Providers are classified by three levels of exposure to the ED patient death: treated, spillover, or control. All providers working in ED h are represented by the set H. Provider $p \in H$ who experiences the in-ED death of patient i with chief complaint c on day d is identified as the treated provider on all days in the interval $(d - 180, d + 180)$. All providers $k \in H_{-p}$ are designated as spillover providers within the same day interval. These providers are likely to have witnessed the patient's death, but were not responsible for the patient themselves. All providers at all EDs outside of ED h are classified as control providers in the interval $(d - 180, d + 180)$. These providers are unlikely to have witnessed the index patient's death, and their treatment behavior should not be affected by it.

To establish trends in provider behavior prior to a patient's death, and flexibly examine changes in provider behavior after the death, I classify patients seen by each category of provider based on the month in which they are seen relative to d. I omit all patients - including the patient who dies in the ED - seen on day d. Month n includes all patients seen within the day interval $(d + 30 \times (n - 1), d + 30(n))$. The reference month 0 is therefore the group of patients seen between 30 and 1 days prior to the index patient death.

I allow the effect of patient i's death on the subsequent treatment received by other patients to vary both by patient i's chief complaint c, and the chief complaints of future patients. This allows me to leverage the rich variety of medical concerns seen in the ED to establish patterns in how providers extrapolate - and perhaps over-extrapolate - from an

[20] As opposed to a freestanding ED or a hospital without emergency services, in which case the provider does not face a decision to admit that can be observed in the data.

Table 3.4: Top 20 Chief Complaints Resulting in ED Death

	N Deaths	P(Death)
Cardiac Arrest	3456	.6874876
Respiratory Arrest	41	.2124352
Alteration of Consciousness	276	.0625283
Hypotension	30	.0116641
Acute Respiratory Failure	37	.009172
Other Respiratory Abnormality	169	.0059467
Altered Mental Status	87	.004744
Shortness of Breath	173	.0029506
Other general symptoms	16	.0028209
Multiple Injury	24	.00269
Syncope	103	.002109
Malaise and Fatigue	53	.0017389
Chest Pain	91	.0007413
Head Injury	21	.000446
Fever	17	.0002005
Chest Pain, Other	9	.0001872
Abdominal Pain, Other	12	.0001663
Abdominal Pain	19	.0001565
Dizziness	7	.000139
Leg Injury	2	.0000347
Total	4643	.0056635
N	819807	

Notes: This table reports the top twenty chief complaints by their likelihood of resulting in ED death, as well as the number of ED deaths they are responsible for. Likelihoods are unadjusted for patient demographics or facility characteristics.

Table 3.5: CHARACTERISTICS OF PATIENTS WHO DIE IN THE ED

	Chest Pains			
	(1) All		(2) Died in ED	
A. Demographics:				
Age (yrs)	48.98	(19.86)	67.61	(17.93)
Male	0.479	(0.500)	0.643	(0.483)
White	0.574	(0.495)	0.686	(0.468)
Black	0.230	(0.421)	0.100	(0.302)
Native American	0.00428	(0.0653)	0	(0)
Asian / Pacific Islander	0.0207	(0.142)	0.0286	(0.168)
Multiracial / Other	0.172	(0.377)	0.186	(0.392)
B. Insurance Type:				
Privately Insured	0.501	(0.500)	0.314	(0.468)
Medicare	0.233	(0.423)	0.457	(0.502)
Medicaid	0.223	(0.416)	0.0857	(0.282)
Uninsured	0.222	(0.415)	0.300	(0.462)
C. Health History:				
Digestive	0.480	(0.500)	0.529	(0.503)
Respiratory	0.435	(0.496)	0.457	(0.502)
Prior Chest Pain	0.432	(0.495)	0.314	(0.468)
Hypertension	0.418	(0.493)	0.586	(0.496)
Prior Ab Pain	0.250	(0.433)	0.157	(0.367)
High Cholesterol	0.301	(0.459)	0.400	(0.493)
Prior Cardiac	0.254	(0.436)	0.357	(0.483)
Metabolic	0.183	(0.386)	0.300	(0.462)
N	123374		70	

Notes: This table plots the patient characteristics of all ED visits for chest pains, and the subset of these visits in which the patient expires in the ED. Patients with chest pain who die in the ED tend to be older, whiter, less insured, and have a higher rate of prior health conditions.

adverse event experienced in chief complaint c to the provider's treatment behavior for all other chief complaints C_{-c}. For example, after experiencing a patient with chest pains die, a provider may become more cautious both with their treatment of chest pain patients and shortness of breath patients. Becoming more cautious in the treatment of leg injury patients, however, would be unlikely. Thus, allowing the effect of a patient death to vary across the full spectrum of chief complaints provides an opportunity to perform a falsification test of the treatment: patients with chief complaints that have no relation to that in which the adverse event occurs should not experience a change in provider behavior.

3.4.4 Identification Assumptions

To quantify the effect of a physician's recent adverse patient outcome on their subsequent treatment behavior, I compare the aggressiveness of admission decisions made by doctors shortly before and after they experience a patient's death in the ED. The key identifying assumption underlying this approach is that the potential treatment decisions made by doctors - that is, the treatment decisions a provider would have made had he not experienced a patient death - are not systematically different between patients being seen by the index provider shortly before and after the adverse event. Differences in the composition of their discharged patients are due to increased caution or aggressiveness of treatment, and not due to a change in the composition of patients they are receiving.

I discuss three potential violations of this assumption and my approach to addressing them. The first potential violation would occur if the composition of patients arriving at the provider's ED changes after the ED experiences a patient death. This could happen if patients become aware of the adverse event, and switch the ED that they use differentially based on their prior health or their awareness, which may correlate with their prior health. The discharged patient population of the treated provider and his peers would then appear to change after the adverse event, through no behavioral change on the part of the providers. This is unlikely to be driving my results because, due to the emergent nature of conditions seen in the ED, patients generally visit the ED that is geographically closest to them. Patients are also unlikely to be made aware of adverse events within a timeframe amenable to an immediate response. I empirically validate this claim by showing that the prior health and demographic characteristics of patients visiting an ED does not change shortly after the ED experiences a patient death.

The second potential violation concerns the allocation of patients to physicians within an ED. This allocation could change in the wake of an adverse event. A physician's discharged patient population could appear to change, not due to a change in their decision to admit, but due to a change in the patients they are assigned. I argue that this does not drive my results, because if my results are driven entirely by the reallocation of patients with in the ED, changes in the discharged patient population of one doctor would be offset by symmetric changes for other providers. Indeed, both index and peer providers' discharged patient populations adjust in the same direction after an adverse event, indicating that behavioral changes on the part of providers - and not a reallocation of patients - appears to be driving the results.

Third, physicians who experience a patient death may be different from those who do not - especially in terms of experience, caution, or diagnostic skill. I rely on a strategy that differences out trends in behavior by comparing a provider's choices in the months after an adverse event to their own behavior in the months just prior to the event. Allowing for differences in overall levels of admission and assuming that treated and untreated providers have similar trends in treatment behavior over the twelve-month observation period around each event allows me to plausibly isolate changes in treatment behavior created by an adverse patient event.

3.4.5 Estimating Equation

I estimate the causal impact of physician bandwidth on treatment choices and patient outcomes using the following estimation equation. I estimate this equation separately for chest pain visits and abdominal pain visits. For instance, for all patients arriving in the ED with a chief complaint of chest pain, the treatment choices and outcomes for patient i who arrives at facility f at date-hour h in year y is

$$
\begin{aligned}
y_{iph,d|c(i)=c} = & \ \delta(\mathbb{1}_{m(d)} \times \text{treated}_p^k \times \text{gender}_p) \\
& + \zeta(\mathbb{1}_{m(d)} \times \text{treated}_p^k) + \theta(\text{treated}_p^k \times \text{gender}_p) + \xi(\mathbb{1}_{m(d)} \times \text{gender}_p) \\
& + \eta_{m(d)} + \gamma \text{treated}_p^k + \beta \text{gender}_p \\
& + \alpha_{\text{cal. month}(d)} + \delta_{h,y(d)} + \epsilon_{iph,d}
\end{aligned}
$$

$y_{iph,d}$ represents a health-related patient characteristic, like the patient's age, at the time of the patient's visit. $\delta_{h,y(d)}$ represents facility-by-year fixed effects, $\alpha_{\text{cal. month}d}$ represents fixed-effects for the twelve calendar months. $\epsilon_{iph,d}$ represents the error term.

β captures time-invariant average differences in the patients that female physicians choose to discharge. This parameter captures both differences in their decisions to admit patients, as well as differences in the patients they serve, which may be driven by the facility they work in, or the hours at which they work. Table 3.6 describes the average differences in training and experience between male and female physicians. γ captures time-invariant differences in control, peer and treated physician discharge behavior. This parameter captures differences in the diagnostic skill or caution that may lead a physician to experience or avoid the death of a patient in the ED. θ further captures time-invariant differences in each of these physician categories by gender. Table 3.7 summarizes the differences in patient characteristics for treated, peer and control physicians.

Table 3.7: Balance on Observable Patient Characteristics

	Control		Peer		Treated	
	(1)	(2)	(3)	(4)	(5)	(6)
	Male Dr.	Female Dr.	Male	Female	Male	Female
A. Demographics:						
Age (yrs)	46.33	45.52	46.94	46.32	47.38	47.28

(continued on next page)

(continued from previous page)

	(19.62)	(19.40)	(19.83)	(19.43)	(19.68)	(19.55)
Male	0.411	0.399	0.411	0.386	0.402	0.388
	(0.492)	(0.490)	(0.492)	(0.487)	(0.491)	(0.488)
White	0.610	0.521	0.600	0.490	0.710	0.646
	(0.488)	(0.500)	(0.490)	(0.500)	(0.454)	(0.479)
Black	0.204	0.250	0.192	0.239	0.121	0.239
	(0.403)	(0.433)	(0.394)	(0.426)	(0.326)	(0.428)
Native American	0.00397	0.00427	0.00210	0.00274	0.00266	0.00478
	(0.0629)	(0.0652)	(0.0458)	(0.0523)	(0.0515)	(0.0692)
Asian / Pacific	0.0186	0.0234	0.0183	0.0214	0.0173	0.0144
Islander	(0.135)	(0.151)	(0.134)	(0.145)	(0.130)	(0.119)
Multiracial /	0.164	0.201	0.188	0.248	0.149	0.0957
Other	(0.370)	(0.401)	(0.391)	(0.432)	(0.356)	(0.295)
B. Insurance Type:						
Privately	0.473	0.495	0.481	0.540	0.422	0.560
Insured	(0.499)	(0.500)	(0.500)	(0.498)	(0.494)	(0.498)
Medicare	0.180	0.173	0.167	0.183	0.129	0.148
	(0.384)	(0.378)	(0.373)	(0.386)	(0.335)	(0.356)
Medicaid	0.185	0.197	0.147	0.172	0.0824	0.163
	(0.388)	(0.398)	(0.354)	(0.377)	(0.275)	(0.370)
Uninsured	0.267	0.232	0.283	0.184	0.430	0.234

(continued on next page)

(continued from previous page)

	(0.442)	(0.422)	(0.450)	(0.388)	(0.495)	(0.425)
C. Health History:						
Digestive	0.475	0.479	0.463	0.472	0.415	0.435
	(0.499)	(0.500)	(0.499)	(0.499)	(0.493)	(0.497)
Respiratory	0.430	0.443	0.412	0.444	0.374	0.364
	(0.495)	(0.497)	(0.492)	(0.497)	(0.484)	(0.482)
Prior Chest Pain	0.354	0.368	0.354	0.389	0.289	0.359
	(0.478)	(0.482)	(0.478)	(0.488)	(0.453)	(0.481)
Hypertension	0.349	0.352	0.350	0.364	0.322	0.349
	(0.477)	(0.478)	(0.477)	(0.481)	(0.467)	(0.478)
Prior Ab Pain	0.278	0.291	0.281	0.308	0.249	0.282
	(0.448)	(0.454)	(0.450)	(0.462)	(0.433)	(0.451)
High Cholesterol	0.232	0.233	0.237	0.245	0.202	0.254
	(0.422)	(0.422)	(0.425)	(0.430)	(0.402)	(0.436)
Prior Cardiac	0.199	0.197	0.200	0.218	0.165	0.215
	(0.399)	(0.398)	(0.400)	(0.413)	(0.371)	(0.412)
Metabolic	0.185	0.185	0.173	0.176	0.138	0.211
	(0.389)	(0.389)	(0.378)	(0.381)	(0.345)	(0.409)
N	249342	92829	8559	3647	752	209

Notes: This table reports the average health characteristics of patients seen by control, peer, or treated ED physicians who are either male or female. "Treated" physicians experienced the death of a patient in the ED. "Peer" physicians work in the same ED as a treated physician, but did not themselves

Table 3.6: Characteristics of Male and Female ED Physicians

	Physician Gender	
	(1) Male	(2) Female
Experience:		
# Yrs Postgrad Training	3.317	3.287
	(1.682)	(1.659)
# Postgrad Training Positions	1.594	1.385
	(0.897)	(0.796)
# Yrs Practicing	12.11	9.097
	(9.987)	(8.481)
Specialty:		
Pediatrics	0.133	0.267
	(0.340)	(0.443)
Urology	0.00716	0.00246
	(0.0843)	(0.0496)
Neurology	0.00443	0.00246
	(0.0664)	(0.0496)
Internal Medicine	0.347	0.275
	(0.476)	(0.447)
OB/GYN	0.168	0.217
	(0.374)	(0.413)
Surgery	0.0961	0.0351
	(0.295)	(0.184)
Hematology/Oncology	0.147	0.171
	(0.355)	(0.376)
Emergency Medicine	0.609	0.578
	(0.488)	(0.494)
Psychiatry	0.0174	0.0234
	(0.131)	(0.151)
Cardiology	0.0119	0.00431
	(0.109)	(0.0655)
N	3404	1945

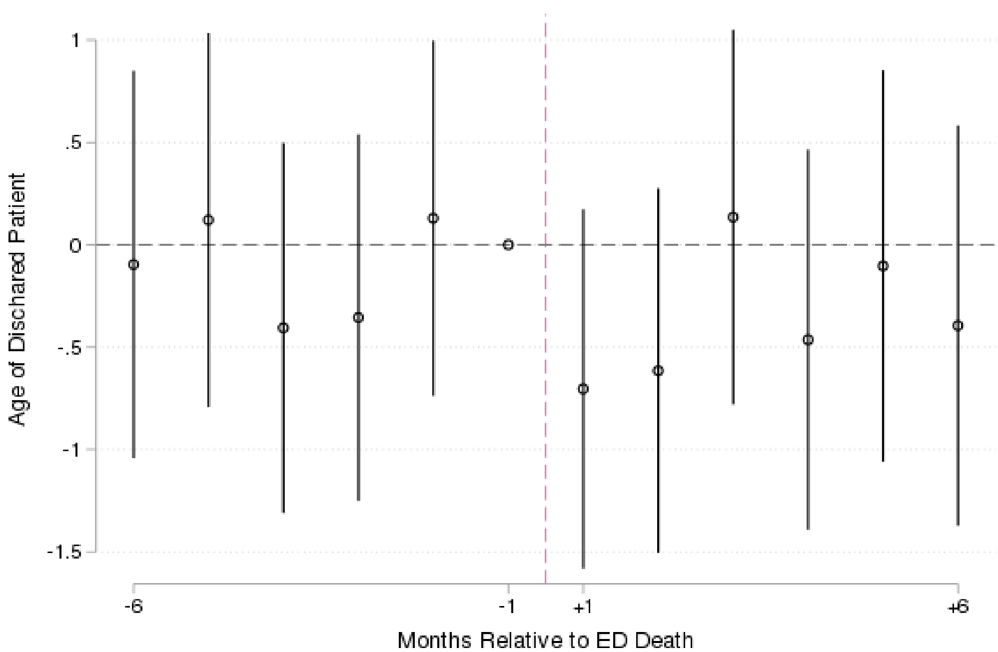
Notes: This table summarizes the average training and specialization characteristics of male and female physicians separately.

treat a patient who died in the ED. All remaining physicians are classified as untreated controls.

$\eta_{m(d)}$ is a set of twelve dummies for the six months prior to, and the six months after, an ED death. ξ captures average differences overall by physician gender. δ, the parameter of interest, captures differences in the characteristics of discharged patients six months before and after an adverse event, for all three types of physicians, by physician gender.

3.5 Results

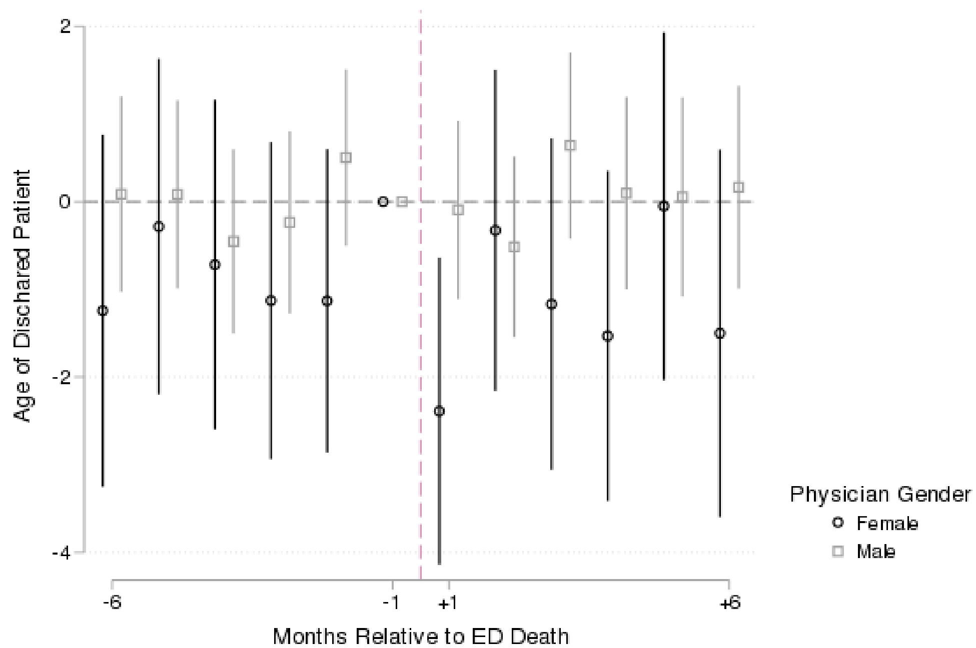
In Figure 3.3, I show that in the months following the in-ED death of a chest pain patient, the average age of a discharged chest pain patient declines from 50 to 45 years old. This affect appears to attenuate in the months following the death, but discharged chest pain patients are still younger several months after the adverse event.



Notes: This figure plots the average age of chest pain patients discharged by physicians who experienced an ED chest pain patient death in the six months prior to, and after, the death. The patient who died, and all other patients treated on the same day, are omitted from the sample.

Figure 3.3: Age of Discharged Patients of Treated Physicians

In Figure 3.4, I plot the relationship between month relative to the adverse event, and the average age of discharged chest pain patients, for male and female physicians separately. Nearly all of the effect observed in Figure 3.3 is driven by female physicians.

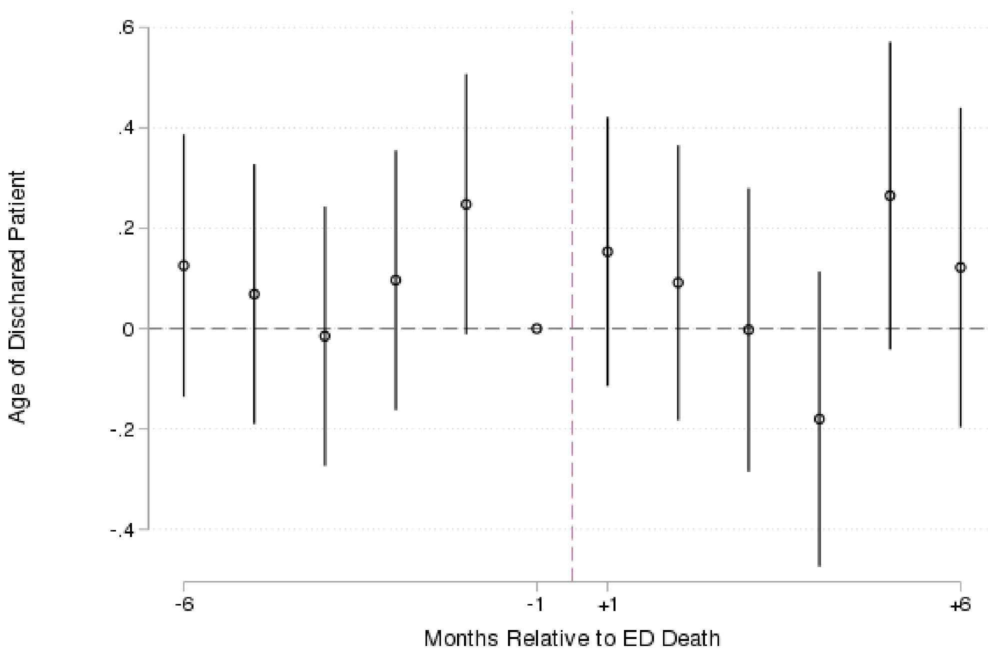


Notes: This figure separately plots the average age of chest pain patients discharged by male and female physicians who experienced an ED chest pain patient death in the six months prior to, and after, the death. The patient who died, and all other patients seen on the same day by the treated physician, are omitted from the sample.

Figure 3.4: Age of Discharged Patients of Treated Physicians by Physician Gender

In Figure 3.5, I plot the relationship between the month relative to an adverse chest pain event and the average characteristics of discharged chest pain patients for the peers of the physician who experienced the adverse event. When aggregated, there appears to be no change in the discharge decisions of peer providers after witnessing an adverse event.

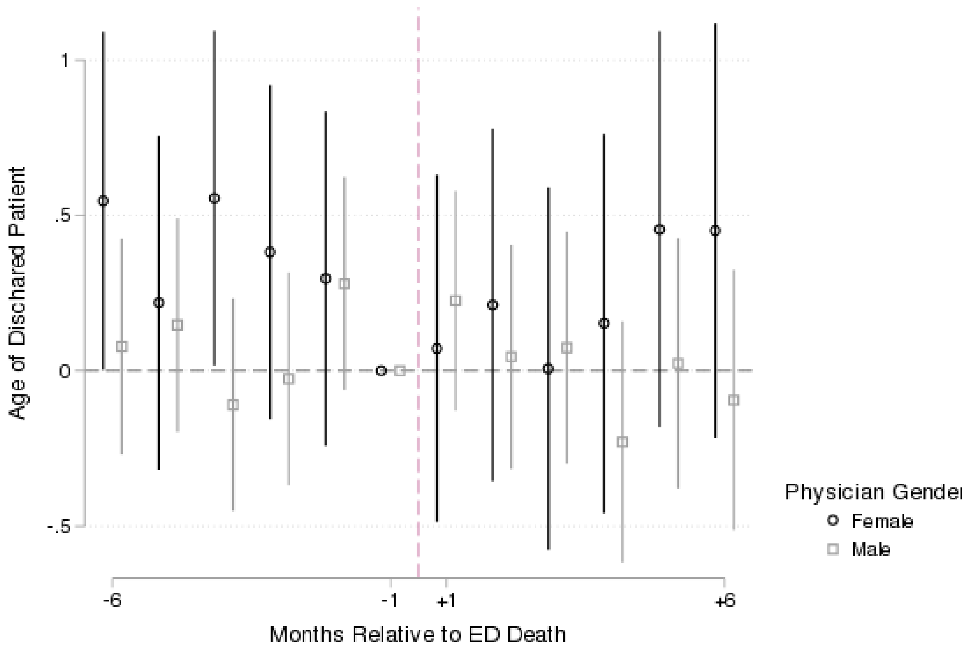
In Figure 3.6, I disaggregate male and female peer physicians. The patients that female physicians choose to discharge after the adverse event are significantly younger, while the patients male physicians choose to discharge are slightly older, indicating that both of the following effects occur after the ED death: 1) reallocation of patients among male and female



Notes: This figure separately plots the average age of chest pain patients discharged by peer physicians - providers who work in an ED that experienced the death of a chest pain patient that they themselves did not treat - in the six months prior to, and after, the death. All patients seen on the same day as the patient who died are omitted from the sample.

Figure 3.5: Spillover Effects on Peer Physicians

members of the ED team and 2) increases in the rate at which female physicians admit patients.



Notes: This figure separately plots the average age of chest pain patients discharged by peer physicians - providers who work in an ED that experienced the death of a chest pain patient that they themselves did not treat - in the six months prior to, and after, the death. The average age is plotted separately for the patients of male and female physicians. All patients seen on the same day as the patient who died are omitted from the sample.

Figure 3.6: Spillover Effects on Peer Physicians by Physician Gender

3.6 Ruling Out Alternative Mechanisms

In order to attribute changes in the composition of discharged patients to changes in provider behavior in reaction to a perceived mistake, I consider empirical strategies for ruling out four alternative mechanisms. First, one can rule out that sorting of patients to facilities is driving changes in the composition of discharged patients by testing that the patients arriving at the ED are similar immediately before and after an adverse event.

Second, I rule out the differential sorting of patients to physicians within a given ED by examining the composition of discharged patients across the entire ED team. If changes in the composition of a treated doctor's discharged patients are being caused by reallocations of the patients across the ED team, these changes should be offset by opposite-signed changes to the rest of the team. If they are not, the changes can be attributed to changes in the treated provider's admission behavior.

Third, I consider whether malpractice claims that have disparate impacts on male versus female physicians could be driving the results. Information on malpractice complaints filed against physicians is maintained by the New York State Department of Health, and cases that make it to court are available through court records. These two sources of data combined can be used to control directly for the effects of malpractice proceedings. However, the delayed nature of these proceedings is unlikely to explain the short-term, dissipating effects of these adverse events. Furthermore, the fact that female physicians are less likely to be the target malpractice suits suggests that this channel does not drive the larger female response to adverse patient events.

Lastly, non-legal, within-hospital consequences could play a role in creating different responses to patient events by physician gender if, for example, supervising doctors are harsher with female physicians than with male physicians after a patient dies. I aim to collect data on how patient deaths are handled internally to quantify the effect of this specific channel.

3.7 Conclusion

I show, using administrative data on high-stakes decisions made in the emergency room, that male and female physicians react differently to similar, rare mistakes. In particular, I focus on how male and female physicians modify their hospital admission behavior after a chest pain patient under their care expires in the emergency room. I find that female physicians become significantly more cautious in the months following a patient death, not just for other patients with chest pains, but for patients with other medical concerns. Male physicians exhibit no change in behavior after the adverse patient event.

These findings suggest that the way mistakes are perceived, communicated, and learned from in the workplace is an important channel through which gender gaps in performance emerge and persist. Policies which help high-stakes decision-makers learn optimally and accurately from these events can not only improve performance, but reduce gender disparities.

References

Abadie, A., Athey, S., Imbens, G. and Wooldridge, J. (2017). *When Should You Adjust Standard Errors for Clustering?* Tech. rep., National Bureau of Economic Research, Cambridge, MA.

Abaluck, J., Agha, L., Kabrhel, C., Raja, A. and Venkatesh, A. (2016). The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care. *American Economic Review*, **106** (12), 3730–3764.

Abraham, L. and Stein, A. (2020). Words Matter: Experimental Evidence from Job Applications.

Albonetti, C. A., Hauser, R. M., Hagan, J. and Nagel, I. H. (1989). Criminal justice decision making as a stratification process: The role of race and stratification resources in pretrial release. *Journal of Quantitative Criminology*, **5** (1), 57–82.

Alexander, D. and Schnell, M. (2019). Just what the nurse practitioner ordered: Independent prescriptive authority and population mental health. *Journal of Health Economics*, **66**, 145–162.

Allegranzi, B. and Pittet, D. (2009). Role of hand hygiene in healthcare-associated infection prevention. *Journal of Hospital Infection*, **73** (4), 305–315.

Alsan, M., Garrick, O. and Graziani, G. (2018). Does Diversity Matter for Health? Experimental Evidence from Oakland.

American College of Emergency Physicians (2011). ED Facility Level Coding Guidelines.

Apgar, V. (1953). A Proposal for a New Method of Evaluation.of the Newborn Infant. *Current Researches in Anesthesia and Analgesia*.

Bailey, M., Cao, R., Kuchler, T. and Stroebel, J. (2018). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, **126** (6), 2224–2276.

Besedeš, T., Deck, C., Sarangi, S. and Shor, M. (2012). Decision-making strategies and performance among seniors. *Journal of Economic Behavior and Organization*, **81** (2), 524–533.

Blau, F. D. and Kahn, L. M. (2007). The gender pay gap: Have women gone as for as they can? *Academy of Management Perspectives*, **21** (1), 7–23.

Bolotnyy, V. and Emanuel, N. (2019). Why Do Women Earn Less Than Men? Evidence from Bus and Train Operators.

Bundorf, M. K., Polyakova, M., Stults, C., Meehan, A., Klimke, R., Pun, T., Chan, A. S. and Tai-Seale, M. (2019). Machine-Based Expert Recommendations And Insurance Choices Among Medicare Part D Enrollees. *Health Affairs*, **38** (3), 482–490.

Burke, G. and Paradise, J. (2015). Safety-Net Emergency Departments: A Look at Current Experiences and Challenges.

Bynum, T. S. and Paternoster, R. (1984). Discrimination revisited: An exploration of frontstage and backstage criminal justice decision making.

Calvet, L. E., Campbell, J. Y. and Sodini, P. (2009). Fight or flight? portfolio rebalancing by individual investors. *Quarterly Journal of Economics*, **124** (1), 301–348.

Camacho, N., Donkers, B. and Stremersch, S. (2011). Predictably Non-Bayesian: Quantifying salience effects in physician learning about drug quality. *Marketing Science*, **30** (2), 305–320.

Carroll, A. (2015). To Be Sued Less, Doctors Should Consider Talking to Patients More - The New York Times.

Centers for Medicaid and Medicare Services (2010). Specifications Manual for Joint Commission National Quality Core Measures (2010A1).

Chan, D. (2016). *Informational Frictions and Practice Variation: Evidence from Physicians in Training*. Tech. rep., National Bureau of Economic Research, Cambridge, MA.

Chang, C. Y. and Obermeyer, Z. (2020). Association of Clinical Characteristics With Variation in Emergency Physician Preferences for Patients. *JAMA network open*, **3** (1), e1919607.

Chen, W., Linthicum, B., Argon, N. T., Bohrmann, T., Lopiano, K., Mehrotra, A., Travers, D. and Ziya, S. (2019). The effects of emergency department crowding on triage and hospital admission decisions. *American Journal of Emergency Medicine*.

Chief Administrator of the Courts (2017). *2017 Annual Report*. Tech. rep., New York State Unified Court System.

Choi, J. J., Laibson, D., Madrian, B. C. and Metrick, A. (2009). Reinforcement learning and savings behavior. *Journal of Finance*, **64** (6), 2515–2534.

Clair, M. and Winter, A. (2016). How Judges Think About Racial Disparities: Situational Decision-Making in the Criminal Justice System. *Criminology*, **54** (2), 332–359.

Coffman, K. B. (2014). Evidence on Self-Stereotyping and the Contribution of Ideas. *The Quarterly Journal of Economics*, **129** (4), 1625–1660.

— and Klinowski, D. (2020). The impact of penalties for wrong answers on the gender gap in test scores. *Proceedings of the National Academy of Sciences*, p. 201920945.

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.

Cutler, D., Skinner, J., Stern, A. D. and Wennberg, D. (2013). *Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending*. Tech. rep., National Bureau of Economic Research, Cambridge, MA.

Danziger, S., Levav, J. and Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America*, **108** (17), 6889–92.

De Vries, E. N., Ramrattan, M. A., Smorenburg, S. M., Gouma, D. J. and Boermeester, M. A. (2008). The incidence and nature of in-hospital adverse events: A systematic review. *Quality and Safety in Health Care*, **17** (3), 216–223.

Deaux, K. and Emswiller, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality and Social Psychology*, **29** (1), 80–85.

— and Farris, E. (1977). Attributing causes for one's own performance: The effects of sex, norms, and outcome. *Journal of Research in Personality*, **11** (1), 59–72.

Dhar, R. (1997). Context and task effects on choice deferral. *Marketing Letters*, **8** (1), 119–130.

Duseja, R., Bardach, N. S., Lin, G. A., Yazdany, J., Dean, M. L., Clay, T. H., Boscardin, W. J. and Dudley, R. A. (2015). Revisit Rates and Associated Costs After an Emergency Department Encounter. *Annals of Internal Medicine*, **162** (11), 750.

Esteves-Sorenson, C. and Snyder, J. (2012). The gender earnings gap for physicians and its increase over time. *Economics Letters*, **116** (1), 37–41.

Etaugh, C. and Brown, B. (1975). Perceiving the causes of success and failure of male and female performers. *Developmental Psychology*, **11** (1), 103.

Exley, C. and Kessler, J. (2019). *The Gender Gap in Self-Promotion*. Tech. rep., National Bureau of Economic Research, Cambridge, MA.

Farley, E. J. (2014). *A Retrospective Analysis of Medical Malpractice Litigation in Three New York Counties*. Tech. rep.

Feather, N. (1968). Change in confidence following success or failure as a predictor of subsequent performance. *Journal of Personality and Social Psychology*, **9** (1), 38–46.

Ferreyra, M. and Kosenok, G. (2011). Learning about New Products: An Empirical Study of Physicians' Behavior. *Economic Inquiry*, **49** (3), 876–898.

Fiedler, M. (2013). *Essays on Provider Behavior in Health Care Markets*. Phd thesis, Harvard University.

Foley, M., Kifaieh, N. and Mallon, W. K. (2011). Financial impact of emergency department crowding. *Western Journal of Emergency Medicine*, **12** (2), 192–197.

Fordyce, J., Blank, F. S. j., Pekow, P., Smithline, H. A., Ritter, G., Gehlbach, S., Benjamin, E. and Henneman, P. L. (2003). Errors in a busy emergency department. *Annals of emergency medicine*, **42** (3), 324–33.

Frakes, M. and Gruber, J. (2019). Defensive Medicine: Evidence from Military Immunity. *American Economic Journal: Economic Policy*, **11** (3), 197–231.

Gallagher, T. H., Mello, M. M., Levinson, W., Wynia, M. K., Sachdeva, A. K., Sulmasy, L. S., Truog, R. D., Conway, J., Mazor, K., Lembitz, A., Bell, S. K., Sokol-Hessner, L., Shapiro, J., Puopolo, A. L. and Arnold, R. (2013). Talking with patients about other clinicians' errors. *New England Journal of Medicine*, **369** (18), 1752–1757.

Ganguli, I., Simpkin, A. L., Lupo, C., Weissman, A., Mainor, A. J., Orav, E. J., Rosenthal, M. B., Colla, C. H. and Sequist, T. D. (2019). Cascades of Care After Incidental Findings in a US National Survey of Physicians. *JAMA Network Open*, **2** (10), e1913325.

Gennaioli, N., Shleifer, A. and Vishny, R. (2015). Neglected risks: The psychology of financial crises. In *American Economic Review*, American Economic Association, vol. 105, pp. 310–314.

Gilboy, N., Tanabe, T., Travers, D. and Rosenau, A. (2011). Emergency Severity Index (ESI): A triage tool for Emergency Department Care. *AHRQ Publication*, **Version 4** (Implementation Handbook 2012 Edition), 69.

Goldberg, R. M., Kuhn, G., Andrew, L. B. and Thomas, H. A. (2002). Coping with medical mistakes and errors in judgment. *Annals of Emergency Medicine*, **39** (3), 287–292.

Gong, Q. (2018). Physician Learning and Treatment Choice: Evidence from Brain Aneurysms.

Gonzalez Morganti, K., Bauhoff, S., Blanchard, J. C., Abir, M., Smith, A., Vesely, J., Okeke, E. N., Kellermann, A. L. and Iyer, N. (2013). *The Evolving Roles of Emergency Departments*. Tech. rep.

Gottfredson, M. and Gottfredson, D. (). *Decision-Making in Criminal Justice: Toward the Rational Exercise of Discretion*.

Grossmann, F. F., Zumbrunn, T., Frauchiger, A., Delport, K., Bingisser, R. and Nickel, C. H. (2012). At risk of undertriage? Testing the performance and accuracy of the emergency severity index in older emergency department patients. *Annals of Emergency Medicine*, **60** (3), 317–325.

Gruber, J., Hoe, T. and Stoye, G. (2018). *Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers*. Tech. rep., National Bureau of Economic Research, Cambridge, MA.

Guardado, J. (2017). *Medical Liability Claim Frequency Among U.S. Physicians*. Tech. rep.

Handel, B. and Schwartzstein, J. (2018). Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care? *Journal of Economic Perspectives*, **32** (1—Winter), 155–178.

Harris, J. E. (1987). Defensive Medicine: It Costs, but Does it Work? *JAMA: The Journal of the American Medical Association*, **257** (20), 2801–2802.

Helo, S. and Moulton, C. A. E. (2017). Complications: Acknowledging, managing, and coping with human error.

Hilfiker, D. (1984). Facing Our Mistakes.

Hinson, J. S., Martinez, D. A., Schmitz, P. S., Toerper, M., Radu, D., Scheulen, J., Stewart de Ramirez, S. A. and Levin, S. (2018). Accuracy of emergency department triage using the Emergency Severity Index and independent predictors of under-triage and over-triage in Brazil: a retrospective cohort analysis. *International Journal of Emergency Medicine*, **11** (1).

Hoffman, M., Kahn, L. B. and Li, D. (2018). Discretion in Hiring. *The Quarterly Journal of Economics*, **133** (2), 765–800.

Hogan, K. A., Fothergill-Bourbonnais, F., Brajtman, S., Phillips, S. and Wilson, K. G. (2016). When Someone Dies in the Emergency Department: Perspectives of Emergency Nurses. *Journal of Emergency Nursing*, **42** (3), 207–212.

Huang, B. (2011). Lightened Scrutiny. *Harv. L. Rev.*, **124**.

Hwang, U., Weber, E. J., Richardson, L. D., Sweet, V., Todd, K., Abraham, G. and Ankel, F. (2011). A research agenda to assure equity during periods of emergency department crowding. *Academic Emergency Medicine*, **18** (12), 1318–1323.

Iyengar, S. S. and Lepper, M. R. (2000a). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, **79** (6), 995–1006.

— and — (2000b). When Choice is Demotivationg: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology*, **79** (6), 995–1006.

Jagsi, R., Griffith, K. A., Stewart, A., Sambuco, D., DeCastro, R. and Ubel, P. A. (2012). Gender differences in the salaries of physician researchers. *JAMA - Journal of the American Medical Association*, **307** (22), 2410–2417.

Jha, A. K., Larizgoitia, I., Audera-Lopez, C., Prasopa-Plaizier, N., Waters, H. and Bates, D. W. (2013). The global burden of unsafe medical care: Analytic modelling of observational studies. *BMJ Quality and Safety*, **22** (10), 809–815.

Johnson, K. D. and Winkelman, C. (2011). The effect of emergency department crowding on patient outcomes: A literature review.

Jones, S. S., Allen, T. L., Flottemesch, T. J. and Welch, S. J. (2006). An Independent Evaluation of Four Quantitative Emergency Department Crowding Scales. *Academic Emergency Medicine*, **13** (11), 1204–1211.

Kaldjian, L. C., Forman-Hoffman, V. L., Jones, E. W., Wu, B. J., Levi, B. H. and Rosenthal, G. E. (2008). Do faculty and resident physicians discuss their medical errors? *Journal of Medical Ethics*, **34** (10), 717–722.

KANZE, D., HUANG, L., CONLEY, M. A. and TORY HIGGINS, E. (2018). We ask men to win and women not to lose: Closing the gender gap in startup funding. *Academy of Management Journal*, **61** (2), 586–614.

KELLER, B. (2015). Inmate. Parolee. Felon. Discuss. *The Marshall Project*.

KHERA, R., WANG, Y., BERNHEIM, S. M., LIN, Z. and KRUMHOLZ, H. M. (2020). Post-discharge acute care and outcomes following readmission reduction initiatives: National retrospective cohort study of Medicare beneficiaries in the United States. *The BMJ*, **368**.

KROLL, L., SINGLETON, A., COLLIER, J. and REES JONES, I. (2008). Learning not to take it seriously: Junior doctors' accounts of error. *Medical Education*, **42** (10), 982–990.

LARSEN, G. Y., MECHAM, N. and GREENBERG, R. (2011). An Emergency Department Septic Shock Protocol and Care Guideline for Children Initiated at Triage. *PEDIATRICS*, **127** (6), e1585–e1592.

LEAPE, L. L. (1994). Error in Medicine. *JAMA: The Journal of the American Medical Association*, **272** (23), 1851.

LECATE, J. (2014). Perception Spillovers Between Bundled Goods: A Model of Physician Learning from Chemotherapy Choice.

LO SASSO, A. T., RICHARDS, M. R., CHOU, C. F. and GERBER, S. E. (2011). The $16,819 pay gap for newly trained physicians: The unexplained trend of men earning more than women. *Health Affairs*, **30** (2), 193–201.

LOXTERKAMP, D. (2014). Staying ahead of getting behind: Reflections on "scarcity". *BMJ (Online)*, **348**.

MA, D. S., CORRELL, J., WITTENBRINK, B., BAR-ANAN, Y., SRIRAM, N. and NOSEK, B. A. (2013). When Fatigue Turns Deadly: The Association Between Fatigue and Racial Bias in the Decision to Shoot. *Basic and Applied Social Psychology*, **35** (6), 515–524.

MAKARY, M. A. and DANIEL, M. (2016). Medical error-the third leading cause of death in the US. *BMJ (Online)*, **353**.

MALMENDIER, U. and NAGEL, S. (2011). Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?*. *The Quarterly Journal of Economics*, **126** (1), 373–416.

MANI, A., MULLAINATHAN, S., SHAFIR, E. and ZHAO, J. (2013). Poverty Impedes Cognitive Function. *Science Magazine*.

MCCARTHY, B. D., BESHANSKY, J. R., D'AGOSTINO, R. B. and SELKER, H. P. (1993). Missed diagnoses of acute myocardial infarction in the emergency department: Results from a multicenter study. *Annals of Emergency Medicine*, **22** (3), 579–582.

MCCOLLISTER, K. E., FRENCH, M. T. and FANG, H. (2010). The cost of crime to society: New crime-specific estimates for policy and program evaluation. *Drug and Alcohol Dependence*, **108** (1-2), 98–109.

Mirilas, P. and Skandalakis, J. E. (2002). Benign anatomical mistakes: incidentaloma. *The American surgeon*, **68** (11), 1026–8.

Molitor, D. (2018). The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration. *American Economic Journal: Economic Policy*, **10** (1), 326–356.

Mullainathan, S. and Obermeyer, Z. (2019). *Who is Tested for Heart Attack and Who Should Be: Predicting Patient Risk and Physician Error*. Tech. rep., National Bureau of Economic Research, Cambridge, MA.

Neprash, H. T. and Barnett, M. L. (2019). Association of Primary Care Clinic Appointment Time With Opioid Prescribing. *JAMA Network Open*, **2** (8), e1910373.

Newman, M. C. (1996). The Emotional Impact of Mistakes on Family Physicians. *Archives of Family Medicine*, **5** (2), 71–75.

NHDOC (). *Prison Release Program Information*. Tech. rep., New Hampshire Department of Corrections.

NY SPARCS (2014a). *SPARCS Inpatient Output Data Dictionary*. Tech. rep., SPARCS.

NY SPARCS (2014b). *SPARCS Outpatient Output Data Dictionary*. Tech. rep., NY SPARCS.

Obermeyer, Z., Cohn, B., Wilson, M., Jena, A. B. and Cutler, D. M. (2017). Early death after discharge from emergency departments: Analysis of national US insurance claims data. *BMJ (Online)*, **356**.

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, **37** (2), 187–204.

Pines, J. M. (2017). What Cognitive Psychology Tells Us About Emergency Department Physician Decision-making and How to Improve It. *Academic Emergency Medicine*, **24** (1), 117–119.

Rao, S. and Nyquist, A.-C. (2018). The Power of the Nudge to Decrease Decision Fatigue and Increase Influenza Vaccination Rates. *JAMA Network Open*, **1** (5), e181754.

Rising, K. L., Victor, T. W., Hollander, J. E. and Carr, B. G. (2014). Patient Returns to the Emergency Department: The Time-to-return Curve. *Academic Emergency Medicine*, **21** (8).

Robertson, J. J. and Long, B. (2018). Suffering in Silence: Medical Error and its Impact on Health Care Providers. *Journal of Emergency Medicine*, **54** (4), 402–409.

Rottman, B. M. (2017). Physician Bayesian updating from personal beliefs about the base rate and likelihood ratio. *Memory and Cognition*, **45** (2), 270–280.

Sabbatini, A. K., Nallamothu, B. K. and Kocher, K. E. (2014). Reducing Variation In Hospital Admissions From The Emergency Department For Low-Mortality Conditions May Produce Savings. *Health Affairs*, **33** (9), 1655–1663.

Saketkhou, B. B., Conte, F. J., Noris, M., Tilkemeier, P., Miller, G., Forman, D. E., Cannistra, L., Leavitt, J., Sharma, S. C., Garber, C. and Parisi, A. F. (1997). Emergency department use of aspirin in patients with possible acute myocardial infarction. *Annals of internal medicine*, **127** (2), 126–9.

Sarsons, H. (2019). Interpreting signals in the labor market: evidence from medical referrals.

Schnell, M. and Currie, J. (2017). *Addressing the Opioid Epidemic: Is There a Role for Physician Education?* Tech. rep., National Bureau of Economic Research, Cambridge, MA.

Shanmugam, P. (2019). Economies of Scale for High-Stakes Choices.

Shoenberger, J. M., Yeghiazarian, S., Rios, C. and Henderson, S. O. (2013). Death notification in the emergency department: Survivors and physicians.

Simon, H. A. (1955). *A Behavioral Model of Rational Choice*. Tech. Rep. 1.

Singh, M. (2020). What Passed Is Past? The Role of Recent Adverse Events in Physician Treatment Decisions.

Sinitsky, D. M., Gowda, S. B., Dawas, K. and Fernando, B. S. (2019). Morbidity and mortality meetings to improve patient safety: A survey of 109 consultant surgeons in London, United Kingdom. *Patient Safety in Surgery*, **13** (1).

Smythe, J. G. (2002). Reputation, public information, and physician adoption of an innovation. *European Journal of Health Economics*, **3** (2), 103–110.

Thakral, N. and To, L. (2018). Daily Labor Supply and Adaptive Reference Points. *Working Paper*.

Tintinalli, J. E., Stapczynski, J. S., Ma, O. J., Cline, D., Cydulka, R. and Meckler, G. (2011). *Tintinalli's emergency medicine : a comprehensive study guide*. McGraw-Hill, 7th edn.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185** (4157), 1124–1131.

Venkatesh, A. K., Agha, L., Abaluck, J., Rothenberg, C., Kabrhel, C. and Raja, A. S. (2018). Trends and Variation in the Utilization and Diagnostic Yield of Chest Imaging for Medicare Patients With Suspected Pulmonary Embolism in the Emergency Department. *American Journal of Roentgenology*, **210** (3), 572–577.

Vigil, J. M., Alcock, J., Coulombe, P., McPherson, L., Parshall, M., Murata, A. and Brislen, H. (2015). Ethnic Disparities in Emergency Severity Index Scores among U.S. Veteran's Affairs Emergency Department Patients. *PLOS ONE*, **10** (5), e0126792.

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M. and Tice, D. M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology*, **94** (5), 883–898.

Wadhera, R. K., Maddox, K. E., Kazi, D. S., Shen, C. and Yeh, R. W. (2019). Hospital revisits within 30 days after discharge for medical conditions targeted by the Hospital Readmissions Reduction Program in the United States: National retrospective analysis. *The BMJ*, **366**.

Wears, R. L. and Wu, A. W. (2002). Dealing with failure: The aftermath of errors and adverse events. *Annals of Emergency Medicine*, **39** (3), 344–346.

Wennberg, J. E. (2002). Unwarranted variations in healthcare delivery: implications for academic medical centres. *BMJ (Clinical research ed.)*, **325** (7370), 961–4.

Woodworth, L. (2019). Swamped: Emergency department crowding and patient mortality. *Journal of Health Economics*, p. 102279.

Yang, C. S. (2015a). Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing. *The Journal of Legal Studies*, **44** (1), 75–111.

— (2015b). Free at Last? Judicial discretion and racial disparities in federal sentencing. *Journal of Legal Studies*, **44** (1), 75–111.

—, Cohen, A., Dobbie, W., Farber, H., Kaplow, L., Katz, L., Malani, A., Miles, T., Morrison, E., Rappaport, J., Sakoda, R., Shavell, S., Spamann, H., Spier, K. and Starr, S. (2016). Resource Constraints and the Criminal Justice System: Evidence from Judicial VacancieS. *American Economic Journal: Economic Policy*, **8** (4), 289–332.

Yee, P. (2002). Brain cramp: The emergency physician's worst nightmare. *Annals of Emergency Medicine*, **39** (3), 329–330.

Zeltzer, D. (2020). Gender Homophily in Referral Networks: Consequences for the Medicare Physician Earnings Gap. *American Economic Journal: Applied Economics*, **12** (2), 169–197.

Zibulewsky, J. (2001). The Emergency Medical Treatment and Active Labor Act (Emtala): What It Is and What It Means for Physicians. *Baylor University Medical Center Proceedings*, **14** (4), 339–346.

Appendix A

Appendix to Chapter 1

A.1 ED Vignette

A 21-year-old college freshman is brought to the emergency room by his roommate. The roommate reports that the patient has not left his dormitory room for 2 weeks and his room is in disarray. He describes the patient as being "normal" until about 3-4 months ago. He states that he noticed that the patient stopped going to social activities and spent most of his time in his room. He also states that the patient sometimes makes odd comments. He has stopped going to his classes and his grades have been declining. He also states that for about a week the patient has stopped eating and drinks only canned beverages and insists on keeping the shades down on the windows and has said that he is worried that someone is watching him. The patient denies using alcohol or any illicit drugs. His temp is 37 F, pulse is 92/min, and blood pressure is 140/80 mm Hg. On mental status exam he appears distracted and repeatedly stops answering your questions mid-sentence. He describes hearing two voices that are telling him to stop eating. He is oriented to place, person and time.

Which is the most likely diagnosis?

1. Anorexia Nervosa

2. Delirium

3. Delusional Disorder

4. Major Depressive Disorder with psychotic features

5. Schizophrenia

Which is the most appropriate pharmacotherapy?

1. Buspirone

2. Clonazepam

3. Clozapine

4. Sertraline

5. Risperidone

A.2 SPARCS Provider IDs

SPARCS data provides up to three "Provider IDs" per visit record. The Provider ID is the healthcare provider's six-digit sate license number, assigned by the New York State Department of Education upon successful board completion.

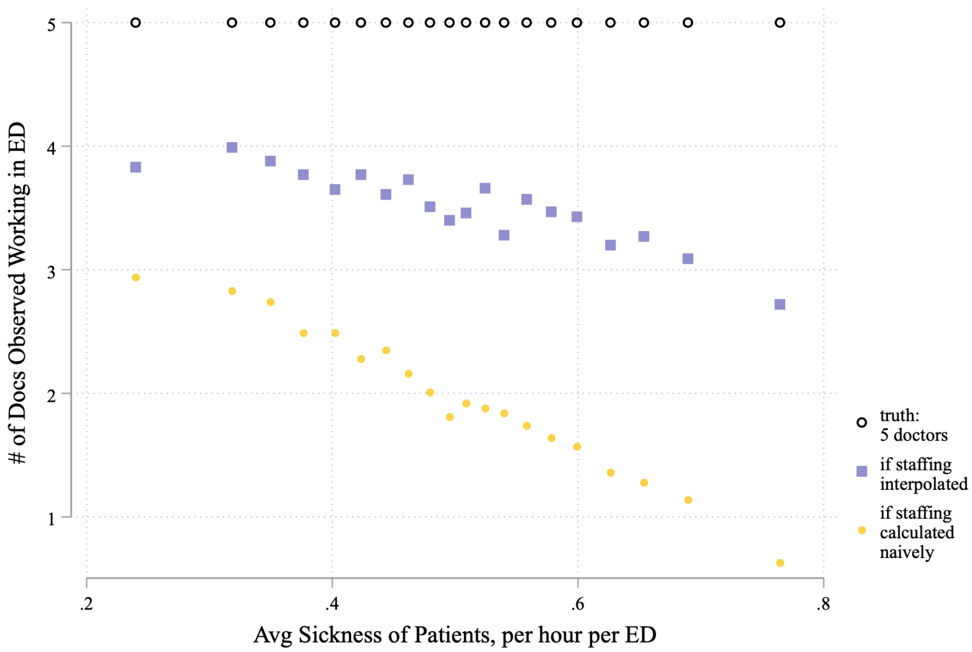
- Attending Provider: Used to identify the physician or other health care professional primarily responsible for the care of the patient. In some instances the health facility's policy may dictate that an Attending Provider or chief of service may be assigned to any number of patients who may not have a primary care giver.
- Operating Provider: Used to identify the physician or other health care professional who performed the principal procedure. Hospital policy may dictate which physician license number will be used for this data element.
- Other Provider: Used to identify the physician or other health care professional who was involved in the patient's care or treatment (i.e., consulting physician, second operating physician, nurse/midwife, etc.)

Each Provider ID also contains a two-digit "license type" prefix, which broadly categorizes the healthcare professionals associated with the visit.

Code(s)	Provider Type
00	Physician
10	Dentist
20	Podiatrist
30, 40	Limited Permit "L" or "P"
50	Psychologist
60	Nurse/Midwife
90	Other Licensed Healthcare Professional

A.3 Measuring ED Doctor Shifts

When a patient being seen in the ED is admitted into the hospital, the doctor IDs that appear on the visit record only reflect the hospital doctors associated with the patient, and not the ED physicians who saw the patient prior to admission. The omission of ED doctor IDs from records in which a patient is admitted into the hospital create the appearance of a relationship between a patient's potential health and ED staffing patterns, even when there is no actual relationship between these two variables.



Notes: This figure shows the relationship between the average sickness of patients visiting the ED in a given hour and the number of doctors that appear to be working in the ED at the time, using simulated data. The true number of ED doctors is 5; however, in an hour where patients are on average sicker due to random variation, fewer doctors appear to be working in the ED due to the deletion of ED doctor IDs for admitted patients. The problem is only partially alleviated when ED doctor working schedules are interpolated.

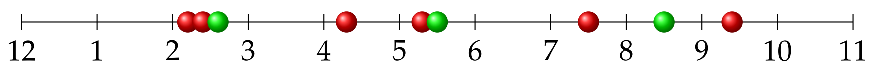
Figure A.1: Patient Sickness and (Apparent) ED Staffing

For this reason, I do not scale ED crowding by the number of ED doctors who appear to be working at a given time. I discuss three other potential issues related to constructing doctor shifts using timestamped doctor IDs.

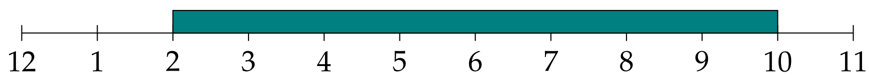
ED Doctors. Most doctors who work primarily in the Emergency Department arrive will be observed in the data as having been assigned to patients randomly throughout their shift. If a physician is not assigned to any patients for six or more hours, I mark this as the end of a shift.

= Discharged = Admitted

Observed data:

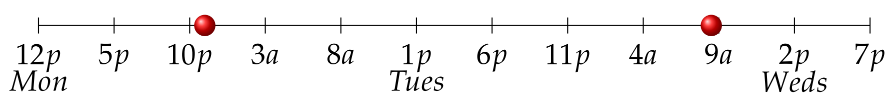


Interpolated Schedule:

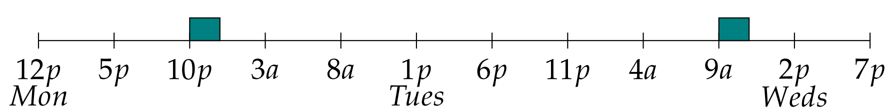


Doctors Who Consult in the ED. Some doctors are asked to consult on complex cases in the Emergency Department from other parts of the hospital. These doctors are not ED doctors and do not participate in the repeated decision-making setting I aim to isolate. In the data, these doctors will appear to handle ED cases at random intervals with interpolated “shifts" of an hour or less. I remove these doctors from the sample by subsetting to shifts of greater than five hours in length.

Observed data:

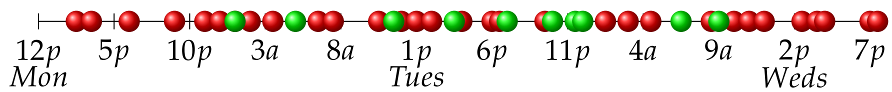


Interpolated Schedule:

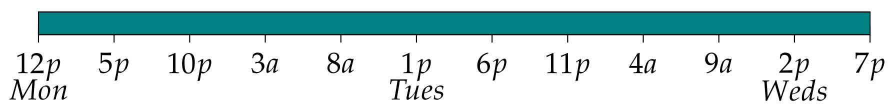


ED Supervisors Instead of ED Doctors. In a handful of facilities, the ED supervisor will be listed as the primary physician. This person oversees or is accountable for the entire ED, but is not making decisions on cases individually. Such a doctor will appear to be on every single record in the entire ED, implying the doctor has worked hundreds or thousands of hours without rest. I remove these false-positive work hours by limiting my sample to shifts of thirty hours or less.

Full schedule:



Interpolated Schedule:



A.4 ED Service Levels

Medical billing codes are used to indicate the intensity of ED services utilized during a patient's visit. A medical coding specialist takes all of the information from the visit and assigns an "ED Service Level", 1 through 5, for each visit. Below are the exact definitions of Level 1 and Level 5 ED visits (American College of Emergency Physicians, 2011).

Level 1:

- A problem-centric history.
- A problem-centric exam.
- Very simple medical decision making.
- Low urgency or self-treatable symptoms, requiring little to no immediate medical care

Level 5:

- A high detailed problem-centric history.
- A high detailed problem-centric exam.
- Highly complex medical decision making.
- Symptoms of highest severity, posing an immediate significant threat to life or physiologic function

A.5 Average Rates of Billed Procedures

Table A.1: Average Rates of Billed Procedures

	(1) Chest Pain	(2) Chest Pain, Other	(3) Ab Pain	(4) Ab Pain, Other
Any Medication	34.07	32.27	43.33	40.58
	(47.39)	(46.75)	(49.55)	(49.10)
Any Lab Testing	77.33	80.06	82.32	83.95
	(41.87)	(39.96)	(38.15)	(36.70)
X-Ray	92.09	89.84	37.10	28.97
	(53.62)	(49.14)	(63.32)	(57.85)
CT Scan	21.46	21.16	53.60	60.49
	(54.58)	(51.76)	(76.17)	(79.15)
Cardiology	31.53	35.97	1.660	0.900
	(75.43)	(77.55)	(14.73)	(10.83)
Electrocardiogram	94.99	94.51	18.87	13.70
(EKG)	(62.68)	(64.04)	(41.75)	(37.22)
Ultrasound	3.617	3.824	23.12	23.22
	(19.14)	(19.55)	(51.49)	(53.70)
Admitted	24.69	33.37	13.76	7.516
	(43.12)	(47.15)	(34.45)	(26.37)
Observations	2441502	963370	2416394	1425590

Notes: This table provides summary statistics for binary indicators of whether or not various procedures were billed as part of the index ED visit, for selected chief complaints. Whether or not a line-item appears on the bill is independent of who pays for the visit. One observation represents one complete ED visit, including any procedures givenafter a patient is admitted from the ED into the hospital for inpatient care. Appendix A.6 describes the NUBC Revenue Codes, from which these procedure indicators are derived, in greater detail.

A.6 NUBC Revenue Codes

I create broad categories for whether or not a patient was billed for a certain service based on whether or not a line-item for that service appears on the patient's record. For example, I assign a patient as having received a medication if they were billed for revenue code 025X (Pharmacy). This code includes the following sub-codes:

- 0250: General
- 0251: Generic Drugs
- 0252: Nongeneric Drugs
- 0253: Take-home Drugs
- 0254: Drugs incident to Other diagnostic services
- 0255: Drugs incident to radiology
- 0256: Experimental drugs
- 0257: Nonprescription
- 0258: IV solutions
- 0259: Other

Listed below are the NUBC revenue codes used to construct various billed procedure indicators:

- Coronary care: 021X
- IV: 026X
- X-Ray: 032X
  - Chest X-Ray: 0324
  - Coronary Angiogram: 0321

  - Coronary Arteriogram: 0323
- Ultrasound: 0402
- Pharmacy: 025X, 063X
- Lab Test: 030X, 031X
- CT Scan: 035X
- Cardiology Unit: 048X
  - Cardiac Catheterization Lab: 0481
  - Cardiac Stress Test: 0482
  - Echocardiogram: 0483
- EKG: 073X
- MRI: 061X
- Nuclear Medicine: 0341, 0343
- Blood: 038X
- Pulmonary Unit: 046X
- Cardiac Rehab: 0943
- GI Unit: 075X
- Preventive Services: 077X

A full list of NUBC Revenue Codes is available at
`https://med.noridianmedicare.com/web/jea/topics/claim-submission/revenue-codes`

A.7 Average Intensity of Diagnostic and Therapeutic Care

Table A.2: Average Intensity of Diagnostic and Therapeutic Care

	(1) Chest Pain	(2) Chest Pain, Other	(3) Ab Pain	(4) Ab Pain, Other
N Procedures	3.654	3.389	3.947	4.273
	(2.812)	(2.829)	(2.704)	(2.567)
N Diagnostics	3.098	2.859	3.341	3.671
	(2.432)	(2.448)	(2.411)	(2.270)
N Therapeutics	0.506	0.478	0.596	0.593
	(0.744)	(0.743)	(0.696)	(0.615)
Total Charges	7991.1	9361.9	6614.2	5046.4
($)	(19054.3)	(21494.6)	(18890.0)	(14320.6)
N Anc Diagnoses	3.246	3.505	2.002	1.840
	(3.819)	(3.884)	(3.128)	(2.674)
No New Diagnosis	37.06	61.63	33.56	36.58
	(48.30)	(48.63)	(47.22)	(48.17)
Observations	2441502	963370	2416394	1425590

Notes: This table summarizes the average diagnostic and therapeutic intensity of care received by patients presenting in the ED with selected chief complaints. One observation represents one ED visit. The number of diagnostic tests and therapeutic treatments is calculated by summing the number of procedures that appear on a patient's record that are classified as either "diagnostic" or "therapeutic" by the HCUP Procedure Classes tool, which is described in Appendix A.8. The table summarizes the number of diagnostics (tests), therapeutics (treatments), number of new diagnoses, whether or not the patient received a diagnosis that was different from their chief complaint, and the total charges for the visit.

A.8 Converting ICD-9-CM and CPT Codes to Clinical Categories

I use the Healthcare Cost and Utilization Project (HCUP) Clinical Categorization Software to convert over 14,000 ICD-9 procedure codes, as well as thousands of HCPCS-CPT service codes, into broad, comparable, clinically relevant categories.

For example, the ICD-9-CM contains separate codes to define a malignant tumor in nine different parts of the stomach:

- malignant neoplams of the cardia
- pylorus
- pyloric antrum
- fundus of stomach
- body of stomach
- lesser curvature of stomach (unspecified)
- greater curvature of stomach (unspecified)
- other specified site of stomach
- other stomach site (unspecified)

The CCS simply groups these categories together as “stomach cancer" for the purposes of empirical analysis. Complete crosswalks between CPT, ICD-9-CM, and CCS codes are available at

`https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp`

A.9 Common Primary Diagnoses for Chest Pain Patients

Table A.3: Common Primary Diagnoses for Chest Pain Patients

	(1) Pct
Chest Pain	0.35
Chest Pain, other	0.19
Painful Respiration	0.04
Coronary Atherosclerosis	0.04
Contusion of Chest Wall	0.03
Non ST Elevated Heart Attack	0.02
Tietze Syndrome	0.02
Unstable Angina	0.01
GERD	0.01
Pneumonia	0.01

Notes: This table lists the most common primary diagnoses given to patients and their frequencies for patients arriving in the ED with a chief complaint of chest pains. The primary diagnosis is the doctor's conclusion as to the main cause of the patient's initial chief complaint. Each ED visit is given one primary diagnosis. Diagnoses are recorded using ICD-9-CM codes.

A.10 Common Primary Diagnoses for Abdominal Pain Patients

Table A.4: Common Primary Diagnoses for Abdominal Pain Patients

	(1) Pct
Abdominal pain	0.38
Gastroenteritis and Colitis	0.05
Constipation	0.04
Calculus of kidney	0.03
Abdominal pain, other	0.03
Gastritis	0.03
Unspecified renal colic	0.02
Calculus of ureter	0.02
Diverticulitis	0.02
Acute Pancreatitis	0.02

Notes: This table lists the most common primary diagnoses given to patients and their frequencies for patients arriving in the ED with a chief complaint of abdominal pains. The primary diagnosis is the doctor's conclusion as to the main cause of the patient's initial chief complaint. Each ED visit is given one primary diagnosis. Diagnoses are recorded using ICD-9-CM codes.

A.11 Patient Demographics

Table A.5: Patient Demographics

	(1) Chest Pain	(2) Chest Pain, Other	(3) Ab Pain	(4) Ab Pain, Other
Demographics:				
Age (yrs)	49.01	50.04	38.17	38.91
	(19.84)	(19.19)	(20.81)	(19.31)
Male	0.479	0.466	0.339	0.351
	(0.500)	(0.499)	(0.473)	(0.477)
Race:				
White	0.573	0.506	0.541	0.540
	(0.495)	(0.500)	(0.498)	(0.498)
Black	0.230	0.262	0.238	0.205
	(0.421)	(0.440)	(0.426)	(0.404)
Native American	0.00431	0.00472	0.00416	0.00433
	(0.0655)	(0.0686)	(0.0644)	(0.0656)
Asian / Pac Island	0.0208	0.0212	0.0219	0.0222
	(0.143)	(0.144)	(0.146)	(0.147)
Multiracial / Unknown	0.172	0.205	0.195	0.228
	(0.378)	(0.404)	(0.396)	(0.420)
Insurance Type:				
Private	0.495	0.517	0.490	0.524
	(0.500)	(0.500)	(0.500)	(0.499)
Medicare	0.232	0.239	0.140	0.121
	(0.422)	(0.426)	(0.347)	(0.326)
Medicaid	0.223	0.252	0.233	0.217
	(0.416)	(0.434)	(0.423)	(0.412)
Other Insurance	0.0561	0.0614	0.0610	0.0529
	(0.230)	(0.240)	(0.239)	(0.224)
Uninsured	0.222	0.215	0.252	0.238
	(0.415)	(0.411)	(0.434)	(0.426)
Observations	2441502	963370	2416394	1425590

Notes: This table reports the average patient characteristics for a set of chief complaints. Frequencies by race sum to 1, as each patient can report one race. One observation represents one ED visit, and patients can visit the ED multiple times in my sample. Insurance types sum to greater than 1 as each visit indicates up to six forms of payment.

A.12 Forms of Payment

SPARCS provides up to six fields in which methods of payment for the ED visit are recorded. Methods of payment are recorded using the following codes.

Code	Form of Payment
A	Self-Pay
B	Workers' Compensation
C	Medicare
D	Other Federal Program
E	Medicaid
F	Insurance Company
G	Blue Cross
H	CHAMPUS
I	Other Non-Federal Program
J	Disability
K	Title V
L	Unknown

I create the following broader categories of patient insurance type based on the presence of certain codes across all six Forms of Payment fields.

Code(s)	Constructed "Insurance Type"
C	Medicare
E	Medicaid
G, F	Private Insurance
B, H, I, J, K, L	Other Insurance
A only	No Insurance

A.13 Characterizing Ex-Ante Patient Health

Table A.6: Characterizing Ex-Ante Patient Health

	(1)	(2)	(3)	(4)
	Chest Pain	Chest Pain, Other	Ab Pain	Ab Pain, Other
Cancer	0.0751	0.0761	0.0738	0.0686
	(0.264)	(0.265)	(0.262)	(0.253)
Metabolic	0.182	0.184	0.189	0.170
	(0.386)	(0.387)	(0.392)	(0.375)
Hypertension	0.417	0.435	0.283	0.279
	(0.493)	(0.496)	(0.450)	(0.448)
High Cholesterol	0.299	0.309	0.178	0.180
	(0.458)	(0.462)	(0.383)	(0.385)
Thyroid	0.0976	0.0973	0.0755	0.0740
	(0.297)	(0.296)	(0.264)	(0.262)
Diabetes	0.202	0.214	0.140	0.136
	(0.402)	(0.410)	(0.347)	(0.343)
Endocrine	0.0294	0.0305	0.0304	0.0286
	(0.169)	(0.172)	(0.172)	(0.167)
Obesity	0.128	0.133	0.106	0.108
	(0.334)	(0.340)	(0.307)	(0.311)
Prior Heart	0.0596	0.0591	0.0182	0.0178
Attack	(0.237)	(0.236)	(0.134)	(0.132)
Anemia	0.180	0.186	0.165	0.155

(continued on next page)

(continued from previous page)

	(0.384)	(0.389)	(0.371)	(0.362)
White Blood Cell	0.0386	0.0403	0.0406	0.0368
	(0.193)	(0.197)	(0.197)	(0.188)
Dementia	0.00993	0.0108	0.00704	0.00613
	(0.0991)	(0.103)	(0.0836)	(0.0780)
Mental Health	0.105	0.104	0.102	0.0927
	(0.307)	(0.306)	(0.302)	(0.290)
Alcohol Dependence	0.0635	0.0653	0.0492	0.0412
	(0.244)	(0.247)	(0.216)	(0.199)
Drug Dependence	0.0633	0.0659	0.0598	0.0519
	(0.243)	(0.248)	(0.237)	(0.222)
Multiple Sclerosis	0.00439	0.00424	0.00369	0.00376
	(0.0661)	(0.0650)	(0.0606)	(0.0612)
Respiratory	0.435	0.438	0.419	0.416
	(0.496)	(0.496)	(0.493)	(0.493)
Digestive	0.480	0.485	0.500	0.485
	(0.500)	(0.500)	(0.500)	(0.500)
Prior Cardiac	0.256	0.258	0.150	0.144
	(0.436)	(0.438)	(0.357)	(0.351)
Prior Chest Pains	0.431	0.442	0.307	0.314
	(0.495)	(0.497)	(0.461)	(0.464)
Prior Abdominal	0.248	0.249	0.362	0.359

(continued on next page)

			(continued from previous page)	
Pains	(0.432)	(0.433)	(0.480)	(0.480)
Nutrition	0.0380	0.0401	0.0400	0.0351
	(0.191)	(0.196)	(0.196)	(0.184)
Kidney	0.0913	0.0958	0.0655	0.0599
	(0.288)	(0.294)	(0.247)	(0.237)
Observations	2441502	963370	2416394	1425590

Notes: This table reports the summary statistics of prior health descriptors for a set of chief complaints. One observation represents one ED visit. Note that although some health conditions are transient, the indicators described above are ever or never indicators. That is, a patient is recorded as having a given health condition if they have ever been given a diagnosis code for that condition in any visit in my sample that occurred prior to their index visit.

A.14 Decision-Making Aids

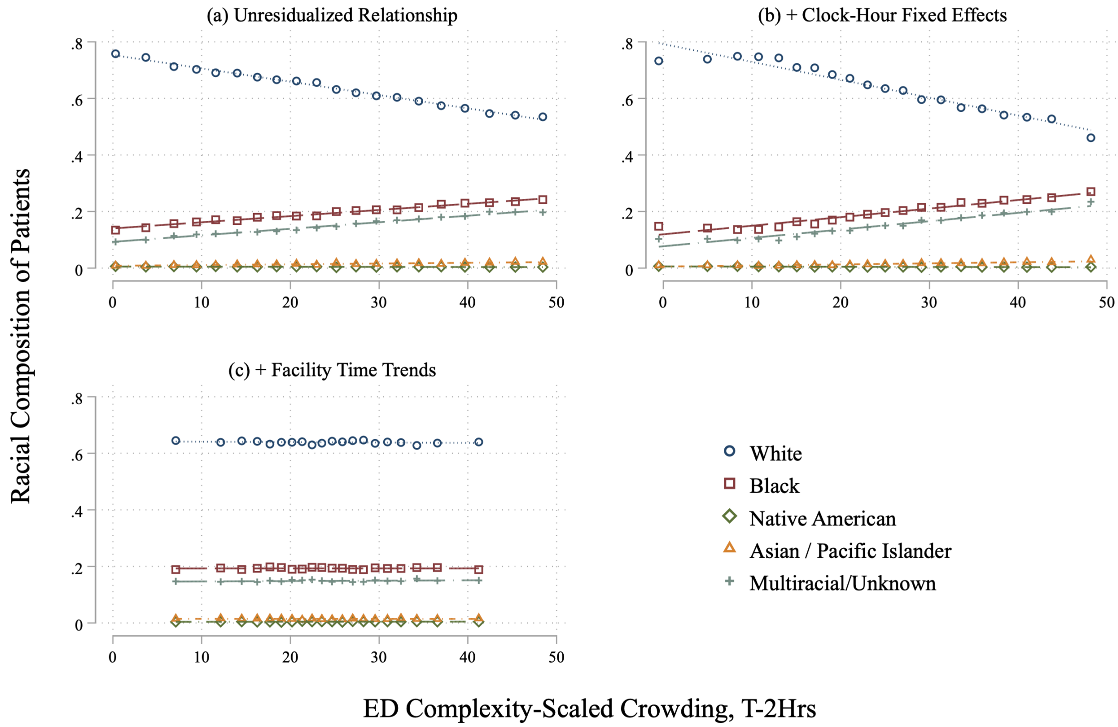
Several decision-making aids have been created in the last decade to assist doctors in identifying patients who present in the ED with chest pains who may be at an elevated risk for a future adverse health event. Below is a detailed explanation of the HEART Score, the most recent of these aids.

HEART Score. Five key components make up the HEART Score diagnostic checklist: **H**istory, **E**KG, **A**ge, **R**isk factors, and **T**roponin. The HEART Score was developed in 2008 using a cohort of only 122 patients presenting with chest pain. The Score was designed to predict four Major Adverse Cardiac Events (MACE) endpoints: acute myocardial infarction (AMI), percutaneous coronary intervention (PCI), coronary artery bypass graft (CABG), and death. The HEART Score has since been validated across various patient populations. Patients identified as low-risk have a $<$2% risk of a MACE within one year of their index visit. Moderate risk patients have a 12% risk, and high-risk patients have over a 60% risk.

- **H**istory: the doctor surveys the patient's health history, including previous cardiac issues or ongoing chest pains. 1 or 2 points are given for moderate or highly suspicious health histories, respectively.
- **E**KG: 1 or 2 points are given for EKG readings that suggest non-specific repolarization disturbance or significant ST depression, respectively.
- **A**ge: 1 or 2 points are given for patients at or over 45 and 65 years respectively.
- **R**isk factors: 1 or 2 points are given for 1-2 or 3+ Major Adverse Cardiac Event (MACE) risk factors: hypertension, high cholesterol, diabetes, obesity, past or current smoking, family history, prior heart attack, atherosclerotic disease, arterial disease.
- **T**roponin: 1 or 2 points are given for troponin levels that are 1-3x, or greater than 3x, the normal limit.

For scores less than 3, discharge is recommended. Scores between 3 and 12 suggest admission. Scores greater than 12 suggest admission and early invasive measures.

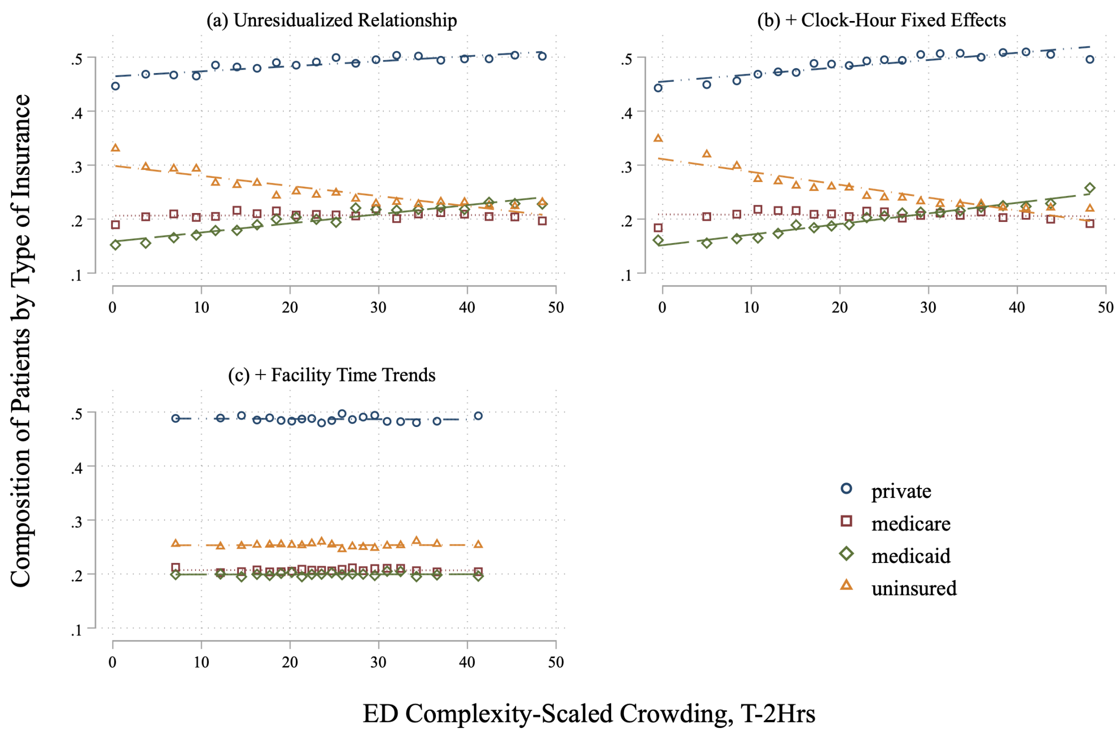
A.15 Patient Race and ED Crowding



Notes: This figure shows the nonparametric relationship between ED complexity-scaled crowding and patient race. One observation represents one ED visit. Panel (a) shows the raw relationship between this measure of crowding and the composition of patients by race. Panel (b) shows the same relationship after the variables have been residualized on fixed-effects for each hour of the day. Panel (c) shows the relationship after further residualizing on facility-by-year fixed-effects (facility-specific time trends, in effect.)

Figure A.2: PATIENT RACE AND ED CROWDING

A.16 Patient Insurance and ED Crowding



Notes: This figure shows the nonparametric relationship between ED complexity-scaled crowding and patient insurance type. One observation represents one ED visit. Panel (a) shows the raw relationship between this measure of crowding and the composition of patients by type of insurance. Panel (b) shows the same relationship after the variables have been residualized on fixed-effects for each hour of the day. Panel (c) shows the relationship after further residualizing on facility-by-year fixed-effects. Insurance types may sum to greater than 1, as each visit indicates up to six forms of payment.

Figure A.3: Patient Insurance and ED Crowding

A.17 Construction of Risk Scores

Table A.7: Construction of Risk Scores

	1Yr Death	Admission	Charges ($)	30d Revisit	30d Death
Insurance Type:					
Private	0.553***	15.50***	4113.2***	-0.0186***	0.181*
	(0.151)	(0.294)	(131.8)	(0.00234)	(0.0800)
Medicaid	0.935***	17.42***	4586.8***	0.00848***	0.278***
	(0.150)	(0.291)	(130.7)	(0.00232)	(0.0793)
Medicare	0.260	17.49***	4881.9***	-0.0126***	0.189*
	(0.146)	(0.284)	(127.3)	(0.00226)	(0.0772)
Other Insurance	0.499	16.95***	5841.7***	-0.0156***	0.232
	(0.257)	(0.500)	(224.1)	(0.00399)	(0.136)
Uninsured	-0.238	-10.67***	669.5***	-0.00570	-0.221*
	(0.195)	(0.379)	(170.0)	(0.00302)	(0.103)
Race:					
White	0.298*	-0.138	359.8**	0.0149***	0.116
	(0.140)	(0.271)	(121.8)	(0.00217)	(0.0739)
Black	0.122	-1.005***	-12.05	0.0124***	0.000911
	(0.151)	(0.292)	(131.1)	(0.00233)	(0.0795)
Native American	-0.895	-0.178	107.1	0.00264	-0.396
	(0.604)	(1.173)	(526.2)	(0.00936)	(0.319)
Asian / Pac	0.250	2.319***	1375.6***	-0.00749	0.256
Island	(0.348)	(0.676)	(303.3)	(0.00539)	(0.184)

(continued on next page)

(continued from previous page)

Health Factors:					
Cancer	7.594***	0.501	713.0***	-0.00262	1.935***
	(0.155)	(0.301)	(135.3)	(0.00241)	(0.0820)
Metabolic	1.444***	1.742***	462.7***	0.0339***	0.392***
	(0.122)	(0.236)	(106.1)	(0.00189)	(0.0644)
Hypertension	-0.731***	1.100***	243.3*	0.00370*	-0.180**
	(0.115)	(0.223)	(100.1)	(0.00178)	(0.0607)
High Cholesterol	-1.449***	-0.123	-113.0	-0.00116	-0.495***
	(0.122)	(0.236)	(106.0)	(0.00189)	(0.0643)
Thyroid	-0.485***	-1.434***	-397.7**	0.000270	-0.0692
	(0.143)	(0.277)	(124.5)	(0.00221)	(0.0755)
Diabetes	0.859***	2.598***	467.7***	0.0107***	0.221***
	(0.123)	(0.239)	(107.0)	(0.00190)	(0.0649)
Endocrine	0.296	-1.139*	93.39	0.0389***	0.0395
	(0.234)	(0.453)	(203.4)	(0.00362)	(0.123)
Obesity	-0.521***	1.469***	469.1***	0.0151***	-0.236***
	(0.134)	(0.260)	(116.5)	(0.00207)	(0.0707)
Prior Heart	3.022***	4.409***	1241.5***	0.0102**	0.900***
Attack	(0.210)	(0.409)	(183.3)	(0.00326)	(0.111)
Anemia	1.977***	1.878***	822.3***	0.0167***	0.385***
	(0.126)	(0.244)	(109.5)	(0.00195)	(0.0664)
White Blood Cell	0.713***	2.019***	1016.4***	0.0546***	0.158

(continued on next page)

(continued from previous page)

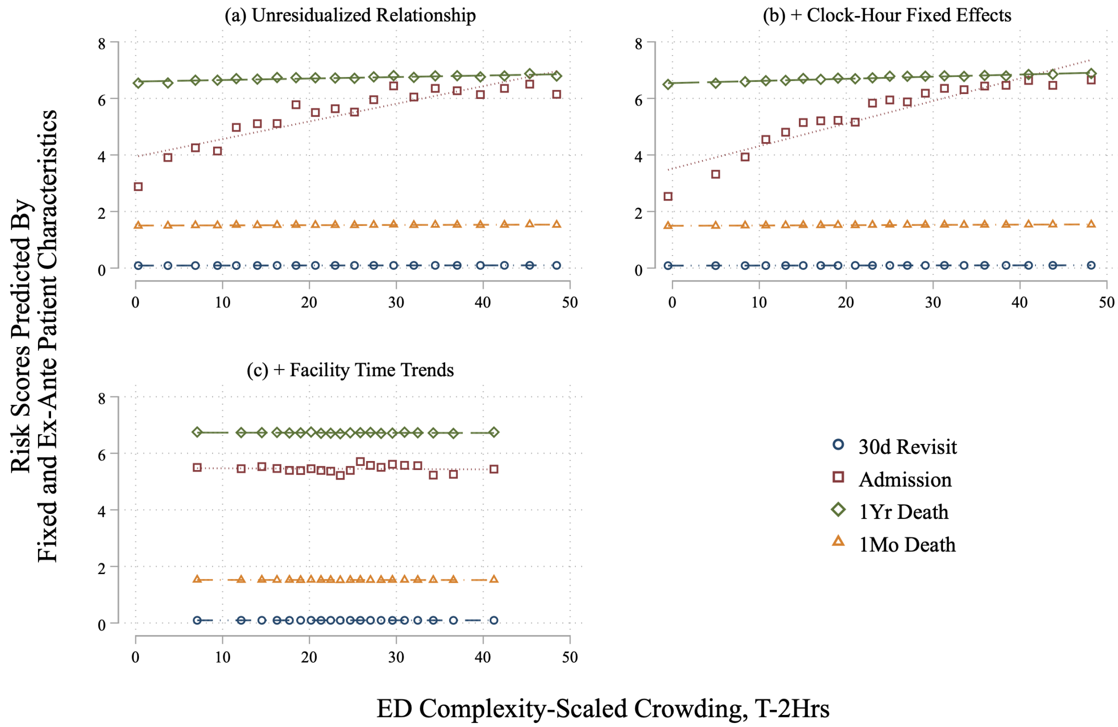
	(0.211)	(0.410)	(184.0)	(0.00327)	(0.112)
Dementia	4.658***	-2.885***	-1266.5***	-0.0260***	0.447
	(0.433)	(0.841)	(377.5)	(0.00671)	(0.229)
Mental Health	-0.561***	-3.403***	-867.7***	0.0582***	-0.252**
	(0.147)	(0.285)	(128.1)	(0.00228)	(0.0777)
Alcohol	0.636**	1.587***	446.8**	0.0463***	0.388***
Dependence	(0.198)	(0.385)	(172.6)	(0.00307)	(0.105)
Drug Dependence	-0.268	-1.315***	-854.5***	0.119***	-0.354***
	(0.194)	(0.376)	(168.9)	(0.00300)	(0.102)
Multiple	-0.616	-1.925	-466.9	0.0183*	-0.0684
Sclerosis	(0.588)	(1.142)	(512.4)	(0.00911)	(0.311)
Respiratory	1.021***	-0.850***	-25.83	0.000876	0.260***
	(0.0959)	(0.186)	(83.57)	(0.00149)	(0.0507)
Digestive	-0.128	0.350	-42.08	0.00608***	-0.0568
	(0.0989)	(0.192)	(86.17)	(0.00153)	(0.0523)
Prior Cardiac	0.756***	1.467***	343.1**	0.0110***	0.147*
	(0.122)	(0.236)	(105.9)	(0.00188)	(0.0642)
Prior Chest	-0.406***	-2.035***	-775.2***	0.0197***	-0.248***
Pains	(0.0995)	(0.193)	(86.64)	(0.00154)	(0.0526)
Prior Abdominal	-0.293**	-1.892***	-817.6***	0.0416***	-0.238***
Pains	(0.101)	(0.196)	(88.06)	(0.00157)	(0.0534)
Nutrition	1.411***	0.302	50.68	0.0249***	0.533***

(continued on next page)

(continued from previous page)					
	(0.214)	(0.416)	(186.7)	(0.00332)	(0.113)
Kidney	4.205***	3.434***	1562.7***	-0.00634*	0.874***
	(0.171)	(0.333)	(149.4)	(0.00266)	(0.0906)
R-Squared	0.119	0.305	0.151	0.100	0.038
Observations	198859	198859	198859	198859	198859

Notes: This table presents the regression results used to construct patient risk scores. Risk scores are constructed by regressing the indicated outcome (such as hospital admission or 1-year mortality) on the full set of ex-ante health condition indicators, patient race, insurance type, facility-specific time trends and clock-hour fixed effects. Fitted values are constructed using ex-ante health condition indicators, patient race, and insurance type. Patient health indicators are constructed as described in Appendix **??**

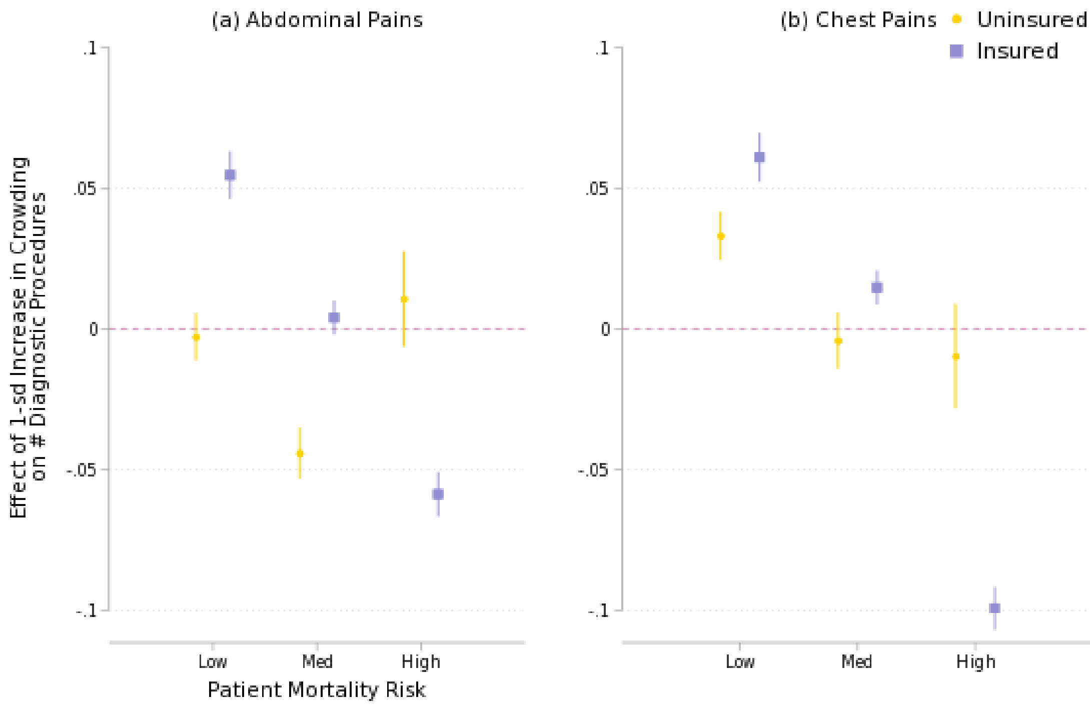
A.18 Patient Risk and ED Crowding



Notes: This figure shows the nonparametric relationship between ED complexity-scaled crowding and various patient risk scores. One observation represents one ED visit. Panel (a) shows the raw relationship between crowding and patient risk. Panel (b) shows the same relationship after the variables have been residualized on fixed-effects for each hour of the day. Panel (c) shows the relationship after further residualizing on facility-by-year fixed-effects.

Figure A.4: Patient Risk and ED Crowding

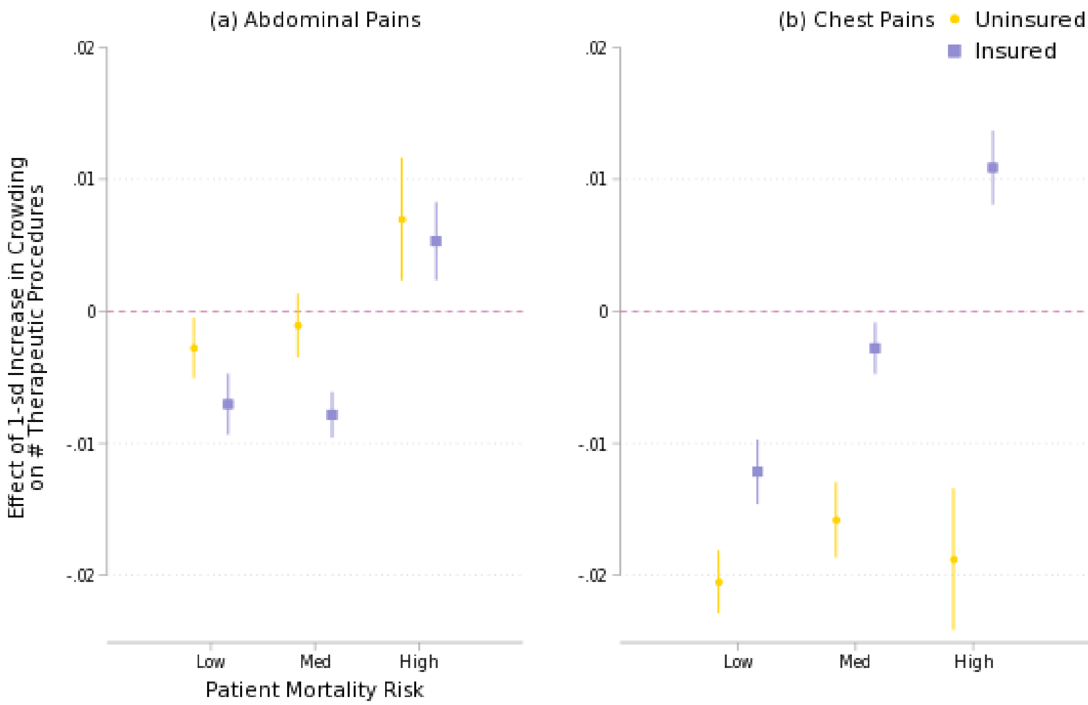
A.19 Effect of ED Traffic on Rate of Diagnostic Testing



Notes: This figure plots the coefficient estimates of the causal impact of a 1-sd increase in ED complexity-scaled two-hour traffic on the diagnostic testing rates of abdominal pain patients, in Panel (a), and chest pain patients, in Panel (b). Effects are plotted separately for insured and uninsured patients. An increase in ED traffic causes large decreases in diagnostic testing for high-risk patients. These effects are more pronounced for chest pain patients than for abdominal pain patients.

Figure A.5: Effect of ED Traffic on Rate of Diagnostic Testing

A.20 Effect of ED Traffic on Rate of Therapeutic Treatment



Notes: This figure plots the coefficient estimates of the causal impact of a 1-sd increase in ED complexity-scaled two-hour traffic on the diagnostic testing rates of abdominal pain patients, in Panel (a), and chest pain patients, in Panel (b). An increase in ED traffic causes large increases in therapeutic treatment for high-risk patients. These effects are more pronounced for chest pain patients than for abdominal pain patients.

Figure A.6: Effect of ED Traffic on Rate of Therapeutic Treatment

A.21 Effect of ED Traffic on Diagnostics for Chest Pain Patients

Table A.8: Effect of ED Traffic on Rates of Diagnostic Testing for Chest Pain Patients

	Diagnostics (Tests)					Admission
	(1) EKG	(2) Any Lab	(3) # Labs	(4) Chest X-Ray	(5) CT	(6)
Crowding X Uninsured, Low	2.058***	-0.828***	-0.000765	-1.269***	-0.107	0.999***
	(0.122)	(0.0969)	(0.00978)	(0.0917)	(0.111)	(0.0664)
Uninsured, Mid Risk	1.902***	-0.332**	0.0327**	-0.836***	0.122	2.608***
	(0.139)	(0.101)	(0.0110)	(0.0972)	(0.130)	(0.0853)
Uninsured, High Risk	1.872***	-0.774***	0.0855***	-0.730***	1.026***	4.151***
	(0.263)	(0.166)	(0.0213)	(0.166)	(0.242)	(0.145)
Insured, Low Risk	0.0902	-0.748***	0.00327	-0.623***	-0.318**	-1.781***
	(0.110)	(0.0815)	(0.00902)	(0.0822)	(0.117)	(0.0729)
Insured, Mid Risk	-0.389***	-0.153**	-0.00587	-0.214***	-0.0667	-0.956***
	(0.0752)	(0.0504)	(0.00610)	(0.0509)	(0.0803)	(0.0537)
Insured, High Risk	-1.836***	0.000444	-0.127***	0.298***	0.550***	0.817***
	(0.0903)	(0.0522)	(0.00765)	(0.0569)	(0.102)	(0.0714)
$\overline{Y}_{\text{low risk}}$	9.175	87.63	6.055	8.077	64.71	4.879
$\overline{Y}_{\text{mid risk}}$	15.62	87.03	5.773	11.65	67.30	9.853
$\overline{Y}_{\text{high risk}}$	43.09	89.21	6.018	27.94	76.60	27.44

Notes: This table reports estimates of the causal impact of a 1-sd increase in ED traffic on the rate of provision of various diagnostic tests for chest pain patients. One observation represents one ED visit. Each outcome variable is a binary indicator for whether or not the patient received the specified diagnostic test based on NUBC revenue codes, as described in Appendix A.6. The average rates of testing for low, medium and high risk patients are reported for each diagnostic test.

A.22 Effect of ED Traffic on Therapeutics for Chest Pain Patients

Table A.9: Effect of ED Traffic on Therapeutic Treatment for Chest Pain Patients

	Therapeutics (Treatments)						
	(1) Rx	(2) Pulm	(3) IV	(4) Stress Test	(5) Cath	(6) Coronary	(7) Preventive
Crowding X Uninsured, Low	0.573***	1.167***	0.396***	-0.163**	-0.786***	1.252***	-0.00424
	(0.0897)	(0.0710)	(0.0739)	(0.0509)	(0.0595)	(0.0402)	(0.0117)
Uninsured, Mid Risk	0.397***	0.988***	-0.0585	0.335***	-0.375***	1.254***	0.000698
	(0.105)	(0.0816)	(0.125)	(0.0629)	(0.0915)	(0.0435)	(0.0148)
Uninsured, High Risk	1.351***	1.191***	0.171	0.605***	-0.483**	1.263***	0.0702**
	(0.196)	(0.161)	(0.146)	(0.105)	(0.174)	(0.0996)	(0.0246)
Insured, Low Risk	-1.056***	0.322***	0.134	-0.607***	-0.395***	-0.307***	-0.0367*
	(0.0941)	(0.0669)	(0.0968)	(0.0611)	(0.0531)	(0.0396)	(0.0146)
Insured, Mid Risk	-0.592***	-0.242***	-0.131*	-0.0836	0.0249	-0.369***	-0.0210
	(0.0630)	(0.0473)	(0.0665)	(0.0453)	(0.0462)	(0.0302)	(0.0118)
Insured, High Risk	-0.198*	-1.413***	-0.379***	0.0739	0.637***	-0.609***	0.0660***
	(0.0772)	(0.0637)	(0.0832)	(0.0562)	(0.0561)	(0.0496)	(0.0167)
$\overline{Y}_{\text{low risk}}$	45.10	9.163	31.93	0.0255	0.00361	0.0639	0.161
$\overline{Y}_{\text{mid risk}}$	51.29	8.795	37.46	0.0702	0.0206	0.171	0.195
$\overline{Y}_{\text{high risk}}$	56.31	11.18	36.53	0.364	0.124	1.213	0.404

Notes: This table reports estimates of the causal impact of a 1-sd increase in ED traffic on the rate of provision of various therapeutic treatments for chest pain patients. One observation represents one ED visit. Each outcome variable is a binary indicator for whether or not the patient received the specified treatment based on NUBC revenue codes, as described in Appendix A.6. The average rates of treatment for low, medium and high risk patients are reported for each treatment.

A.23 Effect of ED Traffic on Diagnostics for Ab Pain Patients

Table A.10: Effect of ED Traffic on Diagnostic Testing for Abdominal Pain Patients

	Diagnostics (Tests)						Admission
	(1) Any Lab	(2) # Labs	(3) CT	(4) Ab X-Ray	(5) EKG	(6) Ultrasound	(7)
Crowding X Uninsured, Low	-0.802*** (0.0777)	-0.0385*** (0.00852)	-0.814*** (0.186)	-0.760*** (0.125)	0.130 (0.0714)	0.493*** (0.110)	0.784*** (0.0491)
Uninsured, Mid Risk	-0.879*** (0.0833)	-0.0596*** (0.00858)	-0.445* (0.193)	-0.447** (0.142)	0.0641 (0.0848)	0.0490 (0.0944)	1.199*** (0.0556)
Uninsured, High Risk	-0.265 (0.144)	0.0435** (0.0167)	1.819*** (0.364)	0.818** (0.302)	0.909*** (0.207)	-0.521*** (0.137)	2.375*** (0.100)
Insured, Low Risk	-0.264*** (0.0646)	-0.00206 (0.00777)	-0.472** (0.160)	-0.948*** (0.113)	-0.397*** (0.0751)	0.345** (0.120)	-0.943*** (0.0584)
Insured, Mid Risk	-0.462*** (0.0463)	-0.0490*** (0.00528)	-0.183 (0.108)	-0.860*** (0.0847)	-0.511*** (0.0571)	-0.172* (0.0690)	-0.551*** (0.0446)
Insured, High Risk	0.135* (0.0538)	-0.0590*** (0.00712)	0.953*** (0.127)	0.0219 (0.121)	-0.508*** (0.0868)	-0.685*** (0.0732)	0.269*** (0.0683)
$\overline{Y}_{\text{low risk}}$	87.63	6.055	64.71	23.47	9.175	24.55	4.879
$\overline{Y}_{\text{mid risk}}$	87.03	5.773	67.30	31.52	15.62	19.08	9.853
$\overline{Y}_{\text{high risk}}$	89.21	6.018	76.60	65.97	43.09	15.37	27.44

Notes: This table reports estimates of the causal impact of a 1-sd increase in ED traffic on the rate of provision of various diagnostic tests for abdominal pain patients. One observation represents one ED visit. Each outcome variable is a binary indicator for whether or not the patient received the specified diagnostic test based on NUBC revenue codes, as described in Appendix A.6. The average rates of testing for low, medium and high risk patients are reported for each diagnostic test.

A.24 Effect of ED Traffic on Therapeutics for Ab Pain Patients

Table A.11: Effect of ED Traffic on Therapeutic Treatment for Abdominal Pain Patients

	Therapeutics (Treatments)						
	(1) Rx	(2) IV	(3) Preventive	(4) Pulm	(5) Cardio	(6) GI	(7) Coronary
Crowding X Uninsured, Low	-0.307**	0.365*	-0.00802	0.496***	0.146***	0.284***	0.0975***
	(0.0979)	(0.153)	(0.00894)	(0.0521)	(0.0199)	(0.0167)	(0.0126)
Uninsured, Mid Risk	-0.276**	0.252	-0.00706	0.304***	0.144***	0.282***	0.111***
	(0.103)	(0.189)	(0.00951)	(0.0558)	(0.0235)	(0.0189)	(0.0143)
Uninsured, High Risk	0.682***	0.397*	0.0259	0.309**	0.489***	0.373***	0.124***
	(0.182)	(0.197)	(0.0168)	(0.113)	(0.0579)	(0.0310)	(0.0328)
Insured, Low Risk	-1.057***	0.185	-0.0255**	0.0837	-0.164***	-0.109***	-0.0123
	(0.0984)	(0.145)	(0.00945)	(0.0448)	(0.0202)	(0.0214)	(0.0112)
Insured, Mid Risk	-0.929***	-0.258**	-0.0122	-0.0603	-0.126***	-0.108***	-0.00176
	(0.0659)	(0.100)	(0.00728)	(0.0325)	(0.0186)	(0.0168)	(0.0119)
Insured, High Risk	0.00609	-0.708***	0.0144	-0.552***	0.0455	0.0592*	-0.159***
	(0.0799)	(0.121)	(0.0114)	(0.0469)	(0.0418)	(0.0292)	(0.0280)
$\overline{Y}_{\text{low risk}}$	45.10	31.93	0.161	9.163	0.265	0.509	0.0639
$\overline{Y}_{\text{mid risk}}$	51.29	37.46	0.195	8.795	0.748	0.989	0.171
$\overline{Y}_{\text{high risk}}$	56.31	36.53	0.404	11.18	4.939	2.884	1.213

Notes: This table reports estimates of the causal impact of a 1-sd increase in ED traffic on the rate of provision of various therapeutic treatments for abdominal pain patients. One observation represents one ED visit. Each outcome variable is a binary indicator for whether or not the patient received the specified treatment based on NUBC revenue codes, as described in Appendix A.6. The average rates of treatment for low, medium and high risk patients are reported for each treatment.

A.25 Adjusting for Facility-Specific Hourly Abdominal Pain Trends

Table A.12: Adjusting for Facility-Specific Hourly Abdominal Pain Trends

	Admission		# Diagnostics		# Therapeutics		1-Yr Mortality		Tot Chgs ($)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Crowding X Uninsured, Low	0.784***	0.600***	-0.00293	0.00625	-0.00276*	-0.00332**	0.167***	0.154***	-118.3***	-169.4***
	(0.0491)	(0.0513)	(0.00434)	(0.00446)	(0.00117)	(0.00121)	(0.0221)	(0.0234)	(18.74)	(20.81)
Uninsured, Mid Risk	1.199***	0.999***	-0.0443***	-0.0369***	-0.00105	-0.00194	0.142***	0.124***	-9.907	-62.04**
	(0.0556)	(0.0572)	(0.00463)	(0.00473)	(0.00124)	(0.00128)	(0.0278)	(0.0288)	(22.00)	(23.34)
Uninsured, High Risk	2.375***	2.139***	0.0105	0.0190*	0.00699**	0.00630**	-0.288*	-0.313**	427.2***	374.0***
	(0.100)	(0.101)	(0.00869)	(0.00873)	(0.00238)	(0.00240)	(0.115)	(0.115)	(60.94)	(60.17)
Insured, Low Risk	-0.943***	-1.109***	0.0545***	0.0575***	-0.00702***	-0.00670***	0.0158	0.00218	-375.4***	-418.2***
	(0.0584)	(0.0594)	(0.00431)	(0.00439)	(0.00118)	(0.00121)	(0.0202)	(0.0213)	(23.60)	(24.80)
Insured, Mid Risk	-0.551***	-0.757***	0.00391	0.00847**	-0.00783***	-0.00827***	0.0232	0.00915	-207.1***	-256.2***
	(0.0446)	(0.0456)	(0.00306)	(0.00316)	(0.000883)	(0.000915)	(0.0199)	(0.0208)	(20.53)	(21.43)
Insured, High Risk	0.269***	-0.00555	-0.0589***	-0.0528***	0.00531***	0.00412**	-0.130**	-0.156**	711.4***	633.0***
	(0.0683)	(0.0687)	(0.00403)	(0.00411)	(0.00151)	(0.00152)	(0.0481)	(0.0484)	(45.01)	(45.31)
Fac X Hr FEs	N	Y	N	Y	N	Y	N	Y	N	Y

Notes: This table reports regression results on the effects of ED patient traffic on hospital admission, diagnostic care, therapeutic care, 1-year mortality, and total visit costs for abdominal pain patients. The left column for each dependent variable reports these results from the original regression specification 1.3, and the right column reports results with the addition of facility-by-hour fixed effects, which capture across-facility variation in hourly staffing patterns which may not fully be captured by separate facility fixed-effects and hour-of-day fixed effects. The patterns of reallocation of care remain unchanged.

A.26 Adjusting for Facility-Specific Hourly Chest Pain Trends

Table A.13: Adjusting for Facility-Specific Hourly Chest Pain Trends

	Admission		# Diagnostics		# Therapeutics		1-Yr Mortality		Tot Chgs ($)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Crowding X Uninsured, Low	0.999***	0.851***	0.0330***	0.0392***	-0.0205***	-0.0202***	0.279***	0.273***	-450.7***	-489.7***
	(0.0664)	(0.0687)	(0.00439)	(0.00449)	(0.00122)	(0.00127)	(0.0237)	(0.0249)	(19.04)	(20.80)
Uninsured, Mid Risk	2.608***	2.401***	-0.00422	0.00430	-0.0158***	-0.0153***	0.183***	0.176***	-145.5***	-185.5***
	(0.0853)	(0.0867)	(0.00511)	(0.00519)	(0.00146)	(0.00150)	(0.0349)	(0.0357)	(24.42)	(25.71)
Uninsured, High Risk	4.151***	3.820***	-0.00978	0.000359	-0.0188***	-0.0188***	-0.634***	-0.641***	188.5***	155.2**
	(0.145)	(0.146)	(0.00948)	(0.00950)	(0.00273)	(0.00275)	(0.139)	(0.139)	(52.47)	(53.19)
Insured, Low Risk	-1.781***	-1.986***	0.0609***	0.0652***	-0.0121***	-0.0116***	0.201***	0.198***	-602.4***	-622.0***
	(0.0729)	(0.0738)	(0.00443)	(0.00450)	(0.00125)	(0.00128)	(0.0216)	(0.0226)	(22.59)	(23.91)
Insured, Mid Risk	-0.956***	-1.182***	0.0146***	0.0207***	-0.00279**	-0.00199	0.137***	0.134***	-186.6***	-204.5***
	(0.0537)	(0.0548)	(0.00309)	(0.00319)	(0.000988)	(0.00102)	(0.0217)	(0.0224)	(22.37)	(23.44)
Insured, High Risk	0.817***	0.532***	-0.0994***	-0.0920***	0.0109***	0.0118***	-0.369***	-0.376***	872.7***	850.1***
	(0.0714)	(0.0723)	(0.00390)	(0.00398)	(0.00143)	(0.00145)	(0.0490)	(0.0494)	(41.31)	(41.78)
Fac X Hr FEs	N	Y	N	Y	N	Y	N	Y	N	Y

Notes: This table reports regression results on the effects of ED patient traffic on hospital admission, diagnostic care, therapeutic care, 1-year mortality, and total visit costs for chest pain patients. The left column for each dependent variable reports these results from the original regression specification 1.3, and the right column reports results with the addition of facility-by-hour fixed effects, which capture across-facility variation in hourly staffing patterns which may not fully be captured by separate facility fixed-effects and hour-of-day fixed effects. The patterns of reallocation of care remain unchanged.

Appendix B

Appendix to Chapter 2

B.1 Incarceration Custody Levels

An IP's custody level, and facility in which they are housed, is determined by the following factors.

Institutional Risk Factors:
- Escape history.
- Protective custody issues.
- Inmate cooperativeness.
- Drug test results/drug test refusals.
- Previous probation and/or parole performance.
- Drug trafficking history.
- Community stability.
- Membership in a security threat group (such as a gang).
- Prior institutional history.

Public Risk Factors:
- Violence of confining offense.

- Use of a weapon in confining offense.
- Confinement history.
- Sentence length.
- Presence of detainers from other jurisdictions.
- Substance abuse history.
- Escape history.
- Sex offenses.

Appendix C

Appendix to Chapter 3

C.1 Processing Provider Name Data

The NYSDOE provides physician licensing data which includes the provider's entire name. I use the following protocol to identify the provider's first name:

1. Physician names are stored in string format, including spaces. The name is separated into components demarcated by the spaces.
2. Classify component 1 as the last name, unless either of the following two conditions apply, in which case classify component 2 as the last name.
   - The component is one letter in length
   - The component is De, Di, Mc, El, St, La, Al, or Van
3. Repeat step #2 until the last name is identified as component n.
4. Classify component $n + 1$ as the first name, unless the component is one letter in length, in which case classify component $n + 2$ as the first name.

Using this method, both “Smith Adam", “Hoover J. Edgar" and “De La Hoya Oscar" will identify the first names as “Adam", “Edgar" and “Oscar", respectively.

C.2 Assigning Genders to Provider Names

I use data from the Social Security Administration on the number of male and female babies born with any name given more than five times total in any given year from 1874 tothe present. This dataset is available at `https://www.ssa.gov/oact/babynames/limits.html`. I assign the name as male if more than 50% of the name recipients in the data are male.

The Genderize API, available at `https://genderize.io/`, assigns names similarly but derives the labels from social media accounts from all 195 countries. The API takes into account the ethnicity of the last name when inferring the gender of the first name: "Kim Watson" is an American name and therefore more likely female, but "Kim Yong" is Korean and therefore more likely male.

Of the 95,084 unique providers in my dataset, I am able to match 85,386 as either male or female. Of the matched providers, 35% are female. Emergency medicine tends to be male-dominated relative to other fields of medicine, in part due to the long, inflexible work hours that characterize the ED.

The unmatched names are primarily East and South Asian. Gender-labeled datasets that draw more heavily from these populations can be incorporated into my analysis to improve the rate of gender-matching. For now, my analysis is limited to the effects of adverse events on physicians to whom gender can confidently be assigned on the basis of first name. Thus, the results are largely driven by white, black and Hispanic physicians.

C.3 Creating Patient Health Indicators

I group ICD-9-CM diagnosis codes into broad indicators of prior health conditions. A patient is classified as having a given health condition if they have ever received a diagnosis code for that condition in any prior visit to an ED, hospital, urgent care clinic or ambulatory surgery center.

- Hypertension: 401.X
- High Cholesterol: 272.X
- Diabetes: 250.X
- Obesity: 278.X
- Smoking: 304.X
- Prior Myocardial Infarction (Heart Attack): 410.X
- Prior Cardiac Problem: 42X.X-45X.X
- Prior Chest Pains: 786.5X
- Prior Abdominal Pains: 789.X
- Poor Nutrition: 26.X
- Immune Disorder: 27.X