Task Allocation and Training in Nonprofit Emergency Departments

(Job Market Paper – Click here for the latest version)

Bryan P. Chu*

November 4, 2023

Abstract

Nonprofit teaching hospitals contribute almost half of Health Care and Social Assistance GDP and educate more than 90% of all future physicians. Existing models of nonprofit hospitals do not account for their teaching endeavors, thereby missing an important trade-off between the short-term delivery of health services and the long-term benefits of physician training. I leverage unusually detailed electronic health record and audit log data from the emergency department of a large, urban teaching hospital to characterize the static costs of training across a range of granular patient outcomes and process measures. Using panel data variation in patient assignment to residents, I find that the primary cost of training is length of stay for complex patients, which decreases about 9.4% over the four-year program. Then, to understand how the hospital trades off throughput costs today with future benefits of more intense physician training, I develop and estimate a dynamic model of training and care quality. I find that policies that increase the shadow cost of training, such as a change from FFS to PPS payments, could have negative effects on the career outcomes of graduating physicians 17 times larger than the savings for current patients. Feasible remedies such as increasing the staffing of attendings by 5% reduce the future cost by 81%.

*Contact: bryan_chu@berkeley.edu. I thank Ben Handel, Jon Kolstad, and Nano Barahona for their continued guidance and support. I also thank Robert Thombley, Julia Adler-Milstein, and the UCSF team for their work in assembling, deriving, and helping me interpret the data, as well as Jaskirat Dhanoa, Ryan Lichtarge, and Katrina Stime for discussion of institutional features. I am also grateful to Matt Backus, Zarek Brot-Goldberg, Kaveh Danesh, Jonathan Holmes, Yuki Ito, Yikun Jiang, Carolyn Stein, and Audrey Tiew, as well as participants of the Berkeley Industrial Organization Workshop and Seminar for their valuable comments.

1 Introduction

Healthcare spending is far higher in the United States than in other developed countries, yet Americans experience worse health outcomes.[1] In order to control spending growth and improve care quality, both public and private insurers are increasingly turning to payment reform. Payments are transitioning from the traditional system where providers are paid for each procedure performed to one where they are paid a fixed amount for each patient or paid based on quality measures.[2] The basis for these reforms is that they align the incentives of the providers, patients, and payers: in these systems, care providers are financially incentivized to improve quality and reduce cost.

However, when providers are not pure profit-maximizers and have multiple objectives, it is less clear what the impact of such policies are. A large portion of the health care system may fit this description: private, nonprofit teaching hospitals contributed 45% of all Health Care and Social Assistance GDP in 2019.[3] These organizations have the dual role of treating patients and training the next generation of physicians. While policy changes may induce teaching hospitals to take steps to increase care quality, they may also induce them to reduce teaching, which could have serious consequences for future patients. Understanding how private, nonprofit teaching hospitals trade off the quantity and quality of patient care with resident training is crucial in order to properly assess the full impact of such policy changes.

In this paper, I study how nonprofit teaching hospitals allocate patients of varying complexity to residents (trainees) and attendings (teaching faculty who can also work independently) in order to trade off care quantity, quality, and training. First, by leveraging detailed electronic health record and audit log data, I characterize the static costs of training: the differences in outcomes and process measures for patients seen by residents with varying experience. Then, to understand how the hospital trades off training costs today with benefits realized in the future, I develop and estimate a dynamic model of physician training and care quality. I use the model to quantify the long-run consequences for future patients caused by changes in physician training that may result from changes in incentives for current patient care quality.

I focus on emergency medicine (EM) residents at the University of California, San Francisco (UCSF). The UCSF EM Residency's day-to-day operations are typical of EM Residency programs. The majority of patients are seen by a single resident, the trainee, who is supervised by an attending physician, a faculty member. The remaining patients are seen by attendings working independently. Residents choose patients with assistance and guidance from attendings. Via their patient allocation decisions, attending physicians execute the trade-off between training and care quality that the hospital and residency director have determined.

The granularity of my data allow me to examine resident learning in great detail. I observe resident and attending identifiers for each distinct, disaggregated action, which allows me to attribute not only patient-level outcomes and decisions but also the dozens of individual decisions and actions for each patient to specific physicians. Timestamps for each action are unmasked, which allow me to not only correctly order patients and actions during each resident's history of work, but also to examine how

[1]In 2021, the United States spent 17.8% of GDP on healthcare, compared to the OECD average of 9.6%, but life expectancy was 77.0 years compared to the OECD average of 80.4 (Gunja, et al., 2023).

[2]Prominent examples include the Medicare's prospective payment system (PPS) for hospitals first enacted in 1983 and the 2010 Affordable Care Act's value-based payment initiatives. In recent years, private insurers have also begun pushing towards value-based care (Sokol, 2020). See McClellan (2011) for an overview on provider payment reforms and their welfare implications.

[3]The author's calculations using data from the BEA and AAMC. Across specialties, between 83.1% and 96.6% of residency programs were affiliated with nonprofit institutions in 2021 (Lassner, et al., 2022a, Lassner, et al., 2022b).

time duration to important actions evolves with experience. This combination of granularity in actions, physician identifiers, and unmasked timestamps is rare in health data.

I begin by using the panel structure of the data to examine the ways in which residents learn by doing over the course of their residency program. I find no evidence of statistically or economically significant changes in the 14-day readmission rate or in the number of signed diagnostic and therapeutic orders.[4] This stability is despite significant resident learning. For instance, the median fourth-year resident is about 20% faster at signing the first batch of orders compared to the median first-year. The speed gains only accrue to complex patients, which I define as those who are ex-ante predicted to require inpatient admission: length of stay for the median fourth-year's patients is 9.4% lower compared to length of stay for the median first-year's patient. There is no change for the "simple" patients who do not need inpatient admission. On the other hand, residents become much more productive in terms of total patients seen per shift. By managing additional patients simultaneously, they go from seeing three patients per eight-hour shift when they enter the program to seeing almost eight patients per shift prior to graduation.

Many of the areas of improvement are driven by or only present in complex patients. Therefore, I conclude that residents are learning how to treat complex patients and getting faster as a result of their increased skill. However, due to attending supervision, the main difference experienced by patients is additional time in the ED–there is no change in readmission rates or number of orders signed. There is relatively little learning for simple patients. This means that the hospital can choose the amount of training it provides by changing the allocation of complex patients to residents, and it optimizes this by considering the trade off between training and patient throughput.

While the throughput costs of training are paid today, the benefits accrue in the future so a model of the hospital's trade-offs must incorporate dynamics. I develop a discrete-time, infinite-horizon model where the hospital allocates complex patients to residents of different cohorts and attendings working alone to maximize a combination of patient throughput and training, subject to a budget constraint. This builds upon models of nonprofit hospital behavior by Newhouse (1970), Lakdawalla and Philipson (1998), and others. My contributions are to add teaching and the necessary dynamics to the hospital's objective function and to estimate the parameters empirically. I use the estimates to simulate counterfactual training and patient quality outcomes if the shadow price of training is increased, for instance via payment reform, or if training is disrupted, similar to what occurred during the Covid-19 pandemic.

I find that an objective function where the hospital maximizes training with respect to a lower bound of patient length of stay can rationalize the observed patient assignment shares during the academic year. I apply this model and the estimated parameters to two counterfactual exercises. The first is one where the hospital, perhaps spurred by payment reform, decides to make the care quality bound 2% stricter in terms of throughput in order to generate more revenue. The second concerns the impact of a one-time, unexpected training disruption, such as the one induced by Covid-19. In both scenarios, I compare the impact to patient care and resident training without further adaptations to alternatives where the hospital takes a mitigating action, such as loosening the care quality constraint, increasing the speed of the attendings working independently, and increasing the speed of resident learning.

In the first counterfactual, I consider the hospital's response to a policy change that increases the shadow cost of training–

[4] An "order" is any diagnostic or therapeutic procedure that is prescribed for the patient. Diagnostic orders are primarily for gathering information and include procedures such as blood tests, echocardiograms (ECGs), and imaging (CT scans, X-Rays, etc.). Therapeutic orders are primarily for treating and stabilizing the patient, and include pain medication, antibiotics, and surgical procedures.

for instance, a change in reimbursements from a fee-for-service (FFS) model to one based on capitated, prospective payments (PPS). If hospital administrators decide to improve throughput by 2%, then they will be forced to decrease training, resulting in less-skilled graduating residents. Assuming that residents go on to a 30-year career, this would result in decreases in future career outcomes 17 times larger in present-value than the hospital's gains. However, investing in attending speed–most simply by staffing additional attending physicians so that the supervisory burden is distributed among additional attendings–such that their speed increases 5% would greatly reduce the impact on training. With this remedy, the cost to career outcomes due to the throughput increase decreases by 81%.

In the second counterfactual, I consider the impact and potential responses to a disruption in training. Specifically, I assuming that the disruption causes affected residents enter their next year of training with half of the steady-state skill. I find that although the hospital returns to the steady-state the following period such that there is no impact to the entering cohort, the skill of the affected cohort is reduced by 2%. A one-period increase in attending speed by 2.5% allows the hospital to train sufficiently to recover 88% of the career costs compared to if there was no change. In both counterfactuals, outcomes for future patients are a large negative externality from the hospital's choices or the shock to the environment, but straightforward and feasible actions can greatly mediate the reduction in training.

This work contributes to several strands of literature. First, it extends the literature modeling the objectives of private, nonprofit hospitals that began with the seminal theoretical contributions of Arrow (1963), Newhouse (1970), Feldstein (1971), and Pauly and Redisch (1973). Since then, the bulk of the theoretical literature has consisted of models where the hospital maximizes the weighted sum of profits and quality or quantity of care (cf. Lakdawalla and Philipson (1998); see Gaynor and Town (2012) for an overview). These models have the appealing feature that nonprofit hospitals have similar objective functions to their for-profit counterparts, but with a lower marginal cost for quality or quantity (Gaynor, 2006), and this is consistent with subsequent empirical findings. For instance, nonprofit and for-profit hospitals are very similar in their responses to financial incentives (Duggan, 2000), CEO compensation incentives (Brickley and Van Horn, 2015), pricing behavior with regard to competition (Gaynor and Vogt, 2003), and provision of charitable care (Capps, et al., 2017).

My contributions are to add a teaching objective to the utility function and to estimate the parameters of the theoretical model. In doing so, I quantify the extent to which the hospital adjusts teaching in response to counterfactual policies that affect its revenue. In my model, the hospital acts like a profit-maximizing firm in that it desires to increase throughput in response to a decrease in revenue per patient, but a new and measurable mechanism it employs to accomplish that goal is to decrease teaching. I compare the costs borne by future patients and future employers of the graduating residents to the savings the teaching hospital gains with the throughput increase. In doing so, I address an issue first raised by Reder (1965): "Still further complications exist: hospitals produce not only current treatment but also train personnel for the production of future treatment. The costs and benefits of this training to the hospitals providing it are not well known."

Because I focus on patient allocation as the hospital's primary choice variable, I also contribute to the literature on task allocation. It has been shown that task allocation to heterogeneous workers may have large implications for productivity (Adhvaryu, et al., 2023) and that productivity differences within sector can be large and persistent (Syverson, 2011). Much of the literature concerns allocation of heterogeneous tasks to workers of fixed skill. In this framework, the firm desires to find each worker's comparative advantage, as in Adhvaryu, et al. (2023), Bergeron, et al. (2022), Cheng (2019), Cowgill, et al. (2023), Dahlstrand (2023), and Kasy and Teytelboym (2022). However, the benefits of task allocation can be especially important when the environment is dynamic, such as when workers learn by doing (Minni, 2023). Like Minni (2023), I consider a setting where worker skill is not only task-specific but is also endogenous to the history of tasks that have been assigned. The granularity of my data allow me to go beyond differences in job title and department and allow skill to vary and improve based on the exact patient types each resident is assigned.

This is also not the first work that incorporates learning by doing, nor is it the first that considers learning from peers and managers. The learning by doing literature is vast. In addition to work investigating manufacturing (such as Benkard, 2000 and Levitt, et al., 2013), there is a long literature documenting and examining learning by doing in a variety of ways and settings within medicine. For instance, there is work on resident learning in internal medicine (Chan, 2021), learning about match values of patients to procedures (Gong, 2018) and to medications (Currie and MacLeod, 2020), and learning to work in teams (Chen, 2021 and Reagans, et al., 2005). There is also a strand of literature concerning learning from coworkers and managers, such as Bloesch and Weber (2023), Jarosch, et al. (2021), and Jovanovic (2014). I study learning by doing by emergency medicine residents from other residents and attending physicians in granular detail across a variety of outcomes and process measures, incorporate how learning is affected by task allocation, and relate the findings to static costs in quantity of patients seen and quality of care. Like Gong (2018), residents in my context learn how to treat specific patient types by doing, and like Bloesch and Weber (2023) and Jovanovic (2014), my dynamic model has an overlapping-generations structure.

My findings also relate to the literature on academic hospitals and value-based payments. Kocher and Wachter (2023) find that academic hospitals tend to do poorly on value-based payments, which they attribute in part to over-specialization and organizational frictions due to teaching responsibilities. McClellan (2011) describes a range of potential payment reforms. I contribute to this policy debate by quantifying the long-run costs of payment reform at teaching hospitals and consider the benefits of offsetting changes. Specifically, decreasing teaching hospital revenue would induce them to reduce training, which would have deleterious effects on future care quality and efficiency over the course of each resident's career, but increasing attending staffing would drastically reduce the cost.

Finally, my findings add to the literature studying cohort turnover. I corroborate Hughes (2017), Wei, et al. (2019), and the recent literature that finds an absence of a significant drop in quality in July, when the most senior residents graduate and are replaced by new medical school graduates.[5] I extend the literature by showing that not only patient outcomes but also many process measures related to productivity are unchanged across July 1. I add to the findings of Song, et al. (2016) and Hausknecht and Trevor (2011) and describe another method the hospital uses in order to smooth the transition. Notably, this method, strategic task allocation, is a choice rather than an investment in infrastructure.

The rest of the paper proceeds as follows: Section 2 gives more details on residency in general and emergency medicine residency at the teaching hospital from which I obtain data. Section 3 describes the electronic health record and audit log data. Section 4 presents empirical results that documents the ways in which residents learn by doing. Section 5 introduces the dynamic framework. Section 6 discusses estimation, and Section 7 provides results. Section 8 motivates and presents counterfactual exercises that explore separately the hospital's response to a change in the shadow cost of training and to a one-time disruption to resident training, as well as the effectiveness of mitigating actions it could take. Section 9 concludes.

[5] In the United States, this is colloquially known as the "July Effect." In the United Kingdom, it is known as "killing season."

2 Medical Residency Background

In the United States, graduates of medical school are required to complete a residency program in order to practice medicine independently. Residency is in a specific predetermined specialty (for example, Radiology, Dermatology, Obstetrics and Gynecology, and Emergency Medicine); medical school students apply to and are accepted to a single program-specialty.[6] Matching residents to program-specialties is done centrally and is a well-known application of the Gale-Shapley algorithm. Programs last between three and seven years, depending on the specialty, and some medical students choose complete a fellowship after their residency ends to further specialize and become for example Cardiologists and Oncologists, or to sub-specialize, for instance in Pediatric Critical Care or Cardiothoracic Surgery. Notably, residency training is not only for learning facts but also for developing "habits, behaviors, attitudes, and values that will last a professional lifetime" (Ludmerer, 2014).

The focus of this study is Emergency Medicine Residency at the University of California, San Francisco (UCSF). At UCSF, EM Residency is a four-year program.[7] The setup of the program and the day-to-day routine is typical of EM Residency Programs. The majority of patients are seen by a single resident, the trainee, who is supervised by an attending physician, a faculty member. The remaining patients are seen by attendings working independently. Work is shift-based, meaning that once physicians are off-shift, they are no longer responsible for the patients they cared for during their shift. At UCSF, both residents and attendings work eight-hour shifts. The schedule is determined prior to the beginning of the academic year and determined exogenously. All residents and attendings will work day, night, and weekend shifts; there is no sense that seniority or other factors permit attendings or residents to avoid working less-desirable shifts. Teams–groupings of attendings and residents–are ad-hoc, meaning that they change from shift to shift, and throughout the course of the year, all residents will work with all other residents and all attendings.

To be clear on terminology, I will use "resident" to refer to the emergency medicine physician trainees who are the focus of this study. At any point of time, EM residents at UCSF must belong to one of four different cohorts–this is defined as the year that they enter the program. Consistent with nationwide averages, I do not observe any attrition or leaves of absence.[8] "Attendings" or attending physicians are faculty members of the medical school, typically on the tenure track, who both supervise residents and see patients independently. I will use the terms "physician" and "provider" interchangeably to refer to residents, attendings, and nurse practitioners, who are also seeing patients independently but do not have supervisory responsibilities. I will use the term "care team" to refer to all providers, nurses, and other medical and non-medical staff (e.g. social workers) who interact with the patient. An "order" is any diagnostic or therapeutic procedure that the care team prescribes for the patient. Diagnostic orders are primarily for gathering information and include procedures such as blood tests, echocardiograms (ECGs), and imaging (CT scans, X-Rays, etc.). Therapeutic orders are primarily for treating and stabilizing the patient and include pain medication, antibiotics, and surgical procedures.

The typical workflow in the ED is depicted in Figure 1. When a patient arrives, a triage nurse will interview them, take their vital signs, and estimate their acuity using a five-point scale called the Emergency Severity Index (ESI). This is done independently from the physicians who will subsequently care for the patient. The patient will then return to the waiting

[6] Students apply to multiple programs but typically a single specialty: among students who successfully match, the average number of specialties ranked is 1.2 (AMA, 2019)

[7] Most EM Residency programs are three years; four-year programs tend to be located at prestigious and highly-ranked programs such as Johns Hopkins, Massachusetts General Hospital (Harvard Medical School), UCLA, and the University of Washington.

[8] The median EM Residency attrition rate from 2010-2020 is 0.83% (Wang, et al., 2022)

room. A resident who is available will select a patient from the waiting room with guidance from the supervising attending. The resident will independently examine the patient and present their findings and plan of care to the attending. The attending will examine the patient, also typically independently, and confer with the resident. An agreement on the plan of care is reached and a set of diagnostic and therapeutic orders are signed. Order results are reviewed, typically independently, and if necessary, predetermined follow-up orders are sent and additional examinations and revisions to the plan are agreed upon and executed. The resident and attending will then make a disposition decision: admit the patient to the hospital for additional care or discharge them home. In the event the patient was discharged home, there is a chance they will return to the ED within 14 days. This is called an ED Readmission and is suggestive that the physicians overlooked something important.

Care quality in the ED is multi-dimensional. Broadly, once the patient is stabilized, the goal is to quickly and efficiently assess the patient's condition. The disposition decision is the crucial element of patient condition: are they healthy enough to send home, or do they need to remain in the hospital for further care? Therefore, the primary measure of ED quality is the accuracy of the disposition decision. A common measure used to evaluate this decision is the 14-day readmission rate (cf. Chan, 2018): among patients who were deemed healthy enough to discharge, at what rate did they return to the ED within 14 days? A second category of quality relates to speed. Doing things faster with no loss in accuracy is also important. Speed is utility-enhancing for patients because they spend less time suffering from their complaint and being in the hospital, and is also efficient because it frees up the examination room for the next patient, thereby increasing patient throughput. Important measures of speed I will consider include process measures such as time to first order and patient length of stay in the ED. Finally, I will consider resource utilization as a measure of efficiency. Resources are both costly orders ("materials") as well as labor in the form of supervision and consults by specialists, and being able to achieve the same patient outcomes with fewer orders or consults represents higher efficiency.

3 Data and Sample Construction

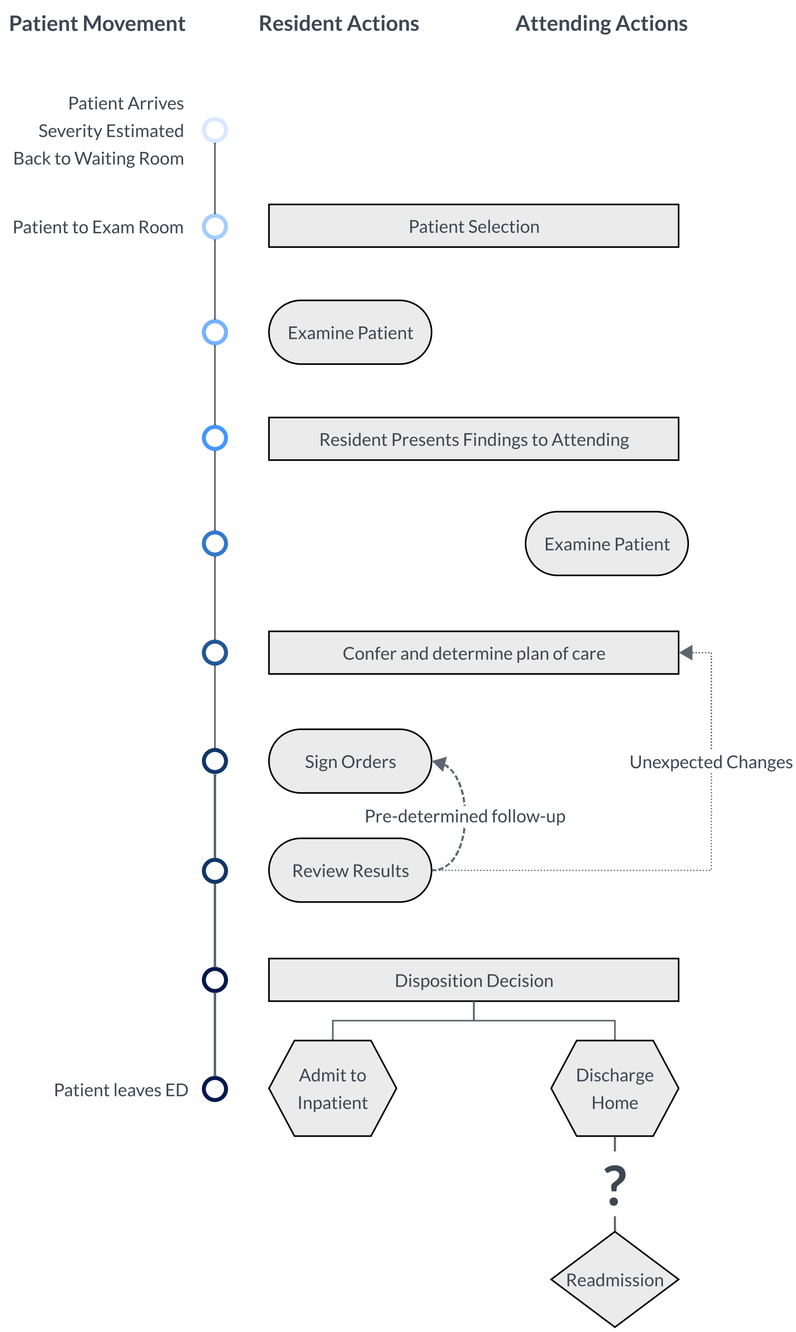
3.1 Data

This research leverages highly granular electronic health record and audit log data from UCSF. This data cover the universe of ED arrivals for patients ages 18-90 over a 24 month period from 2017 to 2019. In total, there are 85,990 patient encounters.[9] In essence, these data record every interaction the physician has with a computer, which is used for gathering information (reading past clinical notes and order results), producing a diagnosis and treating and stabilizing the patient (sending, revising, and canceling orders), and recording information (writing the clinical note summarizing the patient's condition and what was done in the ED).

I observe every instance that any provider interacts with an order. For each of these order actions, I observe patient and encounter identifiers, actual, unmasked timestamps for when each order was signed, completed (or canceled), and results became available (when applicable). I also observe identifiers for both the physician who signed the order (typically a resident) and the physician who authorized it (must be an attending). These are an unusual features of the data. In terms of physician

[9] The unit of observation is an encounter rather than a patient because the same patient may visit the ED multiple times during the sample period. When this occurs, they are assigned a new encounter_id for each visit but retain the same patient_id.

Figure 1: Workflow in the Emergency Department



Notes: This flowchart illustrates of the typical workflow in the emergency department. Actions and outcomes are divided into three categories. Left of the timeline are patient movement. To the right of the timeline, actions are classified into those done by residents (left side), attendings (right side), or together (spanning the width of the section). The dotted arrows originating from Review Results indicate that these actions are done only when deemed necessary. Finally, after the disposition decision is made, if and only if the patient is discharged home, they may feel it is necessary to return to the ED within 14 days, which is called an ED Readmission.

identifiers, most medical datasets only contain the data of the attending physician, as they are the entity who is financially and legally responsible. As for timestamps, most data either only have the date of the encounter or have detailed but de-identified data that preserves the time between actions but scrambles the start dates. Both of these elements are crucial for this analysis as otherwise I would not be able to attribute residents to patients in the correct order and would be greatly limited in the number of process measures I could examine.

I also observe the consumption and production of information. Specifically, I observe the time, duration, and provider for each order result view (e.g. reading the radiologist's report for an MRI; viewing the numerical results of a blood test) and the same information for clinical notes that contain other physicians' impressions of the patient.[10] I also observe the time and duration of edits to the patient's clinical note from the current encounter, as well as the character length of the note. I do not observe any note content.

For patients, in addition to typical covariates such as age, gender, race, and diagnosis codes, I also observe a set of characteristics that I call "ex-ante" characteristics. These are characteristics that are exogenous to the care team who will subsequently care for the patient. Examples include the patient's chief complaint that induced the ED visit, the acuity level assigned to them by the triage nurse, and indicators for abnormal vital signs upon entry to the ED (ex. abnormal pulse). Contrast these with measures such as the final diagnosis, ED disposition, or patient's length of stay in the hospital, which may be endogenous to the composition of the care team and most crucially, resident experience.

For providers, I observe basic covariates. I observe the role of all providers: resident, attending, nurse practitioner, etc. I observe the specialty for attendings and NPs only and infer the specialty of residents based on the specialties of the attendings who most frequently authorize their orders, which I take to be their most frequent supervisors. Residents use different templates in the system if they are in their first two years compared to years thereafter. I also observe their start and end dates if they occur within the sample period; with these two pieces of information I am able to infer the cohort and specialty of each resident.

Separately, I have the administrative schedules for both providers and residents for calendar year 2018. I use this data to validate my sample construction and to provide some sample statistics on the number of shifts worked by EM and non-EM residents. I am unable to match the names in the schedule with the provider identities in the EHR data.

3.2 Sample Construction

I focus on EM Residents and attendings. These providers make up a minority of physicians who ever work in the ED but work a majority of the shifts and see a majority of the patients, especially among the more acute patients. The reason I restrict the analysis to EM Residents is because the ED may have other learning objectives for the residents from other specialties who make short rotations through the ED as part of their training. For instance, Internal Medicine residents complete a three-week rotation in the ED. Not only is this time period is too short for the ED to significantly benefit from training them in the future, they also may have a different set of baseline skills. Therefore, the incentives and constraints for training other residents may differ from those for training ED residents and I exclude them from my analysis.

During the sample period, there were 15 residents in each cohort of EM residents. I am unable to identify them based

[10] Both order result and note views can be from "historical" visits outside the sample period.

on names or identifiers so I do so based on an algorithm that classifies them based on the total number and fraction of their actions that were in the ED. Table 1 shows the breakdown of the residents who work in the ED in terms of the number of individuals, shifts worked, and patients seen. I slightly under-match, identifying 83 residents instead of the expected 90 in the six cohorts in my data. In calendar year 2018, where I am able to validate my resident selection by comparing shift summary statistics with administrative shift data, I also under-match slightly, identifying 67 of 75 residents. Perhaps as a result, I find that they work 60% of the shifts rather than the 69% as suggested by the 2018 administrative data. As expected, the majority of patients are seen by EM residents: almost 70% in the two years of EHR data.

Table 1: Sample Selection: Residents

	Residents	Shifts (EHR data)	Shifts (admin)	Patients (EHR data)
All Residents	610	9,340		54,217
EM Residents	83	5,802		37,463
EM Residents (%)	14%	62%		69%
2018 Residents	389	4512	4,012	26775
2018 EM Residents	67	2725	2,765	18044
2018 EM Residents (%)	17%	60%	69%	67%

Notes: This table shows basic sample statistics on the set of residents who work in the emergency department. I focus on EM Residents, who make up 14% of all residents who work in the ED during the two-year sample based on my classification. They work 62% of the shifts worked by residents and see 69% of all patients seen by residents. I compare the share of shifts with the share of shifts in the administrative data that cover one calendar year and find that EM residents worked 69% of all shifts worked by residents, which compares favorably to the 60% I classify in the data.

Table 2 Panel (a) shows sample selection for patient encounters. Over the two years of data, there are a total of 85,990 patient encounters. I first exclude encounters where the patient left early or against medical advice, or passed away in the ED, so that I can be sure that I capture the full extent of the physician's process rather than some interrupted version. These total roughly 7.7% percent of all encounters. Then, I exclude the patients who the triage nurse categorized upon arrival as being the most urgent (Emergency Severity Index category 1) or the least urgent (ESI 5), who together represent about 2.3% of all arrivals. This is because the ESI 1 patients represent "codes" where the entire ED team contributes to the patient's care, so it is an exception to the usual resident-attending pairing and may not represent cases where the resident is directing care. ESI 5 patients are the other extreme: they are cases where the patient does not need urgent medical care, such as patients with a chief complaint of "Medication Refill" and also do not represent resident learning about urgent patients. Next, among the remaining encounters, I am unable to identify the physician in charge ("Primary MD") for 6.6% of the patients. The next step results in our first sample of interest: EM Residents and Attendings see a total of 65.3% of all patients. Finally, EM Residents see 40.4% of all patients, or about 62% of the patients assigned to EM Residents or Attendings. Panel (b) reveals that the 62% of patients are not evenly distributed among patient types: residents see a greater share of complex patients (about 77%) relative to simple patients (about 58%) by two measures of ex-ante patient complexity.

These tables show that EM residents are doing a plurality of work in the ED and a majority of the work for complex patients. It is not the case that they are only seeing low-risk patients they know how to manage and leaving the complex ones for attendings to care for. In the following section I show how patient outcomes, process measures, and the allocation of complex and simple patients vary with resident experience.

Table 2: Sample Selection: Encounters

(a) Encounter Selection

	Number or Percent of Patients
All ED Arrivals	85,990
Did not Leave Early	92.4%
Did not Pass Away	92.3%
Triage Nurse ESI 2, 3, or 4	90.0%
Primary MD Identified	83.6%
Seen by Attending or EM Resident	65.3%
Seen by EM Resident	40.4%

(b) Resident Encounters by Complexity

		by Predicted Admission		by Triage Nurse ESI	
	All Patients	Complex	Simple	Complex	Simple
All with Primary MD identified	71,892	17,916	53,976	14,935	56,957
Seen by Attending or EM Resident	78.1%	73.1%	79.7%	74.1%	79.1%
Seen by EM Resident	48.3%	56.0%	45.8%	58.0%	45.8%
Percent EM / Attending or EM	61.8%	76.6%	57.5%	78.3%	57.9%

Notes: This table shows the sample selection of patient encounters. Panel (a) shows the steps of sample selection. Patients who Leave Early are those who have leave without being seen, against medical advice, or pass away in the ED. Triage Nurse ESI 2, 3, or 4 are the three middle categories of the triage nurse's assigned Emergency Severity Index. The two excluded categories are extremely severe cases ("codes") where the entire ED team contributes to the patient's care, or cases where the patient does not need urgent care, such as patients with a chief complaint of "Medication Refill." Primary MD Identified means I was able to identify who the primary provider for the patient was. Panel (b) shows the breakdown of the last three steps of Panel (a) by two ex-ante measures of "complex" and "simple" patients. The first is the primary measure I use in the paper: by a LASSO prediction of inpatient admission using only ex-ante and immutable patient characteristics. The second is by the triage nurse's evaluation: ESI category 2 vs. 3 and 4. The bottom row of Panel (b) shows the percent of patients of each patient type seen by EM residents relative to the patients seen by EM Residents and Attendings and reveals that residents see a greater share of complex patients than of simple patients.

4 Documenting Resident Learning

I begin by documenting the ways in which residents improve during their four-year tenure. I first show the improvement graphically via binned scatterplots (Figure 2) and then present additional specifications in regression form (Table 3).

The binned scatterplots are constructed with the set of encounters managed by residents. Each figure shows 24 bins of patient outcomes and process measures on the vertical axis against the residual of fractional years in the program by the resident in charge of the patient on the horizontal axis. The residuals are after removing selected patient covariates X_i. The slope and standard error of the regression line displayed correspond via Frisch-Waugh-Lovell[11] to the coefficient on experience in the regression given by

$$Y_i = \beta \text{Experience}_{j(i)} + X_i'\gamma + \varepsilon_i \tag{1}$$

In this regression, i indexes encounters, and $j(i)$ is a function that returns the experience of resident j in charge of patient

[11]Recall that it is not necessary to also residualize the dependent variable. An alternative interpretation is that I have plotted the residuals of both the dependent variable and resident experience against each other, but have added the dependent variable mean back to the residuals of the dependent variable.

i at that moment in their career. In the binned scatterplots, I select the ex-ante and immutable patient characteristics X_i by hand. The covariates include fixed effects for 10-year bins of patient age, the Charlson comorbidity index, Medicaid status, nonwhite, an interaction of broad chief complaint category and triage nurse assigned emergency severity index, an interaction of indicators for if the encounter began on a weekday and during business hours, and continuous ex-ante predictions of patient complexity and its square from Chu, et al. (2023). The residency program lasts four years, but because my data span two years, I observe each resident for a maximum of two years. Hence, the data is an unbalanced synthetic panel.

Figure 2 breaks down resident learning into various components. I first examine improvements in disposition decision quality as measured by the 14-day ED Readmission Rate and in resource efficiency as measured by costly orders signed. We observe in Panel (a) that there does not appear to be a statistically or economically significant decrease in 14-day ED Readmissions, suggesting that conditional on patient observables, the accuracy of the disposition decision for inexperienced and experienced residents is similar. Next, in Panel (b), we see that the number of costly diagnostic and therapeutic resources similarly does not exhibit a statistically or economically significant change associated with experience. Therefore, I conclude that neither patient outcomes nor costly resource utilization improve with experience.

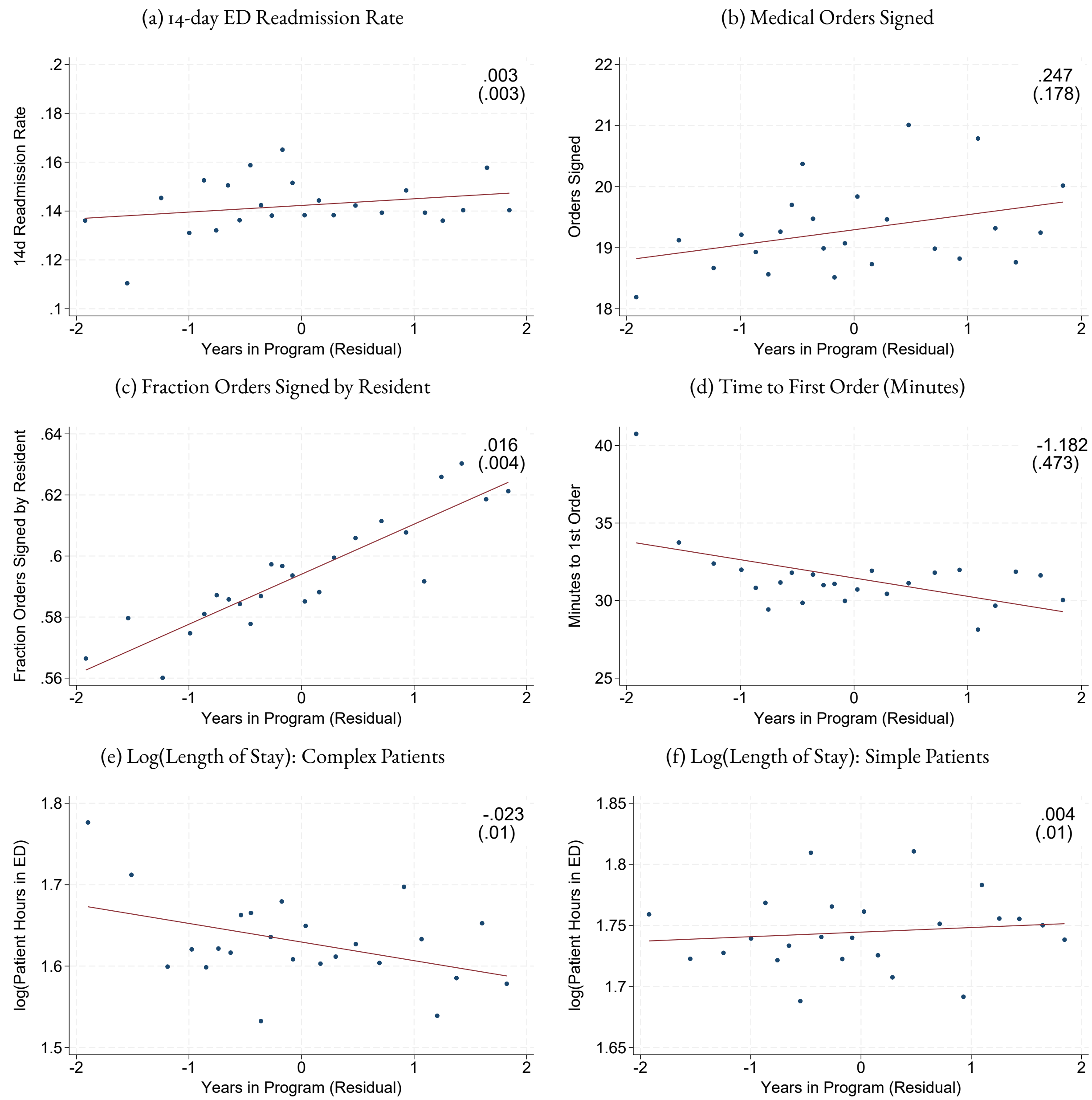
But does that mean that residents do not learn? Panels (c) and (d) refute this. Panel (c) plots the fraction of orders signed by the resident, rather than the supervising attending, other attendings such as consulting physicians from other specialties, nurses, or other residents assisting. This measure increases linearly with experience, and the magnitude over four years is approximately 10% of the mean of 59.5%. Panel (d) plots a measure of speed: how long does it take providers to sign the first order from the time the patient enters the examination room? Over four years, this decreases approximately 15% of the mean of 34 minutes. Taken together, these results imply first that residents gain independence and are less apt to leave out important orders. Second, they become faster at discerning the patient's state and determining which set of orders are appropriate for treatment and for refining the working diagnosis. In other words, they become faster at the initial evaluation of the patient and are able to send more complete batches of orders with that information. I cannot disentangle whether the order batches are more complete due to a better independent evaluation of the patient or because the resident is better able to recall the set of appropriate orders given the initial evaluation but either way, both margins represent important progress.

Curiously, the increases in independence and speed only flow through to the total length of stay for complex patients. Panels (e) and (f) show the evolution of the natural logarithm of the number of hours the patient spends in the ED with resident experience. The patient sample is split by whether they were ex-ante predicted via LASSO[12] to be admitted to the hospital ("complex") or whether they were predicted to be discharged home ("simple"). While there is no change for simple patients, there is is a significant and meaningful improvement for complex patients. Under the assumption of linear learning, the four-year improvement of 9.2 log points is almost 25% of the standard deviation of log(length of stay) conditional on patient covariates, and is relative to a mean length of stay of about 6.6 hours.[13] Therefore, experience greatly increases the throughput for complex patients but has limited effects for simple patients.

[12] This ex-ante prediction uses only information observable to the physicians at the time they are selecting the patient and has high predictive power: the AUC is approximately 0.97. See Appendix A for additional details on the construction of the prediction.

[13] The reason the mean for complex patients is less than the mean for simple patients is because for admitted patients, I end the the length of stay at the moment the patient is confirmed for inpatient upgrade. At that moment, the patient may not leave the ED, but the ED care team's involvement has concluded and the patient is now the responsibility of the admitting department, whether it be cardiology, surgery, hospital medicine, or something else. Unfortunately, there is no consistent analogous marker for discharged patients (discharge orders are inconsistently signed and disappear entirely midway through the sample period). The slope is similar if I use total time in the ED for both sets of patients.

Figure 2: Learning over Time: Binned Scatterplots



Notes: These figures are binned scatterplots with 24 bins of patient outcomes and process measures of interest on the residual of years in the program by the resident in charge of the patient. The sample is all sample patients seen by EM residents. Residuals are after removing selected patient covariates. The coefficient and standard error, clustered by physician, are displayed. The 14-day ED readmission rate is the rate at which patients who are discharged home from the ED have a repeat visit within 14 days. Medical Orders signed is the sum of diagnostic and therapeutic orders signed in the ED. Fraction Orders Signed by Resident is the fraction of medical orders that are signed by the resident, rather than the attending, nurses, or other residents assisting. Time to First Order is the time between the moment the patient is moved from the waiting room to an exam room and the time that the first medical order is signed. Log(Length of Stay) is the natural logarithm of the hours the patient spent in the ED under the care of EM providers. It is split into "complex" and "simple" patients based on an ex-ante prediction of inpatient admission. See text for more details.

These results are confirmed in the regressions in Table 3. The regression specification differs slightly from the binned scatterplots in Figure 2. First, I include resident fixed effects in order to focus on within-resident learning. With these fixed effects, estimates are not subject to bias from individuals in earlier cohorts (e.g. starting residency in 2015) being inherently "better" or "worse" than individuals in later cohorts (e.g. starting residency in 2018). Second, I compare the results using years in the program as a measure of experience to a different measure: cumulative fraction of complex patients seen. This measure is valid and interesting under two assumptions. First, it assumes that learning is only possible with complex patients, meaning residents do not learn from seeing simple patients. Second, it assumes that complex patient arrivals each quarter are identical, so the fraction seen is sufficient to describe experience with patients. This measure only updates at the beginning of each calendar quarter and represents the cumulative fraction of patients that the resident has seen each patient. It is exact for the residents who begin the program during the sample period, and is inferred using the length of stay for complex patients for residents who begin the program prior to the start of the sample period.[14] I am interested in this measure because I will examine how the hospital trades off throughput and learning via the allocation of complex patients, and the fractional years that the resident has been in the program at the time they saw the patient is only a "reduced-form, equilibrium" version of that measure. In other words, I assume learning is task-specific and must be experiential ("learning by doing"), meaning that residents cannot learn by simply observing.

In Table 3, the coefficients displayed are from four separate regressions. The experience variables of interest are the only change within a column, and they are cumulative fraction of patients and years in program ("tenure") entering linearly (top grouping) and the same experience proxies entering as natural logarithms (second grouping). All regressions include physician fixed effects. I present two versions of each regression. The first of each pair includes physician fixed effects but does not include any additional controls. The second includes patient controls, but unlike in the binned scatterplots, I now select them using the post-double-selection LASSO method of Belloni, et al. (2014). Inspection of the covariates chosen by the algorithm reveal that they are more sparse than the set that I manually selected, and tend to include indicators for the number of abnormal vital signs upon entry, which I did not include in the binned scatterplots.

In this table, we first see that the qualitative results of Figure 2 are generally robust to the more sophisticated selection of patient covariates and the inclusion of user fixed effects. There are sometimes differences in the statistical significance of the linear and logarithmic specifications. I rely on graphical analysis via Figure 2 in order to assess which set of results fits better. For instance, the fraction of orders signed by the resident appears to improve linearly with experience (Panel (c)). If the figures are not abundantly clear, such as Time to First Order (Panel (d)), then I tend to prefer the logarithmic specification as we generally believe that learning exhibits diminishing returns (cf. Benkard, 2000 and Levitt, et al., 2013).

Table 4 splits the other process measures by patient complexity. Note that by definition, all ED Readmissions were for discharged patients, so a breakdown is not appropriate here. For medical orders, we see that the null result earlier masked offsetting effects for complex patients and simple patients. One potential explanation, supported by Appendix Table A2, is that with experience, residents obtain less diffuse priors when diagnosing complex patients, but that they also substitute effort with costly resources for simple patients in order to save time.[15] Next, we observe that the increase in fraction of orders signed

[14] Essentially, I regress log(length of stay) on log(patient share) for the residents who begin the program during the sample period and infer the log(patient share) for the residents who begin the program prior to the sample period using the estimated coefficients and their average patient length of stay in the first quarter I observe them in. See Appendix B for additional details.

[15] This potentially facilitates the increase in managing additional patients simultaneously with experience: see Figure 3.

by the resident is primarily driven by complex patients, but that decreases in the minutes to the first order are proportionally similar for complex patients compared to for simple patients. All together, these results suggest that the bulk of the learning that occurs during the residency relate to learning how to treat complex patients and that there is relatively little learning for simple patients.

The main threat to these analyses is that they are biased by unobserved patient heterogeneity. Specifically, if more experienced residents are assigned patients who are unobservably more complex, my estimates will be biased towards zero. Similarly, if they are assigned unobservably less complex patients because they are seeing additional patients simultaneously, then my estimates will be larger in magnitude than the true improvement with experience. I believe this is relatively unlikely in my setting for two reasons. First, providers observe a limited amount of information when allocating patients, and I am able to control for almost all of these covariates. The main thing I do not observe are the patient's appearance and answers to brief questions, but to the extent that is captured in the triage nurse's evaluation and estimation of the patient's severity, I do control for it. Second, other than for the first six months of the program, observable patient severity averages per patient are stable across the four years of experience as can be seen in Appendix Figure A1, Panel (a). Therefore, in terms of ex-ante patient assignment patterns, I believe I sufficiently control for selection on observables, and that unobservables are of limited importance.

It is not entirely straightforward to confirm this formally. I would like to perform the test proposed in Oster (2019) and Altonji, et al. (2011), but that requires the use of a measure of model fit such as R^2. Because I use LASSO to select covariates, the reported R^2 is not correct because it does not take into account uncertainty in covariate selection. I proceed regardless of this limitation and use the R^2 as if there was no uncertainty, which means test results will be biased towards rejecting the null of no treatment effect because I will be overestimating the improvement in model fit from including observable covariates. With these caveats in mind, results suggest that the size of omitted variables bias in this context are small. For instance, when considering length of stay for complex patients, the improvement in R^2 from going from a specification with only physician fixed effects to the covariates chosen by post-double-selection is from 0.0206 to 0.1057, and the coefficient on experience decreases in magnitude from -0.055 to -0.048. If I assume that the maximum R^2 that can be explained by the model is 0.3 (in other words, outside "randomness" such as congestion explains the other 0.7), then we would need the omitted variables to have 3.21 times the amount of selection as the observable factors to obtain the results I have if the true effect was zero. If I assume the maximum R^2 is 0.5, then the omitted variables would need to have 1.66 the amount of proportional selection, whereas if the maximum R^2 is 1, then the omitted variables would need to have 0.75 the amount of proportional selection to obtain the results I have if the true effect is zero. Based on the qualitative arguments based on the context I outlined previously, I find these magnitudes to be unlikely.

The previous analyses were all at the encounter level, which represents within-patient learning. However, there is one additional margin of importance: the ability to manage multiple patients simultaneously. Figure 3, Panel (a) shows growth in "multitasking." These are not regressions but instead means for shifts worked by residents of each month of time in the program ("tenure"). Panel (a) shows that residents only average about 1.5 patients per hour when they begin residency and improve such that they are managing about 3 patients per hour by the fourth year of residency. Panel (b) combines this result with the improvement in individual patient throughput and shows the total improvement in productivity, as measured by the number of patients seen per shift. Patients in Panel (b) are split into three groups of ex-ante severity, as assigned by the triage

Table 3: Learning over Time: Regressions

	ED Readmissions		log(Medical Orders)		Frac. Signed by Res.		log(Mins to 1st Order)		log(Length of Stay, Hours)			
Cumul Pt Share	-0.004	-0.004	0.006	-0.004	0.016***	-0.004	-0.067**	-0.004	-0.022	-0.031**	-0.020	-0.007
	(0.003)	(0.004)	(0.015)	(0.004)	(0.006)	(0.004)	(0.028)	(0.004)	(0.018)	(0.015)	(0.012)	(0.009)
Tenure (Years)	-0.004	-0.003	0.010	-0.003	0.016***	-0.003	-0.074***	-0.003	-0.023	-0.031**	-0.018	-0.006
	(0.004)	(0.004)	(0.016)	(0.004)	(0.006)	(0.004)	(0.028)	(0.004)	(0.018)	(0.014)	(0.013)	(0.009)
log(Cumul Pt Share)	-0.003	-0.000	0.021	0.015*	0.009**	0.010***	-0.085***	-0.079***	-0.053***	-0.047***	-0.000	0.004
	(0.003)	(0.004)	(0.015)	(0.008)	(0.004)	(0.003)	(0.020)	(0.012)	(0.013)	(0.011)	(0.011)	(0.007)
Tenure (log Years)	-0.001	0.002	0.039**	0.024**	0.010**	0.013***	-0.117***	-0.101***	-0.055***	-0.048***	0.002	0.004
	(0.004)	(0.004)	(0.016)	(0.011)	(0.005)	(0.004)	(0.028)	(0.016)	(0.015)	(0.015)	(0.011)	(0.009)
Patient Controls		X		X		X		X		X		X
DepVar Mean	0.140		19.307		0.596		39.485		6.669		7.653	
ED Disposition	Discharged		All		All		All		Complex		Simple	
Obs	22,751	22,712	31,610	31,547	31,610	31,547	27,238	27,184	8,877	8,853	22,751	22,712

Notes: Regressions of selected patient outcome and process measures on various measures of resident experience. The sample consists of all patients seen by EM residents. Every regression includes provider fixed effects. Other Patient Controls are chosen from the set of immutable and ex-ante patient covariates using the post-double-selection LASSO method of Belloni, et al. (2014) and differ from the covariates used in the binned scatterplots. Coefficients shown are from four separate regressions. Yearly experience is time in the program. Patient share is based on quarterly cumulative patient shares and only increments each quarter and are inferred using length of stay outcomes for complex patients for those who begin the program prior to the data begins. Both experience measures enter in separate regressions linearly (top section) as well as in logs (bottom section). The 14-day ED readmission rate is the rate at which patients who are discharged home from the ED have a repeat visit within 14 days. By definition, the measure only exists for discharged patients. log(Medical Orders) is the natural logarithm of the sum of diagnostic and therapeutic orders signed in the ED. Frac. Signed by Res is the fraction of medical orders that are signed by the resident, rather than the attending, nurses, or other residents assisting. log(Mins to 1st Order) is the time between the moment the patient is moved from the waiting room to an exam room and the time that the first medical order is signed. This value is missing if the first order is signed prior to being roomed; see Appendix Table A1 for the extensive margin. log(Length of Stay) is the natural logarithm of the hours the patient spent in the ED under the care of EM providers. It is split into "complex" and "simple" patients based on an ex-ante prediction of inpatient admission. Dependent variable means are listed, always in levels. Standard errors are clustered by physician. See text and Appendix A and B for additional details.

Table 4: Learning Over Time Regressions, Split by Patient Complexity

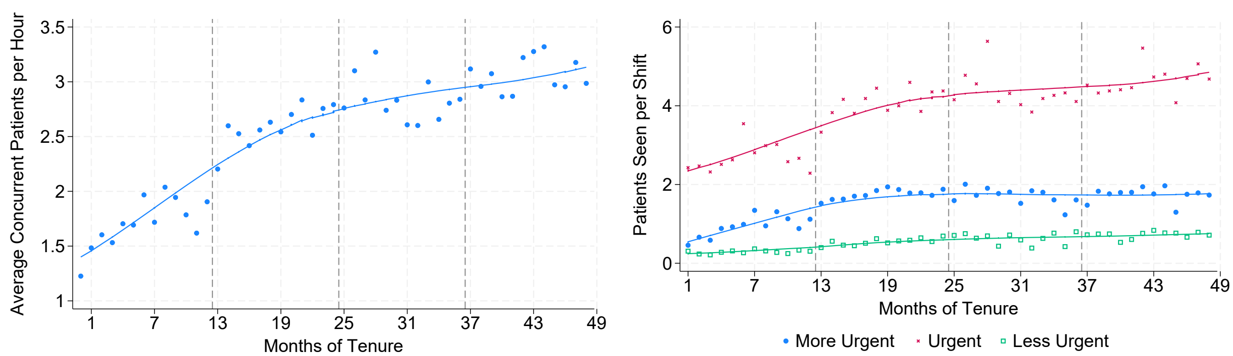
	log(Medical Orders)				Frac. Orders Signed by Resident				log(Minutes to 1st Order)			
Cumul Pt Share	-0.028**	-0.018	0.013	0.040***	0.025***	0.021***	0.014**	0.009**	-0.034	-0.057**	-0.075**	-0.090***
	(0.014)	(0.014)	(0.016)	(0.012)	(0.006)	(0.006)	(0.007)	(0.004)	(0.038)	(0.029)	(0.029)	(0.017)
Tenure (Years)	-0.028*	-0.019	0.014	0.040***	0.025***	0.022***	0.014**	0.010**	-0.064*	-0.081***	-0.072**	-0.087***
	(0.015)	(0.014)	(0.017)	(0.012)	(0.006)	(0.006)	(0.007)	(0.004)	(0.037)	(0.028)	(0.029)	(0.017)
log(Cumul Pt Share)	-0.036***	-0.037***	0.029*	0.039***	0.017***	0.016***	0.008**	0.006*	-0.084***	-0.080***	-0.079***	-0.077***
	(0.012)	(0.012)	(0.016)	(0.010)	(0.006)	(0.005)	(0.004)	(0.003)	(0.025)	(0.024)	(0.023)	(0.014)
Tenure (log Years)	-0.040***	-0.041***	0.047**	0.052***	0.025***	0.024***	0.008	0.007*	-0.129***	-0.118***	-0.102***	-0.094***
	(0.014)	(0.015)	(0.020)	(0.012)	(0.008)	(0.006)	(0.005)	(0.004)	(0.040)	(0.033)	(0.030)	(0.018)
DepVar Mean	29.138		15.467		0.488		0.638		27.626		43.931	
ED Disposition	Complex		Simple		Complex		Simple		Complex		Simple	
Controls		X		X		X		X		X		X
Obs	8,860	8,836	22,750	22,711	8,860	8,836	22,750	22,711	7,421	7,402	19,817	19,782

Notes: Regressions of selected patient process measures on various measures of resident experience, split by ex-ante predicted patient complexity, but is otherwise similar to Table 3. As a reminder, coefficients shown are from four separate regressions. Dependent variable means are listed, always in levels. log(Medical Orders) is the natural logarithm of the sum of diagnostic and therapeutic orders signed in the ED. Frac. Signed by Res is the fraction of medical orders that are signed by the resident, rather than the attending, nurses, or other residents assisting. log(Mins to 1st Order) is the time between the moment the patient is moved from the waiting room to an exam room and the time that the first medical order is signed; see Appendix Table A1 for the extensive margin. This value is missing if the first order is signed prior to being roomed. Standard errors are clustered by physician.

nurse upon patient arrival. The blue circles represent the most urgent patients, and comprise about 20% of all arriving patients. We see growth in these complex patients during the first year that levels off after residents enter their second year. The red X's represent the middle category of urgency, which comprise about 60% of all patients, and growth continues throughout the program. The hollow green squares represent the least urgent patients, who comprise the remaining 20% of patients. Growth in these patients is minimal. Taken together in conjunction with the encounter-level length-of-stay improvements, I conclude that the growth in productivity is driven by the ability to manage additional patients simultaneously, rather than improvements in individual patient length of stay. Second, the growth in patients seen is mostly for the middle category of patient urgency; the growth for complex patients appears to level off after the first year. Finally, EM residents do not see many less urgent patients; these are left for non-EM residents, nurse practitioners, physician assistants, and internal medicine attendings.

Figure 3: Patient Load Breakdown

(a) Simultaneous Patients Managed (b) Patients Seen per Shift



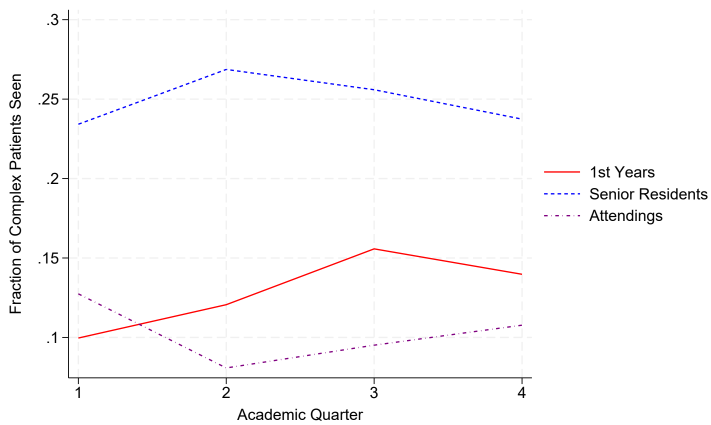
Notes: These figures show the evolution of patient load over the 48 months of the EM residency program. Panel (a) depicts the average number of patients managed per hour of each shift for residents of each month in the program. Panel (b) shows the breakdown of total number of patients managed during the shift, broken down into three ex-ante acuity measures assigned independently by the triage nurse. "More Urgent" patients comprise approximately 20% of all patients; "Urgent" patients approximately 60%, and "Less Urgent" patients the remaining 20%. Comparing the two panels reveals that the growth in patients seen per shift is mainly in the majority category of patients, not the less-urgent, and that it is driven by managing additional patients in parallel rather than large increases in speed per patient.

The preceding analyses suggest that in terms of patient care, EM residents grow in two dimensions. The first is improvement in medical skill: the processes of gathering, synthesizing, and interpreting information about the patient's underlying state and signing the correct set of orders given that information. Tables 3 and 4 show that these improvements are primarily for treating complex patients. The second is improvement in "bandwidth" or cognitive capacity: residents become able to manage additional patients simultaneously, and most of the growth in capacity is for less complex patients. I am primarily interested in the improvement in medical skill for individual patients, which I believe is a more appropriate application of the learning by doing and task allocation frameworks. Because the majority of improvement in medical skill is for complex patients, attendings can affect the trade-off between care quality and training by changing the allocation of complex and simple patients among residents of varying experience and themselves. Are attendings aware of the trade-offs they are making?

The answer appears to be yes. Figure 4 plots the average fraction of complex patients seen by individual providers during each shift across the four quarters of the academic year. This figure illustrates three interesting facts. First, the fraction of complex patients seen increases from 10% to 15% during the first year (solid red line), corroborating the results of Figure 3.

Second, again in line with Figure 3, the share seen by the other three cohorts of residents is relatively stable during the year. Third, it is attendings who "pick up the slack" in July through September and see the patients that the first year residents are unable to treat.[16] I also provide suggestive evidence that attendings are aware of the trade-off on a more micro level via correlational logit regressions. In these regressions, I regress the probability that a first year resident is assigned a complex patient on the number of complex patients currently being seen in the ED, the number of patients in the waiting room, fixed effects for the patient's chief complaint, and other characteristics of the physicians on staff and the index patient. Results are in Appendix Table A4. I find that first year residents are much less likely to be assigned complex patients when there are many patients in the waiting room. As the number of patients in the waiting room increases from the 25th to 75th percentile, the probability that first year residents are assigned a complex patient decreases by 15%. Not shown are the coefficients on patient chief complaint. I find that *ceteris paribus*, first year residents are much more likely to be assigned patients from more common chief complaints (e.g. chest pain, abdominal pain, and shortness of breath) compared to the pooled "less common" category. Together, these findings suggest that the hospital is aware of the costs of teaching because they teach less when the costs are higher due to congestion, and that the hospital is aware of the benefits because they train residents in the patients they are most likely to see first.

Figure 4: Average Fraction of Complex Patients Seen, by Role



Notes: This figure depicts the average fraction of complex patients seen per shift, by role, for each quarter of the academic year. Complex patients are those with the highest values of ex-ante predicted admission. "Senior Residents" are the average shares of residents in years 2-4. This choice is informed by the results in Figure 3, Panel (b) and Appendix Figure A1, Panel (b), where the share of "Most Urgent" and Admitted patients does not continue growing after the first year. The figure shows that first year residents see more patients as they gain experience, but that the patients they are not able to see in quarter 1 are seen by attending physicians rather than other residents. Not shown is that there are no meaningful differences in staffing or arriving patient composition across the academic year.

Broadly, resident progress can be divided into two categories: patient-relevant and not patient-relevant. The main patient-relevant change is complex patients' length of stay in the ED, which decreases by approximately 9.2% over the four-year program. Outcomes, as measured through ED readmissions, are unchanged. The time to first order does decrease, but the average magnitude is only about five minutes, so it is relatively unimportant. Patients are not affected by who signs orders for them, so the fraction of orders signed by the resident is not a patient-relevant outcome. Arguably, they are also relatively insensitive

[16] Indeed, there is no difference in the composition of arriving patients across the year.

to the number of orders signed, insofar as it does not affect their outcomes and the change in out-of-pocket cost is small due to insurance coverage. It is also ambiguous whether the hospital desires a reduction in orders signed as this depends on how the payer will reimburse them. I return to reimbursements in the first counterfactual.

In summary, these facts suggest that learning by doing is most important for learning how to treat complex patients, and that the main patient-relevant cost of training is length of stay for complex patients. Therefore, the training environment can be described as follows: residents are more or less capable at treating simple patients when they begin the residency program. However, they need to learn how to diagnose and treat complex patients, but the only way to learn is to learn by doing: by treating complex patients. Attendings are aware of this, and also of the primary trade-off: inexperienced residents are slower than experienced residents. But because learning is concave, inexperienced residents gain more from seeing each patient, and there is additional time left in the program for the hospital to benefit from their skill compared to more senior residents. Therefore, the hospital strategically allocates complex and simple patients to inexperienced residents, experienced residents, and attendings working independently in order to maximize the discounted sum of its stream of payoffs. I formally describe and quantify the nonprofit teaching hospital's dynamic optimization problem in the next sections.

5 Dynamic Framework

In this section, I present a dynamic model of patient allocation that takes the progress from learning estimated in Section 4 as given. It is necessary to consider dynamics because I am interested in estimating how the hospital allocates patients to trade off current care quantity and quality and future care quantity and quality via training. Unless the hospital acts myopically, a static model cannot capture these trade offs because it does not take into account future benefits of training. In other words, forward-looking hospitals take opportunity costs and future benefits into account when optimizing patient allocation. The dynamic choice model is a discrete-time, infinite-horizon model, where the state-space, resident skill, evolves akin to overlapping-generations models.

Each shift, attendings first observe the the skill of residents who were assigned to work. An infinitesimally divisible unit mass of complex patients arrives and attendings choose a share of patients to assign to each resident and themselves. Attendings help residents see patients and may also see some patients independently. Patient utility, a function of length of stay and therefore a function of resident skill, is realized. At the end of the shift, resident skill increases by the share of patients they saw. Each July 1, 4th years graduate and are replaced by new 1st years with zero skill. Attending skill is fixed.

The hospital's choice of patient allocation share is the stream of allocations that maximizes

$$\sum_{t=0}^{\infty} \beta^t [u(S_t|X_t) + \varepsilon_{St}]$$
$$\text{subject to } X_{t+1} = \begin{cases} (0, x_{t1} + s_{t1}, x_{t2} + s_{t2}, x_{t3} + s_{t3}) & \text{if } t = \text{June 30} \\ X_t + S_t & \text{otherwise} \end{cases} \tag{2}$$

The state variable, X_t, evolves as described above: within the academic year, skill in the next period is simply skill in the current period plus the share of patients seen in the current period. On July 1, the fourth years graduate, the continuing three cohorts are promoted, and the new first years who join enter with zero skill. Flow utility u is a function of the choice of patient allocation shares in period t and is conditional on the state of resident skill X_t and also includes a portion that is observable to the hospital staff but not to the econometrician, ε_{St}. This term is different for every allocation choice S and period t and could reflect things such as congestion or features of the patient that make them particularly suitable or costly for training that I do not observe.

This utility function leads to the standard Bellman equation given by

$$V(X,q) = E_{\varepsilon}\left[\max_{S}\{u(S|X) + \varepsilon_{St} + \beta V(X', q')\}\right] \tag{3}$$

where I explicitly separate out the resident's knowledge state X and the current quarter of the academic year q. I do this to make clear that the value of being in state X differs based on which academic quarter q the current period t is. That the quarter q affects the value of being in knowledge state X is intuitive: consider a state $X = [0, 0]$. In this case, it is far less undesirable for the hospital to be in this state in the first quarter, where it can still train the residents, than it would be for the hospital to be in this state in quarter four, where the training utility for the senior residents are about to be realized.

6 Estimation and Identification

6.1 Simplifying Assumptions and Parameterizations

In order to make progress, I make some simplifying assumptions and consider a set of specifications and parameterizations for the flow utility function. I describe these assumptions and the rationale behind them in this subsection.

The first set of simplifying assumptions is made in order to keep the state-space and action-space manageable. I assume that the program lasts for two years, so that there are only two cohorts: new residents, or "juniors," that I denote with subscript j, and residents who will graduate at the end of the academic year, or "seniors," that I denote with subscript s. I further assume that there is no within-cohort variation in skill. I also take the time period t to be a quarter of the academic year and assume that a unit mass of patients arrives each year, evenly distributed among the four quarters. I also discretize both the state space of resident knowledge and the choice variable of share of patients to assign.

These simplifications are necessary for the following reasons. First, managing a continuous choice of patient share and experience is intractable when taking first-order conditions is not possible, which applies here because the value function is unknown (V in Equation (3)). Therefore, I only allow the hospital to choose a finite set of patient allocations, which also keep the cumulative resident skill on a grid. In practice, a mass of 0.25 patients arrives each quarter, and I allow the hospital to assign one of 52 predetermined values of patients, evenly spaced from 0 to 0.25, and such that the sum of allocations to all providers equals 0.25. I censor the upper bound of the knowledge space to 1.2 because the maximum steady-state allocation for rising senior residents is 1.0 (achieved if every patient in every quarter is assigned to them). The interval of knowledge space $[0.01, 1.2]$ is divided into 250 possibilities. I choose 0.01 as the starting value because residents begin transitioning into the program in June prior to their first year, and they see a share of complex patients that is equivalent to about 0.01 when dividing by the full quarter, and because I need a nonzero value for the natural logarithm of experience. Note that the full state-space must contain one dimension for each cohort, and must have one such array for each time period.[17] Even with just two cohorts and time in quarters, the state space array with my discretization has dimension $[250, 250, 4]$.

The second assumption is made on the steady-state. I assume that the hospital is in the steady-state and that the steady-state is such that training each year is identical to training every other year. This rules out cases where the hospital alternates between training for instance cohorts that enter during even years and ignoring the cohorts that enter on odd years. I make this assumption for two reasons. First, it runs counter to intuition that a teaching hospital would treat all cohorts similarly, and this is borne out in the data in terms of patient assignment (see Appendix Figure A2). Second, it facilitates estimation because it does not require me to infer the skill of the rising seniors when I do not observe their full patient history.

I consider two parameterizations for the hospital's flow utility function u in Equations (2) and (3). The first is one where the hospital maximizes a weighted sum of patient care quality and resident training. The second is one where the hospital maximizes training subject to a lower bound of care quality. I depart slightly from the canonical specifications of Newhouse (1970) and Lakdawalla and Philipson (1998) and do not separately include revenue. I do this because I find in Section 4 that the only meaningful change in patient-relevant outcomes with respect to experience is length of stay. Therefore, length of stay is the measure of care quality the hospital trades off in exchange for greater training. But length of stay is also inextricably linked with revenue: in general, hospitals receive revenue for each additional patient they see, so seeing additional patients, which in the short-run can only be done with shorter length of stay because facility size is fixed, is the only way to increase revenue. To avoid double-counting quality and revenue, I include it in the utility function only once.

The first utility function I consider is where the hospital maximizes a weighted sum of patient care quality and resident training.

$$\max_{S} \sum_{t=0}^{\infty} \beta^t [\phi L(S_t; X_t) + (1-\phi) K(S_t; X_t, q(t)) + \varepsilon_{St}] \tag{4}$$

L is the hospital's utility from length of stay (care quality). This is a function of the allocation of patients S_t and the state variable of resident knowledge X_t. K is the hospital's utility from training ("knowledge"), which depends on the allocation of patients S_t, the state variable X_t, and the result of a function q that maps time t to quarter of the academic year. I write K in this manner in order for it to depend on the quarter. The weight on care quality ϕ is to be estimated.

The second utility function I consider is where the hospital maximizes training subject to a lower bound of care quality.

$$\begin{aligned} \max_{S} \quad & \sum_{t=0}^{\infty} \beta^t [K(S_t; X_t, q(t)) + \varepsilon_{St}] \\ \text{such that} \quad & L(S_t; X_t) \geq L^* \text{ for all } t \end{aligned} \tag{5}$$

Again, L is the hospital's utility from length of stay (care quality). In this specification, it must meet some minimum quality threshold L^* in every period t. The value of L^* is to be estimated. Given L^*, the hospital conditionally maximizes K, the utility from training.

In both specifications, utility from care quality L is the average length of stay utility f for patients, given allocation S and skill X.

[17] This is because the value to the hospital of any level of resident skill is potentially different depending on when in the academic year it is.

$$L(S_t; X_t) = s_j f(x_{tj}) + s_s f(x_{ts}) + (1 - s_{tj} - s_{ts}) f(x_A) \tag{6}$$

This is simply the share of patients allocated to the junior residents, senior residents, and attendings in each period t, times the length of stay utility f for providers of each skill. Utility from training K is the cumulative skill of the graduating senior residents in academic quarter four. That is,

$$K(S_t; X_t, q(t)) = \begin{cases} 0 & \text{if } q(t) \neq 4 \\ f(x_{ts} + s_{ts}) & \text{if } q(t) = 4 \end{cases} \tag{7}$$

In this parameterization, K will only be nonzero in academic quarter four. Note does this does not automatically mean the hospital chooses not to train if it is not in academic quarter four because the value function $V(X, q)$ will generally be increasing in the state X within each quarter: higher levels of skill at any point enable the residents to achieve a higher skill upon graduation. It only means that the hospital does not directly derive utility from resident skill within the academic year–it does not benefit from increased resident skill within the academic year above its impact on length of stay utility L. I will consider three functional forms for f that vary in concavity: linear, quadratic, and log, as delineated in Table 5 Panel (b).

Table 5: Summary of Estimation Parameterizations and Methodology

(a) Parameters to Estimate and Methodology

Category	Unknowns	Estimation Methodology
Learning rate	$\{\alpha_0, \alpha_1\}$	OLS in panel data, "offline"
Attending speed	x_A	Back out from panel data using $\hat{\alpha}$
Discount rate	β	Calibrated; yearly $\beta = \{0.90, 0.95, 0.99\}$
Weight on quality vs. training	ϕ	Dynamic, match patient shares averages
Lower bound on quality	L^*	Dynamic, match patient shares averages

(b) Parameterizations of Length of Stay Utility f

Linear	$f(x) = -\alpha_0 x^{\alpha_1}$
Quadratic	$f(x) = -(\alpha_0 x^{\alpha_1})^2$
Log	$f(x) = \log(C - \alpha_0 x^{\alpha_1})$

Notes: This table enumerates the unknown parameters and the estimation method employed in order to estimate them, as well as the functional forms for the length of stay utility function f. In Panel (a), the first section lists the parameters to be estimated in "offline" in panel data without any dynamics. The middle sections shows that the discount rate is calibrated using various reasonable yearly values, as it is not well-identified in the dynamic model. The final section shows the two parameters that are estimated using the dynamic model and take the offline parameters as fixed; these come from two different utility functions that the hospital may use. Panel (b) lists the three functional forms used for the length of stay utility f. Note that the linear and quadratic parameterizations do differ because the shape parameters α are fixed in the offline estimation. C is a constant chosen to ensure that $C - \alpha_0 x^{\alpha_1}$ is positive for all values of skill x. See text for additional details, as well as Subsection 6.2 for more details on offline estimation and Subsection 6.3 for more details on the dynamic estimation.

Estimation proceeds in two steps in the spirit of Hotz and Miller (1993), Bajari, et al. (2007), and Pakes, et al. (2007). The unknowns and methodology are summarized in Table 5 Panel (a). In the first, "offline," step, I estimate the parameters relating to the learning rate $\{\alpha_0, \alpha_1\}$ using OLS in the panel data, and infer attending speed x_A using the estimates $\hat{\alpha}$. Then, for three candidate values of the discount rate[18] β, I find via iteration the weight on quality vs. training in the utility function given by Equation (4) and the lower bound on quality in the utility function given by Equation (5) that produces optimal patient assignment shares most closely to the observed shares. Note that the linear and quadratic parameterizations of f differ because the shape parameters α are fixed in the offline estimation.

6.2 Step 1: Offline Parameter Estimation

I begin by estimating the learning parameters "offline," outside of the dynamics, via OLS. The goals are to recover how patient length of stay improves with experience with complex patients, as measured with the share of complex patients and to estimate the skill of attendings working alone. Two factors make this not entirely straightforward. First, I must restrict to the subset of residents who begin the program during the sample period because I do not observe the resident's full patient history otherwise. Second, the residency schedule is such that the residents do work at another hospital in the city that I do not have data from so I must infer the total fraction of complex patients seen. I first describe the assumptions I make and then the tests I do in order to test the validity of the assumptions.

The coefficients recovered by OLS are unbiased under the same assumptions on omitted variables as outlined in Section 4. Two additional assumptions are necessary in this setting. First, I assume that the natural logarithm is the correct functional form for resident progress with respect to the cumulative fraction of complex patients seen. Second, I assume that errors in patient assignment inference are also uncorrelated with patient share. In other words, residents who see additional patients relative to their peers at UCSF also see similar "excess" patients at ZSFG. Both assumptions are fundamentally untestable but I offer arguments in favor of accepting them.

In order to test the validity of the functional form assumption, I compare the quarterly patient share results for the subset of residents with the results using years in the program as the measure for experience. The rationale behind this is that years in the program is a "reduced-form" measure of share of complex patients, as much of the variation is in the time-series rather than in the cross-section. If the results for continuous tenure are similar to that of quarterly patient share, then I gain confidence in the validity of my assumptions. Based on the results in Table 3, I believe this to be true and I describe and perform a more detailed analysis in Section 7.

Because UCSF EM residents work in two locations but I only have data from one, I must infer patient assignment at the other location. The assumption I make is that patient assignment at the other location (Zuckerberg San Francisco General Hospital, ZSFG) mirrors patient assignment at the observed location (UCSF). In other words, if a resident sees 12% of patients at UCSF within a period, I assume they are also seeing 12% of patients at ZSFG. Importantly, this means that attendings are not assigning patients in a mean-reverting manner or that observed differences at UCSF are not magnified at ZSFG. I find suggestive evidence for this assumption by leveraging the law of large numbers and showing that patient allocation shares within a cohort are no less dispersed towards the end of the academic year (May and June) compared to the beginning (July and August). Furthermore, I believe factors such as the ad-hoc team structure and variation in congestion and patient arrivals

[18] I choose not to estimate the discount rate, which is a common choice in the dynamic model literature. For instance, Pakes, et al. (2007) writes, "We usually think that the prior information we have on δ [the discount rate] is likely to swamp the information on δ available from estimating an entry model."

make this assumption reasonable, as it is difficult for the rotating attendings to know the resident's history and adjust their assignment instructions accordingly.[19] To lessen the impact of this assumption as well as differences due to exogenous factors such as congestion, I use as the measure of patient-specific experience the average share of complex patients seen during the calendar quarter. This measure has considerably less variation than experience at the two-week level, but still contains some variation, as can bee seen in Appendix Figure A3.

Now, I describe how I infer attending skill with the learning parameters in hand. Conceptually, this is simple as I observe the length of stay for complex patients seen by attending physicians and I know the functional form of learning, so I can just take the inverse of that function. However, the inclusion of physician fixed effects makes this complicated as the concept of the regression constant is not well-defined. Therefore, what I do is I infer attending skill for every estimated physician fixed effect and patient seen by an attending, and use the grand mean of the estimates as the estimate for attending skill.

6.3 Step 2: Dynamic Parameters

The goal of estimation is to find the unknown parameter θ (either ϕ or L^*, depending on the specification) that gives the best fit between the model-predicted optimal quarterly patient shares and the observed quarterly patient shares. The metric of fit used is RMSE, with each quarter receiving equal weight. In other words, I find the value of θ that minimizes:

$$\sum_{t=1}^{4} \sqrt{(s_{tj} - M_m(\theta;\beta)_{tj})^2 + (s_{ts} - M_m(\theta;\beta)_{ts})^2 + (s_{ta} - M_m(\theta;\beta)_{ta})^2} \tag{8}$$

where the subscripts j, s, and a represent the shares assigned to the junior resident, senior resident, and attending. s_t are the observed patient allocation shares, and M_m is the model-predicted steady-state shares given the parameter θ and discount rate β for model m. The two models considered are the weighted sum of patient utility of Equation (4), where $\theta = \phi$ and the constraint maximization of Equation (8), where $\theta = L^*$.

In order to find the value of θ that minimizes Equation (8), I perform a grid search over values of θ. For each value of θ, I first perform value function iteration on Equation (3) in order to solve for $V(X,q)$, the value of being in knowledge state X in academic quarter q. I then use the estimated $V(X,q)$ in conjunction with the flow utility to find the optimal patient allocation choice S for each X and quarter q: $S(X,q)$. Finally, I find the steady-state given the allocation choices S. That is, I search for a value of rising senior knowledge $\sum_{t=t'}^{t'+4} s^*_{tj}$, the cumulative share of patients seen in their first year, such that the optimal training results in the new cohort of junior residents finishes the first year with the same knowledge. In my notation, I search for $\sum_{t=t'}^{t'+4} s^*_{tj}$ that satisfies for all t':

$$\begin{gathered}\sum_{t=t'}^{t'+4} s^*_{tj} = \sum_{t=t'+5}^{t'+8} S(X^*_t, q(t))_j \\ \text{such that } X^*_{t+1} = X^*_t + S(X^*_t, q(t)) \\ \text{and } t' \text{ is the beginning of an academic year}\end{gathered} \tag{9}$$

[19]According to EM residents at UCSF, many decisions regarding progress are made at the cohort level. For example, at the beginning of second year all residents are expected to take on additional patients and there is limited "personalization" of this directive based on individual progress. This is unsurprising because of the ad-hoc team status and because there are 60 EM residents for the various attendings to keep track of.

The left hand side is the starting knowledge of the rising seniors, which is equal to the cumulative share of patients seen in their first year. The right hand side is the sum of patient shares seen by the new juniors in the next academic year, because S is the function that maps knowledge X and time q to a vector of optimal patient assignment decisions and skill X accumulates in the usual way.

As mentioned above, I convert the continuous choice of patient share to a finite set of possibilities. The attendings have a choice of 52 discrete values spanning $[0, 0.25]$, and patient allocation shares map to a grid of knowledge with 250 values spanning $[0, 1.2]$. Because I am interested in the steady-state, it is not useful to extend this further: the maximum possible steady-state is 1 if all patients are assigned to the junior resident each year. I iterate until the average L2-norm (Euclidean distance) between successive elements of the value function is less than 10^{-6}. Although the grid search over possible values of θ is cumbersome, this method has the advantage that I in theory do not risk finding a local minimum rather than the global minimum. In practice, I begin with a relatively coarse grid and perform a finer grid search around the minimum given by the coarse grid.

7 Results

In this section, I first present and discuss estimates of the offline estimation and provide evidence in support of my assumptions. Then, I discuss the results for the dynamic estimations.

Results of the offline OLS estimation of the learning parameters are in Table 6. My preferred estimates are the bolded set in the rightmost column, which are the results using the natural logarithm of quarterly patient share with physician fixed effects and patient controls. Starting from the bottom, the results suggest that the average attending physician has skill similar to a resident with a cumulative experience share of 0.89 patients (recall that a mass of 1 patient arrives each year). While this may be lower than expected, this measure includes interruptions to attending speed due to supervisory duties. Next, in terms of learning speed, the results suggest that for each 1% increase in cumulative quarterly patient share, patient length of stay will decrease by about 0.04%. These results are in-line with the previous results in row three of Table 3, and the remaining columns of Table 6 are a validation exercise.

I gain more confidence in the assumptions underlying these estimates by comparing them to estimates from the full sample of residents and for continuous measures of experience. The first pair of columns presents the results using years in the program as a continuous measure and for the full sample of residents, which represent the baseline. Next, I restrict to full history residents and find minimal changes in the estimated coefficients. Coarsening the years of experience measure to quarterly snapshots results in coefficient estimates of increased magnitude, but the standard errors are large enough that I cannot reject that they are equal to the coefficients from the continuous measures. Similarly, changing the measure of experience to patient share does not create a large difference in the estimates. The estimated attending skill is similar for all specifications other than the full resident sample, which I believe is due to some imprecisely estimated resident fixed effects that have outsize influence on the grand mean of inferred attending skill. Values across the other specifications with patient controls are all similar.

In the dynamic model, I find that the specification where the hospital maximizes training subject to a care quality constraint fits the data far better than the specification where the hospital maximizes a weighted average of care quality and training.

Table 6: Offline Parameter Estimates of Learning Speed and Attending Skill

	log(Patient Length of Stay, Hours)							
Experience Type	Tenure (continuous)				Tenure (quarterly)		Patient Share (quarterly)	
α_1: log("Experience")	-0.030	-0.050	-0.073	-0.059	-0.127	-0.105	-0.054	**-0.043**
	(0.013)	(0.015)	(0.015)	(0.019)	(0.029)	(0.036)	(0.012)	(0.015)
α_0: Constant	1.674	1.692	1.652	1.694	1.665	1.711	1.606	**1.623**
	(0.013)		(0.025)		(0.025)		(0.030)	
Sample	All		Full History Residents					
User ID and Patient Ctrls		Y		Y		Y		Y
Observations	11,520	11,520	1,202	1,202	1,202	1,202	1,202	1,202
Inferred x_A	1.47	1.39	0.86	0.88	1.02	0.77	0.37	**0.89**

Notes: This table shows the results of the offline estimates for the parameters governing learning speed as well as the value of attending skill inferred using these estimates. The bolded estimates in the final column are what are used in the dynamic estimation as they use the desired measure of experience: share of complex patient seen. This measure is only available for residents who begin the program during the sample (the 2018 and 2019 cohorts). In order to test that this sample of residents is not significantly different from the full sample and that discretizing experience to quarterly intervals similarly does not result in different estimates, I begin with the full sample of residents and a continuous measure related to patient share: years in the program. This first pair of regressions is similar to the regressions in Table 3. I then restrict to the 2018 and 2019 cohorts and find that the coefficients do not change much in the specification with physician fixed effects and patient controls. Similarly, I cannot reject equality of coefficients when I only allow tenure to change each quarter, and again when using the quarterly patient share definition of experience. Standard errors are clustered by physician in the first two columns but are the greater of the clustered and robust otherwise because there are fewer than 40 physicians.

It turns out that the weighted average specification of Equation (4) actually does not result in a stable steady state of training. Instead, the steady-state utility-maximizing patient allocation is a two-year cycle where every other cohort is trained. The result holds for all chosen values of the discount rate, as well as all three concavities of utility from length of stay and making the utility from learning more concave than the utility from patient care. This is because for each cohort, training today and training tomorrow are intertemporal complements. A larger amount of training today means that the cost of training tomorrow is decreased because the residents have higher skill. The presence of attending physicians amplifies this feature: instead of distributing training among the two cohorts, it is better to focus it on one cohort and give the remaining patients to the attending to maximize patient utility.

To build intuition for this result, consider a model where there is only one period per academic year. In the first year, if the senior cohort received training in year zero, then it is cheaper to train them further than to train the new cohort, and it is better for patient outcomes if the hospital allocates patients only between the senior cohort and the attendings. In the second year, the senior residents have zero skill because they received no training when they were junior residents. The hospital chooses to ignore them, trains only the new cohort, and gives the remaining patients to the attendings to maximize the patient utility portion of utility. Then, in year three, it trains the senior cohort, because it has already trained them in the previous year, and again ignores the new cohort. Consequently, every other year, both training utility and patient utility are high and we have a two-year cycle.

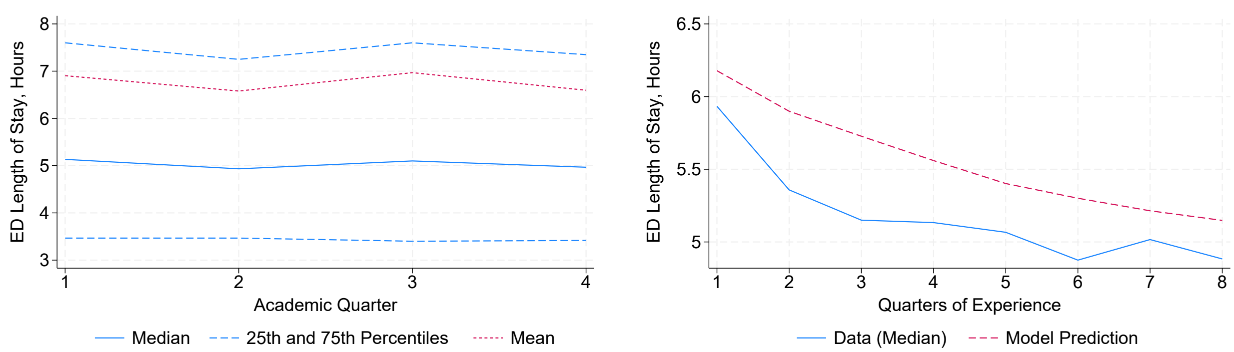
Results for the dynamic model where the hospital maximizes training with respect to a lower bound of care quality are similar across specifications and discount rates. Regardless of the specification or discount rate, the estimated lower bound, converted from utils into hours, is always around 5.33 hours per patient. The model that fits the data best is the specification with quadratic utility from length of stay and a yearly discount rate of $\beta = 0.95$. Full results can be seen in Appendix Table A6. The model fits the qualitative patterns of increasing allocation to first year residents and decreasing allocation to attendings well, but the gradient for both is steeper than observed in the data. This can be seen in Appendix Figure A4.

I next evaluate model fit by examining how well it fits non-targeted moments. Specifically, I examine how the length of stay predictions compare to length of stay averages in the data. First, I examine the average length of stay across the academic year. The estimated quality bound of 5.33 hours is less than the raw mean and median length of stay in the data. The reason it differs is related to the fact that the inferred constant is using the set of first year residents, but the data contains additional residents that are not used in the estimation. Therefore, when assessing model fit, I will ignore the levels and focus instead on matches with changes with over calendar time. The model predicts that length of stay is stable over the academic year, as the quality bound binds in every quarter. Figure 5 Panel (a) shows that the median ED length of stay is very stable with respect to academic quarter, just as the model predicts. The mean shows more movement, but that is driven by the top 25 percent of encounters and potentially related to encounters where patients were in worse condition than expected or the affected by the arrival of a code patient in critical condition who demanded the attention of the entire ED. Next, I examine how average length of stay varies across quarters of experience, and compare it to the model predicted values. Figure 5 Panel (b) shows that see that average length of stay by experience has similar shape as the median length of stay for residents with each level of experience. The similarities between average length of stay in the data and predicted length of stay in the model were not a moment that was targeted in the estimation–only patient share assignment was–and the comparisons give me more confidence in the model.

Figure 5: Model Fit: Non-Targeted Moments

(a) Average Length of Stay

(b) Average Length of Stay by Experience



Notes: These figures show how well the model fits the non-targeted moments relating to patient length of stay. In Panel (a), we see that the median ED length of stay is very stable with respect to academic quarter, just as the model predicts. The mean shows a slightly higher value in the first quarter, but that is driven by the top 25 percent of encounters. In Panel (b), we see that average length of stay by experience has approximately the same shape as the median length of stay for residents with each level of experience, as well as attendings who in the figure have "experience" equal to 9 quarters. All of the empirical averages shown are raw averages without any patient or physician controls.

8 Counterfactuals

I now use the model to assess the consequences of a policy change and of a training disruption on both patient care quality and resident skill. In addition to quantifying the impact of the change, I consider the effectiveness of various remedies that the hospital may enact in order to counteract the effects of the counterfactual changes. First, I consider is an increase in the speed at which attendings see patients independently, which in the model is represented by x_A. This is a feasible action because it does not necessarily require that the hospital hire higher-skilled attendings. Instead, they can simply staff more attending physicians on each shift. This works because x_A includes the supervision portion of the attending's duties. If there are additional attendings working on each shift, then supervisory duties will be split among more physicians, thereby reducing the number of disruptions each attending faces when caring for patients individually. This will reduce length of stay for patients assigned to attendings and increase their effective speed x_A. Second, I consider an increase in the learning rate of residents. This is potentially more difficult to implement because it would likely involve redesigning the curriculum or partnering with additional hospitals so that residents see additional patients.[20] For the first counterfactual, I consider permanent changes, but in the second counterfactual, the changes are for one period only. In the second counterfactual, I also explore the effectiveness of a temporary relaxing of the care quality constraint.

To simulate these counterfactuals, I change the relevant parameters and re-solve the model with other parameters held fixed. I find the new optimal patient assignment function (both counterfactuals) and the new steady-state (first counterfactual only) following the same procedure as outlined in Section 6.3.

The two teaching outcomes I consider are average patient length of stay and the total number of patients seen over the resident's career. For now, I make the extreme assumption that no further learning occurs after the resident graduates from the program.[21] Under this assumption, calculating average patient length of stay is straightforward: it is simply the average length of stay given by the resident's skill upon graduation. This is equivalent to the intensive margin of patient utility: for each patient the resident sees, what is the difference in their length of stay? Estimating the total number of patients seen requires making additional assumptions. I assume that graduates see $\frac{8}{\alpha_0 X^{\alpha_1}}$ patients per shift, where X is the skill they leave residency with, and that they go on to work 18 8-hour shifts per month (AMA, 2017) for 30 years. Differences in total patients seen represent the extensive margin of the change.

8.1 Increasing the Bound on Quality

In the first counterfactual, the hospital decides to increase the lower bound of care quality. In the model, this is governed by an increase in L^*. There are real reasons for why hospital administrators may choose to make this change. The first is that length of stay is an important part of Medicare's Hospital Report Cards.[22] Hospital administrators may care about these ratings both because higher ratings help attract more patients and for intrinsic or reputational concerns (Kolstad, 2013). The second may be due to payment reform, which is a heavily-discussed policy lever to reduce healthcare costs (see McClellan, 2011). I next

[20] EM Residents typically are not up against the ACGME's hours limit so this change is legal, but it ignores other effects, such as the possibility of slower learning due to increased fatigue or a change in selection into specialties (cf. Wasserman, 2023).

[21] In progress is a version where graduating residents learn at half the speed as they did during residency. This approximation takes into account the facts that attendings work fewer shifts per month than residents and they no longer have formal supervision for every patient.

[22] See the "Timely and Effective Care" subsection of Medicare's Care Compare website (accessed October 25, 2023): https://www.medicare.gov/care-compare/

explain how payment reform may interact with the hospital's choice of care quality bound L^*.

Generally, payments have been trending away from the traditional fee-for-service (FFS) system to alternatives such as value-based payments and capitated, prospective payment systems (PPS). Under FFS, providers are paid for every procedure, order, and service they provide to the patient. One of the issues with this system is that providers are not incentivized to reduce utilization or cost and have financial incentives to provide marginally necessary care (cf. Marmor and Gordon, 2021). Two leading alternatives are value-based care and PPS. In value-based care, providers are paid more if they realize better quality outcomes regardless of utilization, for example, for lower rates of complications from surgery or shorter ED length of stay. Because payment is independent of utilization, conditional on care quality, value-based care incentivizes physicians to reduce cost and improve quality. PPS works similarly in that providers are paid the same amount for every patient type regardless of utilization, so again providers have incentives to reduce costs.

Issues may arise for teaching hospitals if the reimbursement rates for value-based care and PPS are set uniformly across hospital types. An example would be if under the two systems, the average hospital's revenue is identical. Teaching hospitals would lose revenue due to a change like this because they tend to do poorly on many typical quality and efficiency metrics (Kocher and Wachter, 2023). Furthermore, teaching hospitals currently have very high FFS reimbursement rates, estimated at 10-20% above FFS payments at non-teaching hospitals, although quality of care is higher for some patient types, which offsets that somewhat (Sloan, 2021). In any case, if the switch from FFS to PPS or value-based care occurs without special accommodations for teaching hospitals, they would stand to lose significant revenue from patient care. This is the spirit of the first counterfactual.

Consider a very simple payment model where instead of being paid for each hour with patients (similar to FFS), hospitals are instead paid a fixed amount for each patient (similar to PPS). Further assume that all hospitals began with the same FFS rates and that the PPS rate is set so that the average hospital in the nation does not experience a change in revenue. My results suggest that UCSF could see at least 5% more patients each day if they did not train at all and instead had attendings see all of the patients.[23] Therefore, under this simple payment structure, they would lose 5% of revenue.[24] The hospital could recover some of the lost revenue by increasing patient throughput via increasing the quality constraint L^*. In the short run, in order to accomplish this, the hospital must reduce training since residents are slower than attendings. However, there are two mitigating actions the hospital can take. First, they could increase the rate of learning α_1 so that residents gain more skill with each patient seen. Second, they could increase the speed of attendings seeing patients independently. In the counterfactual, I assume that the hospital chooses to become 2% faster at caring for complex patients.

Table 7 shows the impact of the increase in the patient quality constraint alone and in combination with mitigating actions. The figures presented compare the new steady-state with the current steady-state, and ignores contributions to revenue, patients seen, and minutes per patient during the transition period. The first row reports current outcomes. Under the assumptions on speed and shifts worked described above, residents see 10,141 complex patients over the course of their career

[23]This is calculated from the model predictions for average length of stay during the academic year given optimal patient assignment under the current parameters, and the inferred value of attending skill. It is a lower bound because the current attending skill measure assumes that attendings also have teaching and supervisory duties, which would be reduced if the hospital reduced teaching.

[24]If the teaching hospital had higher FFS reimbursement rates than non-teaching hospitals prior to the policy change, then it would stand to lose even more revenue. Additionally, while it is true that even in the current world, the hospital could increase revenue by training less, it has chosen not to. This is because the hospital has chosen L^* at the current level from maximizing preferences over care quality, quantity, revenue, and teaching. I take this choice given and do not model it. As long as revenue is a normal good, changes that decrease revenue will cause it to seek ways to increase it.

and spend 307 minutes per patient. If the hospital adjusts training to increase the length of stay by 2% and makes no further changes, then in the new steady-state graduating residents see 520 fewer patients during their career and are almost 17 minutes slower for each individual patient. The 2% gain in teaching hospital revenue from seeing patients faster is paid by the hospital that employs the resident after graduation because its new physicians are slower, and this cost is almost 17 times larger than the revenue gain. This future cost is an externality from the teaching hospital's point of view because it undervalues the future productivity of their graduating residents when making the decision to reduce training.

However, the social planner can induce the teaching hospital to take mitigating actions and reduce the impact on training required by the 2% stricter length of stay requirement. For instance, teaching hospitals can increase the speed that attendings see patients independently by 5%. This is very effective in reducing the difference in counterfactual training from the current training level, making up 81% of the loss relative to when no other actions are taken. The intuition behind this is that the hospital desires to maximize training given a constraint, and increasing the speed at which attendings work in effect makes the constraint less binding. This allows them to increase training while still meeting the quality constraint. On the other hand, increasing the rate of learning by 5% has a much smaller effect, only allowing the hospital to make up 24% of the loss. This is again due to the constrained maximization problem faced by the hospital. Although the benefits of training are increased, the hospital has difficulty taking advantage of this and increasing the fraction of patients allocated to residents because it still must meet the same care quality constraint in every quarter. In other words, the hospital is not permitted to intertemporally substitute decreased speed in earlier quarters due to increased allocation of patients to residents with increased speed in later quarters because residents have gained more skill as it would in the absence of the constraint. Only in the later quarters of the academic year, when residents who learn faster are more skilled can the hospital actually increase patient allocation relative to current levels, and even then it is not by much. Finally, taking both actions actually has the effect of improving future outcomes, as now the hospital is able to take advantage of the increased benefits of faster learning and actually allocate additional patients to residents.

These results show that small changes in training by the teaching hospital can have outsize effects for future patients and future employers of residents. It is important for policymakers to consider these externalities when designing payment systems so that future patients do not end up paying orders of magnitude greater in costs in order to save a little today. Fortunately, there are feasible and straightforward remedies available that can mitigate these costs. The counterfactual shows that increasing the speed at which attendings see patients individually by just 5% can recover 81% of the loss in training resulting from a desire to increase patient throughput by 2%. This can be satisfied by staffing additional attending physicians, which admittedly may be difficult if general equilibrium effects are considered, but is likely far easier and more effective than finding ways to redesign the resident curriculum or having them work additional hours in order to learn faster. Another possibility that I have not discussed is to increase Medicare's Indirect Medical Education Payment to cover non-Medicare patients. Currently, this is a bonus paid to Medicare PPS hospitals, recognizing that teaching hospitals have higher costs than non-teaching hospitals. Increasing the IME Payments would also be effective as it would decrease the hospital's need to gain more revenue from patients due to payment reform, thereby reducing the need to increase patient throughput by decreasing training.

Table 7: Quality Bound Changes: Steady-State Counterfactual Resident Training and Mitigating Factors

Plan	Revenue "Externality"	Lifetime Patients	Minutes per Patient
Current Outcomes		10,141	307
Decrease length of stay 2% and...			
No other changes	-16.7:1	-520	+16.6
Attending Speed $+5\%$	-3.1:1	-97	+3.0
Learning Rate $+5\%$	-12.7:1	-396	+12.5
Learning Rate $+5\%$ & Att $+5\%$	+1.8:1	+57	-1.7

Notes: This table shows the loss for future patients of senior residents if UCSF decides to decrease patient length of stay by 2%, adjust the care quality utility constraint by the corresponding amount, and take the listed mitigating action. The revenue externality captures only the financial cost and is the ratio of the change in the present value of future patient revenue to the current patient revenue increase due to the policy change, assuming no changes in reimbursements. For example, if no other changes are taken, the present value of the cost to future employers of the resident is 16.7 times the additional revenue generated by UCSF by speeding up, because the graduating residents have less skill. This is in essence the discounted difference in lifetime patients seen and assumes that residents go on to work 18 shifts per month, as is typical for EM attendings, for 30 years. Lifetime patients is the total difference in patients seen (extensive margin), and minutes per patient is the length of stay difference for each patient (intensive margin) given graduating resident skill. For now, I make the extreme assumption that no further learning occurs post-residency.

8.2 Unexpected One-Time Training Disruption

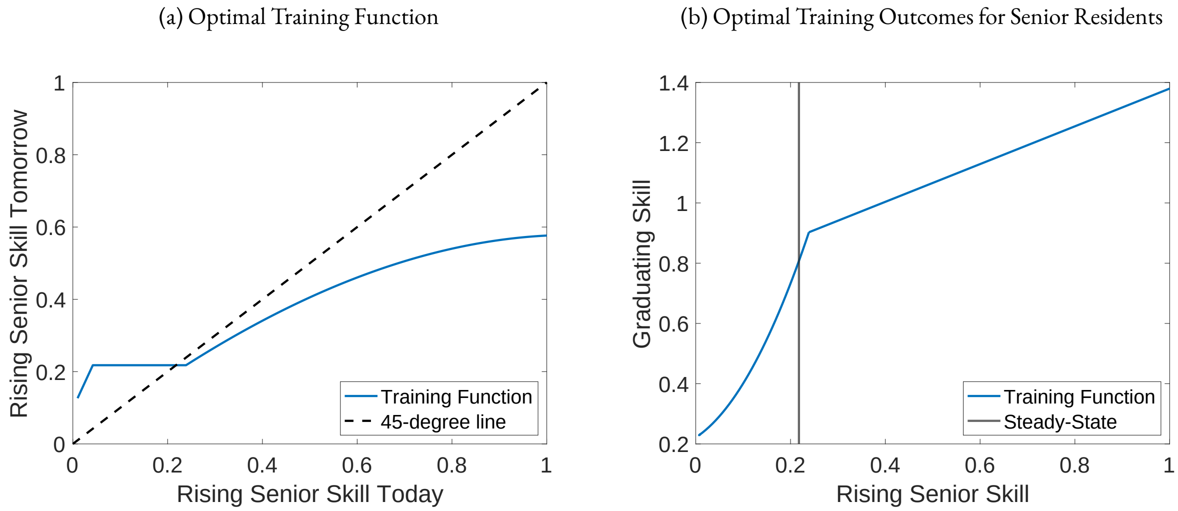
In the second counterfactual, I consider the consequences and effectiveness of policy responses to a one-time, unexpected disruption in training. This scenario resembles disruptions during the Covid-19 Pandemic, which affected residents in at least two general ways. First, the composition of patients who went to the hospital changed as patients delayed and avoided both routine and urgent or emergency care (Czeisler, et al., 2020). This, in addition to the influx of Covid patients, changed the pool of patients that residents could see and learn from, decreasing effective patient share in every period. Second, medical workers were under extreme stress during this period (HHS, 2022), which also likely reduced residents' ability to learn.[25]

I ask three questions. First, how many incoming classes of residents will the one-time disruption affect through spillovers? Second, for affected residents, what are the long-run effects of the disruption on their future careers? Third, what temporary mitigating factors could reduce the impact on residents affected by the training disruption? For simplicity, I assume that future incoming cohorts are equally skilled as their historical counterparts even though this is contrary to evidence (Jhajj, et al., 2022).

Under no additional changes, how long does the hospital take to return to the steady-state of training? This is simply the impulse response function of the system. In this case, it is very simple and easily inferred from the Optimal Training Function presented in Figure 6 Panel (a). The figure illustrates the utility-maximizing choice of patient allocation during the incoming cohort's first year (vertical axis) as a function of the skill of the rising senior (horizontal axis). The steady-state of the system is when rising senior skill today is equal to the resulting rising senior skill tomorrow, which occurs at the point which the optimal training function intersects the 45-degree line (the dashed line in the figure). In the figure, we observe that for most decreases in rising senior skill today, the hospital trains the incoming cohort exactly as much as it would have if there was no disruption,

[25] As with K-12 education, medical student education during this time also suffered (Jhajj, et al., 2022) so it is reasonable to infer that residents were also affected.

Figure 6: Optimal Training Function and Outcomes



Notes: This figure depicts the hospital's optimal choice of total training for the two cohorts given the training the current senior residents received in their first year. Panel (a) depicts the total first-year training for tomorrow's junior residents in their first year given the training the current senior residents received in their first year. The steady-state is where the optimal training choice intersects with the dotted 45-degree line. Panel (a) illustrates how for many values of current rising senior resident skill below the steady-state, the hospital chooses to train the next cohort of junior residents at exactly the steady-state quantity. If today's rising senior resident skill is greater than the steady-state, the hospital increases training in a diminishing way, and skill remains elevated for a few periods as it steps back down to the steady-state. The reason the second kink around 0.25 is not at the steady-state is due to the discount rate: here it is optimal to realize the gains immediately rather than realize small gains for the next few periods. Panel (b) depicts the hospital's optimal choice of total training for senior residents conditional on their skill acquired during their first year. The steady-state is depicted by the vertical line. This figure illustrates that below the steady-state, there is a very steep gradient in training, but the slope is flatter above the steady-state. When combined with Panel (a), one notices that below the steady-state the hospital prioritizes training the junior residents to return to the steady-state and trains the seniors minimally. This is due to the value of training the junior resident in the next period: the higher skilled the junior resident, the more training both cohorts can receive and still meet the quality constraint.

therefore returning to the steady-state the very next period. That means that aside from the senior residents at the time of the disruption, the disruption only affects the continuing cohort of rising seniors and does not affect future incoming residents.

But that does not mean the training disruption has no impact on training: clearly the training of the affected cohort must be reduced as a result in order to maintain the quality bound. This is exactly what we observe in Figure 6 Panel (b), which plots the graduating skill of the senior residents as a function of their skill when they become seniors. Below the steady-state, depicted by the vertical line, there is a very steep gradient in training, but it flattens substantially when rising senior skill exceeds the previous steady-state. Therefore, even though incoming cohorts are not affected by the one-period training disruption, senior residents receive drastically less training.

The intuition behind these finding is as follows. Because the rising seniors are less skilled than usual, the hospital is limited in the total fraction of patients it can assign to residents and maintain the care quality constraint. Therefore, it is very costly for the hospital to train the disrupted cohort because it not only means that the incoming cohort receives less training, but also means that future incoming cohorts would also receive less training. The combination of the yearly discount factor of $\beta = 0.95$ and the concavity of the training utility with respect to patient share means that the hospital is patient enough to sacrifice training in the current period in order to return to the steady-state immediately, rather than spreading the cost of the disruption over multiple periods.

For the counterfactual outcomes, I assume that the result of the disruption is that the affected rising seniors begin the academic year with half of the steady-state skill, but that the incoming cohort is identical to all other incoming cohorts. For the temporary changes, I find the optimal patient allocation with the different model parameters and the lower-than-usual starting value of rising senior skill. Motivated by the previous finding that the hospital returns to the steady-state the very next period, I consider combinations of six one-time, temporary policy changes: relaxing the care quality bound by 0.25% and 0.5% in length of stay, increasing attending speed by 2.5%, increasing the rate of learning by 2.5%, and combining an increase in attending speed with either a relaxation of the lower bound or an increase in the learning speed.

Results are in Table 8. As before, I compare outcomes with the steady-state outcomes in lifetime patients seen and minutes per patient. Those outcomes are represented in the first row of the table. The remaining rows (under the single dividing line) display the outcomes under various one-time changes when the rising senior class begins with half the knowledge as in the steady-state. As we saw in Figure 6, there is a significant training cost in the Status Quo–if no other actions are taken. Under the same assumptions on resident careers as the first counterfactual, residents see 224 fewer complex patients during their career and spend 7 additional minutes on each patient they see. Temporarily relaxing care quality by 0.25% roughly halves both decreases, while temporarily relaxing care quality by 0.5% actually makes the affected cohort slightly more skilled than usual. Attending speed increases again prove to be quite effective as the senior residents recover 88% of the difference in lifetime patients seen under the Status Quo compared to the steady-state. Similar to before, learning rate increases also have limited effect. Curiously, increasing attending speed and lowering the care quality threshold is not particularly effective, but as before, increasing both the attending speed and the learning rate is very effective as now the hospital is able to take advantage of the increased benefits of training.

This counterfactual shows that a disruption in training for residents has large effects on future patients that the teaching hospital may not internalize. But as in the first counterfactual, there exist straightforward remedies, and in this case they only need to be enacted temporarily. If the hospital is willing to temporarily reduce throughput, then a reduction of just 0.5%, or approximately two minutes, is sufficient for these residents to "catch up." If that is not acceptable, a small increase in attending speed of just 2.5% makes up 88% of the difference in training compared to the steady-state. In the event of future disruptions, both local and national, policymakers should consider stepping in to ensure that resident education does not suffer.

9 Discussion and Conclusion

When profit maximization is not the primary goal, how "multi-product" nonprofit firms adjust the production of their various products under changes in revenue is ambiguous. I study nonprofit teaching hospitals, which have the dual role of providing health services and training the next generation of physicians. Because the teaching component in this environment requires learning by doing, the hospital faces a trade-off between care quality and teaching, and I study how the hospital allocates complex patients to residents and attendings to make this trade-off. I find that short-run increases in quality achieved with reducing training are dwarfed by long-run quality decreases because residents see many patients over the rest of their career. Tools that use revenue to incentivize quality improvements to current patients such as value-based care and prospective payment systems,

Table 8: Training Disruption: Counterfactual Resident Training and Mitigating Factors

Counterfactual	Lifetime Patients	Minutes per Patient
Steady-State	10,141	307
Status Quo	-224	+6.9
Relax L^* 0.25%	-113	+3.4
Relax L^* 0.5%	+23	-0.7
Att Speed $+2.5\%$	-26	+0.8
Learning Rate $+2.5\%$	-58	+1.8
A.S. $\uparrow$ & $L^* \downarrow$ 0.25%†	-51	+1.6
A.S. $\uparrow$ & Learning	+21	-1.6

Notes: This table shows the loss for future patients of senior residents given that their training in their first year was disrupted. I assume they enter their second year with half of the knowledge they otherwise would have. In the status quo, the hospital does not adjust anything and behaves as depicted in Figure 6. The other entries show the outcome if the hospital makes a one-time change in behavior as indicated. The outcomes are relative to the steady-state outcome and compare the total number of patients the physician can see after the graduate (extensive margin) and the extra time per patient seen (intensive margin). †The combination of increasing attending speed and decreasing care quality has a benefit not depicted in the table: although outcomes for the affected cohort are not as good as before, the hospital over-trains the incoming cohort in anticipation of the higher steady-state given by the increase in attending skill. See text and notes to Table 7 for more details.

are an increasingly popular policy among both public and private insurers. However, when designing and responding to such policies, policymakers should be aware of potential unintended reductions in teaching, which may result in large long-run costs for future patients orders of magnitude larger than the improvements for current patients.

I examine and quantify these trade-offs in the emergency department of a large, urban teaching hospital. I first investigate at a granular level the potential costs of training and find that despite substantial increases in independence and the ability to manage additional patients simultaneously, there are no differences in patient outcomes or costly resource utilization. However, there are notable differences in patient throughput, but only for complex patients who are predicted to require inpatient admission: the median fourth-year is able to arrive at a disposition decision and complete these patients 9.4 % faster than the median first-year. Next, I embed the throughput, which is related to both revenue and care quality, into a dynamic model of training. I find that hospital behavior can be rationalized by a model where it maximizes training with respect to a lower bound of throughput. In counterfactuals, I find that if hospital administrators decide to increase throughput by 2%, the required reduction in training will result in lower future throughput 17 times larger than the current gains. However, there exist simple and feasible changes that can reduce the externality. For instance, I find that a 5% increase in attending speed would reduce the cost to future patients by 81%.

Although my focus is on the emergency department of a single, top-ranked teaching hospital, I believe that the results may generalize. First, speed and throughput are important quality measures across medical care, even in non-urgent situations. For instance, in the surgical context, it has been shown that increasing operative time is associated with increased odds of complications (Jackson, et al., 2011). Next, although there may be some variation in care correlated with teaching hospital rankings, prior work has shown that the basic production function of health does not differ in outcomes with respect to residency program prestige (Doyle, et al., 2010).

CMS is aware of the increased costs faced by teaching hospitals. The Medicare Prospective Payment System (PPS) includes a bonus paid to academic hospitals, known as the Indirect Medical Education Payment (IME). Although my findings show that patient throughput costs may be significant, they should not be used as the sole basis for determining the size of these payments. I believe that a large part of the reason that patient outcomes and resource utilization do not change with experience is due to the success of attending supervision. Staffing high-quality attendings can be very expensive, especially when they are also spending significant time engaged in research, and the IME Payments should account for this cost. My hypothesis is supported by the workflow, in which attendings and residents confer to determine the plan of care for each patient, as well as the empirical results, which show that resident independence increases most notably and significantly in the first hour of the patient's encounter. Further research exploring the ways in which variation in supervision affects both outcomes and teaching quality could be valuable in improving both care quality and training outcomes.

Finally, while teaching hospitals are a single example of a specialized organization, they constitute an outsize share of both the economy and individual utility. The United States spent 17.8% of GDP on healthcare in 2021, and in 2019, teaching hospitals contributed 45% to Health Care and Social Assistance GDP.[26] Preserving life and increasing the quality of life, the main functions of hospitals, are arguably the most important components of individual utility, well-being, and happiness. That said, the model and empirical strategy can be applied to related settings as many nonprofit institutions are in essence multi-product firms. Most notably, this group includes research universities, which through their research are principal drivers of innovation in the modern economy (Lerner, et al., 2023) yet are also responsible for educating undergraduate, professional, and graduate students. Studying how they make this trade-off and respond to changes in government funding could be a fruitful area for future research.

[26] Gunja, et al. (2023) and the author's calculations using statistics from the BEA and AAMC

References

Adhvaryu, Nyshadham, and Tamayo (2023), “Managerial Quality and Productivity Dynamics,” *Review of Economic Studies*.

Altonji, Elder, and Taber (2005), “An Evaluation of Instrumental Variable Strategies for Estimatingthe Effects of Catholic Schooling,” *Journal of Human Resources*.

American Medical Association (2017), “What it’s like to specialize in emergency medicine: Shadowing Dr. Clem,” AMA-assn.org, accessed October 25, 2023.

American Medical Association (2019), “Applying to more than 1 medical specialty? What you should know,” AMA-assn.org, accessed November 3, 2023.

Arrow, Kenneth (1963), “Uncertainty and the Welfare Economics of Medical Care,” *The American Economic Review*.

Association of American Medical Colleges (2022),“Economic Impact of AAMC Medical Schools and Teaching Hospitals.”

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin (2007), “Estimating Dynamic Modelsof Imperfect Competition,” *Econometrica*.

Belloni, Chernozhukov, and Hansen (2014), “Inference on Treatment Effects after Selection among High-Dimensional Controls,” *Review of Economic Studies*.

Benkard, C. Lanier (2000), “Learning and Forgetting: The Dynamics of Aircraft Production,” *The American Economic Review*.

Bergeron, Augustin, Pedro Bessone, John Kabeya Kabeya, Gabriel Tourke, and Jonathan L. Weigel (2022), “Optimal Assignment of Bureaucrats: Evidence from Randomly Assigned Tax Collectors in the DRC,” *NBER Working Paper 30413*.

Bloesch, Justin and Jacob P. Weber (2023), “Congestion in Onboarding Workers and Sticky R&D,” *Working Paper*.

Brickley, James and R. Lawrence Van Horn (2015), “Managerial Incentives in nonprofit Organizations: Evidence from Hospitals,” *Journal of Law and Economics*.

Capps, Carlton, and David (2017), “Antitrust Treatment of Nonprofits: Should Hospitals Receive Special Care?” *NBER Working Paper 23131*.

Chan, David (2018), “The Efficiency of Slacking Off: Evidence from the Emergency Department,” *Econometrica*.

Chan, David (2021), “Influence and Information in Team Decisions: Evidence from Medical Residency,” *American Economic Journal: Economic Policy*.

Chen, Yiqun (2021), “Team-Specific Human Capital and Team Performance: Evidence from Doctors,” *The American Economic Review*.

Cheng, Yi (2019), “The Unexpected Costs of Expertise: Evidence from Highly Specialized Physicians,” *Working Paper*.

Chu, Bryan, Ben Handel, Jon Kolstad, Filip Matějka, and Ulrike Malmendier (2023), “The Effect of Fatigue and Cognitive Load on Medical Provider Decision-Making and Patient Health Outcomes,” *Working Paper*.

Cowgill, Bo, Jonathan M.V. Davis, B. Pablo Montagnes, and Patryk Perkowski (2023), “Matchmaking Principles: Theory and Evidence from Internal Talent Markets,” *Working Paper*.

Currie, Janet and Bentley MacLeod (2020), “Understanding Doctor Decision Making: The Case of Depression Treatment,” *Econometrica*.

Czeisler MÉ, Marynak K, Clarke KE, et al. (2020), “Delay or Avoidance of Medical Care Because of COVID-19–Related Concerns — United States,” *MMWR Morbidity and Mortality Weekly Report*, June 2020.

Dahlstrand, Amanda (2023), “Defying Distance? The Provision of Services in the Digital Age,” *Working Paper*.

Doyle Jr., Joseph J., Steven M. Ewer, and Todd H. Wagner (2010), "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams," *Journal of Health Economics*.

Duggan, Mark (2000), "Hospital Ownership and Public Medical Spending," *The Quarterly Journal of Economics*.

Feldstein, Martin (1971), "Hospital Cost Inflation: A Study of nonprofit Price Dynamics," *The American Economic Review*.

Gaynor, Martin (2006), "What Do We Know About Competition and Quality in Health Care Markets?" *Foundations and Trends in Microeconomics*.

Gaynor, Martin and William Vogt (2003), "Competition among Hospitals," *The RAND Journal of Economics*.

Gaynor, Martin and Robert J. Town (2012), "Competition in Health Care Markets," *The Handbook of Health Economics*.

Gong, Qing (2018), "Physician Learning and Treatment Choices: Evidence from Brain Aneurysms," *Working Paper*.

Gunja, Munira Z., Evan D. Gumas, and Reginald D. Williams II (2023), "U.S. Health Care from a Global Perspective, 2022: Accelerating Spending, Worsening Outcomes," *Commonwealth Fund*, Jan. 2023.

Hausknecht, John P. and Charlie O. Trevor (2011), "Collective Turnover at the Group, Unit, and Organizational Levels: Evidence, Issues, and Implications," *Journal of Management*.

Hotz, V. Joseph and Robert A. Miller (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*.

Hughes, Emily (2017), *Canadian Medical Association Journal* 2017 August 14;189:E1050-1.

Jackson TD, Wannares JJ, Lancaster RT, Rattner DW, Hutter MM (2011), "Does speed matter? The impact of operative time on outcome in laparoscopic surgery," *Surgical Endoscopy*, 2011 Jul;25(7):2288-95. doi: 10.1007/s00464-010-1550-8. Epub 2011 Feb 7. PMID: 21298533; PMCID: PMC3676304.

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg (2021), "Learning from Coworkers," *Econometrica*.

Jhajj S, Kaur P, Jhajj P, Ramadan A, Jain P, Upadhyay S, Jain R (2022), "Impact of Covid-19 on Medical Students around the Globe," *Journal of Community Hospital Internal Medicine Perspectives*.

Jovanovic, Boyan (2014), "Misallocation and Growth," *American Economic Review*.

Kasy, Maximilian and Alexander Teytelboym (2022), "Matching with semi-bandits," *The Econometrics Journal*.

Kocher, Bob and Robert M. Wachter (2023), "Why is it so Hard for Academic Medical Centers to Succeed in Value-Based Care?" *Health Affairs Scholar*.

Kolstad, Jonathan (2013), "Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards," *The American Economic Review*.

Lassner, Jared W., et al. (2022a), "Growth of For-Profit Involvement in Emergency Medicine Graduate Medical Education and Association Between For-Profit Affiliation and Resident Salary," *AEM Education and Training*.

Lassner, Jared W., et al. (2022b), "Quantifying For-Profit Outcomes in GME: A Multispecialty Analysis of Board Certifying Examination Pass Rates in For-Profit Affiliated Residency Programs," *Journal of Graduate Medical Education*.

Lakdawalla, Darius, and Tomas Philipson (1998), "Nonprofit Production and Competition," *NBER Working Paper 6377*.

Lerner, Josh, Carolyn Stein, and Heidi Williams (2023), "The Wandering Scholars: Understanding the Heterogeneity of University Commercialization," *Working Paper*.

Levitt, List, and Syverson (2013), "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant," *Journal of Political Economy*.

Ludmerer, Kenneth M. (2014), *Let Me Heal: The Opportunity to Preserve Excellence in American Medicine*, New York: Oxford University Press.

Marmor, Theodore R. and Robert W. Gordon (2021), "Commercial Pressures on Professionalism in American Medical Care: From Medicare to the Affordable Care Act," *Journal of Law, Medicine, and Ethics*.

McClellan, Mark (2011), "Reforming Payments to Healthcare Providers: The Key to Slowing Healthcare Cost Growth While Improving Quality?" *Journal of Economic Perspectives*.

Minni, Virginia (2023), "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs," *Working Paper*.

Newhouse, Joseph (1971), "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," *The American Economic Review*.

Reagans Ray, Linda Argote, and Daria Brooks (2005), "Individual Experience and Experience Working Together: Predicting Learning Rates from Knowing Who Knows What and Knowing How to Work Together," *Management Science*.

Reder, M.W (1965), "Some Problems in the Economics of Hospitals," *The American Economic Review*.

Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services (2022), "Impact of the COVID-19 pandemic on the hospital and outpatient clinician workforce: challenges and policy responses (Issue Brief No. HP-2022-13)," May 2022.

Oster, Emily (2019), "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business and Economic Statistics*.

Pakes, Ariel, Michael Ostrovsky, and Steven Berry (2007), "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *RAND Journal of Economics*.

Pauly, Mark and Michael Redisch (1973), "The Not-For-Profit Hospital as a Physicians' Cooperative," *The American Economic Review*.

Sloan, Frank A. (2021), "Quality and Cost of Care by Hospital Teaching Status: What are the Differences?" *The Milbank Quarterly*.

Sokol, Emily (2020), "Private Payers Outpace Public Insurance in Value-Based Care Push," , accessed October 28, 2023.

Song, Hummy, Robert S. Huckman, and Jason R. Barro (2016), "Cohort Turnover and Operational Performance: The July Phenomenon in Teaching Hospitals," *Working Paper*.

Syverson, Chad (2011), "What Determines Productivity?" *Journal of Economic Literature*.

U.S. Bureau of Economic Analysis, "Value Added by Industry" (accessed Thursday, October 19, 2023)

Wang, Xixi, et al. (2022), "Residency Attrition and Associated Characteristics, a 10-Year Cross Specialty Comparative Study," *Journal of Brain and Neurological Disorders*.

Wasserman, Melanie (2023), "Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents," *Review of Economic Studies*.

Wei, Eric, Laura Sarff, and Brad Spellberg (2019), "Debunking the July Effect Myth." *Journal of Patient Safety*.

A Prediction of Inpatient Admission

In this section I provide a brief overview of the ex-ante prediction of inpatient admission. The key feature of this prediction is that only factors that are immutable (e.g. patient age) and determined prior to the physician's involvement (e.g. abnormal vital signs upon entry; chief complaint as recorded by the triage nurse) are included in the prediction. Therefore, it is by construction exogenous to the providers who will subsequently care for the patient.

I use LASSO to select among the large set of ex-ante and immutable patient covariates, with a logit functional form because inpatient admission is a binary outcome. The predictions fit observed patterns of inpatient admission well: the area under the receiver operating characteristic curve (AUC) is around 0.97. One way to interpret AUC is that it is the probability that the model ranks a random positive example more highly than a random negative example. The maximum value is 1, so a value of 0.97 indicates that the model is very successful at predicting the observed outcome.

Similar results are obtained whether the functional form is a linear probability model or a probit.

B Inferring Patient Shares

As mentioned in Section 3, this would be completely straightforward if not for the fact that due to the structure of the data and the program we do not observe the full patient history for any resident.

Because of both noise and unobserved individual fixed effects, some of the imputed values are clearly infeasible: either they are negative or greater than what is possible if the resident received the full share of patients on every shift. The minimum (maximum) allowable values are the cumulative sum of the 5th (95th) percentiles of actual patient shares (or the minimum (maximum) observed value if that is closer to the mean) for residents of the same tenure we do observe.

Then, for any user-quarters that are still missing, I linearly interpolate forwards and backwards using the cumulative experience. If interpolation is not possible, I subtract or add the user-specific academic year average for each quarter.

C Additional Tables

Table A1: Learning Over Time: Immediate Orders Upon ED Admission

	First Order Upon Admission					
Cumul Pt Share	0.011**	0.012***	0.014	0.013*	0.009	0.011***
	(0.005)	(0.004)	(0.010)	(0.008)	(0.006)	(0.004)
Tenure (Years)	0.010**	0.010***	0.013	0.011	0.008	0.011**
	(0.005)	(0.004)	(0.010)	(0.008)	(0.005)	(0.004)
log(Cumul Pt Share)	0.004	0.004	0.012	0.007	-0.000	0.003
	(0.004)	(0.003)	(0.007)	(0.006)	(0.005)	(0.003)
Tenure (log Years)	0.006	0.006	0.014	0.008	0.002	0.005
	(0.005)	(0.004)	(0.011)	(0.008)	(0.005)	(0.004)
DepVar Mean		0.138		0.162		0.128
ED Disposition		All Patients		Complex		Simple
Controls		X		X		X
Obs	31,628	31,565	8,877	8,853	22,751	22,712

Notes: Regressions of a binary indicator for first medical order upon admission to the ED on various measures of resident experience, split by ex-ante predicted patient complexity. This process measure is a complement to the log(Minutes to First Order) outcome in Tables 3 and 4 which is undefined when orders are immediate. The dependent variable is equal to one if the first medical order is signed at or before the patient is moved from the waiting room to an examination room and zero otherwise. As a reminder, coefficients shown are from four separate regressions. Standard errors are clustered by physician. See text and corresponding Table Notes for additional details.

Table A2: Learning Over Time: Diagnostic Orders

	Any Diagnostic Order Signed (Binary)						log(Diagnostic Orders Signed)					
Cumul Pt Share	-0.003	-0.001	-0.001	-0.001	-0.004	0.000	0.013	0.018*	-0.019	-0.013	0.018	0.038***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.013)	(0.010)	(0.014)	(0.014)	(0.015)	(0.012)
Tenure (Years)	-0.001	-0.000	-0.001	-0.001	-0.002	0.002	0.012	0.014	-0.021	-0.014	0.014	0.034***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.014)	(0.010)	(0.016)	(0.014)	(0.016)	(0.012)
log(Cumul Pt Share)	0.002	0.002	-0.001	-0.001	0.002	0.004	0.014	0.007	-0.025**	-0.027**	0.016	0.024**
	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.011)	(0.008)	(0.012)	(0.011)	(0.013)	(0.010)
Tenure (log Years)	0.005	0.004	-0.001	-0.001	0.006	0.007	0.023*	0.010	-0.030**	-0.032**	0.024	0.032**
	(0.004)	(0.003)	(0.002)	(0.002)	(0.006)	(0.004)	(0.013)	(0.011)	(0.014)	(0.015)	(0.016)	(0.013)
DepVar Mean		0.939		0.993		0.918		12.067		19.655		9.103
ED Disposition		All Patients		Complex		Simple		All Patients		Complex		Simple
Controls		X		X		X		X		X		X
Obs	31,628	31,565	8,877	8,853	22,751	22,712	29,705	29,644	8,821	8,797	20,884	20,847

Notes: This table presents regressions of outcomes related to diagnostic orders signed on various measures of resident experience, split by ex-ante predicted patient complexity. It is similar to Table 3. As a reminder, coefficients shown are from four separate regressions. Dependent variable means are listed, always in levels. Diagnostic Orders are medical orders primarily for gathering information about the patient, such as lab tests and imaging, rather than for treating or stabilizing the patient. Any Diagnostic Order Signed is a binary variable equal to one if at least one diagnostic order was signed and zero otherwise. log(Diagnostic Orders Signed) is the natural logarithm of the number of diagnostic orders signed and is undefined when zero orders are signed (for instance, if the patient requires stitches but did not receive an X-Ray prior to the procedure). Standard errors are clustered by physician. See text and notes to Tables 3 and 4 for additional details.

Table A3: When does Supervision Occur and Change for Complex Patients?

	Fraction Orders Signed by Resident			
Experience (Years)	0.047***	0.008	0.010	-0.000
	(0.008)	(0.011)	(0.009)	(0.006)
Period	1st Hour	2nd Hour	Middle	Last Hour
DepVar Mean	0.430	0.505	0.451	0.176
Num Orders	11.8	4.1	11.1	4.3
Obs	9,196	8,632	7,648	9,196

Notes: Regressions of the fraction of orders signed by the resident during various portions of the patient's stay in the ED on the number of fractional years in the program. If the patient stay is less than or equal to two hours, the second hour is counted only as part of the Last Hour. "Middle" includes all hours after hour three and prior to the last hour before inpatient upgrade (for admitted patients) or discharge (for discharged patients). The dependent variable mean is listed, as is the mean number of orders signed during the period. Most of the change occurs in the first hour, which is also when the bulk of the orders are signed. I select a similar set of patient covariates as in the binned scatterplots of Figure 2, but additionally include the number of simultaneous patients managed by the resident and its square. Standard errors are clustered by physician.

Table A4: Allocation of Complex Patients: Congestion

	Patient Assigned to 1st Year Resident			
# Complex Pt in ED	0.976	0.972	0.973	0.973
	(0.017)	(0.018)	(0.018)	(0.018)
# Pt in Waiting Room	0.986	0.971***	0.970***	0.969***
	(0.011)	(0.010)	(0.010)	(0.010)
Likely Handoff				0.421**
				(0.149)
# other EM1 FE	Y	Y	Y	Y
Month FE		Y	Y	Y
Patient Condition FE			Y	Y
Other Controls				Y
Obs	6,903	6,903	6,896	6,896

Notes: Odds ratios reported. Going from the 25th to 75th percentile of patients in the waiting room lowers the probability of assignment to a first year resident by 15 percentage points. # of other EM1 FE are fixed effects for the number of other first year residents on shift at the time of patient allocation: clearly it is more likely to assign a patient to a first year resident when there are more of them working. Not shown are the coefficients for patient condition fixed effects, which include interactions between ex-ante triage nurse estimated severity and chief complaint. Standard errors are clustered by physician.

Table A5: Impact of Excess Patients on Outcomes and Process Measures

	Readmissions	log(LoS)	Any Dx Order	log(Dx Orders)	Frac Orders
EM1 or EM2	-0.017	0.094***	0.007	0.054*	-0.174**
	(0.019)	(0.028)	(0.008)	(0.031)	(0.067)
Excess Pt	-0.025	-0.153***	-0.003	-0.041	-0.055
	(0.018)	(0.024)	(0.007)	(0.033)	(0.062)
EM1 or EM2 X Excess Pt	0.017	0.008	0.004	-0.019	0.129
	(0.023)	(0.039)	(0.012)	(0.042)	(0.080)
Obs	4,306	8,245	8,127	7,784	8,121

Notes: The coefficients of interest are the coefficients on "Excess Patients" and the interaction "EM1 or EM2 X Excess Patients." Excess patients is an indicator variable equal to one when the resident is currently managing three or more patients with high values of the triage nurse assigned emergency severity index. This is the 75th percentile of simultaneous patients. EM1 or EM2 is a binary variable equal to one if the resident is in their first two years of the program. All regressions include resident fixed effects and controls for overall ED congestion, ex-ante and immutable patient characteristics, and business hour and weekend fixed effects. Standard errors are clustered by physician.

Table A6: Dynamic Results: Full Results

Specification	β	L^*	RMSE	L^* Implied Hrs	Graduating Skill (Hrs)
	0.90	-1.335	.0101	5.339	5.112
Linear	0.95	-1.335	.0095	5.338	5.116
	0.99	-1.332	.0092	5.330	5.113
	0.90	-7.144	.0095	5.346	5.112
Quadratic	0.95	-7.144	.0087	5.346	5.112
	0.99	-7.139	.0091	5.344	5.112
	0.90	0.149	.0110	5.363	5.103
Log	0.95	0.149	.0096	5.361	5.116
	0.99	0.149	.0102	5.362	5.108

Notes: This table shows the full estimation results for the three functional forms of hospital utility for patient length of stay and three values of the discount rate β. The first two columns show the estimated lower bound of patient quality L^* in utils, as well as the model's root mean squared error compared to the observed patient assignment shares. L^* implied hours coverts the utils to hours, and I also show the graduating skill of the resident, also in hours per patient. There is not much of a difference between specifications.

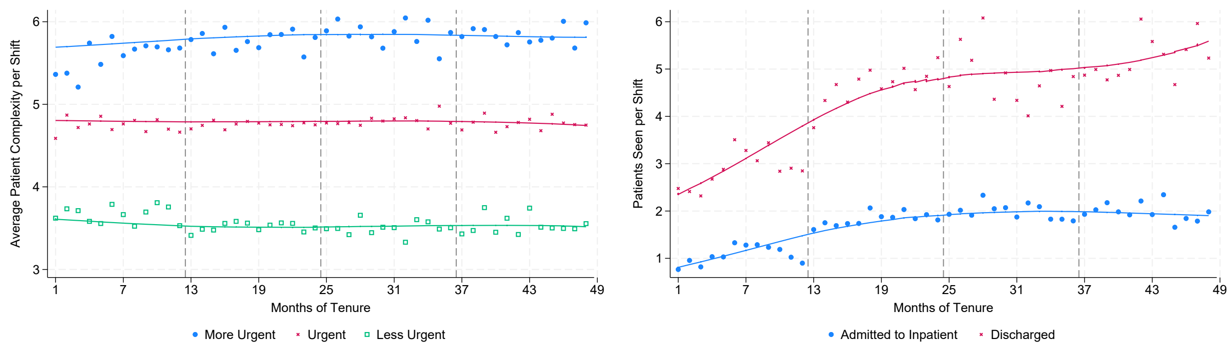
D Additional Figures

Figure A1: Patients Seen Per Shift

Patients per shift broken down by admission and for complexity within ESI category

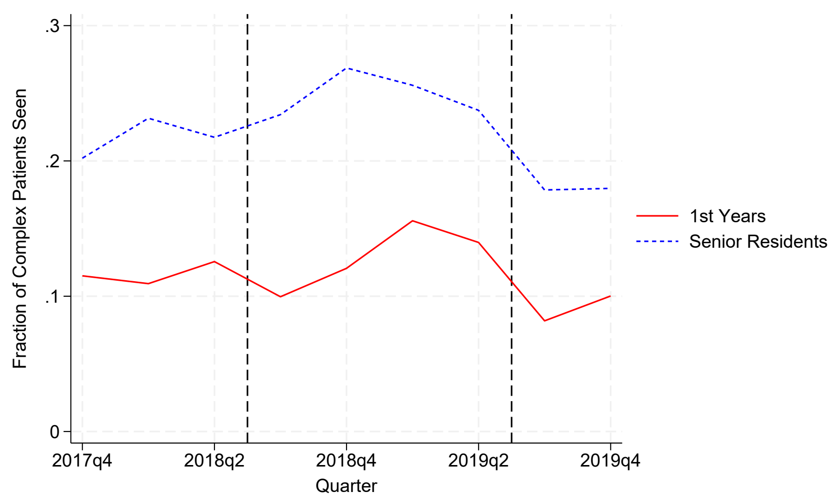
(a) Average Ex-Ante Complexity of Patients

(b) Patients per Shift by Predicted Admission



Notes: Panel (a) shows the average predicted ex-ante complexity of patients assigned to residents in the three median emergency severity categories estimated by the triage nurse. These cover over 97% of all patient arrivals. The complexity measure corresponds to variation in patient severity within ESI category and can be thought of as the "intensive margin" of complexity assignment to residents. The "extensive margin" is shown in Figure 3 Panel (b). "Complexity" is a prediction of patient severity based on ex-ante and immutable patient covariates developed in Chu, et al. (2023). This figure shows that with the exception of residents in the first six months of the program getting slightly simpler patients in the highest complexity category, averages are stable across experience. This means that residents are not assigned less complex patients as they increase the number of patients they see simultaneously with experience (as shown in Figure 3 Panel (a)). Panel (b) shows the number of patients per shift by the other ex-ante severity measure, ex-ante predicted admission. Patterns are similar to those given by emergency severity index as shown in the text.

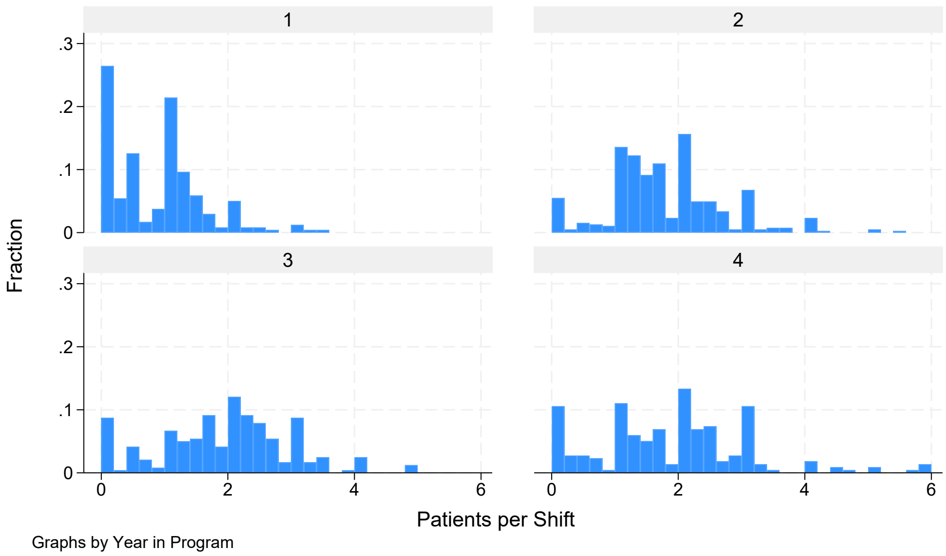
Figure A2: Average Fraction of Complex Patients Seen, by Role in Each Calendar Quarter



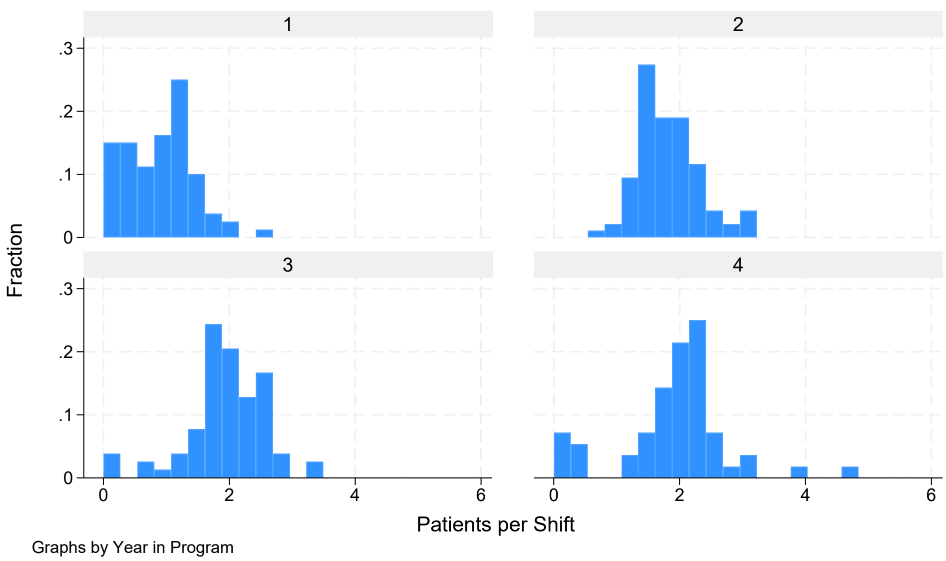
Notes: This figure shows the average fraction of complex patients seen by role for each calendar quarter. The figure shows that the shares are relatively stable across academic years, delineated by the vertical dashed lines, providing evidence supporting the assumption that in the steady-state, the hospital trains the same amount each academic year. See text and notes to Figure 4 for more details.

Figure A3: Cross-Sectional Variation in Average Complex Patients Seen per Shift

(a) Two-Week Averages

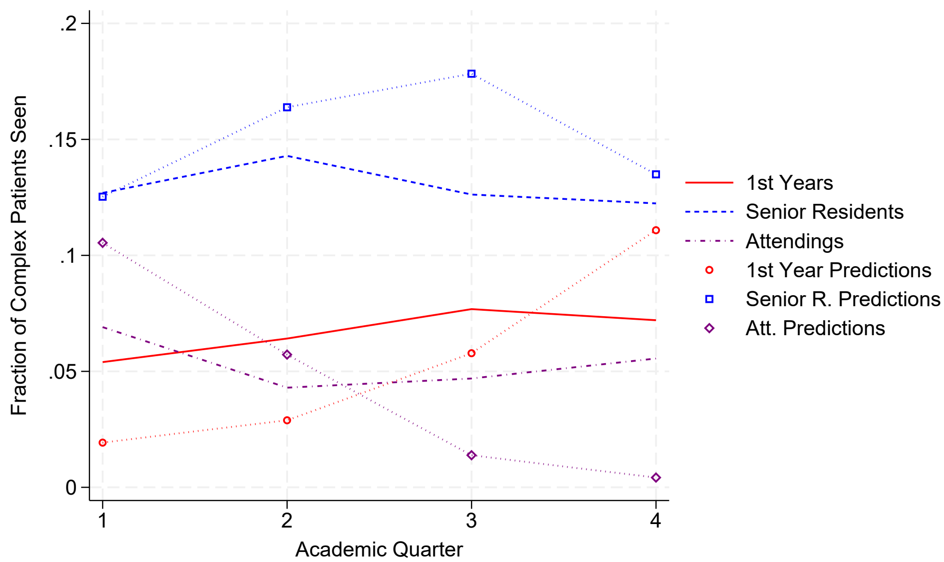


(b) Quarterly Averages



Notes: This figure shows the average number of patients per shift over two different periods of aggregation. Panel (a) shows this over two-week periods, while Panel (b) shows this over calendar quarters. Each observation is a resident-period and variation is shown separately by the resident's year in the program.

Figure A4: Model Fit by Quarter



Notes: This figure plots the actual allocation of complex patients with the model predicted allocation. The scale is different than in Figure 4 and Appendix Figure A2 because I have normalized each quarter to have a mass of 0.25 patients. The data are represented by the heavier lines without markers, and the model predictions are the lighter lines with markers on each quarter. The model fits the qualitative patterns well but the gradient on patients allocated to first year residents and attendings is steeper than in the data. See text for additional details.