Learning and Influence under Uncertainty: Evidence from Physicians in Training

David C. Chan*

January 25, 2015

Abstract

Studying physicians in training ("housestaff") at a large institution, I investigate the role of learning (and endogenous influence) in supporting medical spending variation within organizations. Influence associated with relative experience increases variation in housestaff spending effects, which discontinuously double in magnitude at the end of the first year of training when housestaff relative seniority discontinuously increases within teams. Consistent with a higher rate of learning in information-rich environments, housestaff effects converge in specialist-driven services, especially for diagnoses with guidelines, but not in the generalist-driven service. Compared to these mechanisms of learning and influence, detailed pre-training characteristics associated with preferences and ability determine little variation.

*I am grateful to David Cutler, Joe Doyle, Bob Gibbons, and Jon Gruber for their guidance on this project from an early stage. I also thank Achyuta Adhvaryu, Daron Acemoglu, Leila Agha, Daniel Barron, David Bates, Amitabh Chandra, Michael Dickstein, Grant Miller, and David Molitor for helpful comments. Joel Katz and Amy Miller provided invaluable context to the data. I acknowledge support from the NBER Health and Aging Fellowship, under the National Institute of Aging Grant Number T32-AG000186; the Charles A. King Trust Postdoctoral Fellowship, the Medical Foundation; and the Agency for Healthcare Research and Quality Ruth L. Kirschstein Individual Postdoctoral Fellowship 1-F32-HS021044-01.

1 Introduction

Although an extensive literature has documented substantial variation in medical care between and within regions,[1] little is known about how such variation is rooted in decision-making. This paper examines the simple idea that remaining uncertainty about the best course of action, generally made by teams, supports variation in patient care attributable to providers. Exceedingly few decisions in medicine can ever be based on indisputable evidence. Thus, learning may only be partial, and increases in relative experience within teams and organizations may sustain or even widen individual variation in practice patterns.

I study the effects of internal medicine physicians on medical care decisions as they proceed through training as "housestaff" at a large institution. This study is based on two institutional facts, broadly reflected in modern medicine across the US: First, medical care is provided in teams and within organizations. Utilization and health outcomes are determined by systems of care in which providers directly confer with one another or indirectly influence each other through culture or policies. Second, the amount of knowledge used in patient care varies across medical subspecialties. Specifically, in a large number of academic medical centers in the US, cardiology and oncology care is directed by specialists, whereas patient care in other medical subspecialties (e.g., gastroenterology and pulmonology) is generalist-driven.

Quasi-random assignment of housestaff to other teammates and to patients allows me to estimate the causal effect of each housestaff, in each period of training, on team decisions. Focusing on daily spending, as a summary statistic of numerous medical decisions and an object of particular policy attention, I examine learningand influence as two related mechanisms that determine housestaff spending variation under uncertainty. First, I examine relative influence due to being more informed, by exploiting a natural experiment in team structure. Since patients are cared for by a housestaff team comprised of a first-year "intern" and a second- or third-year "resident," relative experience changes discontinuously across the one-year mark while other physician characteristics (e.g., knowledge and ability) are plausibly continuous across time. I

[1] Papers in this large literature include Wennberg and Gittelsohn (1973); Wennberg et al. (2002, 2004); Fisher et al. (2003a,b). Most of this literature has focused on variation across regions, but more recently, Grytten and Sorensen (2003) and Epstein and Nicholson (2009) have demonstrated variation within regions that is potentially greater than that across regions.

find that spending effects discontinuously double in magnitude, as measured by the standard deviation of the tenure-specific distribution across housestaff, across the first-year mark. The standard deviation among interns is a 15-30% difference in costs, while the standard deviation among residents is a 55-70% difference in costs. Housestaff spending effects are also more serially correlated among residents, which describes greater persistence within housestaff in spending effects despite greater variation across housestaff.

Second, I compare practice patterns over time in the specialist services – cardiology and oncology – with practice patterns by the same housestaff in general medicine. This compares learning in environments differing in the amount of specialized knowledge used in care. I find substantial convergence in spending for residents in specialist services, such that 50-70% of the variation at the beginning of the second year is eliminated by the end of the third year, while the same residents show no reduction in spending variation in general medicine, a difference that is highly significant with systematic placebo tests. I further show that convergence in the specialist services is driven by conditions for which there exists a published guideline. This supports the idea that learning, to the extent required to generate convergence, is greater in environments with greater knowledge. In contrast to the large effects that influence and learning have on the size of variation, detailed housestaff characteristics (e.g., test scores, rank-list positions, and precommitted career choices) predict small, if any, differences in spending. Similarly, comparing second- and third-year residents, I find that housestaff tenure does not significantly shift levels of mean spending and other outcomes, such as readmission and mortality.

This paper connects a large and influential literature on variation in medical care, cited above, to uncertainty and learning within organizational environments. Despite the size of this literature, little is known about the behavioral and organizational foundations of such variation. Recently, Manski (2013) makes the conceptual observation that variation may be justified under uncertainty for reasons of diversification and learning, and Cutler et al. (2013) demonstrate that physician beliefs are correlated with practice patterns. This paper shows that it is the interaction between physician beliefs and relative uncertainty within an organizational environment that matters for observed variation across physicians. While Doyle Jr et al. (2010) show that housestaff physicians influence patient care, this paper is the first to my knowledge that examines the development of practice patterns among physicians in training.[2]

More broadly, this paper is related to a growing literature on productivity variation (Bartelsman and Doms, 2000; Syverson, 2011; Gibbons and Henderson, 2012). While most of this literature has examined variation across firms, little is known about how variation exists across workers within firms. This paper suggests that this variation may depend more on the knowledge (or, equivalently, the lack of uncertainty) embedded in the organizational environment than on fixed differences in worker ability. Finally, this paper is connected to a team-theoretic literature on how decisions are made by multiple agents (Cyert and March, 1963; Radner, 1993). In contrast to recent theoretical work predicting "management by exception," when a given problem is either solved with certainty or not at all (e.g., Garicano, 2000; Garicano and Rossi-Hansberg, 2006), I consider a setting with learning and remaining uncertainty in the proper course of action. In this setting, managers with more experience may have a large influence on *all* decisions, consistent with recent empirical work by Lazear et al. (2014), despite larger spans of control.[3]

The remainder of the paper is organized as follows. Section 2 outlines a simple conceptual framework in which to consider team decisions and learning under uncertainty. Section 3 describes the institutional setting; Section 4 describes the data. Section 5 discusses results on variation across team roles, and Section 6 discusses learning and convergence (or the lack thereof) in different knowledge environments and for conditions with or without guidelines. Section 7 discusses the insignificant effect of traditional explanatory measures, such as housestaff characteristics and tenure, on outcomes. Section 8 concludes.

2 Conceptual Framework

2.1 Influence in Team Decisions

I consider a simple team-theoretic environment of decision-making (e.g., Cyert and March, 1963; Radner, 1993; Garicano, 2000), in which team members use the information they have to

[2] Epstein and Nicholson (2009) and Dranove et al. (2011) have investigated how the place of residency affects subsequent obstetric care and have found that residency history explains relatively little of overall variation.

[3] Of note, Lazear et al. (2014) does not address, and in fact argues for, selection of agents of different ability and characteristics to be bosses vs. workers, and thus they cannot directly study the effect of roles and relative experience.

make the best decision for patient care.[4] The team must take an action a to match an unknown state θ, and will receive utility

$$u(a;\theta) = -(\theta - a)^2. \tag{1}$$

The team responsible for the care of a patient is comprised of two housestaff agents, a first-year "intern" i and a second- or third-year "resident" j. These two physicians in training also operate within a practice environment. This practice environment includes the inputs of other supervising ("attending") and consulting physicians, institutional rules (e.g., they are required to get consultant approval to order expensive tests in certain cases), and known standards of practice at the institution and more broadly.

At the time that the team must make a decision, the intern has a prior subjective distribution of θ, which I assume is normally distributed with mean m_i and precision g_i, defined as the inverse of the variance. The resident also has a normal subjective distribution of θ, with mean m_j and precision g_j. Finally, the team practices in an environment that is equivalent to a normally distributed "prior" with mean 0 and precision G. The action that maximizes expected utility in Equation (1) is then

$$a^* = \frac{g_i m_i + g_j m_j}{g_i + g_j + G}. \tag{2}$$

It is easy to see that, although housestaff h has a subjective belief with mean m_h, her belief is weighted a factor that captures her *influence* on the team and within the practice environment, $g_h / (g_h + g_{-h} + G)$.

The more precise her signal is relative to her teammate and the practice environment, the greater her influence will be. Because teams are always comprised of an intern and a resident, when a housestaff's tenure passes the one-year mark, she will be assigned to a teammate who has one year less experience than her, while she previously worked with a teammate who had at least one year more experience. This discontinuous decrease in g_{-h} results in a discontinuous increase in her influence (and the variation in medical care attributable to newly minted residents relative to seasoned interns), even if m_h and g_h are continuous across time. With respect to the practice

[4] I abstract from heterogeneous preferences or specialization at the individual physician level. However, the intuition should follow in more complicated settings as long as there is a common component to the decision that is agreed upon by both agents, and there is incomplete information about that component.

environment, a housestaff's influence will be lower in a tighter practice environment with higher G. At the extreme, if care were dictated by attending physicians or guidelines, there should be no variation attributable to housestaff.

2.2 Learning in a Knowledge Environment

I next consider the dynamic issue of learning, which depends on the broader environment of knowledge from which there is to learn. The key intuition I wish to capture is that there should be less convergence when there is less knowledge to be learned. In other words, learning does not occur automatically by treating patients but requires accessing broader knowledge and relating it to clinical experience.[5]

Appealing to a broad literature on search theory (see e.g., Rogerson et al. 2005, for a review), I allow physician learning to slow down or stop if the search costs of learning exceed the benefits.[6] I model this in reduced-form as a precision function $g_h = g(\tau_h; \mathcal{K})$ of housestaff h's tenure τ_h, which I consider isomorphic to experience, and the external knowledge environment $\mathcal{K}$. Under classical Bayesian learning, the distribution of subjective means m_h conditional on tenure τ has mean 0 and standard deviation $g(\tau; \mathcal{K})^{-1/2}$.

Restating Equation (2) as

$$a^* = a_i^* + a_j^* = \frac{g_i m_i}{g_i + g_j + G} + \frac{g_j m_j}{g_i + g_j + G}, \tag{3}$$

the standard deviation $\sigma(\tau; \mathcal{K})$ of experience-specific housestaff effects $a_{h,\tau}^*$ can be stated as

$$\sigma(\tau; \mathcal{K}) = \frac{g(\tau; \mathcal{K})^{1/2}}{g(\tau; \mathcal{K}) + g(\tau + \Delta; \mathcal{K}) + G}, \tag{4}$$

where the index cohort $\{h\}$ has tenure τ and the cohort $\{-h\}$ of the other team member has

[5] To be clear, I am referring to learning information that will be useful for future patients. Although θ is known perfectly *ex post* in the setup in Section 2.1, one may consider θ to be imperfectly observed (e.g., observed with some noise), imperfectly remembered, or otherwise uninformative for future patients in the absence of devoting some cost to learning.

[6] An alternative formulation by Acemoglu et al. (2006) allows for a lack of asymptotic agreement if there is sufficient uncertainty in the subjective distributions that map signals onto underlying parameters. Also, Ellison and Fudenberg (1993) show that, under social learning, there will be less convergence if agents observe greater diversity in choices made.

tenure $\tau + \Delta$. At time t relative to the beginning of the academic year, intern tenure is t, and resident tenure is $t + T$ and $t + 2T$, where T is one year, for second- or third-year residents, respectively.

I define convergence as a reduction in $\sigma(\tau; \mathcal{K})$ with time, or as $\partial\sigma(\tau; \mathcal{K})/\partial t < 0$ within academic years. Differentiating Equation (4), this condition is equivalent to

$$\frac{g'(\tau)}{g(\tau)} < \frac{2g'(\tau + \Delta)}{G - (g(\tau) - g(\tau + \Delta))},$$

where I omit reference to $\mathcal{K}$ for simplicity. In the Appendix, restricting to the case in which $g(\tau)$ increases piecewise linearly, I show that significant convergence may occur among housestaff in the senior role of resident ($\tau > T$; $\Delta \in \{-T, -2T\}$), especially for sufficiently large $g'(\tau)/G$, while convergence is insignificant and unlikely for housestaff in the role of intern ($\tau < T$; $\Delta \in \{T, 2T\}$). Finally, if the knowledge environment $\mathcal{K}$ is conceptualized as proportional to G, a sufficient condition for convergence to be increasing in $\mathcal{K}$ is for $g'(\tau)/G$ to be increasing in G, i.e., for $g'(\tau)$ to be convex in G.

3 Institutional Setting

I study physicians in internal medicine training at a large academic hospital. Physicians begin residency after finishing medical school, having seen a handful of patients in internal medicine, while by the end of residency, each housestaff will have personally admitted hundreds and participated in the care of well over a thousand patients. While residency represents a particularly formative time of a physician's career, it also characterizes a broader setting in which physicians practice within organizations and update their knowledge as part of continuing medical education.

3.1 Medical Care by Physicians in Training

Since the Flexner Report in 1910, medical residencies in the United States have largely been standardized to provide supervised training that emphasizes the scientific basis of medical practice (Flexner, 1910; Cooke et al., 2006). Regulated by the American Board of Internal Medicine

(ABIM) and the Accreditation Council for Graduate Medical Education (ACGME), internal medicine residencies are roughly the same in structure. Regulation stipulates the number of clinical weeks, types of rotations, hours per week, and patients per day and per year. Physicians in training, or housestaff, generally train for three years in internal medicine.

Housestaff work in teams and practice within a larger institutional structure. Each patient is cared for by a first-year housestaff ("intern") and a second- or third-year housestaff ("resident"). While there are no regulatory or legal distinctions in job responsibilities between interns and residents, differences in experience lead to endogenous difference in roles on the team.[7] The sole formal distinction is organizational: Residents are usually assigned to two interns at a time and therefore are responsible for twice the number of patients. As a result, absent differences in influence endogenous to experience, interns should have *more* control over their patients than residents do, as they can devote more attention to each patient. Interns are usually responsible for writing orders and communicating with nurses and patients, which reflects a greater first-hand interface with patient care.

Housestaff are supervised by "attending" physicians, who have legal responsibility for overseeing and agreeing to the care provided. In addition, housestaff operate within a broader practice environment, which includes other health care workers (e.g., consulting physicians, pharmacists, and nurses), as well as institutional rules for deciding and implementing care. Institutional rules may include reminders or requirements (e.g., to administer ACE inhibitors for heart failure patients or otherwise explain the rationale for not doing so) that are implemented by information technology and/or organizational design.

In this academic hospital, housestaff from different programs and different "tracks" within a program work together on the same clinical services. For example, a sizeable number of interns only plan to spend one year in the internal medicine residency ("preliminary" interns, as opposed to the standard "categorical" interns), subsequently proceeding to other residency programs, such as anesthesiology, radiology, or dermatology.[8] These plans are committed to prior

[7] More precisely, although I cannot rule out empirically that differences in influence simply could be due to titles, the institutional setting rules out formal differences in responsibility and authority. Further, the overall effect of switching roles on a team is to some extent the policy outcome of interest.

[8] In addition, tracks within a residency program include primary care, "short tracks" to fellowship training, research tracks such as genetics, and medicine-pediatrics or medicine-psychiatry combined programs.

to starting the internal medicine residency. Other residency programs include another internal-medicine residency from a different hospital, as well as obstetrics-gynecology and emergency medicine from the same hospital.

Housestaff schedules are arranged a year in advance to satisfy hospital programmatic requirements and broader regulations. Conditional on program and track, housestaff are exogenously scheduled to rotations. Rotations include intensive care unit (ICU), outpatient, research, subspecialty (mostly outpatient) electives, and ward blocks. This study focuses on inpatient ward rotations, which are comprised of cardiology, oncology, and general medicine services. Per residency administration, preferences are not collected about rotations, and assignment does not consider housestaff characteristics, although housestaff on certain tracks may be unavailable during certain times due to programmatic differences.[9] It is also rare for housestaff to trade blocks, given programmatic and regulatory requirements that must be met for each housestaff, and because scheduling is difficult for administration to redo. Scheduling does not consider the teams of intern, resident, and attending physicians that will be formed as a result. In fact, attending schedules are done independently, and neither housestaff nor attending scheduling is aware of each other's results in advance.

Patients arriving at the hospital are assigned to interns and residents by algorithm, which distributes patients in a rotation among housestaff that are "on-call" and have not reached the maximum number of patients. Patients who remain admitted for more than one day may also be mechanically transferred between housestaff changing rotations. When a housestaff replaces another one, she assumes the care of the entire list of patients from the other housestaff. Because housestaff blocks are generally two weeks in length and staggered for interns and residents, it is not uncommon for a patient to experience a change in either an intern or a resident. In summary, conditional on tracks, housestaff are quasi-randomly assigned teams that include attending physicians and other housestaff, and conditional on rotations, housestaff are also quasi-randomly assigned patients. I present evidence supporting such quasi-random assignment in Section 4 and in the Appendix.

[9] Housestaff are allowed to express preferences about vacation days, although these vacation days are few, about two weeks per year. Senior residents (third-year residents) may also express more general preferences about the timing of non-clinical blocks, such as research electives. For interns, schedules are assigned even prior to their arrival from medical school.

3.2 The State of Medical Knowledge

Inpatient medical care, the focus on this study, is comprised of three services at this institution: cardiology, oncology, and general medicine. Although general medical principles apply to all three services, patients on the subspecialty services of cardiology and oncology differ in their care, relative to patients on the general medicine service, in ways that reflect differences in the state of medical knowledge across medical fields.

In recent decades, by important measures, medical knowledge has progressed in cardiology and oncology to a greater extent than for other diseases. Table A-1 shows the number of original research articles appearing in the *New England Journal of Medicine* in the last ten years according to key disease specialty or subspecialty. Oncology and cardiology research papers lead the pack by a substantial margin. This reflects government, academic, and private industry priorities for producing knowledge, which appear to continue. Table A-2 reports current research funding by National Institute of Health (NIH) Institute or Center. Although Institutes often lump disease categories, the National Cancer Institute (NCI) with current funding of $6.7 billion and the National Heart, Lung, and Blood Institute (NHBLI) with current funding of $3.6 billion occupy the first and third positions for funding out of a list of 27 Institutes and Centers.

Perhaps the most important manifestation of differences in specialized medical knowledge is reflected in the organization of medical care, at both national and local levels. Cardiology and oncology are vastly the most common dedicated inpatient medicine subspecialty services across academic hospitals. Of the 24 residency programs ranked by *US News & World Report* and shown in Table A-3, 22 and 19 programs have dedicated cardiology and oncology services, respectively. Gastroenterology, represented at 5 programs, is the next most common subspecialty service. A similar relationship among subspecialties exists in the universe of internal medicine programs recognized by ACGME (Table A-4). Dedicated subspecialty services by definition are staffed by specialist attending physicians, who have several more years of training after internal medicine. In contrast, generalists are responsible for patients on general medicine services, who may choose to consult a specialist only if they deem it necessary. Finally, in terms of local knowledge, the institution of this study is considered a national leader in cardiology and oncology care according to *US News & World Report* hospital rankings.

As discussed in Section 2, differences in knowledge can affect medical care by housestaff, both statically in decisions and dynamically through the process of learning. Statically, a practice environment that features a high degree of knowledge, embedded in attending physicians, ancillary staff, institutional rules, constrains variation due to physicians in training, even if these housestaff have not yet fully internalized all information available at the institution. Dynamically, a sufficiently rich knowledge environment may allow convergence over time by learning. The fact that physicians need further subspecialty training to assume primary responsibility for cardiology and oncology patients, while no further training is required to treat pneumonia, is consistent with a larger body of knowledge required to care for these patients.

4 Data

This study uses data collected from several sources. First, I observe the identities of each physician on the clinical team – the intern, resident, and attending physician – for each patient on an internal medicine ward service and for each day in the hospital. Over five years, I observe data for 48,185 admissions, equivalent to 220,117 patient-day observations. Corresponding to these admissions are 724 unique interns, 410 unique residents, and 540 unique attendings. Of the housestaff, 516 interns and 347 residents are from the same-hospital internal medicine residency, with the remainder visiting from another residency program within the same hospital or from the other hospital. There is essentially no unplanned attrition across years of residency (i.e., except in two specific cases, housestaff observed only as interns are all "preliminary" interns).

The mean number of admissions for interns on the ward services of interest is 106; this includes numbers of admissions for visiting interns from the other hospital, which are much fewer than same-hospital interns. The corresponding mean number of admissions for all residents, including visiting residents, is 159. Residents see patients over two years (the second and third years of training), while internship is only one year long. Residents have fewer scheduled ward rotations in their third year. Thus the mean number of admissions for second-year residents is 129, while the mean number is 77 for third-year residents. Attending physicians have a much more varied patient load than housestaff, as they can have varying proportions of their work being clinical and can see patients in more than one hospital. The median number of patients seen by an attending over five years is 187, while the 90th percentile attending sees 627 patients over five years.

I observe detailed patient demographic and clinical information. Demographics include patient age, sex, race, and language. Clnical information derives primarily from billing data, in which I observe International Classification of Diseases, Ninth Revision, (ICD-9) codes and Diagnostic-related Group (DRG) weights. I use these codes to construct 29 Elixhauser comorbidity dummies and Charlson comorbidity indices (Charlson et al., 1987; Elixhauser et al., 1998). I also observe the identity of the admitting service (e.g., "Heart Failure Team 1"), within each of which patients are admitted for similar reasons (e.g., heart failure).[10]

I merge housestaff identities with scheduling data to capture residency training histories both on and off the wards, including outside-hospital clinical rotations, ICU rotations, outpatient rotations, and research electives. I also merge detailed residency application information for each housestaff, including demographics, medical school, USMLE test scores, membership in the Alpha Omega Alpha (AOA) medical honors society, other degrees, and position on the residency rank list. USMLE test scores represent a standardized measure of resident knowledge and ability. Position on the residency rank list represents desirability to the residency program; it could be made on a number of different criteria, including objective ability and promise to be a "leader" in a field of medicine. Finally, I observe the track of each housestaff physician, for example whether he is a preliminary or categorical intern, or whether she is from another residency program, as choices committed to prior to starting residency.

Although I will also make use of data on 30-day mortality and 30-day readmission, I focus on test costs as my main outcome measure. Medical spending has been the focus of much of the literature on practice variation (Fisher et al., 2003a,b) and is a key policy focus in its own right (Anderson et al., 2005). Test spending has particularly received increasing attention as the

[10]These admitting services are more narrowly defined than the broad categories of cardiology, oncology, and general medicine. However, even within specific admitting service, attendings may have different types of patients (e.g., a vertically integrated HMO admits to the same service as the hospital's own attendings). Therefore, without hand-coding attendings to practice groups and conditioning on these groups, patients are not quasi-randomly assigned to attendings. Still, as described above, housestaff are quasi-randomly assigned to patients, other housestaff, and attendings.

relative cost of tests has risen and now comprises a significant proportion of overall costs (Bates et al., 1999; Schroeder et al., 1974; Iwashyna et al., 2011). In this academic medical center, test costs comprise 10% of overall costs, which includes costs for physician and nurse salaries and operating costs. In addition to increasing in size, test costs are determined with particular discretion by physicians. Compared to treatment, there is less evidence on best practices for testing. Finally, test costs are daily outcomes, which allows me to exploit variation due to the changing composition of housestaff teams for the same patient. I define test costs as any cost incurred by a radiology (e.g., CT, MRI, nuclear medicine, ultrasound) or laboratory test order. The distribution of daily test costs is heavily right-skewed. I censor daily test cost observations greater than \$800, which comprise 3% of the data; the resulting distribution is shown in Figure A-1.[11] The mean daily test cost is \$124, while the median is \$49 and the 90th percentile is \$337. These daily costs aggregate to overall admission tests costs with a mean of \$714.

A key institutional fact described in Section 3 is that housestaff do not choose most of their learning experiences, at least in terms of their clinical rotations and in what order, peers and supervising physicians, and patients seen on the wards. Table 1 shows that interns and residents, respectively, with high or low spending effects are exposed to similar types of patients. Table 1 also show that housestaff with high or low spending effects are similarly likely to be assigned to coworkers and attendings with high or low spending effects. In the Appendix, I present more formal analyses on the exogenous assignment of housestaff physicians; I cannot reject the null that housestaff identities are jointly unrelated to patients types or other training experiences.

5 Influence under Uncertainty

This section examines the effect of relative influence under uncertainty on the variation of individual physician effects. The main analysis compares variation in spending attributable to housestaff as they progress in their training and discontinuously change roles, from intern to resident, at the end of the first year. As formalized in Equation (3), the effect of housestaff h is $a_h^* = g_h m_h / (g_h + g_{-h} + G)$, a function of both her beliefs and her relative influence under

[11] Results in this paper are robust to this censoring.

uncertainty. While beliefs develop continuously throughout training, I exploit the discontinuous change in her role at the end of the first year of training to estimate the importance of influence, holding beliefs constant, on decision-making. In general, influence could depend both on the relative precision of information (as I have motivated simply) and on titles. However, this institutional setting has the dual advantages of no differences in formal roles that mechanically increase resident influence and no unobserved selection into senior roles.[12] Influence associated with titles is thus likely to be endogenous to the former, fundamental mechanism of information.

For a patient being treated on day t of admission a by intern i, resident j, and attending k, I specify log daily test costs as

$$Y_{aijkt} = \mathbf{X}_a'\beta + \mathbf{T}_t'\eta + \xi_i^{\tau(i,t)} + \xi_j^{\tau(j,t)} + \zeta_k + \nu_{aij} + \varepsilon_{aijkt}. \tag{5}$$

Equation (5) includes a rich set of patient and admissions characteristics $\mathbf{X}_a$ for admission a, described in Section 4, and a set of time categories $\mathbf{T}_t$ for month-year combination, day of the week, and day of service relative to the admission day. I allow for attending fixed effects, ζ_k.[13]

The parameters of interest in Equation (8) are the time-varying effects, $\xi_i^{\tau(i,t)}$ and $\xi_j^{\tau(j,t)}$ for intern i and resident j, respectively, at tenure interval $\tau(\cdot, t)$ that is function of the housestaff and time. I consider $\xi_i^{\tau(i,t)}$ and $\xi_j^{\tau(j,t)}$ as constant within each tenure interval and housestaff, but otherwise impose no structure across tenure intervals for the same housestaff. Given that housestaff are exogenously assigned to patients, attendings, and peers, I conveniently specify housestaff effects as normally distributed random effects. Note that $\xi_i^{\tau(i,t)}$ and $\xi_j^{\tau(j,t)}$ are empirical tenure-specific estimates of $a_{i,\tau}^*$ and $a_{j,\tau}^*$, respectively, in Equation (3), and that tenure-specific standard deviations of $\xi_i^{\tau(\cdot)}$ are empirical estimates of $\sigma(\tau, \cdot)$ in Equation (4). Finally, I allow for shocks at the admission-intern-resident level, ν_{aij}. This reflects that, even controlling for patient

[12] Moreover, as mentioned in Section 3, two interns are usually assigned to a resident, and as a result, interns have more per patient clinical interactions and greater control over orders. These institutional facts suggest that, if information were equal, interns should have *more* influence than residents in the care of a given patient. As such, an observed increase in influence at the first-year mark may be viewed as a lower bound of the effect of more precise information on influence.

[13] To first order in the data, and consistent with the existing literature (Epstein and Nicholson, 2009; Molitor, 2011), the practice patterns of attendings are essentially fixed within a given environment. Further, attending physicians are not of interest in this analysis, and unlike housestaff physicians, they are not randomly assigned patients.

observables, some patients will naturally result in more test costs than others. This specification is more flexible than simply an admission-level shock ν_a but is computationally less burdensome with respect to the maximum-likelihood covariance matrix.

Figure 1 and Table A-5 present results for the estimated standard deviations of the distributions of housestaff effects within each tenure interval τ. In my baseline specification, I consider non-overlapping tenure intervals that are 60-days in length. In addition to the estimated standard deviations of intern and resident effects, I also present the estimated standard deviation for the admission-intern-resident-level shock in Table A-5. This gives a useful benchmark of unobserved variation at the patient (admission) level relative to the contribution of housestaff in terms of daily test spending.[14]

I find large and significant variation in housestaff effects during all intervals of time. A standard-deviation increase in the intern effect, $\xi_i^{\tau(i,t)}$, increases test spending by about 15-30%. A standard-deviation increase in the resident effect, $\xi_j^{\tau(j,t)}$, increases spending by about 55-70%. In comparison, the standard deviation for patient-housestaff-level shocks, ν_{aij}, is 40%. Given the large qualitative heterogeneity across patients that characterizes care across ward services, it is notable that residents alone are responsible for more variation in spending than unobserved patient characteristics. Based on mean test spending of \$714 for an admission, this variation in housestaff effects represents increases of approximately \$70 and \$450 for a standard-deviation increase in intern and resident effect, respectively.

Physician effects are determined by both information and influence, as in Equation (3). However, under the assumption that housestaff beliefs are continuous over time, the discontinuity at the one-year tenure mark identifies the change in influence due to a discontinuous increase in relative tenure, from being at least one year less experienced to being one year more experienced than the teammate. The change in spending-effect variation indeed is discontinuous, more than doubling across the one-year tenure mark. This implies a large effect of influence, due to relative information under uncertainty, on the size of physician spending variation.

[14]In some sense, it provides an overestimate of the patient-level variation because it is actually a shock by the patient interacted with the intern and resident, the distribution of which could be wider. Although most patients are taken care of by only one intern and resident, some patients will be taken care of by more than one intern or resident.

6 Learning under Uncertainty

In this section, I examine housestaff learning, based on two main source of evidence. First, I study the serial correlation of housestaff effects across adjacent time periods, as a measure of persistence. By construction, the correlation coefficient between two sets of spending effects of the same housestaff separated by time should be invariant to changes in scale. The serial correlation therefore measures persistence in a way that is conceptually distinct from changes in influence. Increasing persistence only reflects that physicians are settling on choices similar to their past choices, and these choices may be different from those of other physicians.[15]

Second, I study the convergence of housestaff effects with tenure, separately in the different knowledge environments of specialist and generalist services. Convergence – defined as a decrease in the variation of housestaff effects with tenure – implies that housestaff become more like one another in their effect on care decisions and is a more direct test of learning to practice a similar standard.[16] I compare convergence (or the lack thereof) of housestaff effects in the high-knowledge specialist services and in the low-knowledge general medicine service. I rule out an alternative hypothesis under which differences in learning occur because cardiology and oncology have a higher concentration of diagnoses, by creating pseudo-services in general medicine that have high or low diagnostic concentration. I show that convergence in the specialist services is driven by care for diagnoses for which there exists a published guideline.

6.1 Serial Correlation within Housestaff

In this analysis, I study the serial correlation across estimated housestaff effects across time intervals. I estimate housestaff effects in Equation (5) as distributions of random effects, but I can estimate an empirical Bayes prediction for the effect of each housestaff h during tenure interval τ. Specifically, this is done in a standard manner explained in Searle et al. (1992): First,

[15] In the conceptual framework in Section 2, particularly in Equation (2), this persistence may be most literally thought of as persistence of beliefs. The development of persistent but heterogeneous practices is consistent with housestaff ceasing to learn a common practice. More generally, heterogeneous persistence may represent heterogeneous preferences or skills, which I explore in Section 7.

[16] Maintaining the framework in which housestaff effects represent both beliefs and influence, it is possible that convergence could result from steadily decreasing influence. I show evidence that this is not the case, at least in terms of relative influence between interns and residents, by measuring the influence of interns who work with second- versus third-year residents.

the random effects model parameters are estimated by maximum likelihood. This includes both coefficient estimates (i.e., $\hat{\beta}$ and $\hat{\eta}$) and parameter estimates of the variance-covariance matrix (i.e., the standard deviation estimates $\hat{\sigma}_{\xi}^{\tau}$, $\hat{\sigma}_{\nu}$, and $\hat{\sigma}_{\varepsilon}$). Overall random error terms are imputed by the data and coefficient estimates, and predicted random effects are shrunk by the variance-covariance estimates. The Bayesian shrinkage factor reconciles the fact that in finite samples, the distribution of fixed effects should be wider than the distribution parameter estimated for variance-covariance matrix would imply.

Using these empirical Bayes housestaff-tenure effects, $\tilde{\xi}_h^{\tau}$, I then estimate the serial correlation across intervals. That is, I measure $corr\left(\tilde{\xi}_h^{\tau}, \tilde{\xi}_h^{\tau-j}\right)$, where j is the number of tenure intervals separating effect estimates for the same housestaff.[17] In Figure A-2, I show the estimated correlation coefficient between each tenure interval and the previous interval. In Table A-6, I show corresponding coefficients not only between adjacent intervals but between intervals up to three intervals apart. Recall that each interval is composed of 60 days, so that comparing adjacent intervals is equivalent to comparing housestaff effects at average tenures 2 months apart.

In the first year of training, interns effects are highly unstable. The correlation between housestaff effects in the first two months of training and those in the next two months is not statistically different from 0. However, the serial correlation between housestaff effects increases with tenure. At the beginning of their second year, when interns become residents, the correlation coefficient with the previous interval is greater but still relatively small at 0.15. It then rapidly increases throughout the second year up to 0.80. In the third year, when the volume of patients seen decreases relative to the previous year, the serial correlation between intervals starts to decrease.

The serial correlation of housestaff effects across tenure intervals suggest that practice patterns become more stable with more experience. It also appears that experience as a resident has greater influence in shaping practice patterns, or at least in developing stable ones, than

[17] In addition to this simple calculation of the correlation coefficient, I calculate an alternative measure that adjusts for the number of observations used to calculate the correlation between a given pair of estimated random effects. Using assumptions implicit in Equation (5), I adjust the correlation coefficient downward for fewer observations, given the estimated variances of ν_{aij} and ε_{aijkt}. However, consistent with sufficiently many observations, this exercise does not noticeably change results.

does experience as an intern. The serial correlation of housestaff effects increases slowly during internship but does so much more quickly during residency. By the end of the second year, housestaff effects are remarkably stable. However, stability is not strictly monotonic with training. In the third year, a smaller decrease in serial correlation suggests that practice patterns, at least during residency training, are not permanent. Even after practice patterns have achieved a remarkable degree of stability, they can still become less stable.

Increasing persistence, or stability, of practice patterns is informative of the process of learning. Highly persistent housestaff effects not only suggest that physicians are retaining past information for current medical practice, but also that they are incorporating less new knowledge. A decrease in serial correlation, therefore, is more nuanced: Physicians with decreasing persistence in practice patterns may be incorporating new information to a greater degree, or they may be forgetting past information. In either case, a lack of practice, which occurs with fewer clinical rotations in the third year of residency, may lead to decreased practice-pattern stability.

6.1.1 The Role of Continued Practice

In order to test the hypothesis that stability requires continued practice, I estimate correlation coefficients for month-long tenure intervals that are separated by a month in between and condition on a nearby tenure interval having few or no patient observations. I consider two types of such nearby intervals, for a random effect pair $\tilde{\xi}_h^{\tau}$ and $\tilde{\xi}_h^{\tau-2}$: the interval $\tau - 1$ in between the pair, and the interval $\tau - 3$ prior to the pair.[18] The reason for this conditioning is compare housestaff effects between the same tenure intervals and surrounded by the same relative frequency of patient encounters. If learning is entirely cumulative and does not depend on the continuity of practice, then the relative timing of the nearby tenure interval with few or no patient observations should have little effect; in fact, having many patients in the intervening interval $\tau - 1$ may reduce the correlation between the pair. However, if learned practice patterns need reinforcement through continued practice for stability, then the opposite result

[18] I also consider the interval $\tau + 1$ after the pair, instead of $\tau - 3$ prior to the pair, and results are qualitatively similar.

should obtain: Continued practice during $\tau-1$ should increase the correlation between the pair.

In Figure A-3, I plot the correlation coefficients between random-effect pairs $\tilde{\xi}_h^\tau$ and $\tilde{\xi}_h^{\tau-2}$, where I define each interval τ as lasting a month.[19] A lack of patients in the interval between those corresponding to the random effect pair is associated with a delay, by exactly a month, in the rise in stability after interns become residents. During the later two years of training, correlation also appears to be slightly lower for pairs with few or no patients in between (in interval $\tau-1$ as opposed to interval $\tau-3$). This evidence suggests that continued practice is important for both the initial stabilization of practice patterns, when housestaff gain greater authority, and for the continued stability.

6.2 Knowledge Environment and Convergence

As described in Section 3, I consider specialist-directed services of cardiology and oncology as taking place in a high-knowledge environment, while general medicine represents a low-knowledge environment. As the baseline analysis of convergence, I estimate Equation (5) for each of the three ward services of cardiology, oncology, and general medicine. As in Section 5, this yields the standard deviation of housestaff effect distributions by tenure, now separately for each of the ward services.

In Figure 2, I show each of these profiles of housestaff-effect variation over tenure for cardiology, oncology, and general medicine. Two features distinguish the profiles of the high-knowledge subspecialty services and that of the low-knowledge general medicine service. First, housestaff effects have less variation in cardiology and oncology than in general medicine for all tenure intervals for residents. The highest standard deviation in housestaff effects (at the beginning of the resident role) is a 50-60% spending effect in cardiology and oncology, compared to a 80% spending effect in general medicine. Variation is similar (for oncology) or lower (for cardiology) for interns, as well. Second, housestaff effects significantly converge when housestaff are residents in cardiology and oncology, but there is no evidence of convergence in general medicine.

[19] For the results in Figure A-3, I define "few" patients as 40 patients or fewer, corresponding to the 20th percentile of monthly patient volume conditional on seeing any patients at the main hospital. I omit correlation coefficients calculated with fewer than 10 observations. Results are largely similar using different thresholds for defining few patients, and higher thresholds of observations for calculating correlation coefficients.

The standard deviation of spending variation steadily reduces by the end of training to 15% in cardiology and 30% in oncology.

Consistently lower housestaff-effect variation at all points in tenure supports the prediction that higher G in high-knowledge environments constrain variation. At any given time, especially when housestaff have relatively little information, the external practice environment – including supervising attendings, other health care personnel, and institutional rules – reduces the influence that housestaff have on decisions. The second feature, of convergence in specialist services but not in general medicine, suggests that housestaff learn toward a standard practice in high-knowledge environments (i.e., $g'(\tau)$ is sufficiently large with large G) but that continued variation can remain when the knowledge to be learned is insufficient.[20]

Merging cardiology and oncology services into a single "specialist service," I quantify a rate of convergence in spending effects among residents of about a 14.3 percentage-point decrease in the standard deviation of housestaff effects per year. In other words, given a standard deviation of 47.8% at the beginning of the second year (when interns become residents), this is equivalent to a relative decrease of 59.8% of this standard deviation over the next two years. Randomizing over 10,000 placebo combinations of housestaff-service-months (of about 1.27×10^{970} combinations) yields a range of placebo convergence estimates of $[-0.073, 0.085]$, suggesting that the actual estimate -0.143 is extremely significant (see Figure 3). Details are given in the Appendix.

6.2.1 The Role of Published Guidelines

One measure of knowledge is the existence of published guidelines for managing patients with certain diagnoses. I therefore link primary diagnoses of each patient admission to guidelines for diagnostic practice collected by the US Agency for Healthcare Research and Quality (guidelines.gov). While slightly more cardiology and oncology admissions have guidelines linked to their primary diagnoses, the guidelines are roughly related to about half of the diagnoses in each of the services. This partly reflects the inadequacy of ICD-9 codes both for describing

[20] An alternative explanation, though inconsistent with institutional facts, is that housestaff have steadily decreasing influence during their last two years. Empirically, I show in the Appendix that interns have less influence when working with third-year residents than when working with second-year residents, which supports greater influence with resident tenure.

patients and for describing guidelines. For example, the most common primary ICD-9 code, 786.50 for "Chest pain, not otherwise specified," is the primary diagnosis for 1,465 patients on the cardiology service and 1,655 patients on the general medicine service. Although these two patient groups share the same primary diagnosis, they were triaged to different services based on pre-admission characteristics. Finally, guidelines are an imperfect proxy for true knowledge and often only indicate that clinicians feel a need for guidance on important but unsolved issues.

Nevertheless, I examine differences in the tenure profiles of housestaff-effect distributions between diagnoses with published guidelines *within* each service. I estimate Equation (5) and plot the standard deviation of housestaff effects for each tenure period, separately for each of the services and for diagnoses with and without a published guidelines. Results, in Figure 4, show a significant difference in profiles between diagnoses with and without guidelines in the specialist services. Within the specialist services, resident spending effects have less variation initially for diagnoses without guidelines and little to no convergence over the subsequent two years, while spending effects are larger in magnitude for diagnoses with guidelines and show significant subsequent convergence. This relationship possibly also holds but to a minimal degree within the general medicine service.

The production of guidelines is of course endogenous, and the fact that variation is initially higher for diseases with published guidelines perhaps reflects this. However, the same diagnoses with guidelines subsequently show greater convergence in the specialist services. This is consistent with knowledge as facilitator of learning and suggests an interaction between organizational services driven by specialists and care for which there exists guidelines. However, it is also noteworthy that care under guidelines does not significantly converge in the general medicine service, which suggests that the organization of care (i.e., whether care is driven by specialists) matters, at least as a marker that predicts learning.

6.2.2 Diagnostic Concentration across Services

In comparing specialist services against general medicine, an obvious difference between the two is that cardiology and oncology services care for patients with a greater concentration of primary diagnoses. Naturally, cardiology concerns the treatment of patients with primary cardiology diagnoses. Although inpatient oncology manages a wider variety of primary admission diagnoses, these diagnoses are largely the cause of underlying cancer. An alternative hypothesis is therefore that experience within a set of fewer diagnoses, rather than the greater availability of knowledge, drives convergence in cardiology and oncology as opposed to general medicine.

Table 2 shows summary statistics for patients admitted to cardiology, oncology, and general medicine, as well as measures of concentration in their primary diagnoses. The level of specifying diagnoses is primarily determined by the established system of coding, as reflected by ICD-9 codes, that is accepted by Medicare and other insurers to adequately capture differences in diagnoses that are meaningful for medical evaluation and management. At a higher level, ICD-9 codes are aggregated into Major Diagnostic Categories (MDC), which generally divide diagnoses according to organ system. As expected, cardiology and oncology both have higher diagnostic concentrations than general medicine has, as measured by the Herfindahl-Hirschman Index (HHI) of either patient MDC or primary ICD-9 code.

In order to evaluate the alternative that diagnostic concentration is the distinguishing characteristic of cardiology and oncology that drives convergence, I create two artificial services formed from the pool of general medicine admissions. In the first pseudo-service, which I name "MED1," I select admissions with the three most common MDCs, which correspond to "Circulatory System" (MDC 5), "Respiratory System" (MDC 4), and "Digestive System" (MDC 6).[21] The second pseudo-service ("MED2") is comprised of patients with any of the remaining 22 MDCs. Table 2 shows summary statistics for these two pseudo-services alongside those for the actual cardiology, oncology, and general medicine services. MED1 has, as expected, greater diagnostic concentration than MED2, both in terms of MDC and primary ICD-9 code. MED1 is indeed more diagnostically concentrated than the actual oncology service.

Having constructed these pseudo-services, I then estimate the time profile of housestaff-effect variation, using Equation (5). Figure A-4 shows the standard deviation of the housestaff-effect distribution for both MED1 and MED2. Despite substantial differences in diagnostic concentration, the two pseudo-services have essentially identical tenure profiles of housestaff-

[21] Note that Circulatory System corresponds most closely with cardiology. Including patients with cardiology diagnoses in this pseudo-service creates a stronger falsification test. Results are robust to excluding cardiology and instead using including the fourth most common MDC, "Kidney and Urinary Tract" (MDC 11).

effect variation. The standard deviation in housestaff spending effects remains at 75-80% for both pseudo-service, and in particular, neither pseudo-service shows any evidence of convergence as housestaff gain in tenure.

While this falsification exercise can replicate differences in diagnostic concentration, even separating patients with a primary diagnosis in cardiology into one pseudoservice, patients in both pseudoservices are by definition regarded as appropriate for management by a general medicine attending physician. The larger fund of knowledge in cardiology or oncology, either in terms of published clinical trials or in terms of local expertise at this institution, is the primary institutional determinant that separates certain cardiology and oncology patients on the subspecialty services from the others managed by general medicine. Further, the distinction is not whether there exists subspecialists for certain diagnostic groups. There are pulmonologists and infectious disease specialists who can be consulted for the rare, complicated cases of pneumonia. However, the vast majority of pneumonia cases do not require such consultation, let alone primary management from subspecialists. Finally, convergence as measured in this setting implies learning on the scale of years and not days and is consistent with the fact that cardiologists and oncologists require years of further training. This suggests a greater rate of learning in specialist services *per se*, as opposed to a mechanism depending on the frequency of diagnoses.

7 Mean Effects of Housestaff Characteristics and Experience

I have shown that housestaff-effect variation significantly depends on the external organizational and knowledge environment that governs relative influence and learning under uncertainty. Given the emphasis in the literature on human capital and intrinsic heterogeneity (e.g., ability) (e.g., Doyle Jr et al., 2010; Fox and Smeets, 2011; Bartel et al., 2014), it is natural to compare the magnitude of these mechanisms with the size of predictions in spending levels according to housestaff characteristics and experience.

Using unusually rich data on housestaff characteristics and quasi-experimental variation, I am able to address this question in detail. I find mean effects of numerous housestaff characteristics and measures of experience that are either insignificant or an order of magnitude lower than the effects of relative influence and potential convergence on the standard deviation of housestaff-effect variation. This provides evidence that traditional concepts of intrinsic heterogeneity and human capital are relatively less valuable as predictors of variation than uncertainty is in determining the scale of variation in medical care.

7.1 Housestaff Characteristics

An advantage of this empirical setting is that I observe different types of housestaff undergo the same average training experience.[22] I observe detailed and insightful housestaff characteristics that are likely correlated with differences in preferences and abilities. For example, USMLE scores directly measure medical knowledge as a medical student; position on the residency rank lists reflects overall desirability; and residency tracks such preliminary status reflect important career decisions and lifestyle preferences, such as a commitment to become a future radiologists rather than a primary care physician.

In addition to housestaff in the main residency program, I observe both interns and residents from an internal medicine residency based in another hospital. For these outside-hospital housestaff, I can evaluate the effect of their presence on medical teams, both when they are interns and residents. This effect includes both differences in housestaff characteristics, training experiences, and the interaction between the two: Housestaff from the outside residency are both a selected sample and experience different training by spending most of their time at the outside hospital (the outside residency is nationally recognized but lower ranked, and the outside hospital is known to be more cost-conscious).

For each of these housestaff groups with different characteristics, I perform the following regression:

$$Y_{aijkt} = \alpha_m Characteristic_h^m + \mathbf{X}'_a\beta + \mathbf{T}'_t\eta + \zeta_{-hk} + \varepsilon_{aijkt}, \tag{6}$$

where $HousestaffGroup_i^m$ is an indicator variable that takes the value of 1 if housestaff h had characteristic (or made track choice) m prior to starting residency. The coefficient of interest

[22] Previous studies have investigated the effect of coarse measures of observable physician characteristics (e.g., gender) and training experiences (e.g., place of medical school or residency) in a single regression (e.g., Epstein and Nicholson, 2009). The problem with this approach is that housestaff may select into different experiences.

is α_m, which represents the effect that characteristic m has on daily test spending. By ζ_{-hk}, I include fixed effects for the other housestaff on the team, $-h$, which would be the intern if the index housestaff is the resident or the resident if the index housestaff is the intern, and the attending k.

In Table 3, I show a subset of results for Equation (6). Effects of pre-residency characteristics and track choices are generally small and insignificant. There are two characteristics that predict statistically significant lower spending: male sex and high USMLE test score. Male interns have 2% lower daily spending costs, significant at the 10% level; male residents have 4% lower daily spending costs, significant at the 5% level. A high USMLE score predicts 3% lower daily spending, significant at the 10% level, for residents. Also in Table 3, I show results for the effect of housestaff from different programs, reflecting both differences in selection across programs (i.e., intrinsic heterogeneity) and differences in learning experiences. Interns from other residency programs do not have significantly different mean spending effects. However, medicine residents from the other hospital spend 17% less than residents of the main hospital.

Finally, to study more directly the effect of increasing influence interacted with housestaff characteristics, I estimate this regression, for each housestaff group m:

$$Y_{aijkt} = \sum_{h \in \{i,j\}} \alpha_m^{\tau(h,t)} \mathit{Characteristic}_h^m + \mathbf{X}_a' \beta + \mathbf{T}_t' \eta + \zeta_k + \varepsilon_{aijkt}, \tag{7}$$

in which the effect of being in group m depends on tenure for both the intern and the resident. Results are uniformly insignificant for a wide range of housestaff characteristics. Figure 5 illustrates results for USMLE test scores at each tenure interval; values for the standard deviation of the housestaff-effect distribution at each tenure interval are also shown for reference. Other results are presented in the Appendix.

Overall, these results show that intrinsic heterogeneity, to the extent that it is correlated with characteristics and pre-residency choices, explains relatively little compared to the size of variation that depends on influence and learning. For some characteristics, the effect does seem to increase in magnitude with tenure, which supports the idea of increasing influence.[23] For

[23]Although I do not explicitly consider intrinsic heterogeneity in Section 2, since it is unimportant to first order, intrinsic heterogeneity could be magnified with influence that is purely based on information, as I mention

results regarding housestaff from the outside hospital, the significantly larger effect of residents relative to interns could reflect both increased influence as well as the longer history of learning experiences at the outside hospital. Although I cannot rule out intrinsic heterogeneity (magnified by influence) as contributing to the mean effect of residents from the outside hospital, given that the same medical students are often considered competitive at both programs, intrinsic heterogeneity in this case is also qualitatively less important than learning experiences.

7.2 Housestaff Experience

I consider several measures of cumulative intern experience that consider days on various rotations, patients seen, and supervising physicians worked with (both the number of physicians and their spending effects). For each of these measures, I estimate a regression of the form

$$Y_{aijkt} = \alpha_m InternHx_{it}^m + \mathbf{X}_a' \beta + \mathbf{T}_t' \eta + \zeta_{ijk} + \varepsilon_{aijkt}, \tag{8}$$

where $InternHx_{it}^m$ represents the measures of cumulative experience m for intern i at time t. To evaluate potential effects of intern experiences future practice as a resident, I also consider interactions with resident tenure:

$$Y_{aijkt} = \alpha_m^{\tau(j,t)} InternHx_j^m + \mathbf{X}_a' \beta + \mathbf{T}_t' \eta + \zeta_{ik} + \varepsilon_{aijkt}. \tag{9}$$

I again find no significant effects for any of these analyses; results are presented in the Appendix.

Finally, I consider the effect of resident tenure on outcomes of test daily spending, total daily spending, length of stay, 30-day readmissions, and 30-day mortality for each of the ward services. Because I also control for month-year interactions, I study this as the effect of having a third-year housestaff, as opposed to having a second-year housestaff, as the resident:

$$Y_{aijkt} = \alpha \mathbf{1} \left(\tau \left(j,t \right) > 2 \text{ years} \right) + \mathbf{X}_a' \beta + \mathbf{T}_t' \eta + \zeta_{ik} + \varepsilon_{aijkt}. \tag{10}$$

The coefficient α is small and insignificant for all of these outcomes. Table 4 lists results

in Footnote 4.

along with counterfactuals for switching to a resident one standard deviation above or below in housestaff-effect distribution for the relevant outcome.

Overall, these results indicate that measures of housestaff experience are also poor predictors of outcomes, especially relative to the large variation in outcomes. Rough summary measures of experience, even with rich data, are likely to be impractical representations of the specific experiences that lead to learning. As a rare example of identifiable experiences that have an affect on physician practices, Choudhry et al. (2006) finds that physicians are less likely to prescribe anticoagulants after having a patient with an adverse bleeding event. Such events are not only difficult to identify in the data; they are at best highly challenging to aggregate into measures that predict general spending effects. Furthermore, aggregate experiences in the same training environments should be similar in expectation. However, based on convergence results in Section 6, this study suggests that learning to a common practice pattern requires more than experience but also an environment with sufficient knowledge.

8 Conclusion

In this paper examines the behavioral foundations of persistent variation in medical care, a well-documented but poorly understood phenomenon that is closely related large variation in productivity in other industries (e.g., Chandra et al., 2013). In particular, I study the development of practice patterns summarized by spending among housestaff, physicians who are training during the most formative stage in their careers, within a large institution. This setting allows me to examine two important facts in health care delivery: Medical care is delivered in teams within organizations, and physician practice patterns must be learned within a broader knowledge environment that is often unclear about what is correct course of diagnosis and treatment for a large proportion of decisions. I find evidence for the large effect of team structure on the contribution individual physicians have on medical decisions, and for the importance of the external knowledge environment both in constraining variation and in enabling convergence to a common standard of practice. In my setting, both of these channels dwarf the contribution of intrinsic heterogeneity or human capital as a traditional predictor of differences in outcomes.

Despite the proliferation in new technologies for diagnosis and treatment in medical care in recent decades, information to enable clinicians to choose among these technologies has grown at a much slower rate (Frankovich et al., 2011; Tinetti and Studenski, 2011). Since it seems impossible to design studies to answer questions about the optimal medical course for patients who differ along seemingly countless dimensions, it would be unreasonable to expect that external knowledge will ever inform physicians on the correct decisions to make with infinite precision. Given the costs of learning (Rogerson et al., 2005) and the potential for misinterpreting signals (Acemoglu et al., 2006), uncertainty in the knowledge environment is theoretically crucial in determining whether physicians move toward adopting a common practice pattern. This study shows empirical evidence that substantial convergence across individuals in an institution does occur, depending on the content of knowledge in the practice environment.

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz, "Learning and Disagreement in an Uncertain World," Working Paper 12648, National Bureau of Economic Research October 2006.

Anderson, G. F., P. S. Hussey, B. K. Frogner, and H. R. Waters, "Health spending in the United States and the rest of the industrialized world," *Health Affairs*, 2005, *24* (4), 903–914.

Bartel, Ann P., Nancy Beaulieu, Ciaran Phibbs, and Patricia W. Stone, "Human Capital and Productivity in a Team Environment: Evidence from the Healthcare Sector," *American Economic Journal: Applied Economics*, April 2014, *6* (2), 231–259.

Bartelsman, Eric J and Mark Doms, "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, September 2000, *38* (3), 569–594.

Bates, D. W., G. J. Kuperman, E. Rittenberg, J. M. Teich, J. Fiskio, N. Ma'Luf, A. Onderdonk, D. Wybenga, J. Winkelman, and T. A. Brennan, "A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests," *The American Journal of Medicine*, 1999, *106* (2), 144–150.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Healthcare Exceptionalism? Productivity and Allocation in the U.S. Healthcare Sector," Working Paper 19200, National Bureau of Economic Research July 2013.

Charlson, Mary E., Peter Pompei, Kathy L. Ales, and C. Ronald MacKenzie, "A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation," *Journal of Chronic Diseases*, 1987, *40* (5), 373–383.

Choudhry, N. K., G. M. Anderson, A. Laupacis, D. Ross-Degnan, S. L. T. Normand, and S. B. Soumerai, "Impact of adverse events on prescribing warfarin in patients with atrial fibrillation: matched pair analysis," *British Medical Journal*, 2006, *332* (7534), 141–145.

Cooke, Molly, David M. Irby, William Sullivan, and Kenneth M. Ludmerer, "American Medical Education 100 Years after the Flexner Report," *New England Journal of Medicine*, 2006, *355* (13), 1339–1344.

Cutler, David, Jonathan Skinner, Ariel Dora Stern, and David Wennberg, "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," Working Paper 19320, National Bureau of Economic Research August 2013.

Cyert, Richard and James March, *A Behavioral Theory of the Firm*, Oxford: Blackwell, 1963.

Dranove, David, Subramaniam Ramanarayanan, and Andrew Sfekas, "Does the Market Punish Aggressive Experts? Evidence from Cesarean Sections," *The B.E. Journal of Economic Analysis & Policy*, January 2011, *11* (2).

Elixhauser, Anne, Claudia Steiner, D. Robert Harris, and Rosanna M. Coffey, "Comorbidity Measures for Use with Administrative Data," *Medical Care*, January 1998, *36* (1), 8–27.

Ellison, G. and D. Fudenberg, "Rules of thumb for social learning," *Journal of Political Economy*, 1993, *101* (4), 612–643.

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

Fisher, Elliott S., David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder, "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care," *Annals of Internal Medicine*, February 2003, *138* (4), 273–287.

_ , _ , _ , _ , _ , and _ , "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care," *Annals of Internal Medicine*, February 2003, *138* (4), 288–298.

Flexner, Abraham, *Medical education in the United States and Canada: a report to the Carnegie Foundation for the Advancement of Teaching*, Carnegie Foundation for the Advancement of Teaching, 1910.

Fox, Jeremy T. and Valerie Smeets, "Does Input Quality Drive Measured Differences in Firm Productivity?," *International Economic Review*, November 2011, *52* (4), 961–989.

Frankovich, Jennifer, Christopher A. Longhurst, and Scott M. Sutherland, "Evidence-Based Medicine in the EMR Era," *New England Journal of Medicine*, 2011, *365* (19), 1758–1759.

Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, October 2000, *108* (5), 874–904.

_ and Esteban Rossi-Hansberg, "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, November 2006, *121* (4), 1383–1435.

Gibbons, Robert and Rebecca Henderson, "What do managers do? Exploring persistent performance differences among seemingly similar enterprises," in Robert Gibbons and

John Roberts, eds., *The Handbook of Organizational Economics*, Princeton, NJ: Princeton University Press, 2012, pp. 680–732.

Grytten, J. and R. Sorensen, "Practice variation and physician-specific effects," *Journal of Health Economics*, 2003, *22* (3), 403–418.

Iwashyna, T. J., A. Fuld, D. A. Asch, and L. M. Bellini, "The impact of residents, interns, and attendings on inpatient laboratory ordering patterns: A report from one university's hospitalist service," *Academic Medicine*, 2011, *86* (1), 139.

Jr, J. J. Doyle, S. M. Ewer, and T. H. Wagner, "Returns to physician human capital: Evidence from patients randomized to physician teams," *Journal of Health Economics*, 2010, *29* (6), 866–882.

Lazear, Edward P., Kathryn L. Shaw, and Christopher Stanton, "The Value of Bosses," *Journal of Labor Economics*, 2014, *Forthcoming.*

Manski, Charles F., "Useful Variation in Clinical Practice under Uncertainty: Diversification and Learning," Technical Report, Northwestern University October 2013.

Molitor, David, "The evolution of physician practice styles: Evidence from cardiologist migration," Technical Report, Massachusetts Institute of Technology 2011.

Radner, Roy, "The Organization of Decentralized Information Processing," *Econometrica*, 1993, *61* (5), 1109–46.

Rogerson, Richard, Robert Shimer, and Randall Wright, "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, December 2005, *43* (4), 959–988.

Schroeder, S. A., A. Schliftman, and T. E. Piemme, "Variation among physicians in use of laboratory tests: relation to quality of care," *Medical Care*, 1974, *12* (8), 709–713.

Searle, S. R., G. Casella, and C. E. McCulloch, *Variance Components*, Wiley New York, 1992.

Syverson, Chad, "What Determines Productivity?," *Journal of Economic Literature*, June 2011, *49* (2), 326–365.

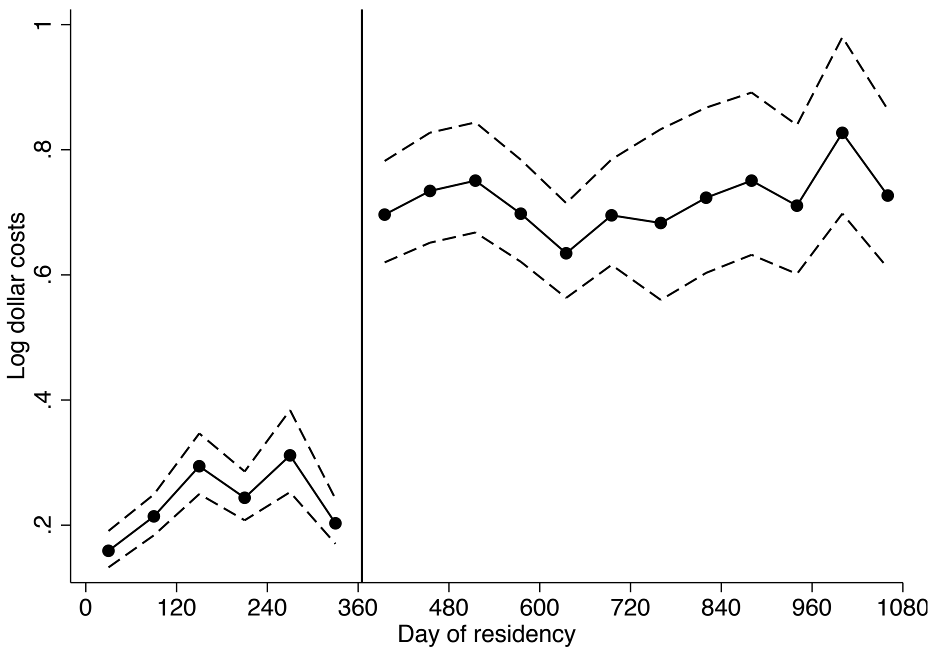
Tinetti, Mary E. and Stephanie A. Studenski, "Comparative Effectiveness Research and Patients with Multiple Chronic Conditions," *New England Journal of Medicine*, 2011, *364* (26), 2478–2481.

Wennberg, J. and A. Gittelsohn, "Small area variations in health care delivery," *Science*, 1973, *182* (4117), 1102–1108.

Wennberg, J. E., E. S. Fisher, and J. S. Skinner, "Geography and the debate over Medicare reform," *Health Affairs*, 2002, *21* (2), W96–W114.

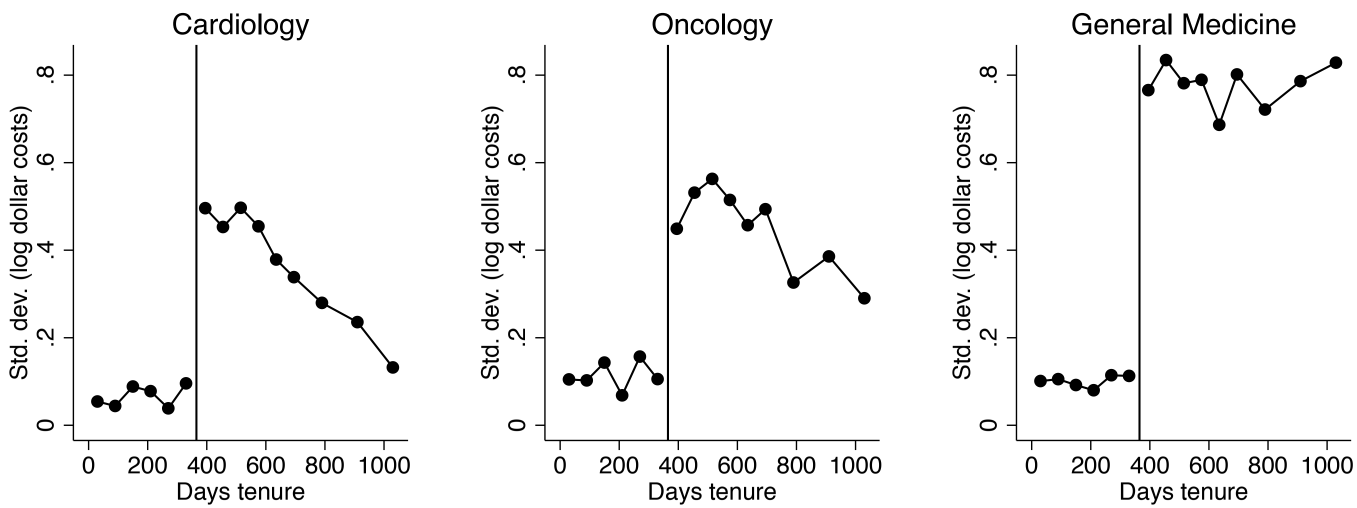
_ , _ , T. A. Stukel, and S. M. Sharp, "Use of Medicare claims data to monitor provider-specific performance among patients with severe chronic illness," *Health Affairs*, 2004, *23*, 5–18.

Figure 1: Standard Deviation of Housestaff Random Effects by Tenure



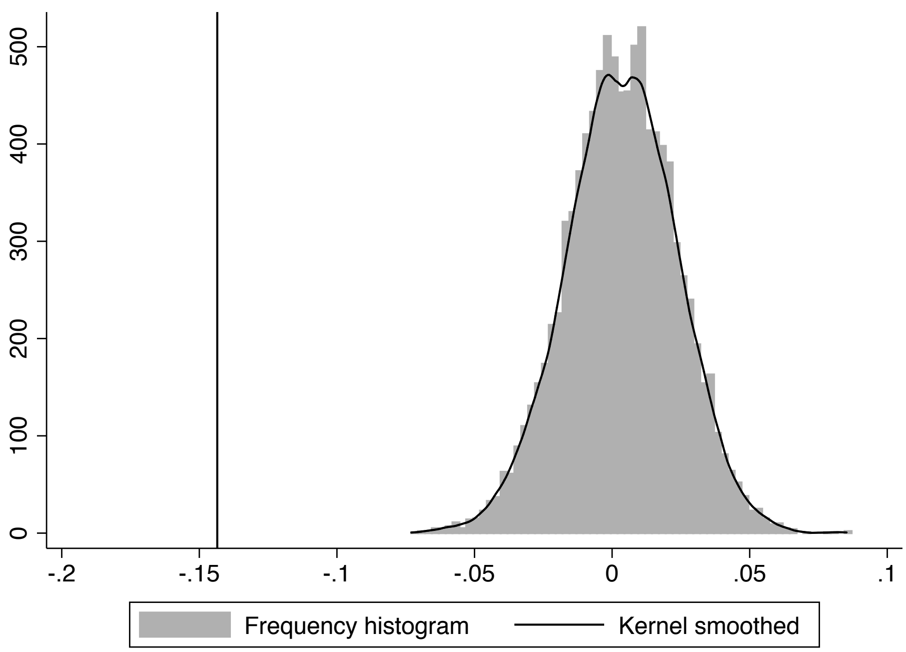
Note: This figure shows the standard deviation in a random effects model of log daily test costs shown in Equation (5) at each non-overlapping two-month tenure interval. Point estimates are shown as connected dots; 95% confidence intervals are shown as dashed lines. The model controls for patient and admission observable characteristics, time dummies (month-year interactions, day of the week), and attending identities (as fixed effects). Patient characteristics include demographics, Elixhauser indices, Charlson comorbidity scores, and DRG weights. Admission characteristics include the admitting service (e.g., "Heart Failure Team 1"). Intern and resident effects are modeled as random effects, as well as an admission-intern-resident interaction random effect in order to account for admission-level unobservables. The standard deviation of housestaff-effect distribution at each tenure interval is estimated by maximum likelihood. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; a vertical line denotes the one-year tenure mark.

Figure 2: Housestaff-effect Variation by Tenure in Each Service



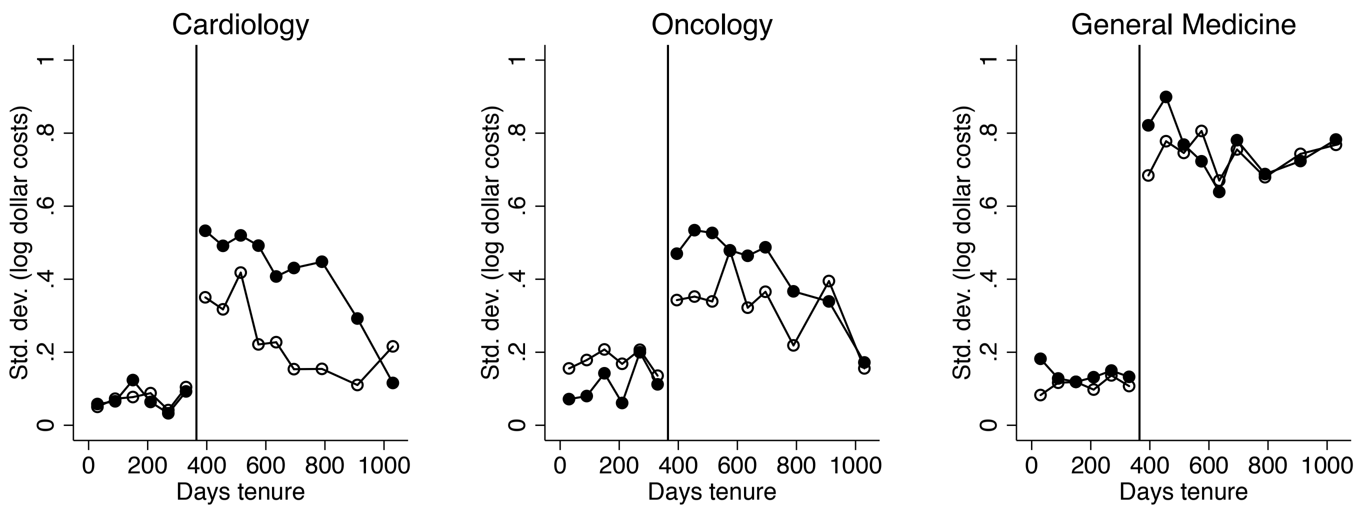
Note: Similar to Figure 1, this figure shows the standard deviation in a random effects model, as in Equation (5), of log daily test costs at each non-overlapping two-month tenure interval but for each service of cardiology, oncology, and general medicine. Controls are the same as those listed in the caption for Figure 1. Intern and resident effects are modeled as random effects, as well as an admission-intern-resident interaction random effect in order to account for admission-level unobservables. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; vertical lines denote the one-year tenure mark.

Figure 3: Systematic Placebo Tests for Specialist-service Convergence



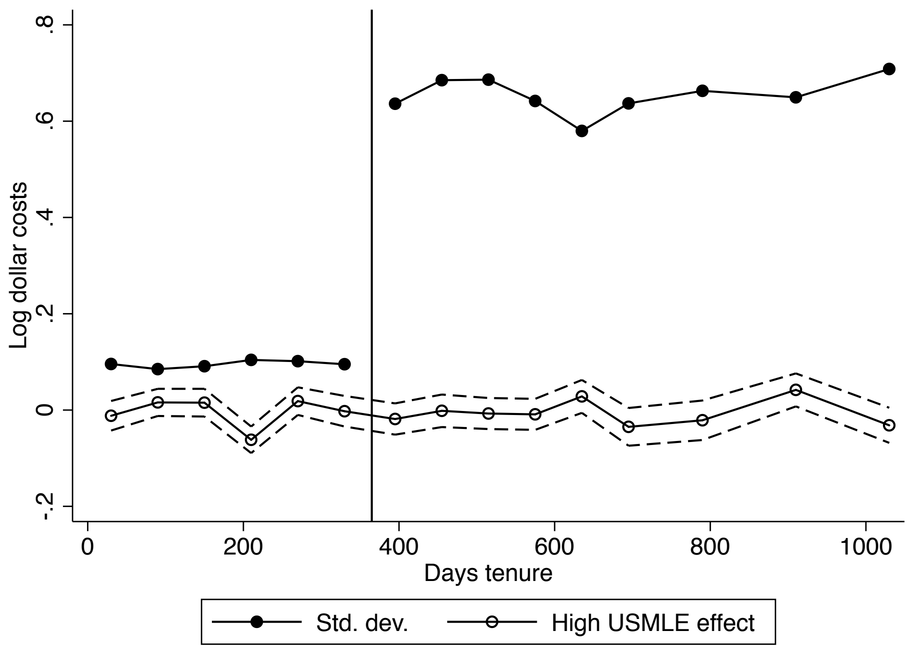
Note: This figure shows 10,000 random placebo tests for convergence in the specialist services. Merging cardiology and oncology yields an actual estimate of -0.143, or a 14.3 percentage point decrease per year in the standard deviation of spending effects of residents over the two years of the resident role, shown by the vertical line. In each of 10,000 placebo tests, I randomize combinations of housestaff-month-service to a placebo specialist service, matching the number of housestaff-month-services assigned to specialist services in each month of tenure. I estimate the same random effects model of log daily test costs shown in Equation (5) for the placebo specialist service and estimate the rate of placebo convergence using estimated housestaff effects in this placebo specialist service. Estimates for convergence are shown as a frequency histogram with a kernel-smoothed overlay.

Figure 4: Housestaff-effect Variation Profiles in Each Service by Guideline Existence



Note: Similar to Figures 1 and 2, this figure shows the standard deviation in a random effects model, as in Equation (5), of log daily test costs at each non-overlapping two-month tenure interval but for each service and for diagnoses with (solid dots) and without (hollow dots) published guidelines. Controls are the same as those listed in the caption for Figure 1. Intern and resident effects are modeled as random effects, as well as an admission-intern-resident interaction random effect in order to account for admission-level unobservables. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; vertical lines denote the one-year tenure mark.

Figure 5: Spending Prediction by High USMLE Scores over Tenure



Note: This figure shows in hollow dots the difference in test spending predicted by high USMLE Step 1 scores (taken during medical school) among housestaff over tenure intervals, as estimated by Equation (7). 95% confidence intervals are shown for the USMLE-score prediction as dashed lines. The profile of standard deviation of test-spending effects over housestaff tenure, estimated as random effects by Equation (5) and shown in Figure 1, is also shown for reference as solid dots. Controls are listed in the caption for Figure 1. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; a vertical line denotes the one-year tenure mark.

Table 1: Exogenous Assignment for Housestaff with Above or Below Average Spending

	Interns		Residents	
	Below-median test spending	Above-median test spending	Below-median test spending	Above-median test spending
Patient characteristics				
Age	62.11	62.13	62.07	62.15
	(16.90)	(16.86)	(16.82)	(16.93)
Male	0.484	0.482	0.489	0.478
	(0.500)	(0.500)	(0.500)	(0.500)
White race	0.706	0.703	0.708	0.702
	(0.455)	(0.457)	(0.455)	(0.457)
Black race	0.161	0.159	0.157	0.162
	(0.367)	(0.365)	(0.364)	(0.368)
Charlson comorbidity index	2.87	2.87	2.84	2.90
	(2.79)	(2.79)	(2.77)	(2.81)
Diagnostic-related Group (DRG) weight	1.25	1.25	1.27	1.24
	(0.86)	(0.84)	(0.85)	(0.84)
Supervising physicians				
Above-median-spending residents	0.500	0.500	N/A	N/A
	(0.501)	(0.501)		
Above-median-spending attendings	0.503	0.502	0.501	0.502
	(0.501)	(0.501)	(0.501)	(0.501)

Note: This table shows evidence of exogenous assignment for housestaff with below-median or above-median averaged spending effects. Averaged spending effects are estimated by a regression of log test spending on patient characteristics and physician (intern, resident, and attending) identities. Lower- and higher-spending interns are identified by their fixed effect relative to the median fixed effect. For each of these groups of interns, this table shows average patient characteristics and spending effects for supervising physicians. Averages are shown with standard deviations in parentheses.

Table 2: Ward Service Summary Statistics

	Actual services			Pseudo-services	
	CAR	ONC	MED	MED1	MED2
Mean admission characteristics					
Patient age	63.71	59.25	62.79	64.76	60.67
DRG weight	2.44	2.24	1.69	1.64	1.75
Test costs	$613.61	$855.38	$687.18	$634.70	$743.75
All costs	$9,703.80	$7,544.00	$5,303.48	$5,071.63	$5,553.42
Length of stay (days)	3.89	4.69	3.66	3.47	3.87
30-day readmission	0.089	0.218	0.090	0.089	0.091
30-day mortality	0.031	0.175	0.034	0.032	0.036
Observations					
Admission count	12,485	22,711	12,989	11,784	10,927
MDC count	23	24	23	3	21
ICD-9 count	440	1101	623	602	897
Concentration					
MDC HHI	0.740	0.117	0.103	0.347	0.101
ICD-9 HHI	0.055	0.019	0.025	0.038	0.013

Note: This table shows summary statistics for actual services – cardiology (CAR), oncology (ONC), and general medicine (MED) – and for "pseudo-services" formed based on Major Diagnostic Categories (MDC) from the general medicine service. The pseudo-service MED1 includes "Circulatory System" (MDC 5), "Respiratory System" (MDC 4), and "Digestive System" (MDC 6); MED2 includes all other MDCs. Summary statistics include mean admission characteristics (patient age, DRG weight) and outcomes (costs, length of stay, readmission, and mortality), counts (Numbers of admissions, MDCs, and ICD-9 codes), and Herfindahl-Hirschman Indices (HHI).

Table 3: Effect of Pre-training Characteristics and Other Hospital Training on Housestaff Spending

	Log daily test costs							
	Interns				Residents			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Characteristic of interest	Male	High USMLE	Highly ranked	Other hospital	Male	High USMLE	Highly ranked	Other hospital
Effect of characteristic	-0.021*	-0.001	0.011	0.002	-0.040**	-0.034*	0.001	-0.169*
	(0.012)	(0.012)	(0.018)	(0.025)	(0.018)	(0.020)	(0.031)	(0.095)
Number of observations	186,186	190,777	131,366	220,074	176,939	190,777	110,898	220,074
Adjusted R^2	0.222	0.165	0.166	0.165	0.251	0.175	0.176	0.178
Sample outcome mean	4.028	4.028	4.043	4.022	4.028	4.028	4.043	4.022
Sample outcome standard deviation	1.341	1.341	1.335	1.353	1.341	1.341	1.335	1.353
Sample characteristic mean	0.596	0.277	0.226	0.049	0.560	0.286	0.205	0.060
Sample characteristic standard deviation	0.491	0.447	0.418	0.215	0.496	0.452	0.404	0.238

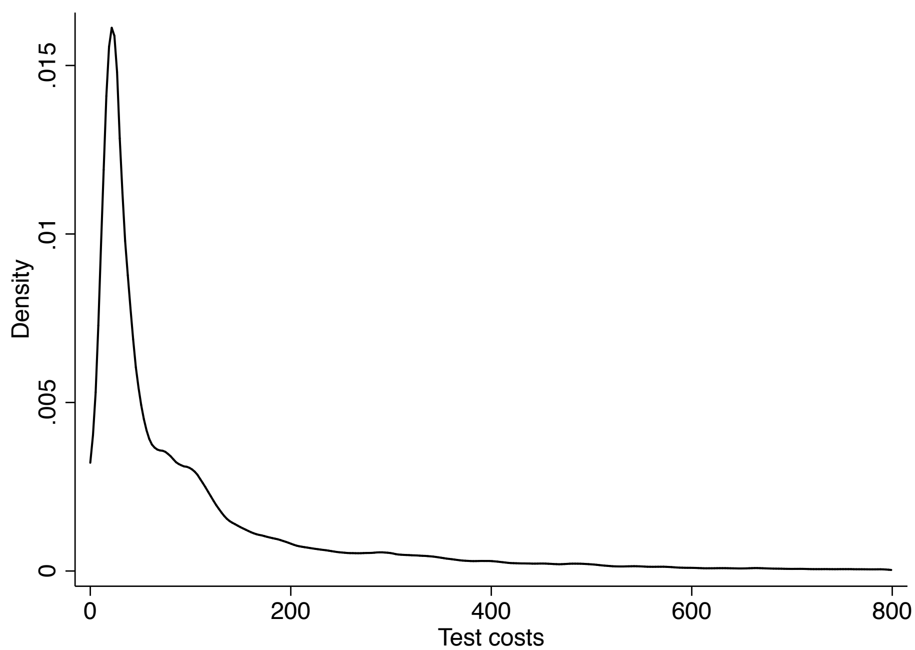
Note: This table reports results for regressions on the effect of the some pre-residency characteristics and of training at another academic hospital. All regressions are of the form in Equation (6), where the coefficient of interest is on a group of housestaff identified by either pre-residency characteristics or whether the housestaff is from the other academic hospital. The effect of many other characteristics of interest (or groups) were estimated and are described in the Appendix; they are all insignificant and omitted from this table for brevity. Models on the left panel show results for groups of interns; models on the right show results for groups of residents. All models control for patient and admission characteristics, time dummies, and fixed effects for attending and the other housestaff on the team (e.g., the resident is controlled for if the group is specific to the intern). Standard errors are clustered by admission. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Mean Effect of Resident Tenure and Variation across Residents

	(1) Daily log test spending	(2) Daily log total spending	(3) Log length of stay	(4) 30-day readmit	(5) 30-day mortality
Mean resident tenure effect regression					
Third-year resident	0.0057	0.0035	0.0072	0.0028	0.0002
	(0.0070)	(0.0042)	(0.0060)	(0.0038)	(0.0027)
Number of observations	219,727	219,727	48,175	47,874	48,175
Adjusted R^2	0.138	0.087	0.271	0.046	0.193
Counterfactual outcomes (none are log)					
Mean outcome	$123.75	$1,279.57	3.996	0.124	0.071
Third-year resident	$124.45	$1,284.07	4.024	0.127	0.071
	($0.87)	($5.35)	(0.024)	(0.004)	(0.003)
1 s.d. increase in resident variation	$210.81	$1,563.90	4.346	0.137	0.084
	($4.08)	($11.62)	(0.019)	(0.003)	(0.003)
1 s.d. decrease in resident variation	$72.64	$1,046.93	3.674	0.113	0.060
	($1.40)	($7.77)	(0.016)	(0.003)	(0.002)

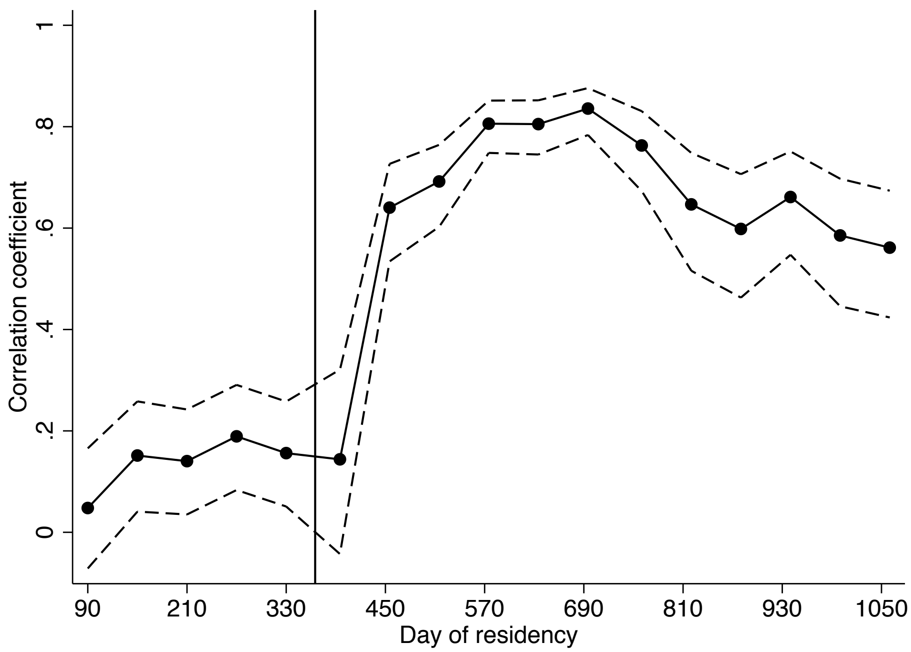
Note: In the top panel, this table reports results of regressions of various outcomes on having a third-year (as opposed to a second-year resident), as defined by Equation (10). Total spending includes imputed costs, such as physician and nurse salaries and operating costs. The third-year coefficient is insignificant in all of the models. In the bottom panel, mean (non-logged) outcomes are reported, a counterfactual for having a third-year resident (assuming that a second-year resident was previously responsible for the mean outcome), and counterfactuals for switching the resident for another one who has a spending effect one standard deviation higher or lower in the relevant outcomes. Distributional counterfactuals are generated by random-effect models. The random effect models are linear for daily log spending and log length of stay; they are logistic for readmissions and mortality. In the logistic models, I do not have attending fixed effects; the distribution of random effects may be overestimated in these models. Random-effect models are estimated for the entire sample, assuming constant spending effects within the two years in the role of resident. Therefore, variation in spending effects is less than in baseline Equation (5) which allows tenure-specific spending effects.

Figure A-1: Distribution of Daily Test Spending



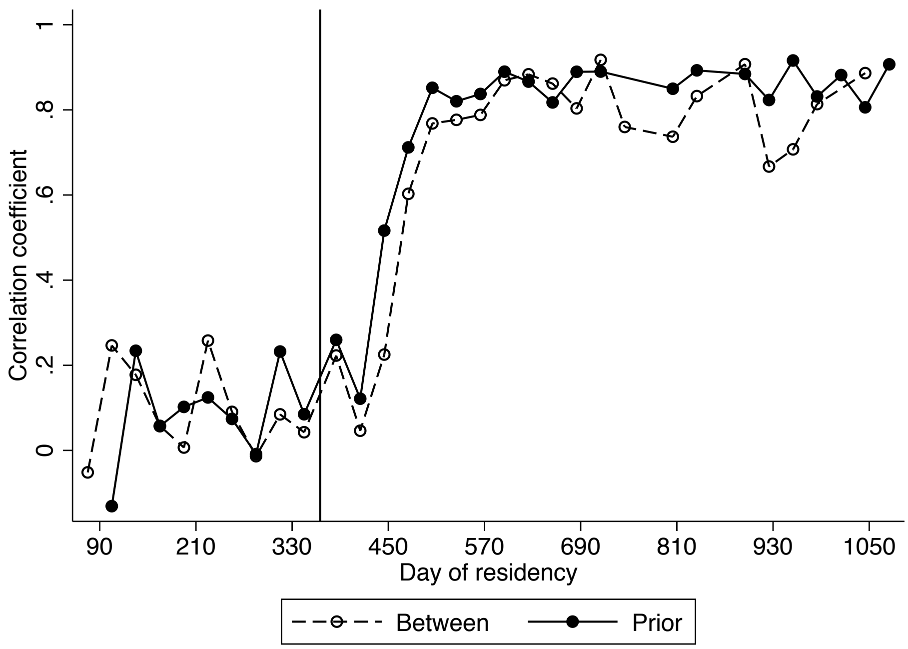
Note: This figure shows the density daily test costs. The distribution is shown up to $800 per day.

Figure A-2: Serial Correlation of Housestaff Random Effects over Tenure



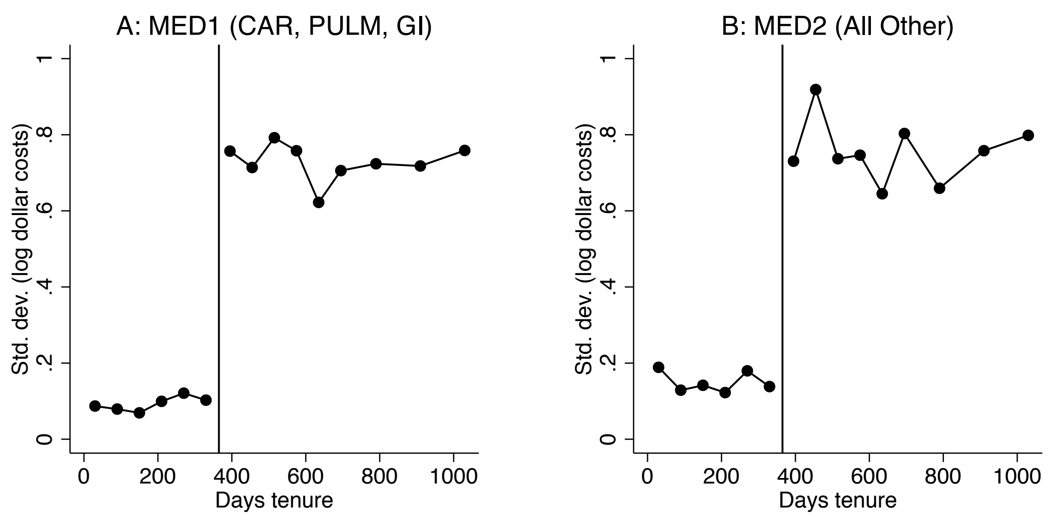
Note: This figure shows the serial correlation between the empirical Bayes random effect of each housestaff during a tenure interval and the random effect of that same housestaff during the previous tenure interval. Correlation coefficients are plotted as a solid line; 95% confidence intervals are plotted as dashed lines. The random effect model of log daily test costs is first estimated as in Equation (5), as described in the notes for Figure 1, using data within each two-month interval. Then the empirical Bayes random effects are calculated in the standard manner, described in Section 6.1, for each housestaff and tenure interval. Correlation coefficients are then calculated within housestaff and across adjacent tenure intervals. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; a vertical line denotes the one-year tenure mark.

Figure A-3: Serial Correlation of Housestaff Random Effects Conditional on a Nearby Interval with Few Patients



Note: This figure shows the serial correlation between empirical Bayes random effects in one-month intervals within housestaff, conditional on a nearby interval with few (40 or fewer, corresponding to the 20th percentile of monthly patient volume) or no patients. Random effects are calculated as described in Section 6.1, similar to Figures 1 and A-2, except using data within one-month rather than two-month intervals. Correlation coefficients are calculated for a housestaff-month random effect and the random effect of the same housestaff corresponding to the month two months prior, i.e., between $\tilde{\xi}_h^{\tau}$ and $\tilde{\xi}_h^{\tau-2}$ in Equation (5). Correlations for random effects with two types of conditions are calculated: those conditional on few or no observations during the tenure interval $\tau - 1$ ("Between," shown with hollow dots), and those conditional on few or no observations during the tenure interval $\tau - 3$ ("Prior," shown with solid dots). Correlations with fewer than 10 observations are omitted. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; a vertical line denotes the one-year tenure mark.

Figure A-4: Housestaff-effect Variation by Tenure in Pseudo-services



Note: This figure shows the standard deviation of test-spending effects over housestaff tenure in two pseudo-services formed from general medicine admissions. These pseudo-services are meant to create a difference in diagnostic concentration. MED1 includes the most common Major Diagnostic Codes (MDCs) of “Circulatory System” (MDC 5), “Respiratory System” (MDC 4), and “Digestive System” (MDC 6), roughly equivalent to cardiology, pulmonology, and gastroenterology; MED2 includes all other MDCs. Summary statistics for these two pseudo-services are given in Table 2. The random effects model is still Equation (5), estimated at non-overlapping two-month tenure intervals. 95% confidence intervals are omitted for simplicity. Controls are the same as those listed in the caption for Figure 1. Intern and resident effects are modeled as random effects, as well as an admission-intern-resident interaction random effect in order to account for admission-level unobservables. Housestaff prior to one year in tenure are interns and become residents after one year in tenure; a vertical line denotes the one-year tenure mark.

Table A-1: *New England Journal of Medicine* Research Articles by Specialty

Specialty / subspecialty	Internal medicine	Article count
Hematology / Oncology	Y	596
Cardiology	Y	562
Genetics	N	476
Infectious Disease	Y	453
Pulmonary / Critical Care	Y	329
Pediatrics	N	285
Endocrinology	Y	283
Gastroenterology	Y	257
Neurology / Neurosurgery	N	245
Surgery	N	228
Primary Care / Hospitalist	Y	179
Nephrology	Y	158

Note: This table reports the number of research papers appearing in the last ten years in the *New England Journal of Medicine*, by specialty or subspecialty as categorized by the journal. Specialties or subspecialties are also categorized as being within internal medicine or not. A training path in clinical genetics is possible from internal medicine, but genetics can also be pursued from pediatrics, obstetrics-gynecology, and other specialties. The *New England Journal of Medicine* has the highest impact factor, 51.7, out of all medical journals; only five other medical journals have double-digit impact factors, with the second-highest of 39.1 belonging to the *Lancet*, and the third-highest of 30.0 belonging to the *Journal of the American Medical Association*. Articles counted as research papers are "scientific reports of the results of original clinical research." Other categories, as defined at http://www.nejm.org/page/author-center/article-types, include reviews, clinical cases, perspective, commentary, and other.

Table A-2: Research Funding by National Institutes of Health (NIH) Institute or Center

NIH Institute or Center	Grants open	Funding (millions)
National Cancer Institute (NCI)	9,872	\$6,670
National Institute of Allergy and Infectious Diseases (NIAID)	7,271	\$5,433
National Heart, Lung, and Blood Institute (NHLBI)	6,294	\$3,591
National Institute of General Medical Sciences (NIGMS)	6,268	\$2,614
National Institute of Diabetes and Digestive And Kidney Diseases (NIDDK)	4,971	\$2,397
Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD)	3,295	\$1,814
National Institute of Neurological Disorders And Stroke (NINDS)	4,639	\$1,753
National Institute of Mental Health (NIMH)	3,650	\$1,500
National Institute on Drug Abuse (NIDA)	2,809	\$1,229
National Institute on Aging (NIA)	2,749	\$1,220
National Institute of Environmental Health Sciences (NIEHS)	1,504	\$1,091
Office of the Director (OD)	820	\$756
National Eye Institute (NEI)	1,798	\$733
National Human Genome Research Institute (NHGRI)	623	\$627
13 Other Institutes and Centers	8,564	\$4,259

Note: This table lists the top fourteen Institutes and Centers of the National Institutes of Health (NIH), ordered by current funding as defined by funds to currently open grants. Grants open and current funding (in millions of dollars) are both listed. For brevity, the thirteen other Institutes and Centers are not listed individually but are aggregated in the last line.

Table A-3: Core Rotations for Most Recognized Internal Medicine Residencies

	Ward rotations							
Residency program	MED	CAR	ONC	GI	PULM	RENAL	ID	RHEUM
Massachusetts General Hospital	✓							
Johns Hopkins University	✓	✓	✓					
Brigham and Women's Hospital	✓	✓	✓					
University of California, San Francisco	✓	✓		✓				
Mayo Clinic	✓	✓	✓	✓	✓	✓	✓	
Duke University Hospital	✓	✓	✓		✓			
Washington University	✓	✓	✓					
University of Pennsylvania	✓	✓	✓					
New York Presbyterian (Columbia)	✓	✓	✓				✓	
Northwestern University	✓	✓	✓	✓				
University of Michigan	✓	✓	✓	✓	✓			
University of Washington	✓	✓	✓					
University of Texas Southwestern	✓	✓						
Cleveland Clinic	✓	✓	✓	✓		✓		
Mount Sinai Hospital	✓							
Stanford University	✓	✓	✓					
Vanderbilt University	✓	✓	✓					
New York Presbyterian (Cornell)	✓	✓	✓					✓
University of Chicago	✓	✓	✓					
Emory University	✓	✓						
UCLA Medical Center	✓	✓	✓					
Beth Israel Deaconess Medical Center	✓	✓	✓					
Yale-New Haven Medical Center	✓	✓	✓	✓		✓	✓	
New York University	✓	✓	✓					
Total Counts (out of 24)	24	22	19	6	3	3	3	1

Note: This table shows core ward organ-based medical rotations for the 24 highly recognized internal medicine residency programs reported by *US News & World Report*, ordered by nominations in a survey of internists and residency program directors. The identities of core rotations were obtained by browsing each residency program's website. Abbreviations: general medicine (MED), cardiology (CAR), hematology/oncology (ONC), gastroenterology (including liver) (GI), pulmonary (PULM), nephrology (RENAL), infectious disease (ID), and rheumatology (RHEUM). I exclude rotations in palliative care and geriatrics, as these are not traditional organ-based subspecialties, and in neurology, as it is a specialty outside of internal medicine. Total counts are shown in the last row.

Table A-4: Core Rotations in Universe of Internal Medicine Residencies

Ward Rotations	Program count
General Medicine (MED)	310
Cardiology (CAR)	131
Hematology / Oncology (ONC)	85
Nephrology (RENAL)	34
Gastroenterology, including Hepatology (GI)	28
Pulmonology (PULM)	27
Infectious Disease (ID)	22
Rheumatology (RHEUM)	7
Endocrinology (ENDO)	3

Note: This table shows core ward medical rotations in the universe of internal medicine residency programs accredited by the American Council for Graduate Medical Education (ACGME), accessed at www.acgme.org. Of the 345 programs listed in the website, 310 programs had curricula detailing core ward rotations. Core ward rotations are defined as required rotations on ward services.

Table A-5: Standard Deviation of Housestaff Random Effects by Tenure

Tenure period	Random effect standard deviation			
(days in year)	Intern	Resident	Admission	Daily
1-60	0.159	0.697	0.398	1.174
	(0.015)	(0.041)	(0.010)	(0.005)
61-120	0.214	0.653	0.427	1.182
	(0.017)	(0.037)	(0.010)	(0.005)
121-180	0.294	0.550	0.450	1.184
	(0.026)	(0.038)	(0.010)	(0.005)
181-240	0.244	0.576	0.442	1.178
	(0.021)	(0.035)	(0.010)	(0.005)
241-300	0.311	0.494	0.432	1.159
	(0.035)	(0.044)	(0.010)	(0.005)
301-365	0.203	0.627	0.422	1.174
	(0.019)	(0.037)	(0.010)	(0.005)

Note: This table reports estimated standard deviations of random effect distributions. As stated in Equation (5), the model controls for patient and admission observable characteristics, time dummies (month-year interactions, day of the week), and attending identities (as fixed effects). Patient characteristics include demographics, Elixhauser indices, Charlson comorbidity scores, and DRG weights. Admission characteristics include the admitting service (e.g., "cardiology team 1"). Intern and resident effects are modeled as random effects, as well as an admission-intern-resident interaction random effect in order to account for admission-level unobservables. This model is estimated during 2-month time windows throughout the year.

Table A-6: Serial Correlation of Housestaff Random Effects over Tenure

Days during	Random effect correlation with previous period		
Training	One period prior	Two periods prior	Three periods prior
61-120	0.048 (-0.071, 0.166)		
121-180	0.151 (0.041, 0.258)	0.061 (-0.079, 0.199)	
181-240	0.140 (0.035, 0.242)	0.057 (-0.057, 0.170)	0.142 (-0.009, 0.286)
241-300	0.189 (0.083, 0.291)	0.203 (0.093, 0.308)	0.031 (-0.089, 0.151)
301-365	0.156 (0.051, 0.258)	0.071 (-0.042, 0.181)	0.147 (0.034, 0.256)
366-425	0.144 (-0.043, 0.321)	0.193 (0.006, 0.367)	0.124 (-0.066, 0.306)
426-485	0.640 (0.534, 0.727)	0.245 (0.076, 0.399)	0.179 (0.007, 0.341)
486-545	0.692 (0.603, 0.764)	0.395 (0.247, 0.525)	0.133 (-0.038, 0.297)
546-605	0.806 (0.748, 0.851)	0.763 (0.691, 0.820)	0.346 (0.195, 0.480)
606-665	0.805 (0.745, 0.852)	0.677 (0.583, 0.753)	0.552 (0.429, 0.656)
666-730	0.836 (0.784, 0.876)	0.782 (0.713, 0.836)	0.699 (0.605, 0.774)
731-790	0.763 (0.673, 0.831)	0.722 (0.617, 0.803)	0.753 (0.650, 0.829)
791-850	0.647 (0.516, 0.748)	0.601 (0.451, 0.718)	0.677 (0.541, 0.778)
851-910	0.598 (0.463, 0.706)	0.524 (0.367, 0.652)	0.438 (0.259, 0.589)
911-970	0.661 (0.547, 0.751)	0.430 (0.266, 0.569)	0.499 (0.342, 0.629)
971-1030	0.585 (0.446, 0.697)	0.616 (0.481, 0.723)	0.494 (0.332, 0.628)
1031-1095	0.561 (0.424, 0.674)	0.580 (0.446, 0.689)	0.538 (0.393, 0.656)

Note: This table reports correlation coefficients between random effects of the same housestaff across periods. As in Equation (5), the random effects model controls for patient and admission characteristics, time dummies (month-year interactions, day of the week), and attendings identities (as fixed effects). It includes random effects for housestaff and random shocks at the admission level. This model is estimated during each non-overlapping 2-month period. Next, empirical Bayes predictions for housestaff random effects are calculated for each housestaff observed in each time period. Finally, correlation coefficients are calculated between time periods within the same housestaff, including across the switch in roles from intern to resident. Correlation coefficients are displayed next to 95% confidence intervals in parentheses.