American Economic Journal: Economic Policy 2021, 13(1): 106–137
https://doi.org/10.1257/pol.20180501

Influence and Information in Team Decisions: Evidence from Medical Residency[†]

By David C. Chan*

I study team decisions among physician trainees. Exploiting a discontinuity in team roles across trainee tenure, I find evidence that teams alter decision-making, concentrating influence in the hands of senior trainees. I also demonstrate little convergence in variation of trainee effects despite intensive training. This general pattern of trainee effects on team decision-making exists in all types of decisions and settings that I examine. In analyses evaluating mechanisms behind this pattern, I find support for the idea that significant experiential learning occurs during training and that teams place more weight on judgments of senior trainees in order to aggregate information. (*JEL* D83, I11, J44, M53, M54)

Scholars have long conceptualized how teams may restructure decision-making. Teams may combine information across individuals in order to address more complex problems (Hayek 1945). Teams may also handle time-sensitive problems arriving at uncertain times by routing decision-making through an organizational structure (Marschak and Radner 1972). Despite the important implications of teams on economic activity, empirical analysis of how teams and organizations may impact decision-making remains scarce.

In health care, a large body of evidence documents wide variation across organizational boundaries in decisions driven by physicians (McCarthy and Blumenthal 2006, Newhouse et al. 2013). Without an understanding of how teams alter decision-making, the scale of practice variation across organizations seems difficult to reconcile with a similar magnitude of variation across individual providers (e.g., Skinner 2011, Van Parys and Skinner 2016). Large numbers of providers practicing independently at each institution should mute any systematic variation across

* Stanford University, 117 Encina Commons, Stanford, CA 94305, and NBER (email: david.c.chan@stanford.edu). Matthew Notowidigdo was coeditor for this article. I am grateful to David Cutler, Joe Doyle, Bob Gibbons, and Jon Gruber for early guidance on this project. I also thank Achyuta Adhvaryu, Daron Acemoglu, Leila Agha, David Autor, Daniel Barron, David Bates, Amitabh Chandra, Wes Cohen, Michael Dickstein, Amy Finkelstein, Matt Gentzkow, Mitch Hoffman, Peter Hull, Emir Kamenica, Pat Kline, Jon Kolstad, Eddie Lazear, Frank Levy, David Molitor, Maria Polyakova, Maya Rossin-Slater, Jon Skinner, Doug Staiger, Chris Stanton, Caio Waisman, Chris Walters, and many seminar audiences for helpful comments. Joel Katz and Amy Miller provided invaluable context to the data. Samuel Arenberg, Atul Gupta, Uyseok Lee, and Natalie Nguyen provided excellent research assistance. 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.

[†] Go to https://doi.org/10.1257/pol.20180501 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

organizations.[1] But if decision-making is concentrated in the hands of fewer providers, then they may drive surprisingly large variation across organizations. Policy implications would then depend on the organizational and informational frictions leading to such concentration.

I address this empirical question in the setting of medical residency, which is well suited for studying team decision-making and the roots of practice variation for several reasons. Each patient is assigned to a well-defined physician trainee team comprising a junior trainee in the first year of training and a senior trainee past the first year of training. Teams are reshuffled weekly so that each physician trainee works with many coworkers throughout training. A large number of patient cases are quasi-randomly assigned to trainees over the course of residency, and trainees take part in dozens of medical decisions per patient-day that are captured in the electronic medical record. Finally, as a separate point of interest, residency training provides a unique window into the evolution of practice variation among providers in health care. By design, residency is an intensive program to impart knowledge to physicians beyond facts, "developing habits, behaviors, attitudes, and values that will last a professional lifetime" (Ludmerer 2014, xi).

Specifically, I follow a group of 799 physician trainees in a large academic hospital and exploit detailed administrative data of physician trainees to teams caring for patients. Team decisions are measured over a five-year period as detailed orders for 3.4 million medications, 3.1 million laboratory tests, and 268,065 radiology tests. I aggregate dozens of physician orders by their costs to form spending summary statistics of team decisions for each of 220,117 patient-days in categories of laboratory testing, radiology testing, medication, blood transfusion, and nursing.

Using random assignment of patients to physician teams and frequent rotation of trainees across teams, I identify the causal trainee effects on team decisions measured by spending at various points in the trainees' tenure. Specifically, I employ a strategy similar to that used in a number of papers, starting with Abowd, Kramarz, and Margolis (1999), which have decomposed joint outcomes into contributions due to workers and firms (Card, Heining, and Kline 2013), workers and managers (Lazear, Shaw, and Stanton 2015), patients and geographic locations (Finkelstein, Gentzkow, and Williams 2016), and physicians and locations (Molitor 2017), among others. A key difference is that I estimate separate trainee effects at different points in their residency training, which is possible because of the frequency of the patient observations and the rotations across teams. As a central object of interest, I define tenure-specific *practice variation* as the standard deviation of the distribution of these trainee effects across trainees in a given tenure period. Given the finite number of observations per trainee in a tenure period, I develop an estimation approach based on random effects in a hierarchical model (Searle, Casella, and McCulloch 1992; Gelman and Hill 2007).

[1] For example, Molitor (2017) finds a lack of any systematic sorting of individual providers to locations by their practices. Perhaps more intriguingly, he also finds that, upon changing locations, providers immediately change their decision-making to match a local practice style, which suggests that physician decision-making is not independent of the local environment.

Next, I use the team structure in medical residency to decompose trainee effects into two components relevant for team decision-making: a trainee's *judgment* (what she would have decided on her own as a single agent) and her *influence* (the extent to which her judgment sways the team decision). I isolate the effect of influence by assessing trainee effects across a discontinuity at the one-year tenure mark: before one year, trainees have relatively less experience than their teammate, while they have relatively more experience than their teammate immediately after their first year. Teams may also induce greater influence due to roles and responsibilities within the team. If trainee judgments (and other individual characteristics) are plausibly continuous across the one-year mark, then a discontinuous change in practice variation across one year reflects the contribution of influence due to any of these team-induced mechanisms.

I find a significant and discontinuous increase in practice variation across the one-year mark of training. Junior trainees before this mark show variation in total spending effects with a standard deviation of 5 percent, while senior trainees beginning their second year show variation in total spending effects with a standard deviation of 24 percent. Subsequent practice variation remains large to the end of training, and there is remarkably little convergence in trainee effects. When I consider two-agent teams (i.e., one junior trainee and one senior trainee are responsible for each patient-day), the senior trainee is responsible for $0.24^2/(0.05^2 + 0.24^2) \approx 96$ percent of the variance in team-level decisions.[2]

The discontinuous change in practice variation at the one-year mark provides strong evidence that teams matter for decision-making. I consider how such a change in influence across roles might reflect three types of "team concerns." First, in classical team theory, teams may address an issue of bandwidth limits among agents by distributing problems to agents. Since senior trainees split their time working with two junior trainees, they have time to attend to fewer problems per patient and should on average have *less* influence on any given case, an idea known as "management by exception" (Marschak and Radner 1972, Garicano 2000). Second, teams may introduce principal-agent problems by their hierarchical nature, inducing herding around senior trainees' beliefs and effectively increasing the influence of senior trainees (Prendergast 1993). Third, teams may aggregate information (DeGroot 2004) by allowing agents to confer with each other before making joint decisions. If senior trainees have more knowledge than junior trainees, they should have more influence. Additionally, unlike other team concerns, information aggregation may blunt convergence in trainee effects, since influence continuously increases with knowledge (unlike discrete hierarchical roles and titles).

In the later half of the paper, I shed some empirical light on potential mechanisms that might explain the combination of (i) a discontinuous increase in practice variation at the one-year tenure mark and (ii) a lack of significant convergence in practice variation even as trainees near completion of residency. First, senior trainees may have greater influence independent of their knowledge, for example, due to institutionalized differences in their job duties or by herding around their beliefs.

[2] Other members of the care team outside of trainees include supervising physicians, nurses, pharmacists, and specialty consultants. I focus on trainee teams because they are the most clearly quasi-randomly assigned.

Relatedly, they may simply hold decision rights over important decisions. Second, influence may reflect systematic differences in knowledge across tenure, acquired by learning during residency. Under the first mechanism, asymmetric influence arises from frictions in the *application* of knowledge. Under the second mechanism, asymmetric influence may represent an efficient application of greater knowledge held by senior trainees; in contrast, this greater knowledge arises from frictions in the *acquisition* of knowledge.

While these mechanisms may coexist, I consider two scenarios representing extreme versions of the two mechanisms. Isolating the first mechanism, I consider the possibility of fixed trainee judgments (i.e., knowledge) across residency. In this scenario, physicians possess all their knowledge from the beginning, and no learning occurs in residency. The increase in influence is thus due only to titles and decision rights unrelated to knowledge. Contradicting the decision-rights hypothesis—based on a team-theoretic Garicano (2000) model that routes decisions to different team members, with more important decisions going to senior trainees—I find that senior trainees have much greater influence over *all* types of decisions, both great and small, and particularly over diagnostic decisions that may be more uncertain (though not particularly expensive). I further rule out the extreme hypothesis of no-learning by showing that trainee practice styles vary over time. Detailed and seemingly important time-invariant trainee characteristics predict only a small portion of practice variation. In addition, the serial correlation between trainee judgments grows weaker over time. In contrast to prior literature that seemingly suggests relatively stable practice styles, this evidence suggests strong learning in the sense that physician practice styles are highly mutable, at least during physicians' early careers.[3]

To isolate the second mechanism, I assume an alternatively extreme scenario in which influence is optimally allocated according to knowledge, as specified by a simple structural model of Bayesian information aggregation in decision-making, but that knowledge must accrue with training. In this model, as trainees learn, their increasing influence on team decisions may dampen or reverse any convergence in their practice styles in teams. This stands in contrast with independent decision-making, in which increasing knowledge will necessarily lead to convergence. Results from this structural model imply substantial learning in the first year of training, relative to any knowledge acquired before residency. Interestingly, the results also suggest much greater learning when trainees become senior and have a larger stake in decision-making, which is consistent with large literatures on *experiential learning*, positing that learning requires active participation and experience.[4]

[3] For example, Epstein and Nicholson (2009) examine practice styles of obstetricians and project changes of other obstetricians practicing in the same hospital on the practice style of each index obstetrician. Molitor (2017) examines practice styles of cardiologist movers and similarly projects changes in the local practice style induced by the move onto the average practice style of moving cardiologists. In both studies, these projections are remarkably stable over time, but they may mask significant evolution of practice styles unrelated to these projections. Doyle, Ewer, and Wagner (2010) study physician trainees from two different residency programs and find systematic differences between trainees of the two programs. However, they abstract from any variation within program or changes within trainee.

[4] Notable contributions in this area include Dewey's (1938) thoughts on progressive education in *Experience and Education*, Montessori's (1948) method of teaching children, Piaget's (1971) constructivist theory of knowing, and Kolb and Fry's (1975) experiential learning. Similar concepts also include problem-based learning (e.g., Wood 2003) and "learning by teaching" (Gartner, Kohler, and Riessman 1971).

Between trainees, I find that deviations from optimal influence are small. However, I also find that, relative to their supervisors, the trainee team has much more influence than is justified by the trainees' knowledge.

This paper contributes to several literatures. First, it contributes to a general literature on decision-making in organizations (e.g., Marschak and Radner 1972, Van Zandt 1998, Garicano 2000). As noted by Hayek (1945, 519),

> The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess.

Despite seminal theoretical contributions in this literature, empirical evidence remains scarce. The evidence in this paper highlights the important team function of aggregating information for a given decision, because the optimal decision may not be known perfectly by any single agent. This stands in contrast to canonical team-theoretic models, notably Garicano (2000), that view the function of organizations as routing problems with known solutions.

Second, this paper sheds new empirical light on the nature of learning, as defined by forming judgments to make decisions. In economics, a large empirical literature on "learning on the job" (Mincer 1962) has mostly relied on wages as a marker of learning, while another empirical literature studying "learning by doing" (Arrow 1962) has measured task performance (e.g., speed or accuracy) attributable to agents or firms gaining experience.[5] Neither approach seems appropriate in this setting: physician trainees are paid fixed salaries, and complicated and potentially high-stakes decisions are made in teams with layers of experience.[6] This paper makes some progress on this problem by developing a notion of learning with empirical implications for team decisions.

Third, as noted above, these results relate to a large literature documenting practice variation in health care (Fisher et al. 2003a, 2003b).[7] Academic and policy discussions on this topic often refer to features of the health care marketplace that insulate providers from competition, but this reasoning assumes that, absent incentives, providers mostly agree on the diagnosis and treatment for any given

[5] Examples in the empirical literature of on-the-job training that focuses on wages include Topel (1991) and Kahn and Lange (2014). Examples in the empirical literature on learning by doing include Benkard (2000); Levitt, List, and Syverson (2013); and Hendel and Spiegel (2014).

[6] Indeed, team-based decision-making may be a key reason why teaching hospitals show no significant decline in patient outcomes in July, when a sudden (but scheduled) influx of fresh physician trainees arrives (Young et al. 2011, Song and Huckman 2018).

[7] In addition to the literature reviewed by Skinner (2011), recent contributions in the economics literature include Doyle et al. (2015); Cooper et al. (2015); Chandra et al. (2016); Finkelstein, Gentzkow, and Williams (2016); and Molitor (2017). Much of this literature focuses on differences among regions or hospitals. See Epstein and Nicholson (2009) as an example of physician-level variation that has generally been difficult to explain. Similar informational frictions can underlie differences across organizations (e.g., Bloom and Van Reenen 2010). Particularly relevant to the setting of residency training is work by Doyle, Ewer, and Wagner (2010) comparing mean practices between two groups of trainees from different programs randomly assigned patients in the same hospital.

patient (Cutler 2010, Skinner 2011). This view is incompatible with survey evidence revealing that experts often widely disagree (Cutler et al. 2019). It is also inconsistent with growing evidence of simultaneous errors of commission and omission among providers in both diagnostic and treatment decisions (Abaluck et al. 2016; Chan, Gentzkow, and Yu 2019). This paper highlights informational mechanisms behind wide practice variation in an intense and highly selective training environment *designed* to create homogeneity. Interestingly, team decision-making and knowledge frictions may concentrate influence behind practice variation into the hands of fewer providers who nonetheless disagree with each other. If so, appropriate policy responses to practice variation should focus on organizational and informational levers.

The organization of this paper is as follows. Section I describes the institutional setting and data. Section II introduces the empirical approach. Section III presents main results and discusses how they relate to team concerns. Section IV investigates mechanisms in greater detail. Section V discusses policy implications for practice variation and concludes.

I. Setting and Data

A. *The Structure of Residency*

I study trainees associated with the internal medicine residency program of a large teaching hospital. The program is highly selective, and the hospital is a source of numerous clinical trials and guidelines. As is standard across internal medicine programs, training takes place over three years in teams organized by experience: each patient is cared for by a first-year junior trainee ("intern") and a second- or third-year senior trainee ("resident").

While each patient is assigned to a team of one intern and one resident, residents split their time between two interns. Thus, interns are assigned half the number of patients as residents. This allows interns to devote more attention to each patient, and they are usually the first to examine a patient and make judgments. Each senior trainee (along with the two junior trainees working with her) is supervised by an "attending" physician who has completed residency. Teams also operate within a broad practice environment that influences decision-making, including institutional rules; information systems; and other health care workers such as consulting physicians, pharmacists, and nurses. Trainees on the same teams may come from different predetermined career tracks, other programs (e.g., obstetrics-gynecology, emergency medicine), or another hospital. A sizable minority of interns plan only to spend one year in the internal medicine residency ("preliminary" versus "categorical" interns), subsequently proceeding to another residency program such as anesthesiology, radiology, or dermatology.

This study focuses on inpatient ward rotations, which comprise cardiology, oncology, and general medicine services. According to interviews with residency administration, trainee rotation preferences are not collected, and assignment does not consider trainee characteristics. Scheduling is done a year in advance and does not consider matches among intern, resident, and attending physicians that will be formed as a result. Supervising physician schedules are created independently, with neither trainee nor supervising physician aware of one another's schedule in advance. Therefore, trainees and supervising physicians are as good as randomly assigned to each other.

Patients admitted to ward services are assigned to interns and residents by a simple algorithm that distributes patients in a rotation among on-call trainees.[8] Patients who remain admitted for more than one day may be mechanically transferred to other trainees as they change rotations. When one trainee replaces another, she assumes the entire patient list of the previous trainee. Because trainee blocks are generally two weeks long and staggered for interns and residents, patients frequently experience a change in either the intern or the resident on the team.

B. *Team Decisions*

As in other small-team settings, formal decision rights are rarely invoked in patient care teams. While senior teammates may influence decisions by their general knowledge or status, junior teammates may acquire more patient-specific knowledge and are usually charged with implementing decisions. A variety of protocols and customs common in residency encourage trainees to function independently and take responsibility for clinical decisions. For example, junior trainees are listed as the first point of contact so that information from patients, nurses, and consultants generally flows through junior trainees before reaching senior trainees or supervising physicians. Similarly, junior trainees are expected to write orders and discuss care plans with patients and other staff so that they are abreast of all decisions made for their patients. While trainees may consult with supervising physicians in real time, they often make and communicate decisions without prior consultation. As a practice, supervising physicians will often delay discussing new patients or new developments until after trainees have evaluated the patient and formulated a treatment plan. Thus, supervisors will often learn about decisions after they are made.

C. *Data*

I collect data from several sources.[9] First, I observe the identities of each physician on the clinical team—intern, resident, and attending physician—for each patient on a ward service on each day that the patient is in the hospital. Over five years, I observe data for 46,091 admissions, equivalent to 220,074 patient-day observations. Corresponding to these admissions are 799 unique trainees and 531 unique attendings. Of the trainees, 516 are from the same internal medicine residency, with the

[8] Depending on the reason for admission, patients may be matched to categories of attending physicians according to the admitting service. Trainees who have reached their capacity may also be taken out of the algorithm for accepting new patients during the remainder of a call day. Conditional on these constraints, patient types are not matched to trainees.

[9] Readers may find more information in the American Economic Association replication repository (Chan 2021).

Table 1—Quasi-Random Assignment for Trainees with Above- or Below-Average Spending

	Interns		Residents	
	Below-median spending	Above-median spending	Below-median spending	Above-median spending
Patient characteristics				
Age	62.04	62.14	62.03	62.14
	(16.91)	(16.85)	(16.92)	(16.83)
Male	0.483	0.482	0.484	0.482
	(0.500)	(0.500)	(0.500)	(0.500)
White race	0.707	0.705	0.703	0.709
	(0.455)	(0.456)	(0.457)	(0.454)
Black race	0.161	0.156	0.156	0.161
	(0.367)	(0.363)	(0.363)	(0.368)
Predicted log total costs	8.477	8.478	8.498	8.477
	(0.142)	(0.139)	(0.140)	(0.140)
Physician teammates				
Above-median-spending residents	0.504	0.495	N/A	N/A
	(0.500)	(0.500)		
Above-median-spending attendings	0.486	0.509	0.484	0.510
	(0.500)	(0.500)	(0.500)	(0.500)

Notes: This table shows evidence of quasi-random assignment for trainees with below-median or above-median average spending effects. Trainee spending effects, not conditioning by tenure, are estimated as fixed effects by a regression of log daily 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, in a regression of admission spending that controls for patient characteristics (race, age, and gender), admission service dummies, and month-year interaction dummies. 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.

remainder visiting from another residency program within the same hospital or from another hospital.[10] There is no unplanned attrition across years of residency.[11]

I collect detailed information for each trainee, including demographics, medical school, US Medical Licensing Examination (USMLE) Step 1 test scores, membership in the Alpha Omega Alpha (AOA) medical honor society, other degrees, and position on the residency rank list. Summary statistics of trainees characteristics are given in online Appendix Tables A-2 and A-3 and are consistent with an elite group of trainees.[12] I also observe precommitted residency tracks for each trainee physician. In addition to trainee characteristics determined prior to residency, I observe each trainee's realized specialty after her training to impute expected yearly future income in the five years immediately following this training based on industry-standard survey data from the Medical Group Management Association. The average above- and below-median future incomes for junior trainees are $424,000

[10] Of the 799 unique trainees, 649 are observed as interns and 407 are observed as residents. Of the 516 trainees from the same-hospital internal medicine residency, 401 are observed as interns and 338 are observed as residents.

[11] In two cases, interns with hardship or illness in the family were allowed to redo their intern year.

[12] For example, trainees in the data are almost three times more likely to be AOA inductees than the national average, a trait that predicts a 6–10 greater odds of matching to a first-choice residency program (Rinard and Mahabir 2010). The mean USMLE Step 1 score is 244, or approximately the seventy-sixth percentile of the national distribution.

Table 2—Summary Statistics of Spending in Categories and Services

	log daily total costs				
	Radiology (1)	Laboratory (2)	Medication (3)	Transfusion (4)	Nursing (5)
Cardiology					
Fifth percentile	0	11	4	0	189
Tenth percentile	0	16	14	0	244
Median	0	34	67	16	658
Mean	54	51	113	33	662
Ninetieth percentile	125	103	233	56	1,075
Ninety-fifth percentile	375	145	417	87	1,212
Oncology					
Fifth percentile	0	3	0	0	192
Tenth percentile	0	13	13	0	256
Median	0	34	94	12	673
Mean	66	58	155	78	682
Ninetieth percentile	248	124	350	204	1,033
Ninety-fifth percentile	423	212	542	411	1,270
General medicine					
Fifth percentile	0	8	2	0	160
Tenth percentile	0	12	10	0	205
Median	0	35	69	14	561
Mean	66	62	99	38	577
Ninetieth percentile	234	139	210	48	959
Ninety-fifth percentile	385	222	286	95	1,130

Notes: This table reports summary statistics of patient-daily spending in categories across columns and in ward services of cardiology, oncology, and general medicine. The statistics are calculated based on 56,780, 66,662, and 96,632 patient-day observations from the cardiology, oncology, and general medicine services, respectively.

and \$269,000, respectively; the respective numbers for senior trainees are \$409,000 and \$249,000.[13]

I use scheduling data and past matches between trainees and with supervising attending physicians. Consistent with Section IA, Table 1 shows that interns and residents with high or low spending effects are exposed to similar types of patients and are equally likely to be assigned to high- or low-spending coworkers and attendings. Online Appendix A-1 presents more formal analyses on conditional random assignment of trainee physicians, including F-tests showing joint insignificance.

Patient demographic information includes age, sex, race, and language. Clinical 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 Charlson comorbidity indices and 29 Elixhauser comorbidity dummies (Charlson et al. 1987, Elixhauser et al. 1998). I also observe the identity of the admitting service (e.g., "Heart Failure Team 1"), which categorizes patients admitted for similar reasons. Patients are *not* randomly assigned to supervising physicians, since supervising physicians within the same service may belong to different practice groups (e.g., HMO, private practice, hospitalist) that I do not explicitly capture and condition on.

[13] The difference in future incomes between junior and senior trainees reflects that the career paths for preliminary interns (e.g., future anesthesiologists, dermatologists, and radiologists) are often more lucrative.

I observe cost information for each patient-day aggregated within 30 cost departments used by the hospital for accounting purposes. I further group these departments into four categories: diagnostic (laboratory and radiology) testing, medication, blood bank, and nursing. Because costs are based on the hospital's accounting of resource utilization due to physician *actions*, not the measures of Medicare reimbursement used in recent studies (Doyle et al. 2015, Skinner and Staiger 2015, Chandra et al. 2016), they provide more direct insight into welfare-relevant resource use.[14] Consistent with prior literature on practice variation, I consider spending as a summary statistic of the many actions involved in patient care. Laboratory costs are based on 3.1 million physician laboratory orders; radiology costs are based on 268,065 tests ordered in CT, MRI, nuclear medicine, and ultrasound; and medication costs are based on 3.4 million medication orders. Table 2 shows distributional statistics of daily spending in each category and in the services of cardiology, oncology, and general medicine.

II. Analysis of Team Decisions

A. *Potential Decisions*

I observe a large set of decisions and the identities of agents on the team responsible for each decision. However, I do not observe an agent's contribution to the team decision, which is a key object of interest. The goal of the empirical approach is thus to decompose a team decision into such contributions made by each agent on the decision and to allow this decomposition to depend on circumstances that may shed light on organizational considerations in team decision-making.

To characterize decision-making on a tractable and continuous scale, I reduce the dimensionality of decisions by aggregating the direct costs of the decisions, observed via the hospital's accounting system, for a given patient-day.[15] Thus, a patient-day, or the combination (i,t) for patient admission i and day t, constitutes a "case" for which a team decision is observed. I denote potential team decisions for patient-day (i,t) assigned to a two-agent team composed of agents $j \in \mathcal{J}_{it}$ and $k \in \mathcal{K}_{it}$ as $Y_{it}(j,k)$. The realized decision is

$$Y_{it} = \sum_{j \in \mathcal{J}_{it}} \sum_{k \in \mathcal{K}_{it}} D_{ijt} D_{ikt} Y_{it}(j,k), \tag{1}$$

where $D_{ijt} \in \{0,1\}$ and $D_{ikt} \in \{0,1\}$ are indicator variables for assignment. Equivalently, since each case is assigned to one pair (j,k), define an assignment

[14] In this prior research, a difficulty in connecting practice variation in health care to the productivity literature is that "spending" input measures are actually government-set reimbursement rates that reflect hospital *revenues* rather than input costs. In large part, the Medicare reimburses inpatient care prospectively based on *diagnoses* rather than social cost of actual utilization.

[15] In principle, given these microdata, I could also study variation at the order level. However, the set of potential orders is large, and many orders are very specific to certain clinical scenarios that may not be observed frequently. Restricting study to certain types of clinical decisions, such as C-sections versus vaginal deliveries (e.g., Currie and Gruber 1996) or interventional treatment of heart attacks versus medical management (e.g., Chandra and Staiger 2007), is an approach used by many influential studies in the literature but does not capture the breadth or complementarity of physician decisions made on a daily basis. Section IVA provides some interesting evidence of such complementarity.

function $j(i,t)$ and $k(i,t)$ such that $D_{ijt} = \mathbf{1}(j = j(i,t))$ and $D_{ikt} = \mathbf{1}(k = k(i,t))$. In this setting, $\mathcal{J}_{it}$ and $\mathcal{K}_{it}$ are disjoint sets for any (i,t) since j is a junior trainee and k is a senior trainee.

B. *Trainee Effects*

Given the potential outcome notation in equation (1), I define trainee effects on team decision-making. For example, the effect of assignment to trainee j instead of j', holding k fixed, is $Y_{it}(j,k) - Y_{it}(j',k)$. Similarly, the effect of assignment to trainee k instead of k', holding j fixed, is $Y_{it}(j,k) - Y_{it}(j,k')$. Because I only observe $Y_{it} = Y_{it}(j(i,t),k(i,t))$, effects for a particular case (i,t) are unobservable.

My goal is to recover expectations of trainee effects by making use of quasi-random assignment of cases to trainees and of trainees to each other, as described in Section I. Specifically, I consider the following conditional independence assumption:

ASSUMPTION 1 (Quasi-Random Team Assignment): *Potential team decisions are independent of team assignments, conditional on clinical service $s(i,t)$ and indicators of time $\mathbf{T}_t$ (e.g., day of the week, month-year combinations)*:

$$\left\{Y_{it}(j,k)\right\}_{(j,k)\in\mathcal{J}_{it}\times\mathcal{K}_{it}} \perp\!\!\!\perp (D_{ijt},D_{ikt})\,\middle|\, s(i,t),\mathbf{T}_t.$$

If case potential outcomes are conditionally independent of team assignments, then trainee treatment effects are also conditionally independent of team assignments. Online Appendix I.A presents evidence of quasi-random assignment of patients to trainees, and online Appendix I.B presents evidence of quasi-random assignment of trainees to each other.

In the main analysis, I wish to capture a trainee's average treatment effect on team decisions, depending on her tenure and the tenure of teammates she could be working with. The timing of residency implies a mechanical relationship between the tenures of the junior and senior trainees. Since trainees all begin residency at the same time of the year, a junior trainee with tenure τ_j will work with senior trainees with one or two more years of tenure, or $\tau_k \in \{\tau_j+1,\tau_j+2\}$. I consider a population of cases defined by a feasible combination of junior and senior trainee tenure periods, or $\mathcal{C} = \{(i,t): \tau(j(i,t),t) = \tau_j, \tau(k(i,t),t) = \tau_k\}$, where $\tau(h,t)$ is a function that maps trainee h at time t to a tenure period.

I then define

$$\mathrm{ATE}(j|\mathcal{C}) \equiv E_{(i,t)\in\mathcal{C}}\left[E_{k\in\mathcal{K}_{it}}\left[Y_{it}(j,k)\right] - E_{(j,k)\in\mathcal{J}_{it}\times\mathcal{K}_{it}}\left[Y_{it}(j,k)\right]\right];$$

$$\mathrm{ATE}(k|\mathcal{C}) \equiv E_{(i,t)\in\mathcal{C}}\left[E_{j\in\mathcal{J}_{it}}\left[Y_{it}(j,k)\right] - E_{(j,k)\in\mathcal{J}_{it}\times\mathcal{K}_{it}}\left[Y_{it}(j,k)\right]\right].$$

Here, $\mathrm{ATE}(j|\mathcal{C})$ is junior trainee j's average effect on team decisions, working with an "average" senior trainee, relative to "average" counterfactual teams of junior and senior trainees; $\mathrm{ATE}(k|\mathcal{C})$ considers a similar object for senior trainee k. In both cases, the "average" teammate and the "average" team are defined by the set of cases, $\mathcal{C}$, that specifies the tenures of the junior and senior trainees.

In these definitions, I exploit the fact that trainee assignment is independent of potential team decisions; this implies that expectations of $Y_{it}(j,k)$, holding j or k fixed, are the same regardless of whether we condition on actual assignment to j or k. Note that $\text{ATE}(j|\mathcal{C})$ and $\text{ATE}(k|\mathcal{C})$ depend not only on the indentity of the trainee j or k, but also on the set of potential teammates implied by $\mathcal{C}$. The same trainee may have different effects on team decisions in different environments, particularly depending on whether they are more or less senior to their teammate.

Assumption 1 implies that I can recover consistent estimates of $\text{ATE}(j|\mathcal{C})$ and $\text{ATE}(k|\mathcal{C})$ by the following regression, performed over a sample of observations (i,t) drawn from the population set $\mathcal{C}$:

$$Y_{it} = \xi^{\mathcal{C}}_{j(i,t)} + \xi^{\mathcal{C}}_{k(i,t)} + \gamma_{s(i,t)} + \mathbf{T}_t\eta + \varepsilon_{it}, \tag{2}$$

where $\xi^{\mathcal{C}}_j$ and $\xi^{\mathcal{C}}_k$ are trainee effects for the junior and senior trainees, $\gamma_{s(i,t)}$ is a fixed effect for the clinical service $s(i,t)$, and ε_{it} is an error term. By construction, $E\left[\varepsilon_{it}\middle| s(i,t),\mathbf{T}_t,\mathbf{D}_{it}\right] = 0$, where $\mathbf{D}_{it}$ is a design vector indicating junior trainee and senior trainee identities. Under Assumption 1, we also have $E\left[\varepsilon_{it}\middle| s(i,t),\mathbf{T}_t,\mathbf{D}_{it}\right] = E\left[\varepsilon_{it}\middle| s(i,t),\mathbf{T}_t\right] = 0$. Thus, regression estimates of $\xi^{\mathcal{C}}_j$ and $\xi^{\mathcal{C}}_k$ are consistent estimators of $\text{ATE}(j|\mathcal{C})$ and $\text{ATE}(k|\mathcal{C})$, respectively.

Because $\mathcal{C}$ is defined by trainee periods τ_j and τ_k for the junior and senior trainees, respectively, I rewrite these effects as $\xi^{\tau_j;\tau_k}_j$ and $\xi^{\tau_k;\tau_j}_k$ to be more explicit about the tenure dependence of the estimated trainee effects. In practice, I perform the following regression with additional controls:

$$Y_{it} = \mathbf{X}_{it}\beta + \xi^{\tau_j;\tau_k}_j + \xi^{\tau_k;\tau_j}_k + \mathbf{T}_t\eta + \gamma_{s(i,t)} + \zeta_{\ell(i,t)} + \varepsilon_{it}. \tag{3}$$

The objects of interest in equation (3) are tenure-specific trainee effects—$\xi^{\tau_j;\tau_k}_{j(\cdot)}$ and $\xi^{\tau_k;\tau_j}_{k(\cdot)}$ for the junior and senior trainees, respectively—that depend on the identity of the trainee and the tenure periods of both trainees. To improve efficiency, I also include fixed effects for the supervising physician, $\ell(i,t)$, and a rich set of patient and admission characteristics, $\mathbf{X}_{it}$.[16] These controls are unnecessary for identification under Assumption 1, and in Section IIIB, I show robustness of results to including none of these controls.

C. *Random Effects versus Fixed Effects Estimation*

As described in Abowd, Kramarz, and Woodcock (2008), two approaches to estimating equation (3) are to treat the trainee effects of interest as "fixed" or as "random." In this subsection, I adopt the random effects approach for three reasons. First, I am interested in measures of practice variation, which are moments of a *distribution* of trainee effects, specifically the standard deviation of trainee effects

[16] Specifically, I control for patient-race dummies, male gender, linear and quadratic age, the Charlson comorbidity score (Charlson et al. 1987), 29 Elixhauser comorbidity dummies (Elixhauser et al. 1998), Diagnostic Related Group (DRG) weights, and day-of-the-patient's-length-of-stay dummies.

across trainees in a given tenure period. Random effect estimation directly focuses on this measure, while fixed effect estimation focuses on individual trainee effects.

Second, relatedly, I observe a finite number of observations for each trainee and in each tenure period. Importantly, this number of observations may vary for different trainees and in different tenure periods. Random effects estimation directly accounts for this by estimating a "prior distribution" of trainee effects by maximum likelihood. Empirical Bayes posteriors may then be obtained for each tenure-specific trainee effect, using the estimated prior and the data for each trainee and in each tenure period. This procedure will "shrink" information from the data toward the prior mean in a way that minimizes prediction errors of trainee effects (Morris 1983; Searle, Casella, and McCulloch 1992). In contrast, OLS estimates of a given trainee (fixed) effect make no use of information on other trainee effects, and naive (unshrunken) fixed effect estimates of trainee effects will overstate practice variation relative to the truth.[17]

Finally, under Assumption 1, the random effects approach is free of any notion of "connected sets" that is required under the fixed effects setup of Abowd, Kramarz, and Margolis (1999). In fixed effects estimation, one junior trainee effect and one senior trainee effect must be dropped within each connected set in order to satisfy a rank condition. Trainees belonging to different connected sets thus cannot be compared. In finite samples, when I consider trainee effects that are tenure specific, connections between trainees will become increasingly sparse.[18] In online Appendix A-2, I compare Assumption 1 with a related fixed effects assumption in Abowd, Kramarz, and Margolis (1999).

D. *Baseline Implementation*

In the random effects approach, I estimate by maximum likelihood underlying *population* moments of trainee effects that would be consistent with the observed data. In the baseline estimation, I focus on the standard deviation of trainee effects, conditional on the trainee's tenure τ and on the teammates tenure τ^{-}, or $\sigma(\tau;\tau^{-})$. I specify discrete tenure periods of 60 days for trainees in their first or second year of residency or periods of 120 days for trainees in their third year of residency, since training in the third year involves fewer days spent on clinical activities.[19]

To improve the robustness of the maximum likelihood estimation, I first form a risk-adjusted measure of log spending, $\tilde{Y}_{it} = Y_{it} - \left(\mathbf{X}_{it}\hat{\beta} + \mathbf{T}_t\hat{\eta} + \hat{\gamma}_{s(i,t)} + \hat{\zeta}_{\ell(i,t)}\right)$,

[17] For example, consider the simple model $Y_i = \xi_{j(i)} + \varepsilon_i$, where $\xi_j \sim N(0,1)$, $\varepsilon_i \sim N(0,1)$, and $E\left[\varepsilon_i \middle| \xi_{j(i)}\right] = 0$. Consider n observations for each agent. The estimated fixed effect for j will be $\hat{\xi}_j = n^{-1}\sum_{i:j(i)=j} Y_i$, which will be measured with error. The standard deviation of estimated agent fixed effects will be $\sqrt{1+1/n}$, which is an overestimate of the true practice variation of 1. Similarly, the difference between the fixed effect for any two agents is on average $E_{j,j'}\left[\hat{\xi}_j - \hat{\xi}_{j'}\right] = \sqrt{4\pi^{-1}(1+1/n)}$, while the difference between true effects should be $\sqrt{4\pi^{-1}}$(Geary 1935). Card, Heining, and Kline (2013) acknowledge this point in Section VB but abstract away from this finite-sample bias using the argument that n is roughly fixed across variances it wishes to compare; this is not the case in our empirical setting.

[18] Trainees switch services every week. So, in the limit, if I were to estimate a fixed effects model using only a week of data in which no trainees switch, then in fact *no* week-specific trainee effects would be identifiable.

[19] I observe approximately half as many patient-days for trainees in the third year, because third-year trainees spend more time in research and electives than in the first two years of training.

where the vector of parameters $\left(\hat{\beta}, \hat{\eta}, \hat{\gamma}_s, \hat{\zeta}_\ell\right)$ is estimated by OLS. This approach is a version of restricted maximum likelihood (REML), which avoids the incidental parameters problem in the later maximum-likelihood stage (Patterson and Thompson 1971). Importantly, as in Chetty, Friedman, and Rockoff (2014), I estimate these OLS parameters using variation *within* interactions of trainee pairs and discrete tenure periods, which allows the remaining trainee effects in $\tilde{Y}_{it}$ to be correlated with the predicted portion of log spending due to $\mathbf{X}_{it}$, $\mathbf{T}_t$, $s(i,t)$, and $\ell(i,t)$.

I then specify a crossed random effects model,

$$\tilde{Y}_{it} = \xi^{\tau_j;\tau_k}_{j(i,t)} + \xi^{\tau_k;\tau_j}_{k(i,t)} + \varepsilon_{it}. \tag{4}$$

Fixing τ_j and τ_k, I aim to simultaneously estimate $\sigma(\tau_j;\tau_k)$ and $\sigma(\tau_k;\tau_j)$ by restricting estimation of equation (4) to the set of observations $\mathcal{C}(\tau_j,\tau_k) = \left\{(i,t) : \tau\left(j(i,t),t\right) = \tau_j, \tau\left(k(i,t),t\right) = \tau_k\right\}$. I can recover the full set of possible standard deviations, $\left\{\sigma\left(\tau;\tau^-\right)\right\}_{(\tau,\tau^-)}$, by considering different sets of observations corresponding to combinations of τ_j and τ_k. In this way, I impose no functional form on the shape of practice variation over time, since practice variation in each pair of junior-senior tenure periods is estimated on a separate sample of observations.

Equation (4) can be stated in matrix form:

$$\tilde{\mathbf{Y}} = \mathbf{D}\mathbf{u} + \varepsilon, \tag{5}$$

where $\tilde{\mathbf{Y}}$ is the vector of differenced outcomes, $\mathbf{D}$ is a selection matrix, and $\mathbf{u}$ is a stacked vector of trainee random effects. Let N be the number of observations, N_J be the number of junior trainees, and N_K be the number of senior trainees in the sample $\mathcal{C}(\tau_j,\tau_k)$. Then the selection matrix $\mathbf{D}$ is $N \times (N_J + N_K)$ and assigns each observation (i,t) to a junior trainee with tenure τ_j and a senior trainee of tenure τ_k. The vector $\mathbf{u}$ is $(N_J + N_K) \times 1$ and contains the stacked effects of the N_J junior trainees and N_K senior trainees.

Assumption 1 implies that junior and senior trainee effects are independent of each other. So the variance-covariance matrix of $\mathbf{u}$ is diagonal:

$$\operatorname{var}\mathbf{u} = \mathbf{G} = \begin{bmatrix} \sigma^2(\tau_j;\tau_k)\mathbf{I}_{N_J} & \mathbf{0} \\ \mathbf{0} & \sigma^2(\tau_k;\tau_j)\mathbf{I}_{N_K} \end{bmatrix}.$$

Further assuming that trainee random effects and the error term are normally distributed, the log likelihood function is

$$\mathcal{L} = -\frac{1}{2}\left\{N\log(2\pi) + \log|\mathbf{V}| + \tilde{\mathbf{Y}}'\mathbf{V}^{-1}\tilde{\mathbf{Y}}\right\}, \tag{6}$$

where $\mathbf{V} = \mathbf{D}\mathbf{G}\mathbf{D}' + \sigma_\varepsilon^2\mathbf{I}_N$. In each sample of data $\mathcal{C}(\tau_j,\tau_k)$, I estimate $\sigma(\tau_j;\tau_k)$ and $\sigma(\tau_k;\tau_j)$ by maximizing equation (6).

The estimated variance components can be treated as empirical Bayes prior distributions. Treating $\tilde{\mathbf{Y}}$ as data, I can obtain empirical Bayes posterior means as

$$\tilde{\mathbf{u}} = \tilde{\mathbf{G}}\mathbf{D}'\tilde{\mathbf{V}}^{-1}\tilde{\mathbf{Y}},$$

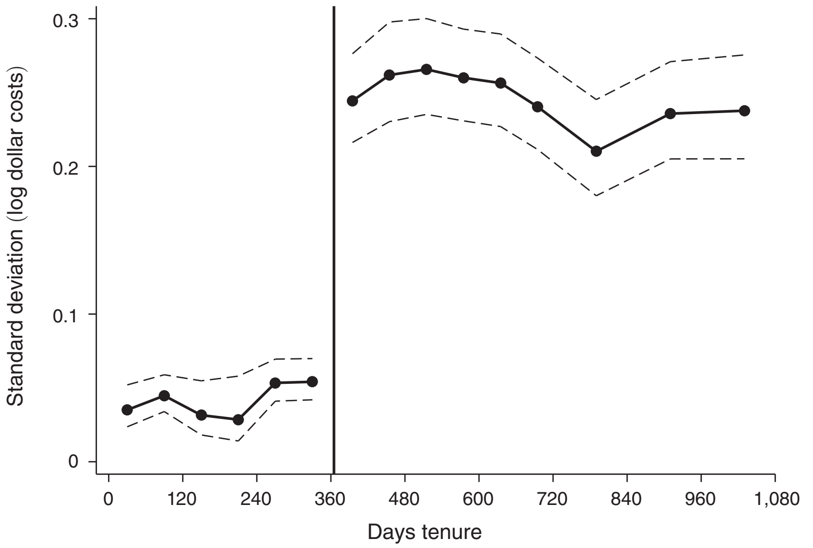

Figure 1. Profile of Practice Variation by Tenure

Notes: This figure shows practice variation, defined as the standard deviation of random trainee effects specified in equation (3), in log daily total costs at each nonoverlapping tenure period. Point estimates are shown as connected dots; 95 percent confidence intervals are shown as dashed lines. Trainees prior to one year in tenure are junior trainees and become senior trainees after one year in tenure; a vertical line denotes the one-year tenure mark. 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"). Estimates for junior trainees are done separately for second-year senior trainees and third-year senior trainees, then subsequently averaged for purposes of presentation. An alternative approach estimating junior-trainee practice variation by pooling observations by junior-trainee tenure yields qualitatively similar results.

where $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{V}}$ are $\mathbf{G}$ and $\mathbf{V}$ with random effects estimates of $\sigma^2(\tau_j;\tau_k)$, $\sigma^2(\tau_k;\tau_j)$ and σ_ε^2 plugged in. These posterior means are also known as "best linear unbiased predictions" or BLUPs (Searle, Casella, and McCulloch 1992).

In online Appendix A-3, I detail two extensions of the baseline model. First, I allow for patient admission random effects, since most patients are admitted for multiple days and may be cared for by multiple trainees. Results are qualitatively unchanged when including patient random effects. Second, I allow for estimation of the correlation between trainee effects of the same trainee in different tenure periods, which I employ in Section IVB.

III. Results

A. *Baseline Results*

Figure 1 presents results for practice variation from the baseline implementation described in Section IID. For each tenure interval τ_h, the figure displays an estimate of practice variation, or the estimated standard deviation of trainee effects among trainees with in the given tenure period.[20] A standard deviation increase in the effect of junior and senior trainees increases daily total spending by about 5 percent and 24 percent, respectively. The difference in practice variation between junior and senior trainees occurs entirely and discontinuously at the one-year tenure mark. Changes in practice variation are otherwise muted. In particular, there exists little convergence in practice styles within either the junior role or the senior role. After the one-year discontinuity, the standard deviation of the trainee effect distribution remains above 20 percent throughout. Including or omitting admission-level random effects for the patient does not qualitatively alter results.

These results suggest that team decision-making is highly concentrated among much fewer agents than would be the case with independent physician practice. One way to quantify this concentration is to consider two-agent teams of one junior trainee and one senior trainee for each case. In this construction, the senior trainee is responsible for $0.24^2/(0.05^2 + 0.24^2) \approx 96$ percent of the variance in team-level decisions across cases. This degree of concentration is even higher when accounting for the fact that a single senior trainee works with two junior trainees: in this case, the practice variation due to each of the two junior trainees are orthogonal to each other, but the common single senior trainee will drive practice variation for all patients under her span of control. Senior agents explain 99 percent of decision-making variance given this construction.

B. *Robustness*

I perform two robustness exercises to evaluate the validity of the baseline results. In the first robustness exercise, I address the institutional fact that a group of junior trainees known as "preliminary interns" who are not scheduled to continue in the same internal medicine residency will instead switch to other specialties (e.g., anesthesiology, dermatology, anesthesiology) after their first year of training. While these trainees make up a minority of the overall sample of trainees, if they as a group have lower practice variation than the remaining group, then their inclusion in the analysis could bias downward an estimate of the practice variation discontinuity at the one-year tenure mark for a *fixed* group of trainees. Since the identities of preliminary interns are known in advance, I exclude preliminary interns and reestimate the practice variation profile. Results of this robustness exercise, shown in panel A of Figure 2, are qualitatively unchanged.

Second, I consider the possibility that patients may not be quasi-randomly assigned to trainee teams. In particular, although online Appendix A-1 supports Assumption 1 in terms of observable patient characteristics, patients may differ along unobservable characteristics across different trainee teams. Since there is only one senior trainee on each team, compared to two junior trainees, systematic sorting of patients across teams would not only bias estimates of trainee effects but would

[20] As described in Section IIB, senior trainees of tenure τ_k only work with junior trainees of a given tenure $\tau_j = \tau_k - \lfloor \tau_k \rfloor$, where τ_j and τ_k are stated in continuous years. Junior trainees of tenure τ_j may work with senior trainees of tenure $\tau_j + 1$ or $\tau_j + 2$. Differences between $\sigma(\tau_j; \tau_j + 1)$ and $\sigma(\tau_j; \tau_j + 2)$ are small and statistically insignificant. I therefore average these two estimates to plot practice variation for tenure τ_j.

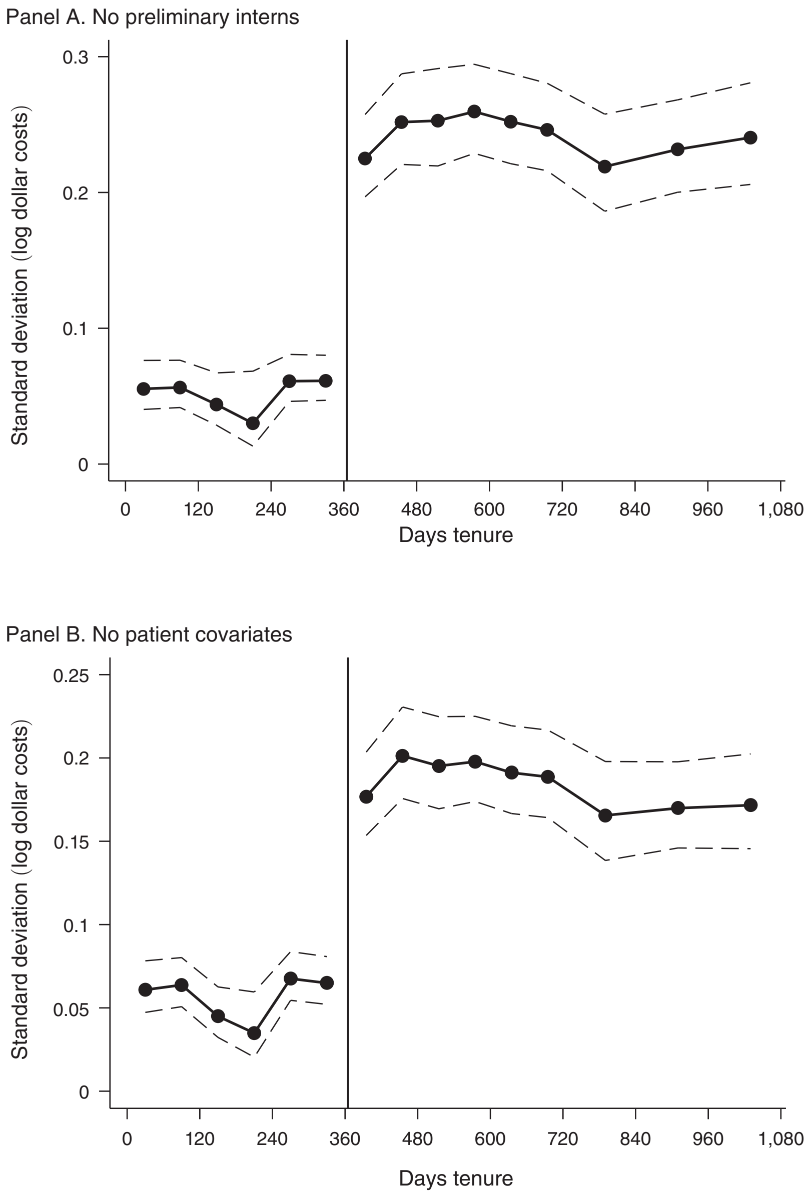

Figure 2. Robustness of Baseline Results

Notes: This figure shows robustness of baseline results shown in Figure 1 along two dimensions. In panel A, I drop “preliminary interns,” or junior trainees who are not scheduled to continue as senior trainees in internal medicine. This leaves only trainees that will continue on as senior trainees in internal medicine. In panel B, I estimate the model with no patient characteristics as covariates. The estimation approach is otherwise the same as for Figure 1. The model estimates practice variation, defined as the standard deviation of random trainee effects specified in equation (3), in log daily total costs at each nonoverlapping tenure period. The model controls are as stated for Figure 1. Point estimates are shown as connected dots; 95 percent confidence intervals are shown as dashed lines. Trainees prior to one year in tenure are junior trainees and become senior trainees after one year in tenure; a vertical line denotes the one-year tenure mark.

Table 3—Team Concerns

	Discontinuity	Convergence
Independent practice	None	Yes
Span of control	Decrease	Yes
Principal-agent, rank	Increase	Yes
Information aggregation	Increase	Depends

Notes: This table summarizes implications of three types of team concerns on two features of the practice variation profile with respect to trainee tenure: (i) the existence and direction of a discontinuity in practice variation as trainees move from junior to senior at the one-year tenure mark and (ii) whether practice variation decreases (i.e., practices "converge") with tenure as trainees learn. Section IIIC discusses team concerns in detail. "Independent practice" represents the benchmark with no team concerns. "Span of control" refers to the team theoretic idea that agents have limited bandwidth, and agents higher in the hierarchy (i.e., senior trainees) assigned to more than one agent lower in the hierarchy (i.e., junior trainees) will attend to fewer problems per case (Marschak and Radner 1972, Garicano 2000). "Principal-agent, rank" refers to models in which "senior" agents may have power over other "junior" agents, which may induce junior agents to suggest decisions that herd around the beliefs of senior agents (Prendergast 1993). "Information aggregation" refers to the possibility of Bayesian information aggregation across agents to make a single decision (DeGroot 2004). Although team concerns may coexist, each row represents one team concern in the absence of the other two team concerns. For example, in "span of control," predictions are for classical team theory, in which there are no principal-agent issues, and decisions are separable across agents and do not aggregate information.

also spuriously induce greater practice variation among senior trainees. To assess this possibility, I reestimate the practice variation profile with no patient controls. As shown in panel B of Figure 2, the practice variation profile from this exercise also remains qualitatively unchanged from the baseline implementation. This shows that including or removing rich patient controls has no qualitative effect on the key moments of practice variation and is consistent with the causal interpretation of trainee effects implied by Assumption 1.

C. *Team Concerns*

Variation in trainee effects may reflect two conceptual objects: (i) differences in trainee *judgments*, if they were allowed to make decisions on their own, and (ii) *influence* on team decisions, or the extent to which trainees may sway team decisions. Judgments may reflect prior knowledge, beliefs, or preferences and exist outside of a team setting. In contrast, if agents make decisions independently, then they will have full and invariant influence. In other words, for influence to matter for practice variation, teams must alter the process of decision-making.

I consider three types of "team concerns." First, in team theory, organizations allocate decisions to individuals under bandwidth constraints (Marschak and Radner 1972, Garicano 2000). Organizations are naturally hierarchical, and higher levels in this hierarchy, where agents have a greater "span of control," handle fewer decisions. Second, in principal-agent models, teams may induce "herding" of decisions around the beliefs of senior agents, simply based on the prestige, rank, or power of senior agents (Scharfstein and Stein 1990, Prendergast 1993). Junior agents may act as "yes men" to further their careers at the cost of worse team decisions. Third, teams may aggregate information, as agents may confer with each other before making team decisions. Joint decisions cannot be fully separated and distributed to individual agents but instead pool input across agents.

What do team concerns imply for the empirical pattern of practice variation with respect to tenure in medical residency? I consider two features of medical residency to answer this question and summarize informal implications in Table 3. The first feature concerns the structure of residency teams relative to the one-year tenure mark. When trainees pass the one-year mark, their span of control, rank, and relative experience all discontinuously increase. Assuming that judgments and preferences are continuous across the one-year tenure mark, any discontinuity in practice variation reflects the impact of influence via team concerns. Limited bandwidth would imply a *decrease* in practice variation at the one-year discontinuity, since senior trainees have greater span of control. However, either career concerns or information aggregation would imply an *increase* in practice variation at this discontinuity. If no team concerns are at play (i.e., physicians practice independently), then there should be no discontinuity in practice variation at the one-year mark.

The second feature concerns implications for practice convergence under learning, since an immense amount of learning occurs in medical residency, at least according to qualitative reports (Ludmerer 2014). If physicians practice independently or if decisions are separable across trainees, learning would imply convergence in practices over time (i.e., practice variation should decrease with tenure).[21] But if teams aggregate information across agents in joint decisions, then influence may grow endogenously as judgments become more precise, and there may be no practice convergence despite dramatic learning in residency.

IV. Mechanisms

In this section, I delve further into mechanisms that may underlie the basic empirical results that (i) influence jumps discontinuously when trainees assume the senior role and (ii) convergence in practice variation is generally muted even as trainees progress in residency. I consider two types of mechanisms introduced in Section IIIC. First, senior trainees may arbitrarily exercise greater influence, regardless of their knowledge. For example, they may hold prestige, rank, or power that is unrelated to knowledge, or they may simply have decision rights in their jobs for "important" decisions. Second, influence in teams may depend on systematic differences in knowledge between teammates with different tenures. This mechanism might allow for differences in knowledge that are simply correlated across tenure groups, as opposed to within tenure groups. Thus, this mechanism may arise even in the case that knowledge in individual cases is not directly observable, if general relationships between tenure and knowledge are known.

The analyses in this section proceed in three parts under the following reasoning. First, if junior and senior trainees simply have different jobs with different decisions rights, then in the Garicano (2000) model, junior trainees could have greater control

[21] In classical team theory, an agent knows how to solve a problem completely or not at all, problems are fully separable across agents, and the organization is structured so that problems are distributed efficiently to the proper agents.

over some types of decisions. On the other hand, if all decisions require knowledge gained with experience, then we should find the same practice variation profile over all types of decisions. I will thus first explore heterogeneity across different types of decisions. Second, in order for the first category of mechanisms to *fully* explain the practice variation pattern in Figure 1, there must be close to no learning, since there is almost no convergence in practice variation. I will therefore examine the extreme proposition of no learning. Third, I will examine the opposite proposition that influence is optimally allocated according to knowledge. While the two mechanisms are not mutually exclusive, these analyses may shed light on the relative importance of each mechanism.

A. *Decision Types*

I reestimate practice variation profiles, using the same approach described in Section II, by subsetting decisions in several ways. First, I consider decisions in the four main clinical cost departments of diagnosis (radiology and laboratory), medication, blood transfusion, and nursing. Rather than aggregating the direct costs of all orders for a given patient-day case (i,t), I only aggregate the direct costs of the subset of orders in a given clinical category. Second, I consider how practice variation profiles may differ by patient severity or whether decisions are early versus late in a patient's stay. Finally, I subset cases $(i,t) \in \mathcal{C}$ according to formal diagnostic codes, grouped by the frequency of the diagnostic code or by whether there exists a formal guideline for the diagnostic code in guidelines.gov (Agency for Health Research and Quality 2015, Chan 2020).

In all of these cases, the practice variation profile is qualitatively similar: variation increases discontinuously at the one-year tenure mark and remains stable to the end of training. Figure 3 shows practice variation profiles across different clinical cost categories. Figures 4 and 5 show virtually identical practice variation profiles across patient severity, patient-days earlier or later in a patient's stay, and patients with different formal diagnoses.[22]

Despite qualitative similarities in Figure 3, the magnitudes of practice variation and its discontinuous increase at the one-year mark do vary meaningfully across clinical cost categories. Diagnostic spending shows the largest increase in practice variation, with a standard deviation of 16 percent to 74 percent before and after the one-year tenure mark. In contrast, medication and nursing spending shows relatively small practice variation, both overall and in the increase at the relative experience discontinuity. These differences may be consistent with greater uncertainty and greater control by trainees of diagnostic and transfusion decisions.[23] On the

[22] Interestingly, practice variation is remarkably similar between formal diagnoses with and without formal guidelines. This possibly reflects the coarseness of formal diagnoses and formal guidelines. For example, "Chest pain, not otherwise specified" is the most common formal diagnostic code both for patients admitted to general medicine and for patients admitted to the subspecialty cardiology service. The coarseness of formal diagnostic codes and a review of the guidelines strongly suggest that very little meaningful clinical information can be formally encoded (Shaneyfelt, Mayo-Smith, and Rothwangl 1999). Another explanation is that, while guidelines may decrease practice uncertainty, diagnoses with more uncertainty may warrant guidelines.

[23] Medication decisions are better described in publicly accessible sources of knowledge, while diagnostic decisions draw more on clinical reasoning that would be difficult to prespecify and reference for trainees who have never

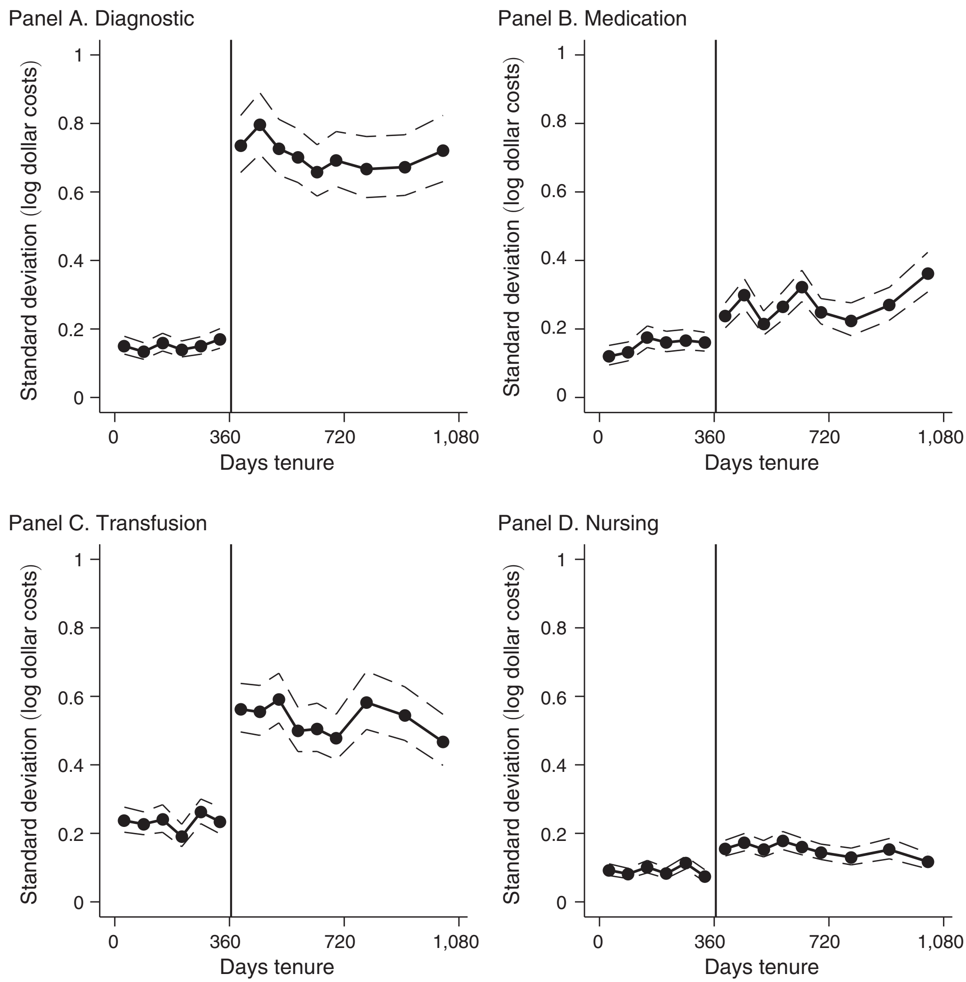

Figure 3. Practice Variation Profile by Spending Category

Notes: This figure shows practice variation, defined as the standard deviation of random trainee effects specified in equation (3), in log daily costs at each nonoverlapping tenure period. Each panel shows a different spending category. The model controls are as stated for Figure 1. Point estimates are shown as connected dots; 95 percent confidence intervals are shown as dashed lines. Trainees prior to one year in tenure are junior trainees and become senior trainees after one year in tenure; a vertical line denotes the one-year tenure mark.

other hand, decisions types with greater proportional increases in influence at the one-year tenure mark do not account for a larger share of total spending (see Table 2 for summary statistics by clinical cost department). Given the magnitudes of trainee effects on overall spending (Figure 1), this suggests spillovers across clinical cost categories, driven by interconnected decisions.

before encountered a patient presentation. Similarly, blood transfusion reflects an important decision with large variation across providers and surprisingly little guidance for how to tailor the transfusion decision to individual cases (Carson et al. 2016). On the other hand, nursing decisions are intuitively outside the scope of most physician decision-making, and it seems intuitive that physician trainees will have little influence on these decisions.

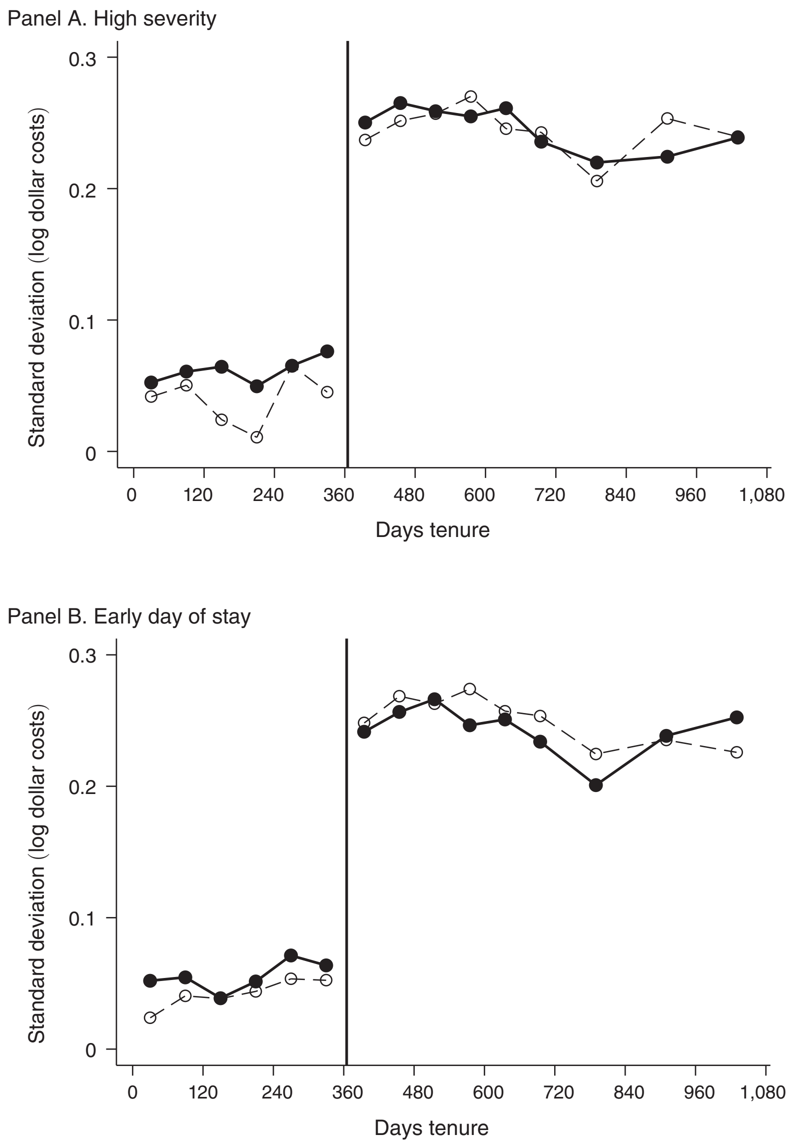

Figure 4. Practice Variation Profile by Patient Severity and Day of Stay

Notes: This figure shows practice variation, defined as the standard deviation of random trainee effects specified in equation (3), in log daily total costs at each nonoverlapping tenure period. Panel A estimates the model separately in two samples of patients with above- (solid dots) versus below-median (hollow dots) expected 30-day mortality. Panel B estimates the model separately in two samples of days before (solid dots) versus after (hollow dots) the middle of each patient's stay (with the middle day, if it exists, randomized between the two groups). The model controls are as stated for Figure 1. Trainees prior to one year in tenure are junior trainees and become senior trainees after one year in tenure; a vertical line denotes the one-year tenure mark.

B. *No-Learning Scenario*

I next evaluate the extreme case in which the influence differential between senior and junior trainees is unrelated to any differential in knowledge. This case is tanta-

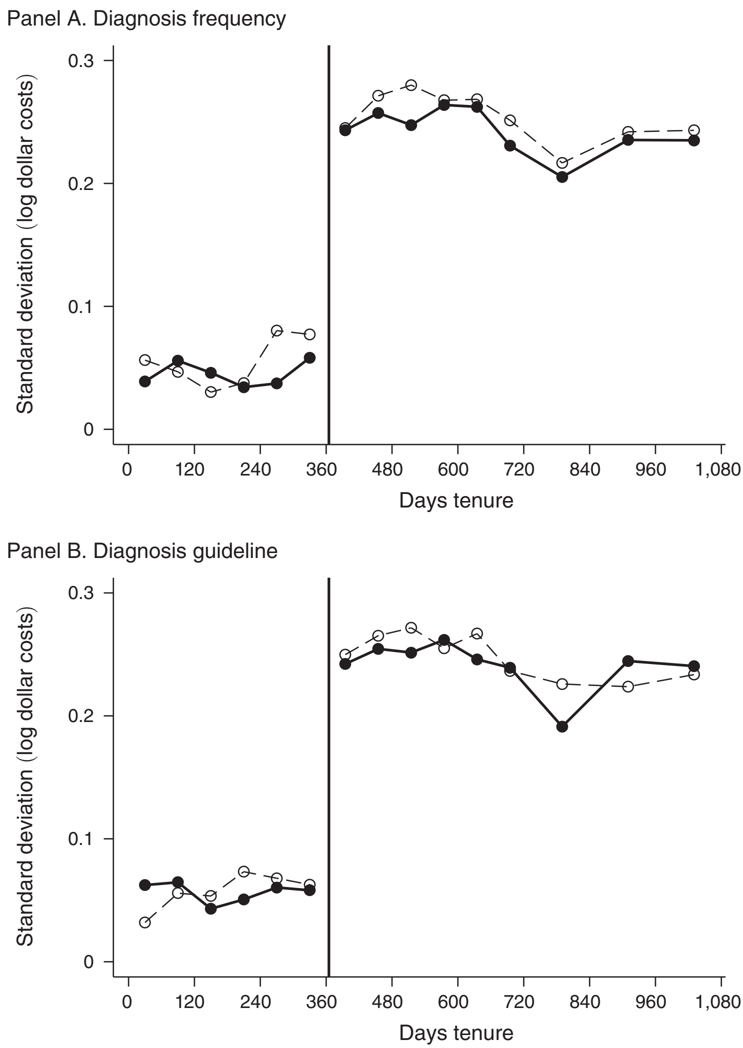

Figure 5. Practice Variation Profile by Diagnosis Type

Notes: This figure shows practice variation, defined as the standard deviation of random trainee effects specified in equation (3), in log daily total costs at each nonoverlapping tenure period. Panel A estimates the model separately in two samples of patients with diagnosis (ICD-9) codes with above- (solid dots) versus below-median (hollow dots) frequency in the data. Panel B estimates the model separately in two samples of patients with diagnosis codes with (solid dots) and those without (hollow dots) published guidelines cataloged by the US Agency for Healthcare Research and Quality (guidelines.gov). The model controls are as stated for Figure 1. Trainees prior to one year in tenure are junior trainees and become senior trainees after one year in tenure; a vertical line denotes the one-year tenure mark.

mount to no learning in residency. Instead, physicians may differ in their judgments, due, for example, to heterogeneous preferences or beliefs, in ways that predate residency and are time invariant during residency. Given the intensity of residency training, this scenario seems unlikely on its face. However, a general version of intrinsic heterogeneity that is relatively stable over time has been invoked in many settings, several of them in health care (e.g., Doyle, Ewer, and Wagner 2010; Fox and Smeets

2011; Bartel et al. 2014; Currie and MacLeod 2017). I therefore evaluate the relative importance of time-invariant heterogeneity in explaining practice variation using two complementary approaches.

In the first approach, I exploit detailed trainee characteristics that should be highly correlated with preferences and ability, including demographics, prior formal degrees, place of medical school, standardized examination scores, position on the rank list, and future income. Indeed, these characteristics are the key summary statistics considered by residency programs in accepting future physicians and may represent important differences in ability and future careers. Empirically, I show that ex ante trainee characteristics strongly predict position on the rank list (i.e., desirability to the residency program) and the probability of higher-than-median future income, which is at least 50 percent greater than the future income below median.[24] However, despite these important relationships between trainee characteristics and career-changing outcomes, I strikingly find that these trainee characteristics are broadly uncorrelated with trainee effects on clinical decisions. In Figure 6, I show that the *distribution* of trainee effects in each tenure period throughout residency is also unchanged regardless of conditioning on trainees rank or future income. I describe these analyses further in online Appendix A-4 and present more exhaustive results in Table 4.[25]

In the second approach, I measure the serial correlation between random trainee effects in two different tenure periods and provide further details of the statistical approach in online Appendix III.B. The conceptual reason for examining serial correlation is as follows: if practice variation reflects intrinsic heterogeneity and no learned beliefs, then effects in different time periods within the same trainee should be constantly and highly correlated, regardless of the time between the time periods. However, if trainees are learning, then adjacent time periods should exhibit higher correlation in trainee effects than do distant time periods. Figure 7 presents averages of serial correlation estimates between trainee effects as a function of the distance between the tenure periods.[26] Serial correlation in trainee effects across two adjacent periods is moderately positive, while the correlation quickly decreases to zero with more distance between the two periods. Interestingly, correlation eventually becomes negative, though statistically indistinguishable from zero, between trainee effects in distant periods. These results strongly suggest that judgments during residency are quite dynamic. In other words, consistent with numerous qualitative accounts, trainees are engaging in active learning during residency.

[24] Trainees with a predictive score one standard deviation above mean are two to three times more likely to be ranked in the upper half of the rank list than those with a predictive score one standard deviation below mean. Trainees with a predictive score one standard deviation above mean are more than three times as likely to obtain above-median future income than those with a predictive score one standard deviation below mean.

[25] In online Appendix A-5, I also explore whether trainee practice styles can be predicted by supervising physicians and senior trainees whom they have worked with in the past. Interestingly, practice variation is orthogonal to the practice styles of these past teammates. This suggestive evidence is consistent with tacit knowledge and experiential learning.

[26] Online Appendix Table A-4 and online Appendix Figure A-4 show results for individual pairs of tenure periods.

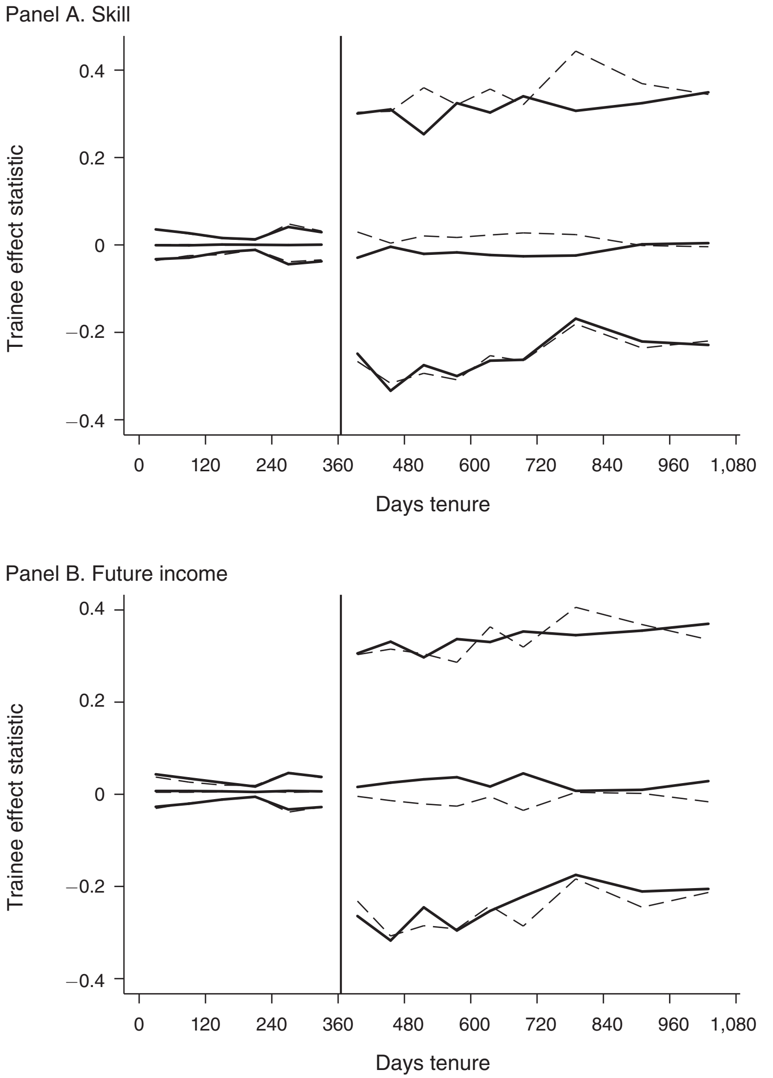

Figure 6. Practice Style Distribution by Trainee Type

Notes: This figure shows the patient-day-weighted ninetieth percentile, mean, and tenth percentile of the practice style (trainee effect) distribution according to trainee type. The unconditional distribution in each tenure period is normalized to have mean zero. Panel A shows the distribution for high-skill trainees (solid lines) relative to low-skill trainees (dashed lines), where "skill" is defined as position on the rank list more favorable than median when defined, and above-median USMLE test score when position on the rank list is missing. Panel B shows the distribution for trainees with above-median expected future income relative (solid lines) to those with below-median future income (dashed lines), where future income is based on known subsequent subspecialty training (if any) and imputed with national average yearly income in the first five years of practice after training. The average yearly future incomes of above- and below-median junior trainees are \$424,000 and \$268,000, respectively; the respective yearly future incomes for senior trainees are \$409,000 and \$249,000 (junior trainees include "preliminary interns," described in Section I, who generally move on to more lucrative specialties). Practice styles are calculated as empirical Bayes posterior means from the random effects model specified in equation (3), where estimated variance components of the random effects model are treated as prior distributions. The model controls are as stated for Figure 1. Trainees prior to one year in tenure are junior trainees and become senior trainees after one year in tenure; a vertical line denotes the one-year tenure mark. Results for other trainee characteristics are shown in online Appendix Tables A-2 and A-3.

TABLE 4—EFFECT OF TRAINEE CHARACTERISTICS ON SPENDING

	log daily total costs					
	Male (1)	High USMLE (2)	Highly ranked (3)	High future income (4)	Other hospital (5)	Overall score (6)
Panel A. Interns						
Effect of trainee with characteristic	−0.001	0.002	0.010	0.007	0.017	0.003
	(0.004)	(0.005)	(0.006)	(0.004)	(0.010)	(0.002)
Observations	186,398	185,201	131,247	215,678	219,727	190,331
Adjusted R^2	0.089	0.089	0.090	0.088	0.087	0.090
Sample characteristic mean	0.596	0.258	0.234	0.415	0.055	N/A
Panel B. Residents						
Effect of trainee with characteristic	−0.013	0.010	−0.004	−0.001	0.013	0.004
	(0.004)	(0.005)	(0.007)	(0.004)	(0.011)	(0.002)
Observations	206,455	199,371	129,281	218,376	219,727	206,455
Adjusted R^2	0.089	0.089	0.083	0.087	0.088	0.090
Sample characteristic mean	0.564	0.235	0.214	0.332	0.060	N/A

Notes: This table reports results for some regressions of the effect of indicators of some trainee characteristics, including other hospital status, and a normalized predictive score (with standard deviation one) based on all observed trainee characteristics. Panel A shows results for interns; panel B shows results for residents. Columns 1 to 5 are regressions of the form in online Appendix equation (A-9), where the coefficient of interest is on an indicator for a group of trainees identified by either preresidency characteristics, whether the trainee is from the other academic hospital, or whether the trainee is expected to have above-median future income based on known subspecialty training following residency. The effects of many other characteristics of interest (or groups) were estimated as insignificant and omitted from this table for brevity. Column 6 reports results where the covariate of interest is a normalized predictive score based on predetermined characteristics of age, sex, minority status, track, rank on matching rank list, USMLE score, medical school rank in *US News & World Report*, indicators for whether the medical school is foreign or "rare," AOA medical honor society membership, and additional degrees at time of residency matriculation. By comparison, a predictive score for being highly ranked (in the top 50 rank positions) based on the same characteristics (except rank) changes the probability of being highly ranked by about 20 percent for both interns and residents. All models control for patient and admission characteristics, time dummies, and fixed effects for attending and the other trainees on the team (e.g., the resident is controlled for if the group is specific to the intern). Standard errors are clustered by admission.

C. *Optimal Influence*

In the other extreme, I consider a simple model of optimal influence by Bayesian information aggregation. In order to optimize the decision at hand, teams allocate influence in proportion to the knowledge of each team member (DeGroot 2004). The more precise the signal from her prior knowledge relative to other sources of information, the greater her influence will be. At the one-year tenure mark, influence discontinuously increases because the knowledge of a trainee's teammate discontinuously decreases.

In this model, as trainees gain knowledge, their judgments will converge, but their influence will increase. These two factors have opposing implications for practice variation, so that practice variation may not always decrease with learning. In contrast, agents who practice independently have constant (full) influence and should always exhibit convergence in their practice styles as they learn. In this way, I use tenure-specific knowledge and Bayesian aggregation in teams to imply tenure-specific practice variation. An assumption of continuous knowledge places restrictions on patterns of influence—and therefore practice variation—over trainees' tenure. A further assumption that supervising physicians possess at least as

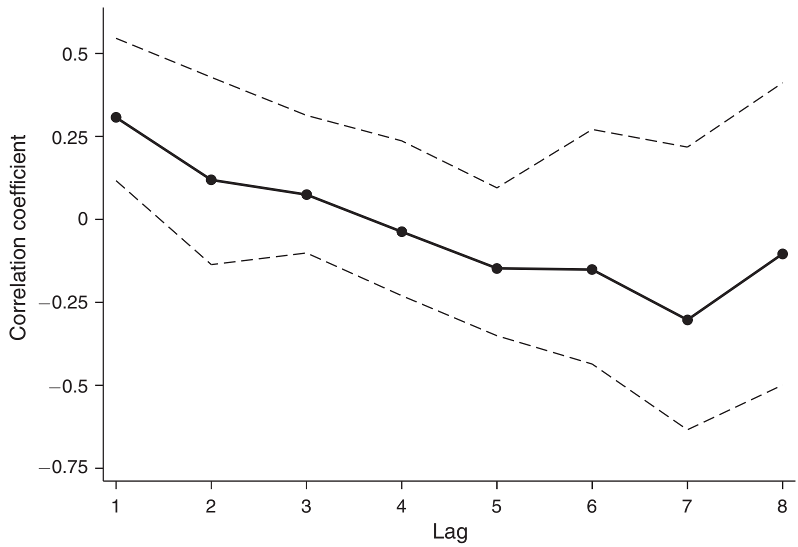

FIGURE 7. AVERAGE SERIAL CORRELATION BY TENURE PERIOD LAG

Notes: This figure shows the average serial correlation in trainee effects between 120-day tenure periods as a function of the lag between the tenure periods. Serial correlation parameters are estimated for each pair of tenure periods by a maximum likelihood method described further in online Appendix III.B. There are a total of nine nonoverlapping tenure periods across the three years of training. The x-axis corresponds to the lag between the tenure periods, such that when the lag is 1, the y-axis displays the average of the serial correlations across the pairs of tenure periods $(1,2),(2,3),\ldots,(8,9)$. In general, for lag L, the y-axis displays an average of the serial correlations computed for $9-L$ tenure periods $(1,1+L),\ldots,(9-L,9)$. Thus, for the lag of 1, the average is across 8 serial correlation cells, while for the lag of 8, the "average" simply contains the serial correlation between tenure periods $(1,9)$. Underlying results for each pair of tenure periods are given in online Appendix Table A-4 and are also shown graphically in online Appendix Figure A-4. Confidence intervals are calculated by bootstrap. The model controls are as stated for Figure 1.

much knowledge as a senior trainee imposes another restriction on the scale of practice variation. The actual pattern and scale of practice variation therefore allow identification of deviations from optimal influence.

In online Appendix A-6, I provide details of the model setup, identification, estimation procedure, and results. In brief, the model uses tenure-specific moments of practice variation from the random effects model in equation (3) to recover underlying primitives of learning (i.e., the rate at which knowledge increases with tenure), as well as potential deviations from optimal influence capturing other team concerns in Section IIIC. I specify influence as divided between the junior trainee, the senior trainee, and "external information," which may be drawn from the supervising physician, any other personnel, or guidelines and protocols.

In results, I find very little knowledge at the beginning of residency compared to learning in the first year. Learning in the second year occurs at a much faster rate than in the first year but appears to cease by the third year.[27] Between junior and

[27] There exists a large theoretical literature on why learning may stop, related to learning costs or knowledge constraints (e.g., Rogerson, Shimer, and Wright 2005; Caplin and Dean 2015), uncertainty in the mapping between beliefs and data (Acemoglu, Chernozhukov, and Yildiz 2006), and social learning (Ellison and Fudenberg 1993).

senior trainees, influence approximates the Bayesian benchmark. In likelihood ratio tests, I cannot reject a model with learning and optimal influence between trainees, compared to a less restrictive model that allows for deviations from optimal influence. However, I find that external information (including the supervising physician) influences decisions by less than half of the influence of a graduating trainee. This suggests that trainees, as a group, are given much more influence than warranted under the Bayesian benchmark. Many of these patterns persist when reestimating the model with practice variation profiles in specific spending categories and types of cases.

Although the model is highly stylized and is based on relatively few empirical moments, the idea that learning increases when trainees become senior and have a greater stake in decision-making is consistent with experiential learning (Dewey 1938). Experiential learning implies a tradeoff in the use of information to make team decisions. While supervisory information improves the quality of decision-making at hand, it may constrain experiential learning by trainees. Perhaps for this reason, external information receives much less weight than it should in a Bayesian framework that optimizes only decisions at hand. Online Appendix A-6 undertakes counterfactual analyses to quantify the welfare consequences of this tradeoff.

V. Discussion and Conclusion

In this paper, I study decision-making in teams within the setting of physician trainees in medical residency. As in other settings involving teams, I observe decisions attributable to teams and the team members at the time of each decision, but I do not directly observe the agents' contributions to the decisions. Building on a "movers literature" starting with Abowd, Kramarz, and Margolis (1999), I develop and apply a method to extract each team member's average contribution to decisions over time, using quasi-experimental variation in the assignment of cases and physician trainees to teams, as well as frequent switches of trainees across teams. By tracking the effects of trainees on team decisions over their tenure, I also shed light on how teams may alter decision-making relative to agents who make decisions "on their own."

In my primary finding, I show that senior trainees explain the vast majority of practice variation across teams. This suggests that differences across organizations in health care and other settings may be driven by a few individuals. Furthermore, by exploiting a discontinuity in team roles at the one-year tenure mark, I show that team dynamics are responsible for this outsize influence of senior trainees. From multiple analytical lenses, I find evidence suggestive of an intriguing interplay between experiential learning during residency training and the allocation of influence in teams. While this evidence on its own may be suggestive, it is consistent with a large body of work, much of it outside of economics, suggesting "tacit knowledge" that is difficult to pass on to others (Polanyi 1966) and "experiential learning" that accrues only through experience (Dewey 1938).

At a minimum, these results suggest the importance of team concerns in analyzing decisions made within organizations. In health care, a large and influential literature has focused primarily on variation across regions or institutions. If decision-making is concentrated in the hands of a few individuals, then understanding microlevel foundations of decision-making will be essential for characterizing variation that has long been noted at more aggregate levels.

Moreover, if an important task for teams is to aggregate information, particularly for complex and consequential decisions, then policymakers may need to focus more on the informational frictions that underlie the skewed concentration of knowledge and the remaining practice variation even among experts with the most knowledge. Such informational levers could be more effective at reducing practice variation (Newhouse et al. 2013) compared to previously proposed policy levers of financial incentives, simple reporting of variation, and patient cost sharing (see Skinner 2011 for a summary). This idea is consistent with a growing literature that suggests that skill, or productivity, plays an important role in practice variation in both diagnostic and treatment decisions.[28] If providers simultaneously under- and overtreat patients, then instituting policy levers to encourage providers to treat either uniformly more or uniformly less will be ineffective. Similarly, imposing a uniform treatment rate can be counterproductive if undertreatment is costlier and if providers who treat more do so because they are less skilled at targeting (Chan, Gentzkow, and Yu 2019). Instead, policy levers need to accomplish better targeting of resources by improving the use of information in decision-making.

Further, if learning requires experience and feedback, then the usual forms of spreading information, such as clinical guidelines, formal instruction (e.g., "continuing medical education"), or formal testing (e.g., board recertification), may do little to change practice or generate consensus (Shaneyfelt, Mayo-Smith, and Rothwangl 1999). While effective policies are beyond the scope of this paper, such policies will likely need to improve the use of existing information within an organization and encourage its spread across team members and organizations. For example, process innovations might invite feedback from peers, or they might explicitly use consensus building to specify nuanced "clinical pathways." These approaches aim to promote learning among "experts"—well beyond residency training—and by organizations themselves (Smith et al. 2012; Bohmer, Edmondson, and Feldman 2013).

REFERENCES

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh. 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review* 106 (12): 3730–64.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2): 251–333.

Abowd, John M., Francis Kramarz, and Simon Woodcock. 2008. "Econometric Analyses of Linked Employer-Employee Data." In *The Econometrics of Panel Data*, edited by Lászió Mátyás and Pat-

[28] A recent literature in economics has begun to directly consider skill in diagnosis, decision-making, and treatment. Abaluck et al. (2016) investigate whether providers decide to test for pulmonary embolisms and find that misallocation of resources has much larger welfare consequences than systematic overuse; Mullainathan and Obermeyer (2019) similarly find over- and undertesting of heart attack. Currie and MacLeod (2017) show variation in allocation of cesarean sections to patients according to their characteristics ("diagnostic skill") that could be as important as variation in procedural skill. Gowrisankaran, Joiner, and Léger (2017) investigate diagnosis and treatment of specific potential conditions in the emergency department. Chandra and Staiger (2017) apply a framework to study variation in spending across hospitals and examine to what extent this variation reflects allocative inefficiency versus comparative advantage.

rick Sevestre, 727–60. Advanced Studies in Theoretical and Applied Econometrics. Heidelberg: Springer Berlin.

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz. 2006. "Learning and Disagreement in an Uncertain World." NBER Working Paper 12648.

Agency for Health Research and Quality. 2015. "Guidelines and Measures." www.guidelines.gov (accessed January 21, 2015).

Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 29 (3): 155–73.

Bartel, Ann P., Nancy D. Beaulieu, Ciaran S. Phibbs, and Patricia W. Stone. 2014. "Human Capital and Productivity in a Team Environment: Evidence from the Healthcare Sector." *American Economic Journal: Applied Economics* 6 (2): 231–59.

Benkard, C. Lanier. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review* 90 (4): 1034–54.

Bloom, Nicholas, and John Van Reenen. 2010. "Why Do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives* 24 (1): 203–24.

Bohmer, Richard M.J., Amy C. Edmondson, and Laura Feldman. 2013. "Intermountain Health Care." Harvard Business School Case 603-066.

Caplin, Andrew, and Mark Dean. 2015. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review* 105 (7): 2183–2203.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128 (3): 967–1015.

Carson, Jeffrey L., Gordon Guyatt, Nancy M. Heddle, Brenda J. Grossman, Claudia S. Cohn, Mark K. Fung, Terry Gernsheimer, et al. 2016. "Clinical Practice Guidelines From the AABB: Red Blood Cell Transfusion Thresholds and Storage." *Journal of the American Medical Association* 316 (19): 2025–35.

Chan, David C. 2021. "Replication data for: Influence and Information in Team Decisions: Evidence from Medical Residency." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.38886/E117862V1.

Chan, David C., Jr., Matthew Gentzkow, and Chuan Yu. 2019. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." NBER Working Paper 26467.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Healthcare Exceptionalism? Productivity and Allocation in the US Health Care Sector." *American Economic Review* 106 (8): 2110–44.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Healthcare: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy* 115: 103–40.

Chandra, Amitabh, and Douglas Staiger. 2017. "Identifying Sources of Inefficiency in Health Care." NBER Working Paper 24035.

Charlson, Mary E., Peter Pompei, Kathy L. Ales, and C. Ronald MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40 (5): 373–83.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79.

Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen. 2015. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." NBER Working Paper 21815.

Currie, Janet, and Jonathan Gruber. 1996. "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women." *Journal of Political Economy* 104 (6): 1263–96.

Currie, Janet, and W. Bentley MacLeod. 2017. "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians." *Journal of Labor Economics* 35 (1): 1–43.

Cutler, David. 2010. "Where Are the Health Care Entrepreneurs?" *Issues in Science and Technology* 27 (1).

Cutler, David, Jonathan Skinner, Ariel Dora Stern, and David Wennberg. 2019. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending." *American Economic Journal: Economic Policy* 11 (1): 192–221.

DeGroot, Morris H. 2004. *Optimal Statistical Decisions*. Hoboken, NJ: John Wiley and Sons.

Dewey, John. 1938. *Experience and Education*. New York: Kappa Delta Pi.

Doyle, Joseph J., Jr., Steven M. Ewer, and Todd H. Wagner. 2010. "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams." *Journal of Health Economics* 29 (6): 866–82.

Doyle, Joseph J., Jr., John A. Graves, Jonathan Gruber, and Samuel A. Kleiner. 2015. "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns." *Journal of Political Economy* 123 (1): 170–214.

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

Ellison, Glenn, and Drew Fudenberg. 1993. "Rules of Thumb for Social Learning." *Journal of Political Economy* 101 (4): 612–43.

Epstein, Andrew J., and Sean Nicholson. 2009. "The Formation and Evolution of Physician Treatment Styles: An Application to Cesarean Sections." *Journal of Health Economics* 28 (6): 1126–40.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics* 131 (4): 1681–1726.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F.L. Lucas, and Etoile L. Pinder. 2003a. "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care." *Annals of Internal Medicine* 138 (4): 273–87.

Fisher, Elliott S., David E. Wennberg, Thérèse A. Stukel, Daniel J. Gottlieb, F.L. Lucas, and Etoile L. Pinder. 2003b. "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care." *Annals of Internal Medicine* 138 (4): 288–98.

Fox, Jeremy T., and Valérie Smeets. 2011. "Does Input Quality Drive Measured Differences in Firm Productivity?" *International Economic Review* 52 (4): 961–89.

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

Gartner, Alan, Mary Kohler, and Frank Riessman. 1971. *Children Teach Children: Learning by Teaching*. New York: Harper and Row.

Geary, R.C. 1935. "The Ratio of the Mean Deviation to the Standard Deviation as a Test of Normality." *Biometrika* 27 (3/4): 310–32.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

Gowrisankaran, Gautam, Keith A. Joiner, and Pierre-Thomas Léger. 2017. "Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments." NBER Working Paper 24155.

Hayek, F.A. 1945. "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519–30.

Hendel, Igal, and Yossi Spiegel. 2014. "Small Steps for Workers, a Giant Leap for Productivity." *American Economic Journal: Applied Economics* 6 (1): 73–90.

Kahn, Lisa B., and Fabian Lange. 2014. "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures." *Review of Economic Studies* 81 (4): 1575–1613.

Kolb, D.A., and R. Fry. 1975. "Toward an Applied Theory of Experiential Learning." In *Theories of Group Process*, edited by C. Cooper, 33–57. London: John Wiley.

Lazear, Edward P., Kathryn L. Shaw, and Christopher T. Stanton. 2015. "The Value of Bosses." *Journal of Labor Economics* 33 (4): 823–61.

Levitt, Steven D., John A. List, and Chad Syverson. 2013. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant." *Journal of Political Economy* 121 (4): 643–81.

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

Marschak, Jacob, and Roy Radner. 1972. *Economic Theory of Teams*. New Haven, CT: Yale University Press.

McCarthy, Douglas, and David Blumenthal. 2006. "Stories from the Sharp End: Case Studies in Safety Improvement." *Milbank Quarterly* 84 (1): 165–200.

Mincer, Jacob. 1962. "On-the-Job Training: Costs, Returns, and Some Implications." *Journal of Political Economy* 70 (5): 50–79.

Molitor, David. 2017. "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration." *American Economic Journal: Economic Policy* 10 (1): 326–56.

Montessori, Maria. 1948. *The Discovery of the Child*. Thiruvanmiyur, Madras: Kalkshetra Publications Press.

Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78 (381): 47–55.

Mullainathan, Sendhil and Ziad Obermeyer. 2019. "A Machine Learning Approach to Low-Value Health Care: Wasted Tests, Missed Heart Attacks and Mis-predictions." NBER Working Paper 26168.

Newhouse, Joseph P., Alan M. Garber, Robin P. Graham, Margaret A. McCoy, Michelle Mancer, Ashna Kibria, Committee on Geographic Variation in Health Care Spending and Promotion of High-Value Core, and Board on Health Care Services, eds. 2013. *Variation in Health Care Spending—Target Decision Making, Not Geography*. Washington, DC: National Academies Press.

Patterson, H.D., and R. Thompson. 1971. "Recovery of Inter-Block Information When Block Sizes Are Unequal." *Biometrika* 58 (3): 545–554.

Piaget, Jean. 1971. *Psychology and Epistemology*. New York: Grossman Publishers.

Polanyi, Michael. 1966. *The Tacit Dimension*. Chicago: University of Chicago Press.

Prendergast, Canice. 1993. "A Theory of Yes Men." *American Economic Review* 83 (4): 757–70.

Rinard, Jeremy R., and Raman C. Mahabir. 2010. "Successfully Matching into Surgical Specialties: An Analysis of National Resident Matching Program Data." *Journal of Graduate Medical Education* 2 (3): 316–21.

Rogerson, Richard, Robert Shimer, and Randall Wright. 2005. "Search-Theoretic Models of the Labor Market: A Survey." *Journal of Economic Literature* 43 (4): 959–88.

Scharfstein, David S., and Jeremy C. Stein. 1990. "Herd Behavior and Investment." *American Economic Review* 80 (3): 465–79.

Searle, Shayle R., George Casella, and Charles E. McCulloch. 1992. *Variance Components*. Wiley Series in Probability and Statistics. New York: John Wiley and Sons.

Shaneyfelt, T.M., M.F. Mayo-Smith, and J. Rothwangl. 1999. "Are Guidelines Following Guidelines? The Methodological Quality of Clinical Practice Guidelines in the Peer-Reviewed Medical Literature." *Journal of the American Medical Association* 281 (20): 1900–1905.

Skinner, Jonathan. 2011. "Chapter 2—Causes and Consequences of Regional Variations in Health Care." In *Handbook of Health Economics*, Vol. 2, edited by Mark V. Pauly, Thomas G. McGuire, and Pedro P. Barros, 45–93. Amsterdam: North-Holland.

Skinner, Jonathan, and Douglas Staiger. 2015. "Technology Diffusion and Productivity Growth in Health Care." *Review of Economics and Statistics* 97 (5): 951–64.

Smith, M., R. Saunders, L. Stuckhardt, J.M. McGinnis, Committee on the Learning Health Care System in America, and Institute of Medicine, eds. 2012. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. Washington, DC: National Academies Press.

Song, Hummy, and Robert S. Huckman. 2018. "Cohort Turnover and Operational Performance: The July Phenomenon in Teaching Hospitals." SSRN Scholarly Paper ID 3037753, Social Science Research Network, Rochester, NY January 2018.

Topel, Robert. 1991. "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority." *Journal of Political Economy* 99 (1): 145–76.

Wood, Diana F. 2003. "ABC of Learning and Teaching in Medicine: Problem Based Learning." *BMJ* 326 (7384): 328–30.

Young, John Q., Sumant R. Ranji, Robert M. Wachter, Connie M. Lee, Brian Niehaus, and Andrew D. Auerbach. 2011. "'July Effect': Impact of the Academic Year-End Changeover on Patient Outcomes: A Systematic Review." *Annals of Internal Medicine* 155 (5): 309–15.

Van Parys, Jessica, and Jonathan Skinner. 2016. "Physician Practice Style Variation: Implications for Policy." *JAMA Internal Medicine* 176 (10): 1549–50.

Van Zandt, Timothy. 1998. "Organizations with an Endogenous Number of Information Processing Agents." In *Organizations with Incomplete Information—Essays in Economic Analysis*, edited by Mukul Majumdar, 239–305. Cambridge, UK: Cambridge University Press.