Online Appendix for

Influence and Information in Team Decisions:

Evidence from Medical Residency

David C. Chan

February 2020

A-1 Random Assignment

This appendix presents two sets of randomization tests for quasi-random assignment, complementing evidence in Table 1. Section I.A presents results regarding the assignment of patients to trainees. Section I.B presents the assignment of trainees to supervising physicians.

I.A Assignment of Patients to Trainees

First, I test for the joint significance of trainee identities in regressions of this form:

$$X_a = \mathbf{T}_{t(a)}\eta + \mu_{s(a)} + \zeta_{j(a)}^{\tau<T} + \zeta_{k(a)}^{\tau>T} + \zeta_{\ell(a)} + \varepsilon_a, \quad \text{(A-1)}$$

where a is a patient admission and X_a is some patient characteristic or linear combination of patient characteristics for the patient in admission a, described in Section I.C. $t(a)$ refers to the day of admission, $s(a)$ is the service of admission, $j(a)$ is the junior trainee, $k(a)$ is the senior trainee, and $\ell(a)$ is the supervising physician. $\mathbf{T}_{t(a)}$ is a set of time categories for the admission day, including the day of the week and the month-year interaction; μ_s is a fixed effect that corresponds to the admitting service s (e.g., "heart failure service" or "oncology service"). $\zeta_i^{\tau<T}$, $\zeta_j^{\tau>T}$, and ζ_k are fixed effects for the intern i, resident j, and attending k, respectively. I do not impose any relationship between the fixed effect of a trainee as an intern and the fixed effect of the same trainee as a resident. I then test for the joint significance of the fixed effects $\left(\zeta_j^{\tau<T}, \zeta_k^{\tau>T}\right)_{j\in\mathcal{J},k\in\mathcal{K}}$.

In Column 1 of Table A-1, I show F-statistics and the corresponding p-values for the null hypothesis that $\left(\zeta_j^{\tau<T}, \zeta_k^{\tau>T}\right)_{j\in\mathcal{J},k\in\mathcal{K}} = \mathbf{0}$. I perform the regression (A-1) separately each of the following patient characteristics X_a as a dependent variable: patient age, a dummy for male gender, and a dummy for white race.[1] I also perform (A-1) using as dependent variables the linear prediction of log admission total spending based on patient age, race, and gender. I fail to find joint statistical significance for any of these tests.

Second, I test for the significance of trainee characteristics in regressions of this form:

$$X_a = \mathbf{T}_{t(a)}\eta + \mu_{s(a)} + \gamma_1 Z_{j(a)} + \gamma_2 Z_{k(a)} + \zeta_{\ell(a)} + \varepsilon_a. \quad \text{(A-2)}$$

Equation (A-2) is similar to Equation (A-1), except for the use of a vector of trainee characteristics $Z_{j(a)}$ and $Z_{k(a)}$ for the junior and senior trainee, respectively, on day of admission to test whether certain types of residents are more likely to be assigned certain types of patients. Trainee characteristics include the following: position on the rank list; USMLE Step 1 score; sex; age at the start of training; and dummies for foreign medical school, rare medical school, AOA honor society membership, PhD or another graduate degree, and racial minority.

Columns 2 and 3 of Table A-1 show F-statistics and the corresponding p-values for the null hypothesis that $(\gamma_1, \gamma_2) = \mathbf{0}$. Column 2 includes all trainee characteristics in Z_h; column 3 excludes

[1] I do not test for balance in patient diagnoses, because these are discovered and coded by physicians potentially endogenous. Including or excluding them in the baseline specification of Equation (3) does not qualitatively affect results.

position on the rank list, since this information is missing for a sizable proportion of trainees. Patient characteristics for dependent variables in (A-2) are the same as in (A-1). Again, I fail to find joint significance for any of these tests.

Third, I compare the distributions of patient age and of predicted total costs across patients admitted to interns and residents with high or low spending. I consider trainee spending effects that are fixed within junior or senior role using this regression:

$$Y_a = \mathbf{X}_a\beta + \mathbf{T}_{t(a)}\eta + \zeta_{j(a)}^{\tau<T} + \zeta_{k(a)}^{\tau>T} + \zeta_{\ell(a)} + \varepsilon_a, \tag{A-3}$$

where Y_a is log total spending for admission a, and other variables are defined similarly as in Equation (A-1). Figure A-1 shows kernel density plots of the age distributions for patients assigned to interns and residents, respectively, each of which compare trainees with practice styles above and below the mean. Figure A-2 plots the distribution of predicted spending for patients assigned to trainees with above- or below-mean spending practice styles. There is essentially no difference across the distribution of age or predicted spending for patients assigned to trainees with high or low spending practice styles. Kolmogorov-Smirnov statistics cannot reject the null that the underlying distributions are different.

I.B Assignment of Trainees to Other Providers

To test whether certain types of trainees are more likely to be assigned to certain types of other trainees and attending physicians, I perform the following regression to examine the correlation between two trainees and between a trainee and the supervising physician assigned to the same patient:

$$\hat{\zeta}_{h(a)}^{r} = \gamma_h \hat{\zeta}_{-h(a)}^{1-r} + \gamma_\ell \hat{\zeta}_{\ell(a)} + \varepsilon_a, \tag{A-4}$$

where $r \equiv \mathbf{1}(\tau > T)$ is an indicator for whether the fixed effect for trainee h was calculated while h was a junior trainee ($r = 0$) or a senior trainee ($r = 1$). As in Equation (A-1), I assume no relationship between $\hat{\zeta}_h^{\tau<T}$ and $\hat{\zeta}_h^{\tau>T}$. Each observation in Equation (A-4) corresponds to an admission a, but where error terms are clustered at the level of the intern-resident-attending team, since there are multiple observations for a given team. $\hat{\zeta}_\ell$ is the estimated fixed effect for attending k.[2] Estimates for γ_h and γ_ℓ are small, insignificant, and even slightly negative.

Second, I perform a similar exercise as in the previous subsection, in which I plot the distribution of estimated attending fixed effects working with trainees with above- or below-mean spending practice styles. In Figure A-3, the practice-style distribution for attendings is similar for those assigned to high- versus low-spending trainees. As for distributions of patient characteristics in Appendix I.A, differences in the distributions are not qualitatively significant, and Kolmogorov-Smirnov statistics

[2] I use two approaches to get around the reflection problem due to the first-stage joint estimation of ζ_j^0, ζ_k^1, and ζ_ℓ (Manski, 1993). First, I perform (A-4) using "jack-knife" estimates of fixed effects, in which I exclude observations with $-h$ and ℓ to compute the $\hat{\zeta}_h{}^r$ estimate that I use with $\hat{\zeta}_{-h}^{1-r}$ and $\hat{\zeta}_k$. Second, I use the approach by Mas and Moretti (2009), in which I include nuisance parameters in the first stage to absorb team fixed effects for (j,k,ℓ).

cannot reject the null that these distributions are different, at least when clustering at the level of the intern-resident-attending team.

A-2 Random-Effects vs. Fixed-Effects Identification

The fixed-effects estimation approach (e.g., Abowd et al., 1999; Card et al., 2013) relies on a version of Assumption 1 that is only slightly weaker:

Assumption 2 (Quasi-Random Team Assignment within Connected Sets (Abowd et al., 1999)). *Potential team decisions are independent of team assignments, conditional on clinical service $s(i,t)$, indicators of time t, and connected sets $g(i,t)$:*

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

As discussed in Abowd et al. (2008), a "connected set" g comprises cases (i,t) such that $j(i,t)\in\mathcal{J}^g$ or $k(i,t)\in\mathcal{K}^g$. $\mathcal{J}^g$ includes any junior trainee who has worked with a senior trainee in $\mathcal{K}^g$, and $\mathcal{K}^g$ includes any senior trainee who has worked with a junior trainee in $\mathcal{J}^g$. Any pair of trainees $(j,k)\in\mathcal{J}^g\times\mathcal{K}^g$, whose observations are in the same connected set, can be "connected" via a chain of trainees that have worked together.

Assumption 1 implies Assumption 2, and if $\mathcal{J}^{g(i,t)}\supseteq\mathcal{J}_{it}$ and $\mathcal{K}^{g(i,t)}\supseteq\mathcal{K}_{it}$, then Assumption 1 is equivalent to Assumption 2. Fixed-effects estimation, under Assumption 2, comes with the cost that the effects of trainees in different connected sets are not comparable: For each g, one junior-trainee effect and one senior-trainee effect need to be dropped from estimation to satisfy the rank condition. Stated differently, to identify *any* trainee effects, the fixed-effects framework requires trainee "movers," who work with more than one teammate. While our setting involves and exploits such movers, this requirement is not strictly necessary in the random-effects approach, under Assumption 1. The sense in which Assumption 2 is weaker than Assumption 1 mostly results from the rank condition and not a necessarily substantive difference in the quasi-experimental design. In finite samples, if we observed fewer cases for the same set of trainees, the sets $\mathcal{J}^{g(i,t)}$ and $\mathcal{K}^{g(i,t)}$ could contain fewer elements, even though $\mathcal{J}_{it}$ and $\mathcal{K}_{it}$ would be unchanged.

A-3 Statistical Model of Trainee Effects

III.A Patient Admission Random Effects

We may augment Equation (4) to allow for patient admission random effects, since the same patient may stay for more than one day and be exposed to different trainees:

$$\tilde{Y}_{it}=\xi^{\tau_j;\tau_k}_{j(i,t)}+\xi^{\tau_k;\tau_j}_{k(i,t)}+\nu_i+\varepsilon_{it}, \tag{A-5}$$

where ν_i is a random effect for the patient admission.[3] Under Assumption 1, $\xi_j^{\tau_j;\tau_k}$, $\xi_k^{\tau_k;\tau_j}$, and ν_i are uncorrelated with one another.

Let N_I be the number of patient admissions in sample $\mathscr{C}\left(\tau_j,\tau_k\right)$. Then in Equation (5), $\mathbf{D}$ is an $N \times (N_J + N_K + N_I)$ selection matrix for junior trainees, senior trainees, and patient admissions. $\mathbf{u}$ is an $(N_J + N_K + N_I) \times 1$ stacked vector of junior trainee, senior trainee, and patient admission random effects. We can then restate the variance-covariance matrix of $\mathbf{u}$ as

$$\operatorname{Var}\mathbf{u} = \mathbf{G} = \begin{bmatrix} \sigma^2\left(\tau_j;\tau_k\right)\mathbf{I}_{N_J} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma^2\left(\tau_k;\tau_j\right)\mathbf{I}_{N_K} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\nu^2\mathbf{I}_{N_I} \end{bmatrix}.$$

The log likelihood function in Equation (6) remains the same, with $\mathbf{V} = \mathbf{DGD}' + \sigma_\varepsilon^2\mathbf{I}_N$. I maximize this log likelihood with respect to $\sigma^2\left(\tau_j;\tau_k\right)$, $\sigma^2\left(\tau_k;\tau_j\right)$, σ_ν^2, and σ_ε^2. Estimates of $\sigma^2\left(\tau_j;\tau_k\right)$ and $\sigma^2\left(\tau_k;\tau_j\right)$ in this augmented model are qualitatively unchanged relative to the baseline implementation in Section II.D.

III.B Correlation of Trainee Effects

I augment models in (4) and (A-5) to estimate the correlation between trainee effects in two separate tenure periods, τ_1 and τ_2, which I denote by $\rho(\tau_1,\tau_2)$. Although I observe each trainee across her entire training, I only observe a subset of these trainees in each period. The number of trainees observed in both tenure periods in the pair (τ_1,τ_2) is even smaller. Because trainees that I do not observe in both τ_1 and τ_2 do not contribute to the estimate of $\rho(\tau_1,\tau_2)$, I include in the estimation sample only observations associated with a trainee observed in both tenure periods. I also redefine tenure periods to be 120 days in order to enlarge the sample of trainees whom I observe in both periods in a tenure-period pair.

Specifically, in place of Equation (4), I consider

$$\tilde{Y}_{it} = \xi_{h(i,t)}^{\tau} + \xi_{-h(i,t)} + \varepsilon_{it}, \tag{A-6}$$

where $\tau \in \{\tau_1,\tau_2\}$ may be one of two tenure periods in a pair.. This specifies that effects of trainees in the tenure periods of interest (τ_1 and τ_2) may be drawn from two separate distributions depending on the tenure period τ_1 or τ_2 corresponding to observation t; I pool the effects of the teammates into a single distribution that does not depend on tenure. Because I focus on the correlation between trainee effects, I am unconcerned about the scale of practice variation and I thus do not specify the tenure of the teammate. The analog for Equation (A-5) is

$$\tilde{Y}_{it} = \xi_{h(i,t)}^{\tau} + \xi_{-h(i,t)} + \nu_i + \varepsilon_{it}. \tag{A-7}$$

[3]This specification requires the use of sparse matrices for estimation. In specifications without the use of sparse matrices, I nest this effect within interns, i.e., I include ν_{ai} as an intern-admission effect. While it is easier to estimate a specification with ν_{ai}, I will describe this specification for ease of explication. In practice, results are materially unaffected by whether I use ν_a or ν_{ai}, or in fact whether I include an admission-related effect at all.

I estimate (A-6) or (A-7) in a sample of observations, which I define as follows: $\mathscr{C}(\tau_1,\tau_2) = \{(i,t,h) : h \in \{j(i,t), k(i,t)\}, \tau(h,t) \in \{\tau_1,\tau_2\}\}$. I require that, for every trainee h in $\mathscr{C}(\tau_1,\tau_2)$, there are observations in the sample in which she has tenure τ_1 and other observations in the sample in which she has tenure τ_2. Otherwise, we cannot use trainee h to estimate the correlation in trainee effects between these two periods.

As above, I can represent both Equation (A-6) and Equation (A-7) in matrix form, as Equation (5). Denote the number of trainees h in $\mathscr{C}(\tau_1,\tau_2)$ as N_H. Denote the number of teammates trainees interacted with their tenure periods as N_H^-. The selection matrix $\mathbf{Z}$ is of size $N \times \left(2N_H + N_H^-\right)$, since it now maps observations onto one of two random effects, depending on whether $\tau = \tau_1$ or $\tau = \tau_2$, for each trainee h observed in both τ_1 and τ_2 tenure periods. The stacked vector of random effects $\mathbf{u}$ is similarly of size $\left(2N_\tau + N_\tau^-\right) \times 1$. The variance-covariance matrix of $\mathbf{u}$ is

$$\operatorname{Var}\mathbf{u} = \mathbf{G} = \begin{bmatrix} \mathbf{G}_H & \mathbf{0} \\ \mathbf{0} & \sigma_{\xi^-}^2 \mathbf{I}_{N_H^-} \end{bmatrix},$$

where $\mathbf{G}_H$ is a $2N_H \times 2N_H$ block-diagonal matrix of the form

$$\mathbf{G}_H = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{A} \end{bmatrix}, \tag{A-8}$$

with each block being the 2×2 variance-covariance matrix $\mathbf{A}$ of random effects within trainee and across tenure periods:

$$\begin{aligned} \operatorname{Var}\begin{bmatrix} \xi_h^{\tau_1} \\ \xi_h^{\tau_2} \end{bmatrix} &= \mathbf{A}, \text{ for all } h, \text{ where} \\ \mathbf{A} &\equiv \begin{bmatrix} \sigma^2(\tau_1) & \rho(\tau_1,\tau_2)\sigma(\tau_1)\sigma(\tau_2) \\ \rho(\tau_1,\tau_2)\sigma(\tau_1)\sigma(\tau_2) & \sigma^2(\tau_2) \end{bmatrix}. \end{aligned}$$

Representing (A-7) as (5) is a similar exercise. The selection matrix $\mathbf{Z}$ is of size $N \times \left(2N_H + N_H^- + N_I\right)$, and the vector of random effects $\mathbf{u}$ is of size $\left(2N_H + N_H^- + N_I\right) \times 1$. The variance-covariance matrix of $\mathbf{u}$ is

$$\operatorname{Var}\mathbf{u} = \mathbf{G} = \begin{bmatrix} \mathbf{G}_H & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_{\xi^-}^2 \mathbf{I}_{N_H^-} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\nu^2 \mathbf{I}_{N_I} \end{bmatrix},$$

where $\mathbf{G}_H$ is the same as in Equation (A-8). The log likelihood is the same as in Equation (6), but using revised definitions of $\mathbf{G}$ that allow for covariance between random effects of the same trainees across tenure periods. The correlation parameter of interest $\rho(\tau_1,\tau_2)$ is constrained to be between -1 and 1.

A-4 Intrinsic Heterogeneity: Trainee Characteristics

The key alternative explanation for persistent variation that I explore in this section is that physicians may intrinsically differ for reasons unrelated to knowledge and learning, such as preferences or ability (e.g., Doyle et al., 2010; Fox and Smeets, 2011; Bartel et al., 2014). To assess the possibility of intrinsic heterogeneity, I first exploit detailed trainee characteristics that should be highly correlated with preferences and ability. For example, USMLE scores measure medical knowledge as a medical student; position on the residency rank lists reflects overall desirability; and specialty tracks, mostly predetermined relative to the beginning of residency, reflect important career decisions and lifestyle preferences, such as a decision to become a radiologist rather than a primary care physician. To capture the variety of future career paths across internal medicine trainees, I impute future yearly incomes after specialty training based on the final specialty choices of trainees. As cited in Section I.C, trainees with above-median future incomes will earn substantially more than their peers with below-median future incomes.

I assess the relationship between each of these characteristics and daily spending totals for either the junior or senior trainee:

$$Y_{it} = \alpha_m \text{Characteristic}^m_{h(i,t)} + \mathbf{X}_i\beta + \mathbf{T}_t\eta + \zeta_{-h(i,t)} + \zeta_{\ell(i,t)} + \varepsilon_{aijkt}, \tag{A-9}$$

where $\text{Characteristic}^m_h$ is an indicator for whether the junior (or senior) trainee h has the characteristic m, ζ_{-h} is a fixed effect for the other senior (or junior) trainee $-h$, and ζ_ℓ is a fixed effect for attending ℓ.[4] The coefficient of interest, α_m, quantifies the predictive effect of a trainee with characteristic m on patient spending decisions. I also evaluate the combined predictive effect of trainee characteristics in two steps. First, I regress outcomes on all direct trainee characteristics, with continuous characteristics like position on rank list entered linearly, along with the other admission and time regressors in Equation (A-9):

$$Y_{it} = \sum_m \alpha_m \text{Characteristic}^m_{h(i,t)} + \mathbf{X}_i\beta + \mathbf{T}_t\eta + \zeta_{-h(i,t)} + \zeta_{\ell(i,t)} + \varepsilon_{it}. \tag{A-10}$$

This yields a predicted score Z_h for each trainee h, $Z_h = \sum_m \hat{\alpha}_m \text{Characteristic}^m_h$, which I normalize to $\tilde{Z}_h = Z_h/\sqrt{\text{Var}(Z_h)}$ with standard deviation 1. Second, I regress daily total spending on this normalized score:

$$Y_{it} = \alpha\tilde{Z}_{h(i,t)} + \mathbf{X}_i\beta + \mathbf{T}_t\eta + \zeta_{-h(i,t)} + \zeta_{\ell(i,t)} + \varepsilon_{it}. \tag{A-11}$$

In addition, I evaluate the predictive power of trainee characteristics more flexibly by allowing splines of continuous characteristics and two-way interactions between characteristics, while assuming an "approximately sparse" model and using LASSO to select for significant characteristics (e.g., Belloni et al., 2014). This approach guards against overfitting in finite data when the number of po-

[4] In principle, I could include trainee characteristics as mean shifters in the baseline random effects model in Equation (3). However, since characteristics are generally insignificant predictors of variation, results of (residual) variation attributable to trainees are unchanged.

tential characteristics becomes large. In total, excluding collinear characteristics, I consider 36 and 32 direct characteristics for interns and residents, respectively, and 285 and 308 two-way interactions, as potential regressors in Equation (A-9).

Table 4 shows results for Equation (A-11) and a subset of results for Equation (A-9). Considering characteristics individually in Equation (A-9), only two characteristics (gender and high USMLE test score) are statistically significant at the 5% level, and no characteristic approaches the one-standard deviation benchmark effect in the trainee effect distribution. Likewise, a standard-deviation change in the overall predictive score has no economically significant effect on spending for either interns or residents. LASSO selected no intern characteristic as significant and selected only resident gender as significant. Although it is possible that there are other unmeasured and orthogonal characteristics that are more relevant for practice variation, this seems *a priori* unlikely given that these are the characteristics on which the residency program bases acceptance decisions,[5] and that they are also highly predictive of future career paths and incomes.

Finally, I investigate the *distribution* of trainee effects as a function of tenure for trainees in different groups. As shown in Figure 6, the distributions of trainee effects throughout training are not meaningfully different between groups of trainees separated by their test scores, rank list positions, or future earnings. This finding implies that trainees who differ significantly along meaningful dimensions still practice similarly not only on average, but also in terms of variation over time. That is, trainees evaluated with higher test scores, more desirable rankings, or higher future earnings do not exhibit lower variation or higher convergence over training.

A-5 Learning by Osmosis: Predictable Learning

Finally, I assess whether trainee practice styles can be predicted by the sequence of observable learning experiences. This evaluation tests two concepts. First, practice styles may predictably change if they reflect acquired skill that may grow with greater experience. Second, trainees may absorb spending patterns from supervising physicians or from a broader practice environment.[6]

To explore the potential effect of learning from others in greater detail, I estimate supervising physician "effects" by shrinking their observed fixed effects, and I similarly calculate best linear unbiased predictions (BLUPs) of senior trainee effects. The standard deviation of shrunken supervising physician effects is 7.3%, and the standard deviation of the senior trainee BLUPs is 16.6% in terms of overall spending. I then form measures of prior exposure to spending due to supervising physicians by averaging spending effects of supervising physicians who have previously worked with a given trainee, weighted by patient-days, at a given point in time. This exposure measure may or may not be

[5] Using the same characteristics to predict whether a trainee was ranked in the upper half on the residency program's rank list (excluding rank as a characteristic) yields a predictive score that with one standard deviation changes the probability of being highly ranked by about 20%.

[6] The related concept of "schools of thought," in which physicians may have systematically different training experiences, has been proposed as a mechanism for geographic variation (e.g., Phelps and Mooney, 1993). This hypothesis is not inconsistent with tacit knowledge and in fact relies partly on it, but it does not by itself explain large variation within the same training program.

restricted to patient-days on the same ward service (e.g., cardiology, oncology, or general medicine). Similarly, the measure may be calculated for all prior patient-days or only for patient-days in the last three months. I also calculate similar measures of exposure to senior trainees for trainees based on their previous team matches when they were junior.

For a given prior exposure measure, I define trainees with above-median measures in a given tenure period as having "high exposure" to spending and trainees with below-median measures as having "low exposure" to spending. Compared to other trainees with the same tenure, these trainees have worked with attending physicians or residents trainees (while they were interns) with higher average spending effects. Table A-5 shows the difference between high-exposure and low-exposure trainees for various spending-exposure measures at different trainee tenure periods. Differences between high and low exposure to supervising-physician spending range from 1.9% to 6.7%. Differences between high and low exposure to senior-trainee spending range from 17.5% to 23.4%.

I then estimate the effect of high exposure to spending over each tenure period of training with a regression of the form

$$\begin{aligned} Y_{it} \quad = \quad & \sum_{\tau:\tau<1} \alpha_\tau \mathbf{1}\left(\tau\left(j\left(i,t\right),t\right)=\tau\right) \cdot \text{HighSpendingExposure}^{m}_{j(i,t),t} + \\ & \sum_{\tau:\tau\geq 1} \alpha_\tau \mathbf{1}\left(\tau\left(k\left(i,t\right),t\right)=\tau\right) \cdot \text{HighSpendingExposure}^{m}_{k(i,t),t} + \\ & \mathbf{X}_i\beta + \mathbf{T}_t\eta + \zeta_{\ell(i,t)} + \varepsilon_{it}, \end{aligned} \qquad \text{(A-12)}$$

where, as in Equation (3), $j(i,t)$ is the junior trainee, $k(i,t)$ is the senior trainee, and $\tau(j(i,t),t)$ and $\tau(k(i,t),t)$ are the relevant tenure periods of the junior and senior trainees at t. The variables $\text{HighSpendingExposure}^{m}_{j,t}$ and $\text{HighSpendingExposure}^{m}_{k,t}$ are indicators for high exposure to spending under measure m for the junior and senior trainee, respectively. The effect of this exposure can vary by τ. Figure A-6 shows results for exposure to spending by supervising physicians, and Figure A-7 shows similar results for exposure to spending by senior trainees. Results among the wide range of exposure measures are broadly insignificant.

More broadly, I also consider several measures of prior experience—including days on ward service, patients seen, and supervising physicians for a given trainee prior to a patient encounter—for either the junior or senior trainee. For each of these experience measures, I estimate a regression of the form

$$Y_{it} = \alpha_m \text{Experience}^{m}_{h(i,t),t} + \mathbf{X}_i\beta + \mathbf{T}_t\eta + \zeta_{-h(i,t)} + \zeta_{\ell(i,t)} + \varepsilon_{it}, \qquad \text{(A-13)}$$

where $\text{Experience}^{m}_{h,t}$ is an indicator for whether trainee h at time t has experienced a measure (e.g., number of days on service, average supervising physician spending effect) above median for the relevant tenure period, where both the measure and the median are calculated using observations prior to the relevant tenure period. In my baseline specification, I control for the other trainee and supervising physician identities, although this does not qualitatively affect results. Results are shown in Table A-6 and are broadly insignificant. A LASSO implementation that jointly considers a larger number of summary experience measures in early or more recent months relative to the patient encounter, as well as two-way interactions between these measures (112 and 288 variables for interns and residents, respectively), also fails to select any measure as significant.

In addition to trainees in the main residency program, I observe visiting trainees based in a hospital with 20% lower Medicare spending according to the Dartmouth Atlas. I evaluate the effect of these trainees on teams, as interns and as residents, using Equation (A-9). This effect includes both differences in selection (i.e., intrinsic heterogeneity) into the different program and in training experiences across the programs. Table 4 shows that visiting trainees do not have significantly different spending effects, either as interns or as residents.[7]

Overall, these results indicate that summary measures of trainee experience are poor predictors of practice and outcomes, especially relative to the large variation across trainees. The results fail to support "learning by osmosis" as a major source of practice variation, at least within an organization with *ex ante* uniform training experiences but nonetheless large practice variation.

A-6 Model of Information Aggregation and Experiential Learning

VI.A Setup

Each decision d can be summarized perfectly by an unknown parameter θ_d. If θ_d were known, then the optimal action would be $a_d = \theta_d$. Each agent has only *partial* knowledge about the correct action, in the form of a Bayesian prior about θ_d. A team decision is made as follows:

1. Each agent $h \in \{j, k\}$ has prior knowledge bearing on the decision; specifically, a Bayesian prior distribution, $\theta_{d,h}$. $\theta_{d,h}$ is a normal distribution and can be summarized by mean $\mu_{d,h}$ and precision $\rho_{d,h}$. One may describe $\mu_{d,h}$ as the *judgment* (due to prior knowledge) that agent h has about d.

2. There may also be external information about d. Some of this knowledge is held by the attending physician, but other sources derive from hospital nurses, consultants, and protocols. Each agent may also collect information about the decision, which I assume to be independent of prior knowledge. I consider external information as a public judgment with mean 0 and precision P_d^*.

3. The team takes an action and derives utility $u = -(\theta_d - a_d)^2$. As in the standard team-theoretic environment, there is no conflict of interest between agents.

Proposition A-1. *The optimal (Bayesian) action for decision d assigned to trainees j and k is*

$$a_d^* = \frac{\rho_{d,j}\mu_{d,j} + \rho_{d,k}\mu_{d,k}}{\rho_{d,j} + \rho_{d,k} + P_d^*}. \tag{A-14}$$

[7] This result of course does not rule out that training programs can matter. Doyle et al. (2010) studies the effect of trainee teams from two different programs and find that trainees from the higher-prestige program spend less. However, this result does suggest that even when trainees come from significantly different hospitals, differences in their mean practice styles can be dwarfed by variation within training program.

This expression aggregates information as a weighted average of judgments in proportion to the precisions of the respective judgments (DeGroot, 2005). Supervisory information, measured by precision P_d^*, reduces the effect of either trainee's judgment on a_d^*.

The weights on judgments in the Bayesian action in Equation (A-14),

$$g_{d,h;-h}^* \equiv \frac{\rho_{d,h}}{\rho_{d,h} + \rho_{d,-h} + P_d^*},$$

have a natural interpretation as the *influence* of trainee h on the action a_d^*. The more precise the signal from her prior knowledge relative to her teammate and any supervisory information, the greater her influence will be. In the limit, if either her teammate's knowledge or external information is perfect (i.e., $\rho_{d,-h} = \infty$ or $P_d^* = \infty$), a trainee would have no influence. On the other hand, if a trainee has perfect knowledge, then she would have full influence. At the one-year tenure mark, influence discontinuously increases because the precision of a trainee's teammate $\rho_{d,-h}$ discontinuously decreases.

Influence may deviate from the Bayesian benchmark due to other team concerns. Career concerns or the "prestige" of senior titles may underweight the knowledge of junior trainees (Scharfstein and Stein, 1990; Prendergast, 1993; Ottaviani and Sorensen, 2001), or trainees may be given more influence than justified by their knowledge if supervisors wish to encourage experiential learning that requires a stake in decision-making (Lizzeri and Siniscalchi, 2008; Ludmerer, 2014). In estimation, I allow for actions that deviate from the Bayesian benchark:

$$\hat{a}_d = \frac{\tilde{\rho}_{d,j}\mu_{d,j} + \tilde{\rho}_{d,k}\mu_{d,k}}{\tilde{\rho}_{d,j} + \tilde{\rho}_{d,k} + P_d}. \tag{A-15}$$

$\tilde{\rho}_{d,h} = \rho_{d,h} + \delta(\tau_h)$ as an effective "precision" that equals the true precision of h's knowledge adjusted by $\delta(\tau_h)$, depending on the tenure of h, τ_h. The influence of trainees with tenure τ_h relative to their peers may receive less influence than the Bayesian benchmark if $\delta(\tau_h) < 0$ or more influence if $\delta(\tau_h) > 0$. Similarly, for external and supervisory information, P_d is an effective "precision": Even though supervising physicians and the broader supervisory structure may have access to information relevant for d with precision P_d^*, this information may be underweighted ($P_d < P_d^*$) or overweighted ($P_d > P_d^*$) in decision-making.

I consider the precision of knowledge as a function of tenure for given class of decisions, c: $\rho_{d,h} = \rho_{c(d)}(\tau_h(t(d)))$. I similarly specify external information as depending on the class of decisions: $P_d = P_{c(d)}$.[8] Effective influence of a trainee with tenure τ_h working with teammate with tenure τ_{-h} is

$$g_c(\tau_h;\tau_{-h}) = \frac{\tilde{\rho}_c(\tau_h)}{\tilde{\rho}_c(\tau_h) + \tilde{\rho}_c(\tau_{-h}) + P_c}. \tag{A-16}$$

Model-predicted practice variation (i.e., standard deviation of trainee effects) for trainees with

[8]In Appendix Figure A-5, I support for this assumption by showing that both the trainee-related variation and the residual variation in spending are relatively constant across July, when old interns transition to residents and new interns begin training.

tenure τ_h, working with teammates with tenure τ_{-h}, is then

$$\sigma_c(\tau_h, \tau_{-h}) = g_c(\tau_h; \tau_{-h}) \sqrt{\kappa_c / \rho_c(\tau_h)}, \tag{A-17}$$

where $\kappa_c \in [0,1]$ reflects the similarity of judgments across different decisions in class c within the same provider. Systematic practice variation across trainees, requires that $\kappa_c > 0$, or that trainees practice similarly across different decisions. While levels of knowledge, learning, and practice variation are scaled by κ_c, ratios comparing different points in training will be unaffected by κ_c.

VI.B Identification

As trainees learn, the precision of their knowledge, or $\rho_c(\tau_h)$, increases with tenure. Greater knowledge increases influence, or $g_c(\tau_h; \tau_{-h})$, holding teammates and external information fixed, while it reduces dispersion in judgments, or $\sqrt{1/\rho_c(\tau_h)}$. Thus practice variation may not always decrease even as trainees learn. In general, the effect of increasing influence on practice variation will tend to dominate when a trainee's influence is relatively low, while when a trainee has relatively high influence, the effect of reducing dispersion in judgments will tend to dominate. In the extreme, agents who practice independently (i.e., they have full influence over their decisions) will show convergence in their decisions as they learn.

VI.B.1 Analytical Evaluation

Consider practice variation—or the standard deviation of trainee effects—under Bayesian-benchmark influence:

$$\begin{aligned} \sigma(\tau_h, \tau_{-h}) &= \frac{g^*(\tau_h; \tau_{-h})}{\sqrt{\rho(\tau_h)}} \\ &= \frac{\sqrt{\rho(\tau_h)}}{\rho(\tau_h) + \rho(\tau_{-h}) + P}, \end{aligned} \tag{A-18}$$

where I assume that $\kappa = 1$ in (A-17) without loss of generality.

As a first observation, note that the discontinuity in practice variation is greater across the one-year tenure mark than it is across the two-year tenure mark.

Proposition A-2. *Define* $\sigma(1^-) \equiv \lim_{\tau \to 1^-} E_{-h}[\sigma(\tau_h, \tau_{-h}) | \tau_h]$, *and* $\sigma(1^+) \equiv \lim_{\tau \to 1^+} E_{-h}[\sigma(\tau_h, \tau_{-h}) | \tau_h]$; *similarly define* $\sigma(2^-) \equiv \lim_{\tau \to 2^-} E_{-h}[\sigma(\tau_h, \tau_{-h}) | \tau_h]$, *and* $\sigma(T^+) \equiv \lim_{\tau \to 2^+} E_{-h}[\sigma(\tau_h, \tau_{-h}) | \tau_h]$. *Then*

$$\frac{\sigma(1^+)}{\sigma(1^-)} > \frac{\sigma(2^+)}{\sigma(2^-)} > 1.$$

Proof. Assume that interns work with second-year residents in λ proportion of the time and work with third-year residents in the remaining $1 - \lambda$ proportion of the time. At the first-year discontinuity,

$$\frac{\sigma(1^+)}{\sigma(1^-)} = \frac{\rho(1) + \lambda\rho(2) + (1-\lambda)\rho(3) + P}{\rho(1) + \rho(0) + P}.$$

At the second-year discontinuity,

$$\frac{\sigma(2^+)}{\sigma(2^-)} = \frac{\rho(2)+\rho(1)+P}{\rho(2)+\rho(0)+P}.$$

Since $\rho(\cdot)$ is increasing in τ, $\rho(0) \leq \rho(1) \leq \rho(2) \leq \rho(3)$, which yields our result. □

Because there is a change in the tenure of the other trainees as new interns arrive at the beginning of each academic year, there is in principle a discontinuous increase in influence (and therefore practice variation) at the beginning of each year. However, the increase at $\tau_h = 1$ is always larger than the increase at $\tau_h = 2$ for two reasons, both related to the monotonic increase in precision with tenure: First, trainees at $\tau_h = 1$ have less precise subjective priors than those at $\tau_h = 2$, so any decrease in the relative tenure of their peer trainee increases their influence by more. Second, the decrease in the relative tenure of the peer is greater at $\tau_h = 1$ (from $\tau_{-h} = 2$ to $\tau_{-h} = 0$) than at $\tau_h = 2$ (from $\tau_{-h} = 1$ to $\tau_{-h} = 0$). I show below in the numerical examples that, within this framework, this difference in the discontinuous increases at $\tau_h = 1$ and at $\tau_h = 2$ can be quite large, and that the discontinuity at $\tau_h = 2$ can be quite trivial.

Second, I consider whether practice variation is likely to increase or decrease with tenure. Since trainees and their teammates gain tenure together, I consider $\tau_{-h} = \tau_h + \Delta$, where Δ is fixed in a continuous portion of practice variation (i.e., not at the one- or two-year discontinuities). Applying the quotient rule to $\sigma(\tau_h, \tau_{-h}) = \sigma(\tau_h, \tau_h + \Delta)$,

$$\begin{aligned}\sigma'(\tau_h) &\equiv \frac{\partial \sigma(\tau_h, \tau_h+\Delta)}{\delta \tau_h} \\ &= \frac{\frac{1}{2}\rho(\tau_h)^{-1/2}\rho'(\tau_h)(\rho(\tau_h)+\rho(\tau_{-h})+P) - \rho(\tau)^{1/2}(\rho'(\tau)+\rho'(\tau_{-h}))}{(\rho(\tau)+\rho(\tau_{-h})+P)^2}.\end{aligned}$$

Focusing on the numerator to determine the sign of $\sigma'(\tau)$, I arrive at the following necessary and sufficient condition for convergence (i.e., decreasing practice variation with tenure, or $\sigma'(\tau_h) < 0$):

Proposition A-3. *Practice variation decreases if and only if*

$$\frac{\rho'(\tau_h)}{\rho'(\tau_h)+\rho'(\tau_{-h})} < 2g^*(\tau_h; \tau_{-h}). \qquad \text{(A-19)}$$

Learning (i.e., $\rho'(\tau_h) > 0$) does not guarantee convergence. Instead, convergence requires that the "share of learning," defined as $\rho'(\tau_h)/(\rho'(\tau_h)+\rho'(\tau_{-h}))$, is smaller than twice the influence. Since this "share" is always less than 1, convergence is guaranteed whenever the trainee has full influence, or $g^*(\tau_h; \tau_{-h}) = 1$, as is the case in a single decision-maker. The larger the trainee's influence, the more likely convergence will occur. Since influence grows with tenure, this also implies that practice variation generally increases and then decreases. Special cases may involve practice variation only increasing or only decreasing, but not decreasing and then increasing with tenure.

VI.B.2 Numerical Examples

Figure A-8 presents a few numerical examples of variation profiles under various learning profiles described by functions of the piecewise linear form in Equation (A-20). The three parameters of interest are ρ_0, or initial knowledge; ρ_1, or the rate of increase in the precision during the first year as a junior trainee; and $\rho_2 = \rho_3$, or the rate of increase during the subsequent two years as a senior trainee. The precision of judgments at the end of training is $\rho(3) = \rho_0 + \rho_1 + 2\rho_2$. I also normalize $P = 1$, so that whether precisions of beliefs are greater than the precision of the supervisory prior simply depends on whether they are greater or less than 1. I consider this normalization as only relevant for the scale of the variation profile, since any scale keeping the same shape over the overall variation profile $\sigma(\tau)$ can be implemented by multiplying ρ_0, ρ_1, ρ_2, and P by some constant.

I discuss each panel of Figure A-8 in turn:

- Panel A considers equal $\rho_0 = \rho_1 = \rho_2 = 0.2$, which are relatively small compared to $P = 1$. The result is broadly non-convergence, as greater experience primarily results in greater influence against a relatively strong supervisory practice environment. The discontinuity in variation is significantly larger at $\tau = 1$ than at $\tau = 2$. Variation increases in intern year and decreases but only slightly in the next two years as resident.

- Panel B imposes no resident learning ($\rho_2 = 0$) and presents the limiting case in which discontinuous increases in variation at $\tau = 1$ and $\tau = 2$ are the same. Variation is still at least as big during the two years as resident as during the year as intern, driven by influence. Variation seems relatively constant over training.

- Panel C generates a similar variation profile as in Panel B with a non-zero ρ_2 by increasing the ratios of ρ_0 and ρ_1 to ρ_2. The scale of variation is smaller than in Panel B, which reflects that precision in trainee beliefs are now larger. A rescaled version with smaller precisions (and smaller P) would reveal larger relative increases in variation at the discontinuities.

- Panel D examines increasing ρ_1 relative to ρ_0, so that more learning occurs in the first year of training compared with knowledge possessed before starting training. Influence more obviously increases in the first year, and increases in variation are sharper at the discontinuities, since intern experience matters more. Note that working with a resident is equivalent to working with an end-of-year intern, and increases in variation at $\tau = 1$ and $\tau = 2$ are the same (as in Panel B).

- Panel E asserts that most of the learning occurs during the role as resident. There is much greater variation across residents than across interns, and the discontinuous increase in variation is much larger at $\tau = 1$, while the increase is negligible at $\tau = 2$. There is significant convergence during the two years as resident.

- Panel F is similar to panel E but shows less convergence during role as resident. The ratio of learning as intern to learning as resident (ρ_1/ρ_2) is similar, but learning during training is reduced relative to knowledge from prior to training (ρ_0) and to supervisory information (P).

VI.C Specification and Estimation

I specify the precision of knowledge as a piecewise-linear function of trainee tenure:

$$\rho_c(\tau) = \begin{cases} \rho_{0,c} + \rho_{1,c}\tau, & \tau \in [0,1]; \\ \rho_{0,c} + \rho_{1,c} + \rho_{2,c}(\tau - 1), & \tau \in [1,2]; \\ \rho_{0,c} + \rho_{1,c} + \rho_{2,c} + \rho_{3,c}(\tau - 2), & \tau \in [2,3], \end{cases} \tag{A-20}$$

where $\rho_{0,c}$ represents the precision of knowledge before starting residency, and $\rho_{1,c}$, $\rho_{2,c}$, and $\rho_{3,c}$ are the yearly rate of learning in the first, second, and third years of residency, respectively, for decisions in class c.

Assuming that knowledge is continuous with tenure, I also identify deviations from efficient influence that come from a step function with respect to years of training. That is, the "effective" trainee precision relevant for influence is

$$\tilde{\rho}_c(\tau) = \rho_c(\tau) + \delta_{1,c}\mathbf{1}(\tau \geq 1) + \delta_{2,c}\mathbf{1}(\tau \geq 2). \tag{A-21}$$

$\delta_{1,c}$ and $\delta_{2,c}$ represent deviations in influence from the efficient benchmark that may result from titles (e.g., "senior trainee") that discontinuously change at years of training, $\lfloor \tau \rfloor$.[9] Finally, I identify deviations from the Bayesian benchmark from the fact that $P_c^* \geq \rho_c(\tau = 3)$: At a minimum, external information must be greater than the knowledge held by a senior trainee, since all supervising physicians have completed training, and since supervisory information includes informational inputs from outside staff (e.g., nursing, consultants), or any information gathered by the trainees themselves.[10] $P_c < \rho_c(3)$ would strongly imply that trainees are granted *more* influence than warranted by their knowledge.

I estimate learning and influence parameters as a two-step process. The first step recovers moments of practice variation, specifically the standard deviation of the distribution of trainee effects, for trainees of tenure τ_h working with teammates of tenure τ_{-h}. These empirical moments, $\hat{\sigma}(\tau_h, \tau_{-h})$, are estimated from the random effects model in Equation (3) and were previously discussed in Section II. The second step takes these moments of practice variation and, from the model in Section IV.C, recovers underlying primitives of knowledge and influence using minimum distance estimation.

For each class of decisions c, I estimate model primitives $\boldsymbol{\theta}_c = \left(\rho_{0,c}, \rho_{1,c}, \rho_{2,c}, \rho_{3,c}, \delta_{1,c}, \delta_{2,c}, P_c\right)$ by minimum distance:

$$\hat{\boldsymbol{\theta}}_c = \arg\min_{\boldsymbol{\theta}_c \in \Theta} \left(\hat{\boldsymbol{\sigma}}_c - \boldsymbol{\sigma}(\boldsymbol{\theta}_c)\right)' \mathbf{W} \left(\hat{\boldsymbol{\sigma}}_c - \boldsymbol{\sigma}(\boldsymbol{\theta}_c)\right),$$

[9] Common title conventions may refer to trainees by their year of training: PGY1, PGY2, and PGY3 use the acronym "PGY" for "post-graduate year"; R1, R2, and R3 simply use "R" for "resident."

[10] While I consider the distribution of this "supervisory" information as having mean 0 in the simple model, this assumption is inconsequential, as it is by definition orthogonal to trainee knowledge. The "judgment" of the supervisory information can be viewed as captured by all terms other than the trainee effects in the regression Equation (3), including the error term.

where $\hat{\boldsymbol{\sigma}}_c$ is the vector of empirical estimates of practice variation corresponding to decisions in class c from the first step, with elements corresponding to $(\tau_h, \tau_{-h}) \in \mathcal{T}$; $\boldsymbol{\sigma}(\boldsymbol{\theta}_c)$ is the corresponding vector of model-implied practice variation from Equation (A-17) given $\boldsymbol{\theta}_c$; and $\mathbf{W}$ is a weighting matrix. Primitives may also be estimated on overall practice variation moments, in which case I omit labels of c.

Consistent with previous reduced-form estimation, I fit the model on $\|\mathcal{T}\| = 18$ moments of practice variation: I divide observations with residents in the second year of training into resident tenure blocks of 60 days, resulting in 6 resident moments and 6 intern moments of practice variation; I also divide observations with residents in the third year of training into resident tenure blocks of 120 days, resulting in 3 resident moments and 3 intern moments of practice variation. If $\sqrt{n}(\hat{\boldsymbol{\sigma}}_c - \boldsymbol{\sigma}(\boldsymbol{\theta}_c)) \xrightarrow{d} N(\mathbf{0}, \Omega_c)$, then the asymtotic variance of $\hat{\boldsymbol{\theta}}_c$ is given by

$$\text{Asy. Var}\, \hat{\boldsymbol{\theta}}_c = \frac{1}{n}\left(\Gamma(\boldsymbol{\theta}_{0,c})' \mathbf{W} \Gamma(\boldsymbol{\theta}_{0,c})\right)^{-1} \left(\Gamma(\boldsymbol{\theta}_{0,c})' \mathbf{W} \Omega_c \mathbf{W} \Gamma(\boldsymbol{\theta}_{0,c})\right) \left(\Gamma(\boldsymbol{\theta}_{0,c})' \mathbf{W} \Gamma(\boldsymbol{\theta}_{0,c})\right)^{-1},$$

where $\boldsymbol{\theta}_{0,c}$ is the true parameter vector, and $\Gamma(\boldsymbol{\theta}_{0,c}) = \text{plim}\, \partial \boldsymbol{\sigma}\left(\hat{\boldsymbol{\theta}}_c\right) / \partial \hat{\boldsymbol{\theta}}_c$ is an 18×7 matrix of analytical derivatives of Equation (A-17) with respect to $\boldsymbol{\theta}_c$, evaluated at $\hat{\boldsymbol{\theta}}_c$. The optimal weighting matrix is $\mathbf{W} = \hat{\Omega}_c^{-1}$, which I obtain from the first-step estimation of practice variation. This yields for inference

$$\widehat{\text{Var}}\, \hat{\boldsymbol{\theta}}_c = \frac{1}{n}\left(\Gamma\left(\hat{\boldsymbol{\theta}}_c\right)' \hat{\Omega}_c^{-1} \Gamma\left(\hat{\boldsymbol{\theta}}_c\right)\right)^{-1}.$$

I also calculate likelihood ratio tests for the joint-significance of learning and influence parameters against a restricted model with no learning but potentially inefficient senior influence via "status" (i.e., only $\rho_{0,c}$, $\delta_{1,c}$, and P_c are non-zero).

VI.D Results

In Table A-7, Column 1, I show baseline parameter estimates based on practice variation in overall spending. In Figure A-9, I show the implied path of practice variation according to the model and estimated parameters, overlaid on reduced-form estimates from Section II. Structural estimates imply very little knowledge at the beginning of residency ($\rho_0 = 0.04$) compared to learning in the first year ($\rho_1 = 0.20$). Learning in the second year occurs at a rate 30 times faster than in the first year ($\rho_2 = 7.5$), but appears to cease by the third year ($\rho_3 = 0$). Between junior and senior trainees, influence approximates the Bayesian benchmark.[11] However, I find that the contribution of external information ($P = 3.7$) is much lower than the knowledge of a graduating trainee ($\underline{P} \equiv \rho(3) \approx 7.74$). Since external information includes knowledge of supervising physicians who have completed training, this suggests that trainees are given much more influence than under the Bayesian benchmark.

[11] I estimate that $\delta_1 = 0.23$. Although this deviation from the Bayesian benchmark for senior trainees is large relative to knowledge at the end of the first year ($\rho_0 + \rho_1 = 0.24$), it is relatively small compared to learning that occurs in the second year ($\delta_1 / \rho_2 \cdot 365$ days = 11 days worth of second-year learning). I also estimate that $\delta_2 = -1.4$, which implies that third-year trainees have *less* influence than under the Bayesian benchmark, although this parameter is imprecisely estimated and small relative to ρ_2.

I also estimate model parameters based on practice variation in spending specific to classes of decisions (Table A-7) and by types of patient-days (Table A-8). Learning is often greatest in the second year of training, regardless of the set of decisions. Decisions broken into components of diagnostic testing, prescriptions, blood transfusions, and nursing orders show somewhat less pronounced learning in the second year, which suggests potential interactions between components that are important for learning.

Based on likelihood ratio tests comparing the baseline model and more restrictive models, I can reject a model with no learning (i.e., $\rho_1 = \rho_2 = \rho_3 = 0$) and only senior prestige (i.e., $\delta_1 > 0$) for overall spending decisions (Column 1 of Table A-7) and for the majority of other outcomes or subsets of the data (Tables A-7 and A-8). On the other hand, if I allow for learning but impose the Bayesian benchmark influence between trainees (i.e., $\delta_1 = \delta_2 = 0$), the restricted model (Panel B of Figure A-10) fits the data quite well and cannot be rejected by the likelihood ratio test. Finally, I can strongly reject a model with strictly Bayesian influence between trainees and supervisors (i.e., $\delta_1 = \delta_2 = 0$, $P \geq \rho_0 + \rho_1 + \rho_2 + \rho_3$); the graphical fit of this model (Panel C of Figure A-10) is obviously problematic.

VI.E Counterfactual Analyses

VI.E.1 Model of Learning

In my baseline results, I find that learning is low as a junior trainee in the first year, high as a senior trainee in the second year, and null in the third year. I interpret the first switch in the rate of learning—from low learning in the first year to high in the second—as due to the effect of influence on learning. $\tau = 1$ serves as an intuitive kink point for this switch.

I interpret the second switch in learning—from high learning in the second year to none in the third—as an indication that trainees have reached "full knowledge," after which learning stops, due to the relative benefits and costs of learning. It is not obvious why this kink in the rate of learning should occur at $\tau = 2$. Thus, the first step in my approach for counterfactual analyses is to specify a more flexible model of trainee learning, in which this kink point occurs at any $\tau = \tau_c \in (1,3)$ during the two years of the senior trainee role. In this model, trainee knowledge takes this form:

$$\rho(\tau) = \begin{cases} \rho_0 + \rho_1 \tau, & \tau \in [0,1]; \\ \rho_0 + \rho_1 + \rho_2 (\tau - 1), & \tau \in [1, \tau_c]; \\ \rho_0 + \rho_1 + \rho_2 (\tau_c - 1) + \rho_3 (\tau - \tau_c), & \tau \in [\tau_c, 3]. \end{cases} \tag{A-22}$$

Estimation of this more flexible model yields similar results to those from the baseline model: $\hat{\rho}_0 = 0.04$, $\hat{\rho}_1 = 0.20$, $\hat{\rho}_2 = 8.01$, $\hat{\rho}_3 = 0$, $\hat{\tau}_c = 1.87$, $\hat{\delta}_1 = 0.21$, $\hat{\delta}_2 = -1.42$, and $\hat{P} = 3.65$.

In counterfactual scenarios of learning, I assume that the rate of learning depends on influence, but that learning continues until full knowledge has been reached. Parameters in Equation (A-22) imply that full knowledge is $\overline{\rho} = \hat{\rho}_0 + \hat{\rho}_1 + \hat{\rho}_2 (\hat{\tau}_c - 1) \approx 7.17$, which I consider as fixed in counterfactual scenarios. For the key relationship that drives learning from influence, I assume that the rates of learning during training, ρ_1 and ρ_2, are piecewise linear functions of the average influence of the trainee during the respective tenure intervals, $T_1 \equiv [0,1]$ and $T_2 \equiv [1,\tau_c]$.

In notation, first define average influence over tenures uniformly distributed in interval T as

$$\overline{g}(T;\boldsymbol{\theta}) \equiv E_{\tau_h}\left[g(\tau_h;\tau_{-h})|\boldsymbol{\theta}\right], \tag{A-23}$$

where influence $g(\tau_h;\tau_{-h})$ is given in Equation (A-16) and depends on $\boldsymbol{\theta} = (\rho_0,\rho_1,\rho_2,\rho_3,\delta_1,\delta_2,P)$. Consider a counterfactual scenario as defined by key parameters of supervisory information or influence, and denote the corresponding set of counterfactual parameters as $\boldsymbol{\theta}^\Delta$. Then a counterfactual rate of learning takes the following form: For $t \in \{1,2\}$,

$$\rho_t^\Delta = \begin{cases} \hat{\rho}_1 \overline{g}\left(T_t;\boldsymbol{\theta}^\Delta\right), & \overline{g}\left(T_t;\boldsymbol{\theta}^\Delta\right) \leq \overline{g}\left(T_1;\hat{\boldsymbol{\theta}}\right), \\ \hat{\rho}_1 + \frac{\hat{\rho}_2-\hat{\rho}_1}{\overline{g}(T_2;\hat{\boldsymbol{\theta}})-\overline{g}(T_1;\hat{\boldsymbol{\theta}})}\left(\overline{g}\left(T_t;\boldsymbol{\theta}^\Delta\right) - \overline{g}\left(T_1;\hat{\boldsymbol{\theta}}\right)\right), & \overline{g}\left(T_t;\boldsymbol{\theta}^\Delta\right) > \overline{g}\left(T_1;\hat{\boldsymbol{\theta}}\right). \end{cases} \tag{A-24}$$

Under estimated parameters $\hat{\boldsymbol{\theta}}$, the implied rates of learning are similar for $\overline{g}\left(T_t;\boldsymbol{\theta}^\Delta\right)$ above and below $\overline{g}\left(T_1;\hat{\boldsymbol{\theta}}\right)$: $\hat{\rho}_1/\overline{g}\left(T_1;\hat{\boldsymbol{\theta}}\right) \approx 13.2$, and $(\hat{\rho}_2 - \hat{\rho}_1)/\left(\overline{g}\left(T_2;\hat{\boldsymbol{\theta}}\right) - \overline{g}\left(T_1;\hat{\boldsymbol{\theta}}\right)\right) \approx 14.6$.

VI.E.2 Counterfactual Scenarios and Outcomes

I consider counterfactual scenarios defined by counterfactual supervisory information (P^Δ) or influence between trainees (δ_1^Δ and δ_2^Δ). A counterfactual scenario implies varying levels of influence along the entire course of training, as given by Equations (A-16) and (A-21). Influence also depends on knowledge, as given by Equation (A-22), which in turn depends on learning via influence, as given by (A-24).

Thus, I must find an internally consistent set of parameters $\boldsymbol{\theta}^\Delta$ that contains P^Δ. In all counterfactual scenarios, I hold fixed $\rho_0^\Delta = \hat{\rho}_0$ and $\rho_3^\Delta = \hat{\rho}_3 = 0$. In counterfactual scenarios involving P^Δ, I also hold fixed $\tilde{\delta}_1^\Delta \equiv \delta_1^\Delta/\left(\rho_0^\Delta + \rho_1^\Delta\right) = \delta_1/(\rho_0 + \rho_1)$, since it is not possible to have $\delta_1^\Delta - \left(\rho_1^\Delta + \rho_0^\Delta\right) < 0$; I similarly hold fixed $\tilde{\delta}_2^\Delta \equiv \delta_2^\Delta/\min\left(\overline{\rho},\rho_0^\Delta + \rho_1^\Delta + \rho_2^\Delta\right) = \delta_2/\min(\overline{\rho},\rho_0 + \rho_1 + \rho_2)$. Conversely, for counterfactual scenarios involving influence between trainees, I vary $\tilde{\delta}_1^\Delta$ or $\tilde{\delta}_2^\Delta$ while holding fixed $P^\Delta = P$. Given these constraints, I identify an internally consistent $\boldsymbol{\theta}^\Delta$ by solving for ρ_1^Δ and ρ_2^Δ in the nonlinear system of two equations implied by Equations (A-16), (A-21), (A-22), (A-23), and (A-24), for $t \in \{1,2\}$.

For each of the counterfactual scenarios, I consider the following outcomes of learning and decision-making information:

1. Time for trainees to acquire full knowledge:

$$\overline{\tau}^\Delta = 1 + \frac{\overline{\rho} - \left(\rho_0 + \rho_1^\Delta\right)}{\rho_2^\Delta}.$$

This calculated time summarizes the counterfactual rates of learning, ρ_1^Δ and ρ_2^Δ. Since learning is always incomplete in the first year of training under all counterfactual scenarios (i.e., $\rho_1^\Delta < \overline{\rho}$), this time is always greater than one year.

2. Average information from trainee knowledge: A trainee can contribute no more information than her knowledge, but she can contribute less if decision-making departs from the Bayesian benchmark. In other words, when working with peers of tenure τ_{-h}, trainees of tenure τ_h contribute precision equal to

$$\underline{\rho}^{\Delta}(\tau_h; \tau_{-h}) = \min\left(1, \frac{g(\tau_h; \tau_{-h})}{g^*(\tau_h; \tau_{-h})}\right) \rho^{\Delta}(\tau_h).$$

Counterfactual knowledge, $\rho^{\Delta}(\tau_h)$, is given by Equation (A-22) using the counterfactual parameters ρ_1^{Δ} and ρ_2^{Δ}; $\tilde{\rho}^{\Delta}(\tau)$, as given by Equation (A-21), may differ from $\rho^{\Delta}(\tau)$ if $\delta_1^{\Delta} \neq 0$ or $\delta_2^{\Delta} \neq 0$. For patients uniformly distributed over the course of an academic year, the average information from trainee teams is then

$$Q^{\Delta} = \int_0^1 \left(\lambda \left(\underline{\rho}^{\Delta}(\tau; \tau+1) + \underline{\rho}^{\Delta}(\tau+1; \tau)\right) + (1-\lambda)\left(\underline{\rho}^{\Delta}(\tau; \tau+2) + \underline{\rho}^{\Delta}(\tau+2; \tau)\right)\right) d\tau,$$

where $\lambda = 0.7$ is the approximate fraction of patients seen by teams with second-year trainees, and $1-\lambda$ is the remaining fraction of patients seen by teams with third-year trainees. The three terms inside the integral represent levels of information contributed by first-, second-, and third-year trainees, respectively.

3. Average total information in decision-making: $P^{\Delta} + Q^{\Delta}$, or the sum of supervisory information and average information from trainee knowledge.

VI.E.3 Discussion of Results

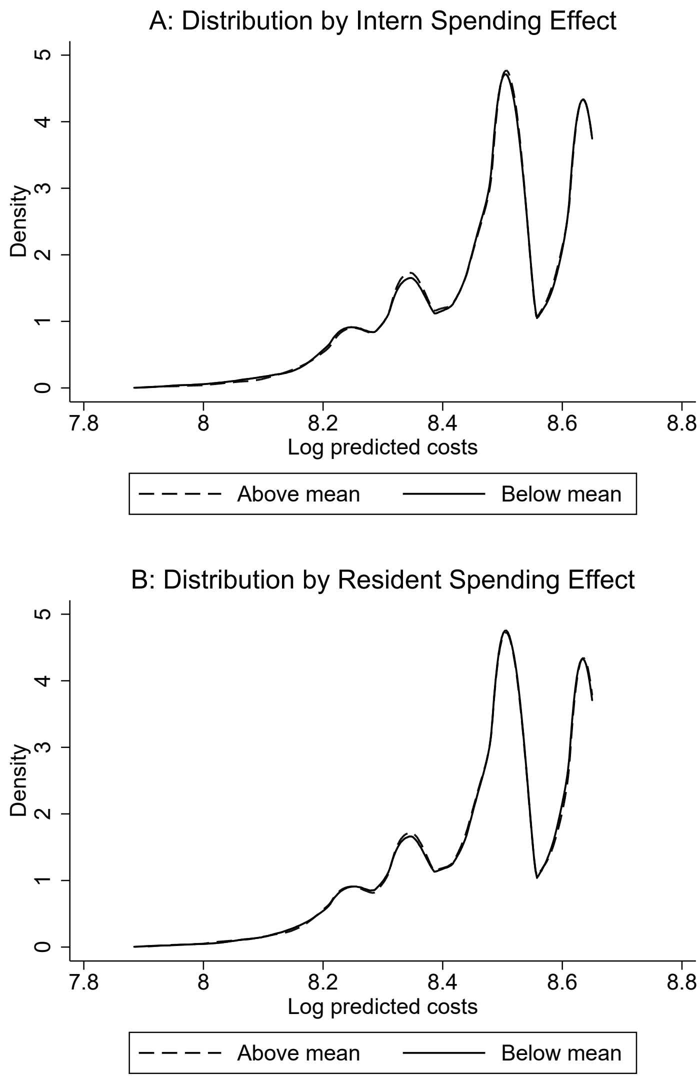
In Figure A-11, I show outcomes under counterfactual scenarios varying P^{Δ} and $\tilde{\delta}_1^{\Delta}$. As expected, increasing P^{Δ} slows the rate of learning and increases the time for trainees to acquire full knowledge. There are direct effects of P^{Δ} in decreasing trainee influence as well as indirect effects, as trainees with less influence acquire less knowledge to contribute to decision-making. Thus, increasing supervisory information decreases the information from trainee knowledge used in decision-making. The gain in total decision-making information is reduced by about 40% by this mechanism of diminishing trainee knowledge. In contrast, there is only limited impact of varying $\tilde{\delta}_1^{\Delta}$ on learning and trainee knowledge over the course of residency, at least in the range of $\tilde{\delta}_1^{\Delta} \in [-1, 1]$. By decreasing $\tilde{\delta}_1^{\Delta}$, trainees gain more knowledge when they are junior but less when they are senior. The effect of influence on learning is slightly steeper for senior trainees, which explains why there are some slight returns to increasing $\tilde{\delta}_1^{\Delta}$ in terms of decreasing years to acquire full knowledge and increasing information from trainee knowledge in the average team decision.

In Figure A-12, I show outcomes under counterfactual scenarios varying $\tilde{\delta}_2^{\Delta}$. The effects of increasing $\tilde{\delta}_2^{\Delta}$ on learning and decision-making information are similar to those of increasing $\tilde{\delta}_1^{\Delta}$: Increasing senior influence speeds up training and increases overall trainee knowledge. The effect range of counterfactual values of δ_2^Δ is larger, since the denominator in $\tilde{\delta}_2^\Delta$ (i.e., $\rho^\Delta(2)$) is larger. Interestingly, around $\tilde{\delta}_2^\Delta = 0$, decreasing $\tilde{\delta}_2^\Delta$ has a larger effect on Q^Δ than does increasing $\tilde{\delta}_2^\Delta$, due to the following intuition: Near baseline parameters, much of the third year involves no learning. Therefore, increasing the influence of third-year trainees does not aid learning for those trainees, and learning among junior trainees will suffer. However, learning indirectly increases for second-year trainees who then work with less knowledgeable junior trainees. Nonetheless, the effects on learning are generally small relative to those for varying P^Δ.

Figure A-1: Patients Age by Housestaff Spending Effect

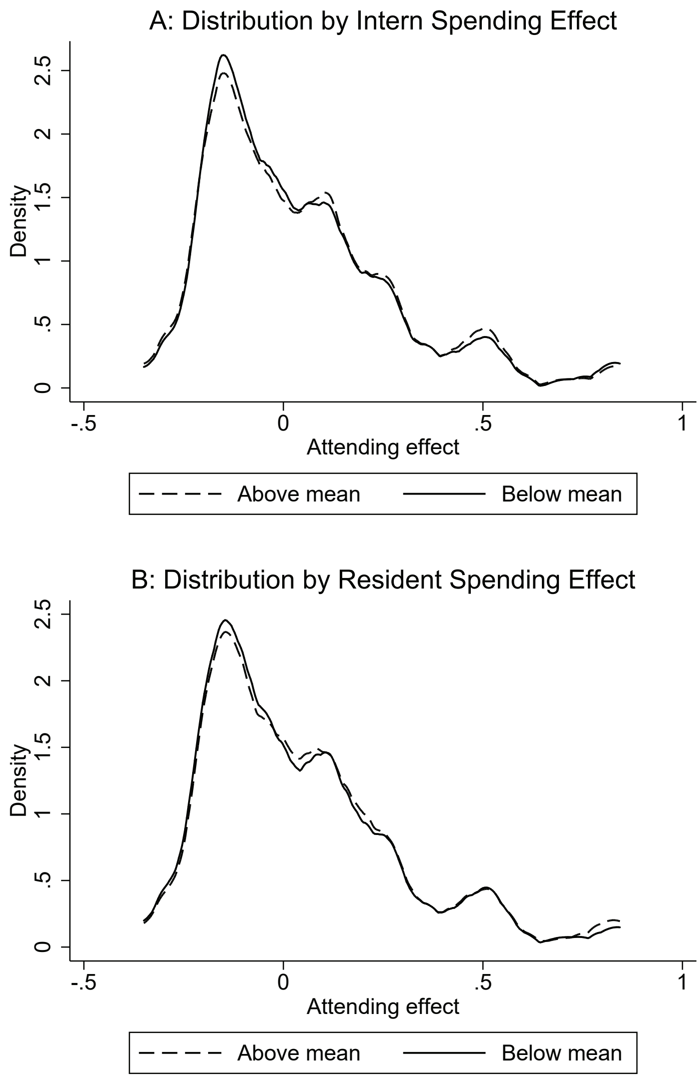
Note: This figure shows the distribution of the age of patients assigned to interns with above- or below-average spending effects (Panel A) and residents with above- or below-average spending effects (Panel B). Trainee spending effects, not conditioning by tenure, are estimated by Equation (A-3) as fixed effects by a regression of log spending on patient characteristics and physician (intern, resident, and attending) identities. Kolmogorov-Smirnov statistics testing for the difference in distributions yield *p*-values of 0.496 and 0.875 for interns (Panel A) and residents (Panel B), respectively.

Figure A-2: Demographics-predicted Spending by Trainee Spending Effect



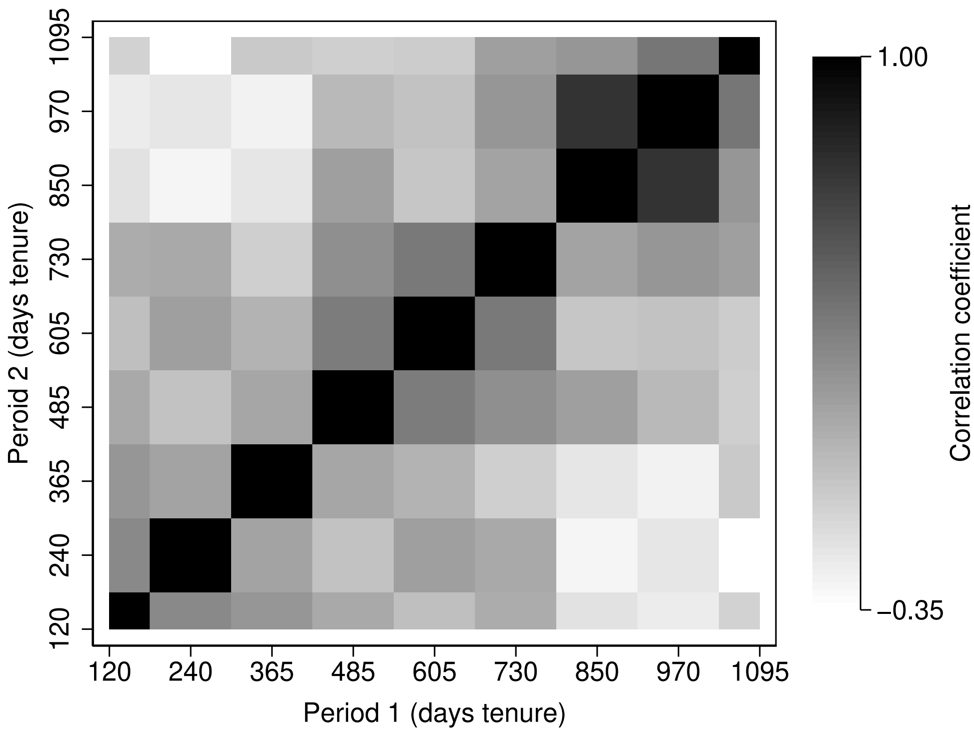
Note: This figure shows the distribution of predicted log costs (based on patient age, race, and gender) for patients assigned interns with above- or below-average spending effects (Panel A) and residents with above- or below-average spending effects (Panel B). Trainee spending effects, not conditioning by tenure, are estimated by Equation (A-3) as fixed effects by a regression of log spending on patient characteristics and physician (intern, resident, and attending) identities. Kolmogorov-Smirnov statistics testing for the difference in distributions yield p-values of 0.683 and 0.745 for interns (Panel A) and residents (Panel B), respectively.

Figure A-3: Attendings Spending Effects by Trainee Spending Effect



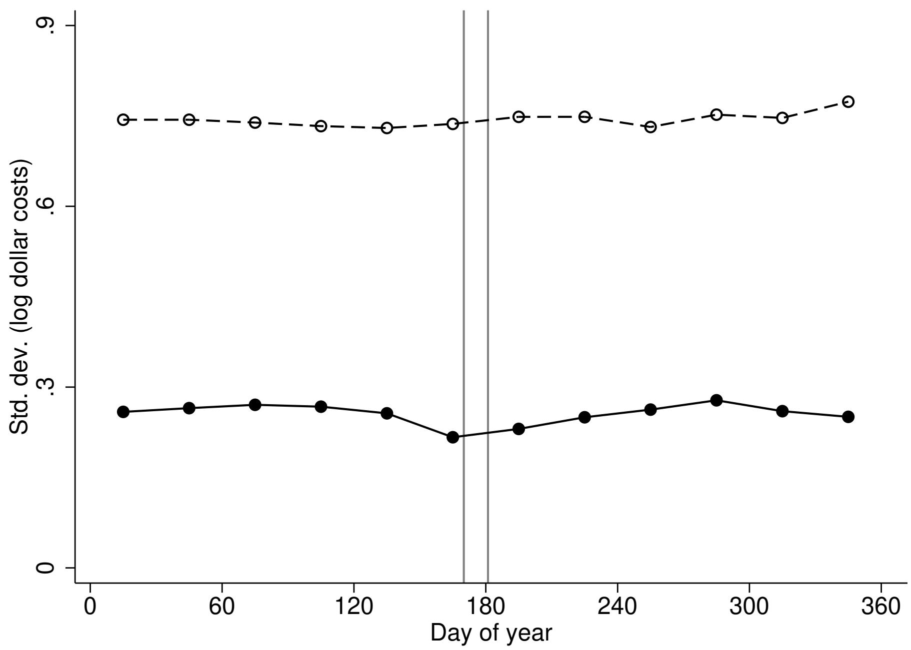
Note: This figure shows the distribution of spending fixed effects for attendings assigned to interns with above- or below-average spending effects (Panel A) and residents with above- or below-average spending effects (Panel B). Trainee and attending spending effects, not conditioning by tenure, are estimated by Equation (A-3) as fixed effects by a regression of log spending on patient characteristics and physician (intern, resident, and attending) identities. Kolmogorov-Smirnov statistics testing for the difference in distributions yield *p*-values of 0.059 and 0080 for interns (Panel A) and residents (Panel B), respectively.

Figure A-4: Serial Correlation of Trainee Random Effects



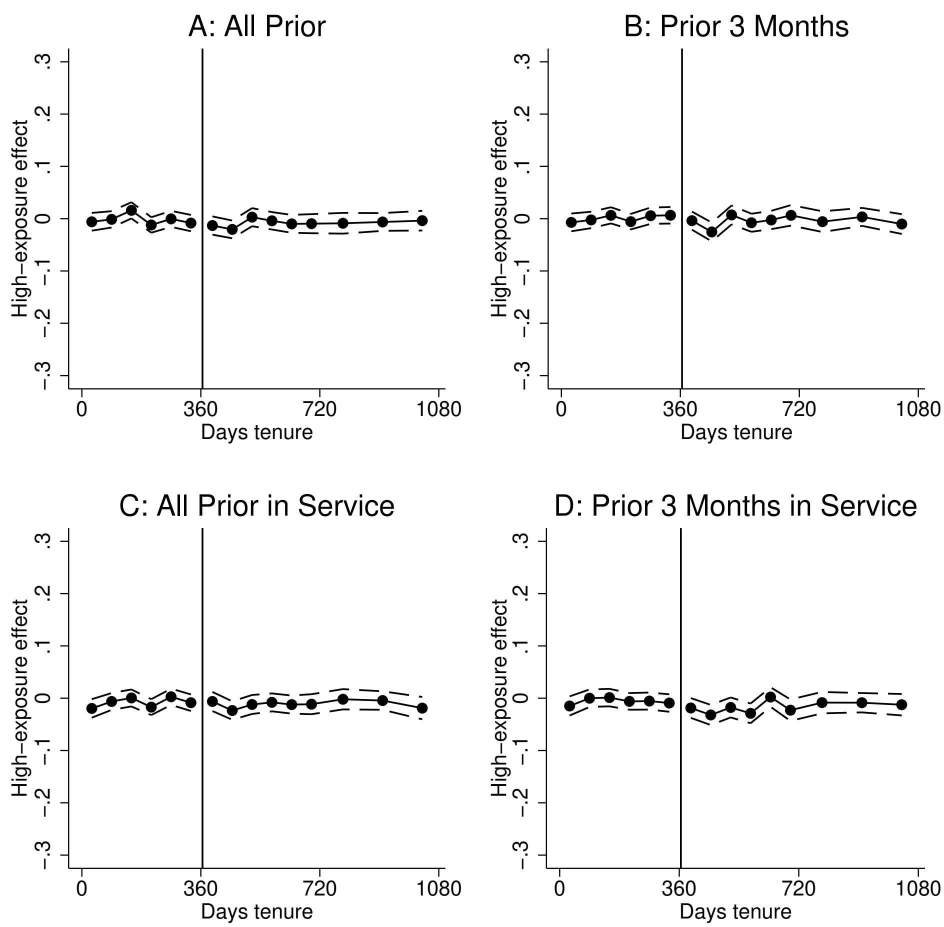
Note: This figure shows the serial correlation between random effects within trainee between two tenure periods. Details of the estimation routine are given in Appendix III.B. The random effect model of log daily total costs is given in Equation (3). The model controls are as stated for Figure 1. Trainees prior to one year in tenure are junior trainees and become senior trainees after one year in tenure. Numerical values and confidence intervals are given in Table A-4.

Figure A-5: Trainee-associated and Residual Variation by Day of Year



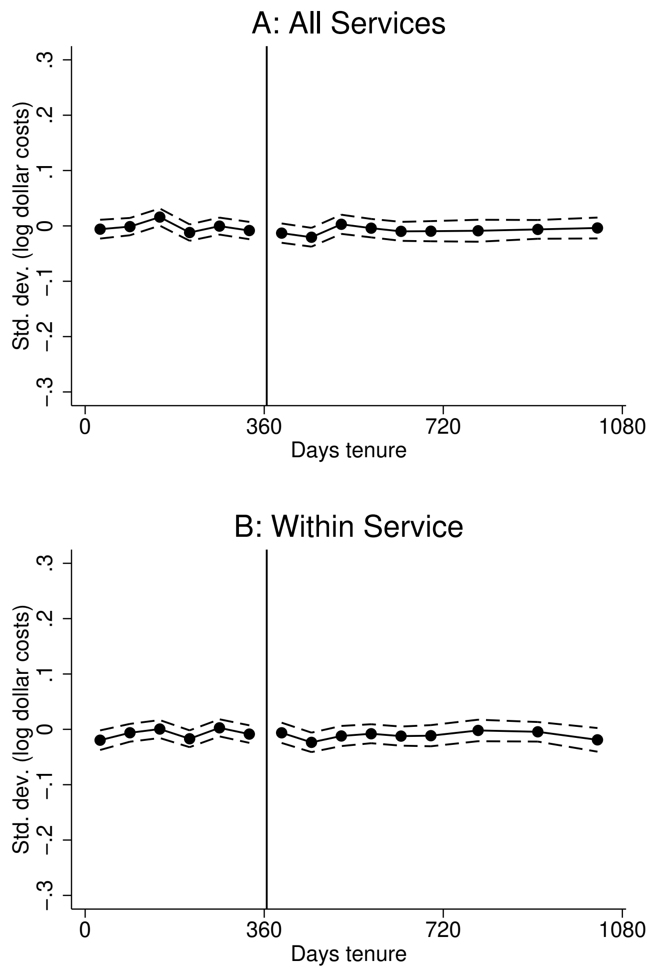
Note: This figure shows the standard deviation of random effects due to junior and senior trainee teams (solid dots) and the standard deviation of the residual (hollow dots) in 30-day periods by day of the year. Residual variation can be interpreted as variation due to independent observation. The two vertical gray lines indicate when new junior trainees begin residency on July 19 and when senior trainees advance a year on July 28 (i.e., becoming a new second-year senior trainee, becoming a third-year trainee, or completing residency). The model is similar to Equation (3), except that a single random effect is modeled for the junior and senior trainee combination, instead of two additively separable random effects for the respective trainees. Controls are given in the note for Figure 1.

Figure A-6: Effect of High Prior Exposure to Spending



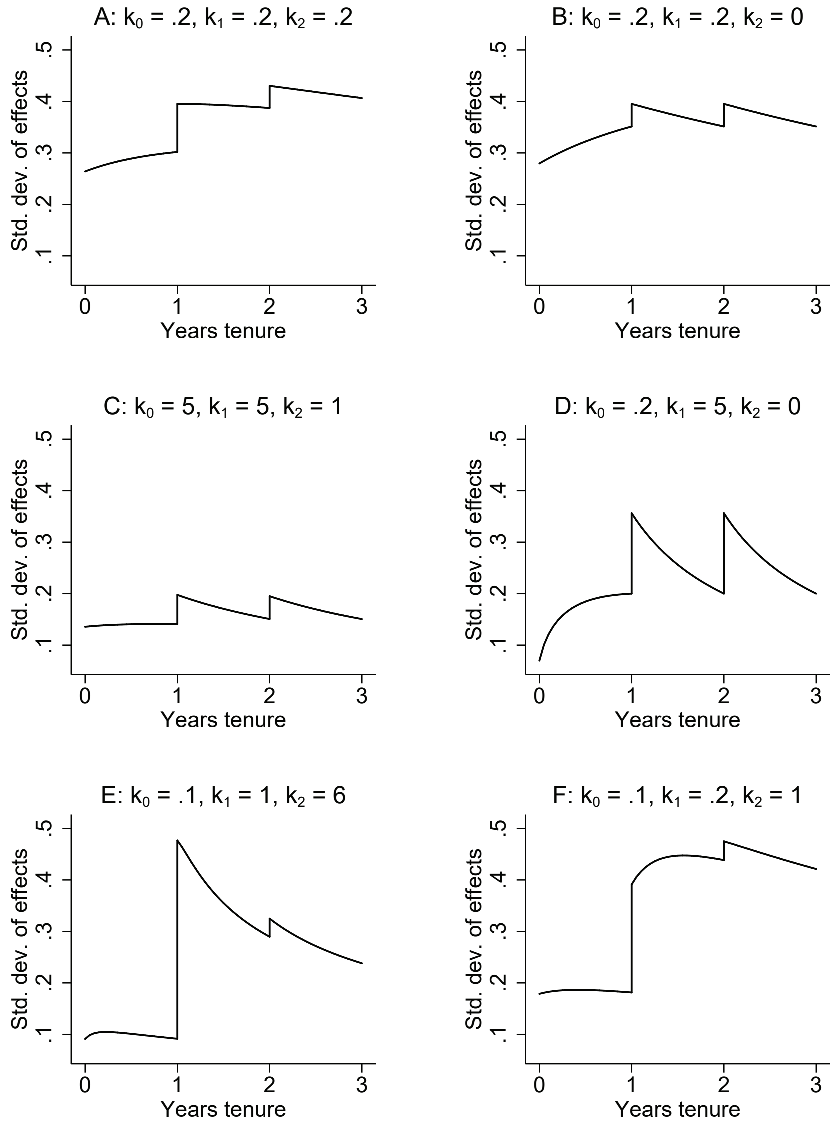
Note: This figure shows the effect of high prior exposure to supervising-physician spending. This exposure measure is discussed in Section and in Table A-5 and reflects the average spending effects of supervising physicians that a given trainee was matched to in the past. The tenure-specific effect of having high prior exposure to spending is estimated as in Equation (A-12). Panel A uses an exposure measure that includes all prior matches, regardless of service (corresponding to Column 1, Panel A of Table A-5). Panels B and D use an exposure measure that includes matches within the last three months with supervising physicians (corresponding to Columns 2 and 4, Panel A of Table A-5). Panels C and D use an exposure measure that is restricted to prior matches on the same service (corresponding to Columns 3 and 4, Panel A of Table A-5). The vertical line indicates the one-year mark of training; trainees are junior prior to this and senior after this. The model controls are as stated for Figure 1. The effect of high prior exposure to senior-trainee spending is shown in Figure A-7. Point estimates are shown as connected dots; 95% 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.

Figure A-7: Effect of High Exposure to Senior-trainee Spending



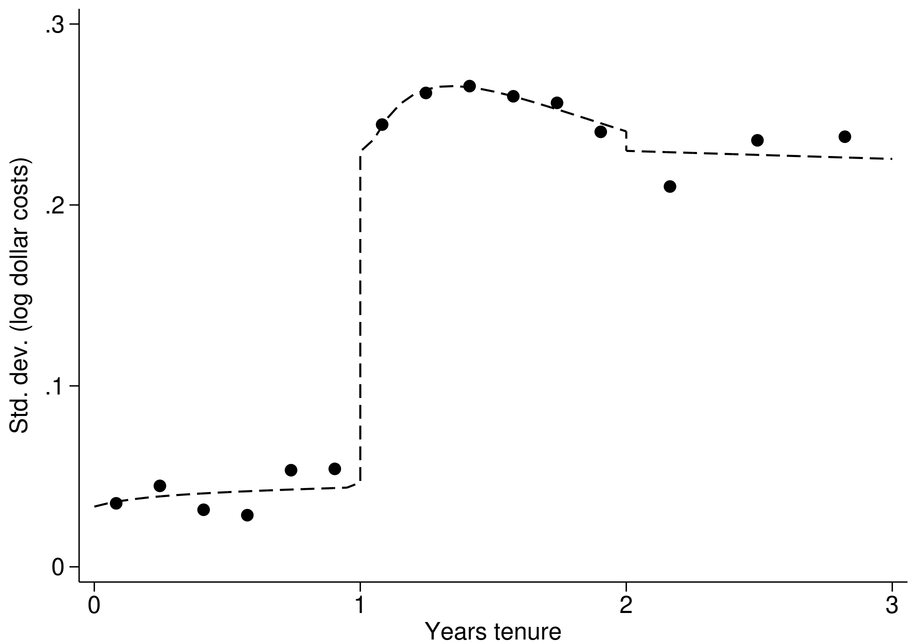
Note: This figure shows the effect of high prior exposure to senior-trainee spending. This exposure measure is discussed in further detail in Appendix A-5 and in Table A-5 and reflects the average spending effects of senior trainees that a given trainee was matched to in the past as a junior trainee. The tenure-specific effect of having high prior exposure to spending is estimated as in Equation (A-12). Panel A uses an exposure measure that includes all prior matches with senior trainees, regardless of the ward service (corresponding to Column 1, Panel B of Table A-5); Panel B uses an exposure measure that is restricted to prior matches on the same service (corresponding to Column 3, Panel B of Table A-5). For tenure periods after the one-year mark (shown as the vertical line), the trainee of interest is senior, and matches with senior trainees all date back to the trainee's first year of training as a junior trainee. The model controls are as stated for Figure 1. The effect of high prior exposure to supervising-physician spending is shown in Figure A-6.

Figure A-8: Numerical Examples of Variation Profiles



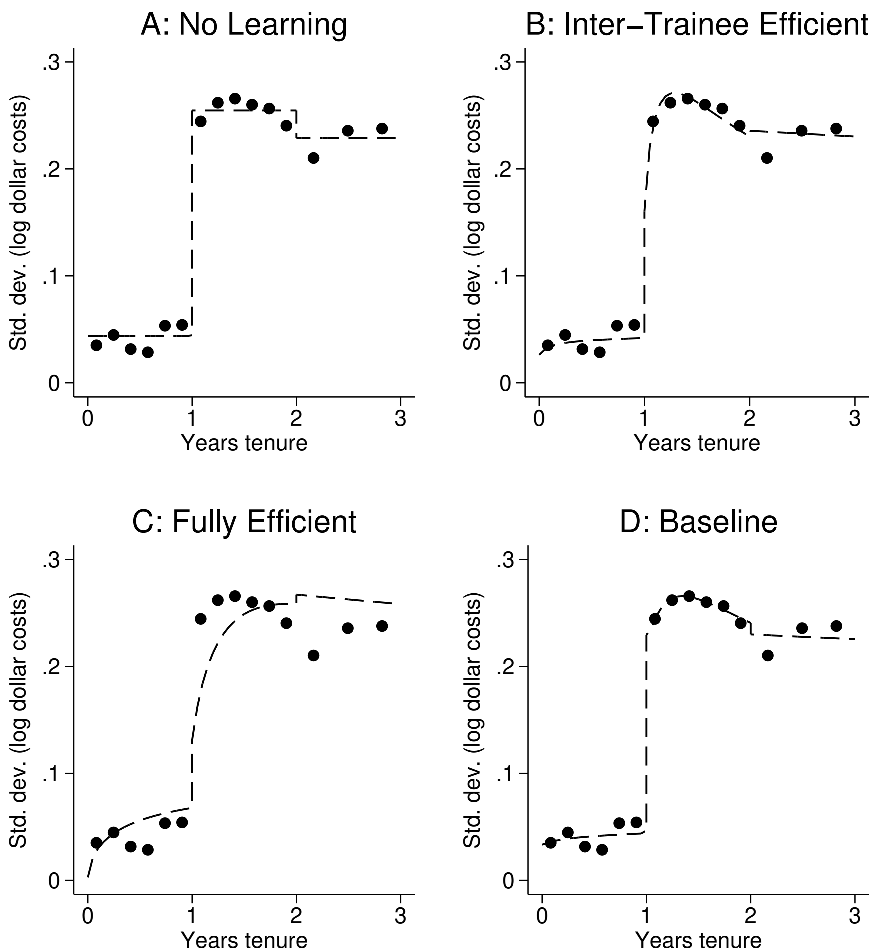
Note: This figure shows variation profiles of the expected standard deviation of trainee effects over tenure, $\sigma(\tau)$, differing by the underlying profile of learning over tenure. Learning is parameterized as a piecewise linear function $g(\tau)$ that describes how the precision of subjective priors increases over tenure. In particular, this figure considers piecewise linear functions of the form (A-20), parameterized by ρ_0, ρ_1, and $\rho_2 = \rho_3$. Each panel considers a different set of parameters of $\rho(\tau)$. Given $\rho(\tau)$, I calculate the expected standard deviation of trainee effects over tenure using Equation (A-18). I assume that interns are equally likely to work with second-year residents and third-year residents. These profiles are discussed further in Appendix VI.B.

Figure A-9: Model Fit to Practice Variation Profile



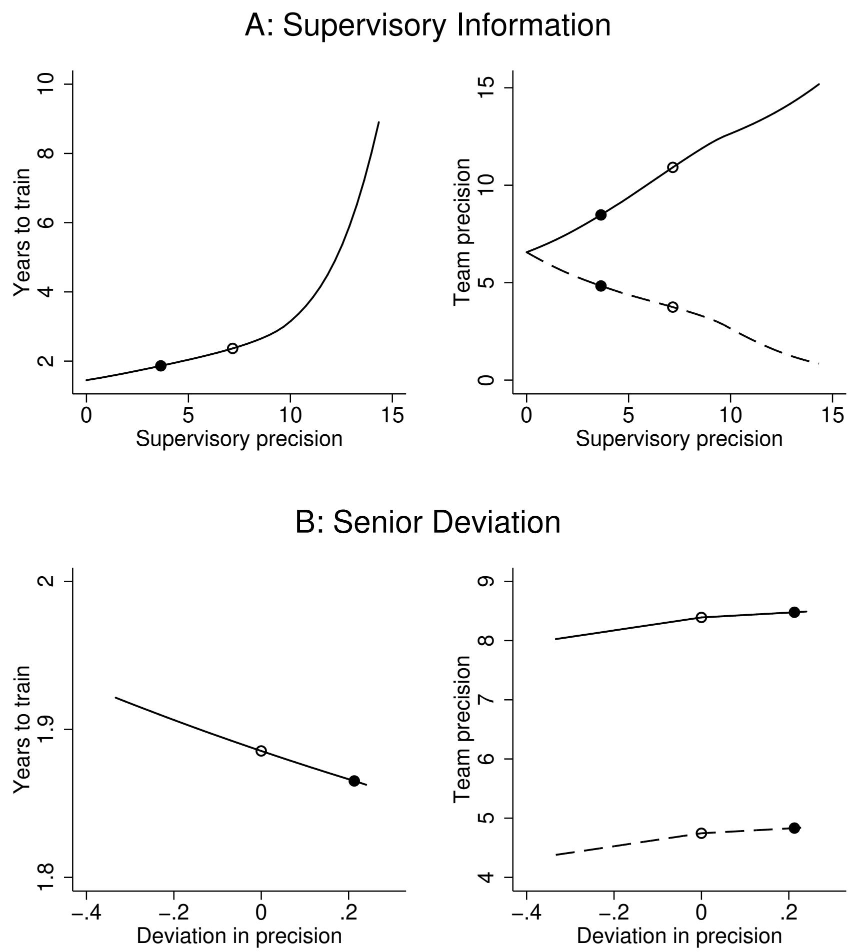
Note: This figure shows practice variation, defined as the standard deviation of random trainee effects specified in Equation (3), in log daily total costs at each non-overlapping tenure period. Trainee prior to one year in tenure are junior trainees and become senior trainees after one year in tenure. Reduced-form estimates of practice variation are shown in dots and are the same as shown in Figure 1. Practice variation implied by the model of learning and influence, specifically Equation (A-17), is shown as a dashed line. Estimation of parameters of this model is described in Section IV.C. The Sargan-Hansen over-identification J-test statistic of the model is $J = 8.60$, which is less than the 95th percentile value of 19.7 the χ^2_{18-7} distribution (the p-value corresponding to $J = 8.60$ is 0.67)

Figure A-10: Model Restrictions



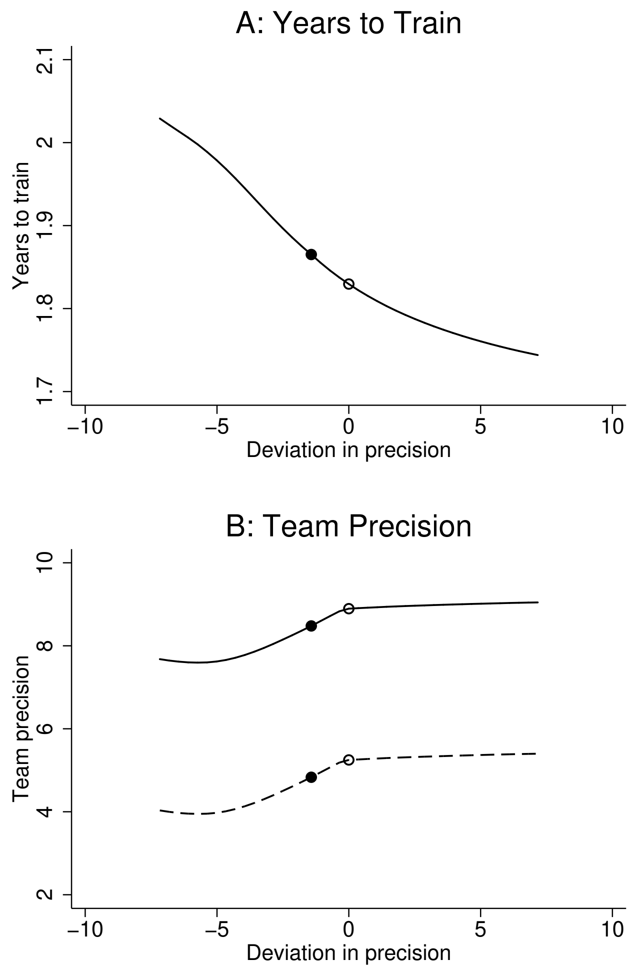
Note: This figure shows the fit of restricted models of learning and influence, with parameters described in Table A-7. Each panel shows the same reduced-form moments of practice variation for each tenure period, which are also the same as those shown in Figure A-9, reproduced in Panel D. Panel A restricts the model to no learning (i.e., $\rho_1 = \rho_2 = \rho_3 = 0$). Panel B restricts the model to the Bayesian benchmark of influence between trainees (i.e., $\delta_1 = \delta_2 = 0$). Panel C additionally restricts the model so that supervisors receive as much influence as warranted by the lower bound of their knowledge (i.e., $\delta_1 = \delta_2 = 0$, $P \geq \rho_0 + \rho_1 + \rho_2 + \rho_3$). The likelihood ratio test comparing a no-learning model (Panel A) with the baseline model (Panel D) rejects the restricted model with a p-value less than 0.01. Likelihood ratio tests for other outcomes or for subsets of the data are also given in Tables A-7 and A-8. Sargan-Hansen over-identification J-test statistics are 15.66 (p-value = 0.405 under χ^2_{18-3} distribution) for Panel A, 13.42 (p-value = 0.416 under χ^2_{18-5} distribution) for Panel B, and 65.97 (p-value < 0.01 under χ^2_{18-4} distribution).

Figure A-11: Counterfactual Training Time and Team Information



Note: This figure shows counterfactual results on time for trainees to acquire "full knowledge" and on information used in decision-making. I consider two types of counterfactual scenarios: In subpanels in Panel A, I alter on the *x*-axes the amount of supervisory information used in decision-making, or P in the model, while holding fixed the relative influence between junior and senior trainees. In subpanels in Panel B, I alter on the *x*-axes the relative influence between junior and senior trainees, or δ_1 in the model, while holding fixed the amount of supervisory information. Appendix Figure A-12 shows results for varying δ_2 in the model. The time for trainees to acquire full knowledge (or "years to train") is measured on the *y*-axes of the left subpanels, and the information used in decision-making is measured on the *y*-axes of the right subpanels. The right subpanels show both information from trainee knowledge (dashed lines) and total information (solid lines) used in decision-making. On each line, I plot a solid dot indicating actual results and a hollow dot indicating counterfactual results under Bayesian-benchmark influence; supervisory influence in Panel A is a lower bound for the Bayesian benchmark that equals full trainee knowledge, or $\underline{P} = \rho_0 + \rho_1 + \rho_2 (\tau_c - 1)$. Lines in Panel A are plotted for counterfactual $P^\Delta \in [0, 2\underline{P}]$; lines in Panel B are plotted for counterfactual $\delta_1^\Delta / (\rho_0^\Delta + \rho_1^\Delta) \in [-1, 1]$. Further details are given in Appendix VI.E.

Figure A-12: Counterfactual Results, Varying δ_2



Note: This figure shows results for counterfactual scenarios in which I vary the additional deviation in effective precision for third-year trainees, or δ_2 in the model and shown in the x-axes of both panels. The y-axis of Panel A plots the time for trainees to acquire "full knowledge" (or "years to train"). The y-axis of Panel B plots information from trainee knowledge (dashed lines) and total information (solid lines) used in decision-making. On each line, I plot a solid dot indicating actual results and a hollow dot indicating counterfactual results under the Bayesian benchmark. Lines are plotted for counterfactual $\delta_2^\Delta/\rho^\Delta(2) \in [-1,1]$. Further details are given in Appendix VI.E.

Table A-1: Tests of Joint Significance of Trainee Identities and Characteristics

	Independent variables		
	Trainee identities	Trainee characteristics	
Patient characteristic	(1)	(2)	(3)
Age	$F(1055, 46364) = 0.98$ $p = 0.661$	$F(22, 14568) = 0.49$ $p = 0.978$	$F(20, 32434) = 0.55$ $p = 0.945$
Male	$F(1055, 46364) = 1.01$ $p = 0.378$	$F(22, 14568) = 1.16$ $p = 0.276$	$F(20, 32434) = 1.07$ $p = 0.376$
White	$F(1055, 46364) = 1.02$ $p = 0.356$	$F(22, 14568) = 0.72$ $p = 0.829$	$F(20, 32434) = 0.77$ $p = 0.756$
Predicted spending	$F(1055, 46364) = 0.98$ $p = 0.685$	$F(22, 14568) = 0.71$ $p = 0.836$	$F(20, 32434) = 1.12$ $p = 0.322$

Note: This table reports tests of joint significance corresponding to Equations (A-1) and (A-2). Column 1 corresponds to Equation (A-1); Columns 2 and 3 correspond to (A-2). Column 2 includes all trainee characteristics: trainee's position on the rank list; USMLE Step 1 score; sex; age at the start of training; and dummies for whether the trainee graduated from a foreign medical school, whether he graduated from a rare medical school, whether he graduated from medical school as a member of the AOA honor society, whether he has a PhD or another graduate degree, and whether he is a racial minority. Column 3 includes all trainee characteristics except for position on the rank list. Rows correspond to different patient characteristics as the dependent variable of the regression equation; the last row is predicted spending using patient demographics (age, sex, and race). F-statistics and p-values are reported for each joint test.

Table A-2: Practice Style Distribution by Junior Trainee Characteristic

Characteristic	Above median	Cases	Trainees	Characteristic Mean	Characteristic Range	Trainee empirical Bayes posterior Mean	Trainee empirical Bayes posterior 10th percentile	Trainee empirical Bayes posterior 90th percentile
Age	Y	92,482	183	30.6	[28.1,39.4]	0.000	−0.025	0.024
	N	92,381	184	26.8	[24.1,28.1]	0.001	−0.025	0.025
AOA honor society	Y	63,864	124	1	[1,1]	0.001	−0.023	0.027
	N	121,429	243	0	[0,0]	0.000	−0.025	0.023
Future Income ($ thousands)	Y	107,080	275	424	[357,850]	0.001	−0.022	0.026
	N	107,222	391	268	[199,357]	0.000	−0.026	0.024
Male	Y	110,316	216	1	[1,1]	0.000	−0.024	0.026
	N	74,977	151	0	[0,0]	0.001	−0.027	0.023
Medical school rank	Y	87,149	175	24.7	[6,84]	0.000	−0.022	0.024
	N	87,254	176	2.2	[1,6]	0.000	−0.027	0.025
Minority	Y	22,564	43	1	[1,1]	0.003	−0.022	0.023
	N	162,729	324	0	[0,0]	0.000	−0.025	0.025
Other advanced degree	Y	62,433	119	1	[1,1]	0.000	−0.024	0.024
	N	122,860	248	0	[0,0]	0.001	−0.026	0.024
Rank list position	Y	65,267	132	80	[51,255]	0.001	−0.024	0.026
	N	65,221	128	26	[1,51]	0.001	−0.024	0.024
USMLE Step 1	Y	91,842	186	256	[246,272]	0.001	−0.024	0.027
	N	92,254	192	233	[184,246]	0.000	−0.025	0.023

Note: This table reports the distribution of empirical Bayes posterior means for junior trainees divided in groups by characteristics. Posterior means are calculated as in the note for Figure 6 and pooled across all tenure periods in the first year. Results for senior trainees are given in Table A-3. Cases denote the number of observations (i,t) such that the junior trainee belongs to the group. The mean, 10th percentile, and 90th percentile of each Empirical Bayes posterior distribution are all calculated over the set of relevant cases. Characteristics are discussed further in Section I.C. Age refers to the trainee age in years at the start of residency. AOA (Alpha Omega Alpha) honor society inducts 13.5% of US medical graduates, and inclusion in AOA is associated with an odds ratio of 6-10 of successful matching into the first-choice residency (Rinard and Mahabir, 2010). Future income is imputed by future observed specialty or subspecialty entry of the trainee. Medical school rank is obtained from the *US News & World Report*. The national mean and 95th percentile of the United States Medical Licensing Examination (USMLE) Step 1 test scores are approximately 229 and 260, respectively.

Table A-3: Practice Style Distribution by Senior Trainee Characteristic

				Characteristic		Trainee empirical Bayes posterior		
Characteristic	Above median	Cases	Trainees	Mean	Range	Mean	10th percentile	90th percentile
Age	Y	98,703	167	30.3	[28,37.9]	−0.020	−0.256	0.321
	N	100,712	169	26.8	[24.1,28.0]	0.050	−0.234	0.359
AOA honor society	Y	70,334	112	1	[1,1]	0.010	−0.243	0.344
	N	129,668	223	0	[0,0]	0.020	−0.247	0.341
Future Income	Y	100,571	161	411	[357,540]	0.047	−0.241	0.351
($ thousands)	N	98,087	187	246	[199,357]	−0.013	−0.256	0.330
Male	Y	112,520	188	1	[1,1]	0.012	−0.260	0.338
	N	87,482	147	0	[0,0]	0.022	−0.228	0.344
Medical school	Y	91,874	149	26.9	[7,84]	−0.005	−0.243	0.330
rank	N	91,722	165	2.7	[1,7]	0.047	−0.244	0.353
Minority	Y	23,108	40	1	[1,1]	0.000	−0.246	0.320
	N	176,894	295	0	[0,0]	0.019	−0.244	0.345
Other advanced	Y	48,765	83	1	[1,1]	−0.009	−0.261	0.324
degree	N	151,237	252	0	[0,0]	0.025	−0.242	0.349
Rank list position	Y	64,025	101	80	[54,246]	0.089	−0.255	0.363
	N	63,967	104	28	[1,54]	0.053	−0.253	0.342
USMLE Step 1	Y	97,224	160	255	[246,275]	0.031	−0.247	0.350
	N	96,916	172	232	[156,246]	0.005	−0.243	0.333

Note: This table reports the distribution of empirical Bayes posterior means for senior trainees divided in groups by characteristics. Posterior means are calculated as in the note for Figure 6 and pooled across all tenure periods in the second and third years. Results for junior trainees are given in Table A-2. Cases denote the number of observations (i,t) such that the junior trainee belongs to the group. The mean, 10th percentile, and 90th percentile of each Empirical Bayes posterior distribution are all calculated over the set of relevant cases. Characteristics are discussed further in Section I.C. Age refers to the trainee age in years at the start of residency. AOA (Alpha Omega Alpha) honor society inducts 13.5% of US medical graduates, and inclusion in AOA is associated with an odds ratio of 6-10 of successful matching into the first-choice residency (Rinard and Mahabir, 2010). Future income is imputed by future observed specialty or subspecialty entry of the trainee. Medical school rank is obtained from the *US News & World Report*. The national mean and 95th percentile of the United States Medical Licensing Examination (USMLE) Step 1 test scores are approximately 229 and 260, respectively.

Table A-4: Serial Correlation in Trainee Effects

Period 1 (days)	Correlation in Trainee Effects Period 2 (days) 121-240	241-365	366-485	486-605	606-730	731-850	851-970	971-1095
0-120	0.27 [−0.02, 0.58]	0.20 [−0.09, 0.58]	0.11 [−0.15, 0.40]	−0.02 [−0.29, 0.27]	0.09 [−0.22, 0.43]	−0.19 [−0.72, 0.51]	−0.25 [−0.61, 0.39]	−0.10 [−0.50, 0.39]
121-240		0.14 [−0.14, 0.54]	−0.03 [−0.32, 0.39]	0.15 [−0.15, 0.44]	0.11 [−0.25, 0.52]	−0.30 [−0.80, 0.28]	−0.21 [−0.58, 0.33]	−0.35 [−0.86, 0.28]
241-365			0.12 [−0.15, 0.43]	0.06 [−0.22, 0.35]	−0.09 [−0.45, 0.29]	−0.21 [−0.67, 0.42]	−0.27 [−0.61, 0.21]	−0.05 [−0.45, 0.40]
366-485				0.33 [0.03, 0.67]	0.24 [−0.12, 0.69]	0.16 [−0.28, 0.59]	0.02 [−0.44, 0.45]	−0.10 [−0.45, 0.27]
486-605					0.36 [0.03, 0.72]	−0.05 [−0.47, 0.42]	−0.03 [−0.31, 0.28]	−0.08 [−0.46, 0.39]
606-730						0.13 [−0.16, 0.46]	0.20 [−0.12, 0.51]	0.15 [−0.27, 0.58]
731-850							0.72 [0.37, 1.00]	0.21 [−1.00, 1.00]
851-970								0.38 [−0.02, 0.75]

Note: This table displays bootstrapped correlation parameters estimated from the model described in Section II.D and Appendix III.B. In each bootstrapped run, observations are drawn with replacement from strata defined by the junior and senior trainees, the admission service, and the tenure periods of each trainee. As described in Section II.D, log spending is residualized by patient characteristics and fixed effects for time categories, clinical service, and supervising physician. These residuals are then used in maximum likelihood to estimate the correlation between trainee effects, as described in Appendix III.B. Each cell in the table corresponds to the correlation between trainee effects in two tenure periods. The mean bootstrapped correlation parameter is shown in the first line of each cell; the 95% confidence interval is shown in the second line in brackets. Mean correlation parameters are also shown in Figure (A-4).

Table A-5: Differences in Prior Exposure to Spending

	Differences Between High and Low Exposure			
	(1)	(2)	(3)	(4)
	All services		Within service	
Tenure period (days)	All prior	Prior 3 months	All prior	Prior 3 months
Panel A: Exposure to Spending by Supervising Physicians				
0-60	5.31%	5.65%	4.62%	4.84%
61-120	5.16%	5.52%	4.81%	5.03%
121-180	4.64%	5.41%	4.39%	4.87%
181-240	4.47%	5.43%	3.85%	4.41%
241-300	4.06%	5.21%	3.85%	4.41%
301-365	3.81%	4.92%	3.31%	4.28%
366-425	3.54%	5.80%	3.87%	5.41%
426-485	3.70%	6.06%	4.05%	6.04%
486-545	3.30%	5.71%	3.31%	4.83%
546-605	3.15%	5.27%	3.67%	5.47%
606-665	3.34%	6.01%	4.05%	6.26%
666-730	3.39%	5.91%	3.44%	5.24%
731-850	3.53%	4.97%	2.22%	3.78%
851-970	3.52%	5.82%	2.56%	4.05%
971-1095	3.03%	3.91%	1.80%	3.02%
Panel B: Exposure to Spending by Senior Trainees				
0-60	19.08%	19.50%	20.82%	20.92%
61-120	19.88%	20.32%	22.89%	23.02%
121-180	19.54%	21.03%	21.51%	23.23%
181-240	19.52%	20.54%	22.12%	23.53%
241-300	19.04%	20.12%	21.95%	23.61%
301-365	17.76%	17.99%	19.88%	20.28%

Note: This table presents differences in average spending effects of supervising physicians (Panel A) and of senior trainees (Panel B) who worked with trainees in the past at each tenure period for the trainees. Columns 1 and 2 include prior team pairings in all services, while Columns 3 and 4 only include prior team pairings within the same service. For example, for an observation in the cardiology service, Columns 3 and 4 only include prior team pairings for a trainee while working in the cardiology service. Columns 2 and 4 further restrict prior team pairings to those within the last three months. The spending effect of the relevant supervising physician or senior trainee is the empirical Bayes posterior mean from a random-effects model of log daily overall spending. Of the set of eligible prior team pairings, the exposure to spending measure is a weighted average (by patient-day) of the spending effects of the relevant matched physician (i.e., either the supervising physician or the senior trainee). Trainees in a given tenure period are categorized as having "high exposure" to spending if this measure is above the median measure for trainees in the same tenure period. The difference in exposure to spending between high and low exposure is simply the average measure for high-exposure trainees subtracted by the average measure for low-exposure trainees in a given tenure period.

Table A-6: Effect of Trainee Experience on Spending

	Log daily total costs				
	(1) Number of days	(2) Number of patients	(3) Number of attendings	(4) Attending spending	(5) Attending spending
Panel A: Interns					
Effect of trainee with measure above median	0.001	0.003	-0.001	-0.010	-0.001
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Observations	182,500	182,500	182,500	156,545	131,654
Adjusted R^2	0.088	0.088	0.088	0.089	0.089
Panel B: Residents					
Effect of trainee with measure above median	0.005	-0.005	-0.001	0.010	0.013
	(0.007)	(0.008)	(0.007)	(0.005)	(0.005)
Observations	200,266	200,266	200,266	182,982	176,086
Adjusted R^2	0.089	0.089	0.089	0.086	0.086
Measure and median within service	Y	Y	Y	N	Y

Note: This table reports results for some regressions of the effect of indicators of trainee experience. Panel A shows results for interns; Panel B shows results for residents. Regressions are of the form in Equation (A-9), where the coefficient of interest is on an indicator for a group of trainees identified whether their measure (e.g., number of days) is above the median within a 60-day tenure interval (across all trainees). The relevant tenure interval is the tenure interval before the one related to the day of the index admission. All columns except for (4) represent measures and medians that are calculated within service (e.g., number of days is calculated separately for a trainee within cardiology, oncology, and general medicine and compared to medians similarly calculated within service). Columns 4 and 5 feature a measure of attending spending, which is the average cumulative effect of attending physicians who worked with the trainee of interest up to the last prior tenure interval. Attending "effects" are calculated by a random effects method that adjusts for finite-sample bias; since patients are not as good as randomly assigned to attending physicians, these effects do not have a strict causal interpretation at the level of the attending physician. Other specifications (e.g., calculating all measures across services, or not conditioning on trainee identity) were similarly estimated as insignificant and omitted from this table for brevity. 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.

Table A-7: Model Parameter Estimates for Overall Spending and by Spending Category

		Spending Category			
	(1)	(2)	(3)	(4)	(5)
	Overall	Diagnostic	Transfusion	Medication	Nursing
Knowledge parameters					
Prior to training (ρ_0)	0.039	0.936	0.225	1.005	0.000
	(0.032)	(0.235)	(0.341)	(0.043)	(0.000)
First year (ρ_1)	0.204	0.296	0.361		4.172
	(0.138)	(0.332)	(0.339)		(0.654)
Second year (ρ_2)	7.542	0.263	0.245		15.357
	(2.307)	(0.140)	(0.343)		(2.556)
Third year (ρ_3)	0.000	0.000			4.501
	(0.000)	(0.000)			(4.344)
Influence parameters					
Deviation after first year (δ_1)	0.231	4.388	0.349	0.725	−2.784
	(0.223)	(0.433)	(0.439)	(0.053)	(0.533)
Deviation after second year (δ_2)	−1.366	−0.682			−10.284
	(0.800)	(0.563)			(2.175)
Supervisory information (P)	3.678	0.000	0.941	3.755	4.326
	(0.503)	(0.000)	(0.150)	(0.181)	(0.427)
Likelihood ratio test p-value	0.003	0.009	0.001	N/A	0.022

Note: This table shows parameter estimates of the model of learning and influence described in Section IV.C and specified in Section IV.C. Parameters are estimated from reduced-form practice variation moments, as shown in Figure 1 overall (Column 1) and Figure 3 for each spending category (the remaining columns). Knowledge parameters represent units of precision as function of tenure, as in Equation (A-20): ρ_0 is precision of knowledge prior to training; ρ_1, ρ_2, and ρ_3 are increases in precision (learning) in the first, second, and third years, respectively. Influence parameters δ_1 and δ_2 are deviations from the Bayesian benchmark benchmark in terms of effective precision as a function of completed years of training, as given in Equation (A-21). Specifically, a trainee who has completed one year of training receives influence that is δ_1 more (if positive) or less (if negative) units of effective precision than the efficient benchmark would imply. Similarly, a trainee who has completed two years of training receives an additional deviation of δ_2 relative to the efficient benchmark. P is the effective precision of supervisory information, including knowledge from supervisors, consultants, rules, or information produced by the trainees. Cells with missing values indicate that the model was estimated with these values constrained to 0, as less-constrained models did not converge. The likelihood ratio test p-value compares the estimated model against a restricted model of no learning (i.e., only ρ_0, δ_1, and P are non-zero). Note that this test is not relevant for the medication model, as the estimated model is in fact a no-learning model. Standard errors are displayed in parentheses.

Table A-8: Model Parameter Estimates by Day of Stay and Patient Severity

	Day of Stay		Patient Severity	
	(1)	(2)	(3)	(4)
	Early	Late	High Severity	Low Severity
Knowledge parameters				
Prior to training (ρ_0)	0.076	0.006	0.091	0.060
	(0.056)	(0.000)	(0.078)	(0.059)
First year (ρ_1)	0.346	0.294	0.371	0.207
	(0.223)	(0.087)	(0.299)	(0.253)
Second year (ρ_2)	6.681	6.655	6.242	7.644
	(2.528)	(1.414)	(2.719)	(3.572)
Third year (ρ_3)	0.000	0.845	0.000	0.000
	(0.000)	(0.007)	(0.000)	(2.000)
Influence parameters				
Deviation after first year (δ_1)	0.271	0.192	0.294	0.204
	(0.288)	(0.198)	(0.315)	(0.300)
Deviation after second year (δ_2)	−0.912	−1.554	−1.347	−0.367
	(0.719)	(0.082)	(0.780)	(1.597)
Supervisory information (P)	3.850	3.495	3.725	3.759
	(0.545)	(0.419)	(0.608)	(0.622)
Likelihood ratio test p-value	0.151	0.000	0.020	0.182

Note: This table shows parameter estimates of the model of learning and influence described in Section IV.C. Columns correspond to models estimated on observations by patient-day: Columns 1 and 2 are for days respectively before or after the middle of each patient's stay; Columns 3 and 4 are for patients with above- or below-median expected 30-day mortality, respectively. Parameters are as described in the note for Table A-7 and are estimated from reduced-form practice variation moments, as shown in Figure 4 for type of patient-day. The likelihood ratio test p-value compares the estimated model against a restricted model of no learning (i.e., only ρ_0, δ_1, and P are non-zero). Standard errors are displayed in parentheses.