Appendix

A-1 Identification under Patient Selection

A-1.1 Quantifying Selection in Terms of Predicted Outcomes

The analyses to evaluate selection have focused on implications on the EOS effect on length of stay, because qualitative differences in observed patient characteristics suggest that slightly easier patients are assigned near EOS. Easier patients would suggest that cost, orders, and admissions should be *lower* near EOS, and therefore to the extent that there is any bias of from selection, estimated effects on these other outcomes should be conservative lower bounds of the true effect. This appendix section formally assesses this intuition and quantifies selection on observable characteristics with respect to predicted outcomes.

As above, consider predicted outcomes $\hat{Y}_{it}^{set}$, where $set \in \{ante, full\}$. I estimate regressions of predicted outcomes, similar to the baseline regression in Equation (1):

$$\hat{Y}_{it}^{set} = \alpha_{\overline{m}(i,t)} + \gamma_{\underline{m}(i,t)} + \eta \mathbf{T}_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_{it}, \tag{A-1.1}$$

leaving out variables in $\mathbf{X}_{it}$ as regressors. I interpret the coefficients $\alpha_{\overline{m}}$ as the amount patient selection in the $\overline{m}^{\text{th}}$ hour prior to EOS (compared to greater than 6 hours prior to EOS) according to the predicted outcome $\hat{Y}_{it}^{set}$.

Figure A-1.1 presents estimates of selection for each set of patient characteristics and for each of the outcomes of length of stay, orders, admission, and costs. To reference magnitude, selection estimates are overlaid onto estimates for the EOS effect from Equation (1) for each respective outcome. Coefficients for selection estimated using the two sets of characteristics are extremely similar, which supports Assumption 1. Selection nearing EOS appears to be in the direction of healthier patients: those expected to have shorter lengths of stay, lower admission probabilities, lower costs, and fewer orders. Predicted length of stay is 5.4% lower in the last hour prior to EOS compared to seven or more hours prior to EOS, about an order of magnitude smaller than effects for actual length of stay. All predicted outcomes show a *decreasing* relationship with proximity to EOS, in contrast to increases in actual admission, costs, and orders.

A-1.2 Assessing Identifying Assumptions

In Section 4.1 and Tables 1 and A-1.1, I assess two identifying assumptions: Assumption 1 is that patient selection occurs only via *ex ante* patient characteristics, which I observe in the data. Other patient characteristics that are correlated with potential outcomes are mostly unknown to physicians before assignment and, under this assumption, excluded from assignment policies. This assumption would be violated if physicians and triage nurses make further assessments of patient severity and use these assessments to assign to physicians according to their time to

EOS. In this setting, such violations are less likely because triage and assignment is specifically supposed to be summarized by the sufficient statistic of ESI, which is an *ex ante* characteristic observable in the data.

I assess Assumption 1 by considering an *ex post* characteristic X^{post} and estimating the following regression:

$$X_{it}^{post} = \alpha_{\overline{m}(i,t)} + \gamma_{\underline{m}(i,t)} + \beta \mathbf{X}_i^{ante} + \eta \mathbf{T}_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_t. \tag{A-1.2}$$

Excludable characteristics (and any linear combination of them, such as an outcome prediction based on them) should be mean independent of time relative to EOS, conditional on *ex ante* characteristics, time categories, pod identities, and provider identities. Therefore, under this assumption, the set of coefficients $\{\alpha_{\overline{m}}\}$ in Equation (A-1.2) should be jointly insignificant. Results are shown in Table 1, Columns 1 and 2, and in Table A-1.1, Panel A.

Assumption 2 is that, conditional on time categories, pod identities, and provider identities, patient potential outcomes are unrelated to specific ED arrival times that drive the propensity for assignment to a physician nearing EOS. Specifically, variation in shift schedules within a time category of ED arrival drives the propensity of being assigned to a physician near EOS but is mean independent of potential outcomes of the arriving patients. This assumption would be violated if patients were to know ED shift schedules and arrive according to their severity and the propensity to be assigned to a physician ending her shift. The assumption would also be violated if the ED changed EOS times to meet changing patient severity within time categories. These violations do not seem likely because schedule changes are primarily driven by changes in physical capacity and physician availability (e.g., new hires).

To assess Assumption 2, I regress *any* patient characteristic or linear combination of characteristics as follows:

$$X_{it} = \sum_{\overline{m}=6}^{1} \alpha_{\overline{m}} P_{\overline{m}}(t) + \eta \mathbf{T}_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_t. \tag{A-1.3}$$

$P_{\overline{m}}(t)$ represents the fraction of patients arriving at time t who are assigned to $\overline{m}$ hours prior to EOS, where the omitted category is 7 or greater hours. Results are in Table 1, Columns 3 to 8, and in Table A-1.1, Panel B. Assumption 2 formally conditions on pod and provider identities because in my preferred specification I use arrival hour at the ED floor as the primary measure of arrival time t. Conditioning on pod and provider identities is of course part of my baseline specification for estimating EOS effects in Equation (1).

A-1.3 Eliminating Selection between Physicians (Chetty et al, 2014)

This appendix section considers additional robustness checks using variation across arrival times, similar to Assumption 2. I only use variation in the overall composition of ED shifts at the patient's time of arrival. Because I control for hour of the day, day of the week, and month-year interactions, correlations between patient arrival and underlying ED shift structure would have to be conditional on these time categories.

This approach is closely related to one used by Chetty et al. (2014). First, I estimate "leave-shift-out" (jackknife) EOS effects specific to shift s, using Equation (1) on all observations except those corresponding to s (Jacob et al., 2010). I denote these estimates as $\left\{\hat{\alpha}_{\overline{m}}^{-s}\right\}$. This method thus excludes any idiosyncratic selection within a shift that would otherwise introduce bias into $\{\hat{\alpha}_{\overline{m}}\}$. Next, I construct hourly patient-weighted averages (at the level of the entire ED) that represent the overall ED shift environment at hour t. That is, for patients i arriving at time $t_i^a = t$, construct the average EOS effect

$$Q_t \equiv \frac{\sum_i \mathbf{1}\left(t_i^a = t\right) \sum_{\overline{m}=1}^{6} \hat{\alpha}_{\overline{m}}^{-s} \mathbf{1}\left(\left\lceil \bar{t}\left(J\left(i,t\right),t\right) - t \right\rceil = \overline{m}\right) \mathbf{1}\left(S\left(J\left(i,t\right),t\right) = s\right)}{\sum_i \mathbf{1}\left(t_i^a = t\right)}, \tag{A-1.4}$$

where $J\left(i,t\right)$ is a physician assignment function for patient i at time t, and $S\left(j,t\right)$ is a shift assignment function for physician j at time t.

A-1.3.1 Forecast Bias

I first evaluate systematic bias in baseline estimates by using only cross-time variation in hourly averages. I construct hourly averages Y_t of (residualized) length of stay $\tilde{Y}_{it}$:

$$Y_t \equiv \frac{\sum_i \mathbf{1}\left(t_i^a = t\right) \tilde{Y}_{it}}{\sum_i \mathbf{1}\left(t_i^a = t\right)} \tag{A-1.5}$$

where

$$\tilde{Y}_{it} \equiv Y_{it} - \left[\hat{\gamma}_{\underline{m}(i,t)} + \hat{\beta}\mathbf{X}_i + \hat{\eta}\mathbf{T}_t + \hat{\zeta}_{p(i)} + \hat{\nu}_{j(i)}\right]. \tag{A-1.6}$$

Coefficients $\hat{\gamma}_{\underline{m}}$, $\hat{\beta}$, $\hat{\eta}$, $\hat{\zeta}_p$, and $\hat{\nu}_j$ are estimated using within-EOS variation from an equation very similar to Equation (1):

$$Y_{it} = \alpha_{\overline{m}(i,t)} + \gamma_{\underline{m}(i,t)} + \beta\mathbf{X}_i + \eta\mathbf{T}_t + \zeta_{p(i)} + \nu_{k(i)} + \varepsilon_{it},$$

where I use attending-physician fixed effect $\nu_{k(i)}$ instead of physician-team fixed effects $\nu_{j(i),k(i)}$ to broaden the number of observations for which I observe an identified residual. This approach, which includes effects for time to EOS, only uses within-EOS-time variation to estimate coefficients and therefore provides consistent estimates even if the covariates are correlated with time relative to EOS.

I compare these residualized outcomes with average EOS effect Q_t from Equation (A-1.4) with the regression

$$Y_t = a + bQ_t + \chi_t. \tag{A-1.7}$$

This regression yields an estimate of "forecast bias" due to systematic selection of patients arriving within t across physicians,

$$\begin{aligned} Bias\left(\hat{\alpha}_{\overline{m}(i,t)}^{-s(j(i),t)}\right) &= \mathrm{Cov}\left(\varepsilon_{it}, \hat{\alpha}_{\overline{m}(i,t)}^{-s(j(i),t)}\right) / \mathrm{Var}\left(\hat{\alpha}_{\overline{m}(i,t)}^{-s(j(i),t)}\right) \\ &= 1 - b, \end{aligned}$$

under the assumption that

$$\mathrm{Cov}\left(Q_t, \chi_t\right) = 0. \tag{A-1.8}$$

This assumption is similar to Assumption 2, that there is no selection of unobservable patient types across average ED times relative to EOS, conditional on time categories, observable characteristics, pod, and physician, but is even less restrictive in the sense that Q_t is formed by leave-shift-out estimates.

Column 1 of Table A-1.2 reports estimates of b from Equation (A-1.7). The point estimate of b is 1.029 with a robust standard error of 0.060 (clustered at each hour of t), which reflects tight estimation indistinguishable from 1 (i.e., I cannot reject the hypothesis of $\hat{\alpha}_{\overline{m}}^{-s} - \alpha_{\overline{m}} = 0$). That is, under the assumption in (A-1.8), I cannot reject the null of no unobservable selection across physicians within time (i.e., Assumption 1). Panel A of Figure A-1.2 plots the relationship between Y_t and Q_t nonparametrically, dividing the data into 20 equal-sized groups ("vigintiles") according to Q_t. This plot nonparametrically represents the expectation function of Y_t conditional on Q_t. The relationship is highly linear, with slope close to 1.

A-1.3.2 Observable Selection between Hours

I use a similar exercise to consider the amount of selection on observables across hours in order to support the assumption in Equation (A-1.8) (and Assumption 2). Similar to the analysis in Appendix A-1.2, I form predictions about length of stay using two sets of patient characteristics. The first is the set of *ex ante* characteristics $\mathbf{X}^{ante}$ that are observable to the physician prior to assignment, while the second set $\mathbf{X}^{full}$ is a superset (simply referred to by $\mathbf{X}$ in the main text) that also includes *ex post* clinical characteristics that generally unobserved to the physician until after assignment. I average predictions for all patients within a given hour, again eliminating selection across physicians within hour. This exercise therefore evaluates the degree of selection (on observable characteristics) remaining across hours.

For each variable in $\mathbf{X}^{full}$, I form residualized variables obtained after subtracting predictions of each variable based on time categories, $\mathbf{T}_t$, and indicators for hours relative to shift beginning, $\underline{m} = \lfloor t - \underline{t}(j,t) \rfloor$. Using residualized characteristics in each set, I construct respective predictions $\hat{Y}^{ante}$ and $\hat{Y}^{full}$. Similar to Equation (A-1.5), I average these predictions over all patients arriving at a given hour:

$$\hat{Y}_t^{set} \equiv \frac{\sum_i \mathbf{1}\left(t_i^a = t\right) \hat{Y}_{it}^{set}}{\sum_i \mathbf{1}\left(t_i^a = t\right)}, \tag{A-1.9}$$

where t denotes an hour, and $set \in \{ante, full\}$.

The regression

$$\hat{Y}_t^{set} = a + b^{set} Q_t + \chi_t \tag{A-1.10}$$

quantifies the degree of selection across hours, as predicted by characteristics $\mathbf{X}^{set}$: Under (A-1.8), $b^{set} = \text{Cov}\left(E\left[\hat{\alpha}_{\overline{m}}^{-s}|t\right], \hat{Y}_t^{set}\right) / \text{Var}\left(E\left[\hat{\alpha}_{ms}|t\right]\right)$ for $\overline{m} = \lceil \overline{t}(j,t) - t \rceil$ and set $s = S(j,t)$. Although the assumption in Equation (A-1.8) is not directly testable, a lack of observable selection (b^{set} is indistinguishable from 0) supports this assumption.

Columns 2 and 3 of Table A-1.2 report of estimates b^{ante} and b^{full}, respectively, from Equation (A-1.10). Both estimates are small and indistinguishable from 0: The point estimate of b^{ante} is 0.029 (robust standard error 0.025), and the point estimate of b^{full} is 0.024 (robust standard error 0.026). Panels B and C of Figure A-1.2 show corresponding nonparametric expectations of $\hat{Y}_t^{ante}$ and $\hat{Y}_t^{full}$, respectively conditional on Q_t, where the data is again divided into vigintiles of Q_t. The relationship is again linear, but consistent with the regression results, there is no relationship between length of stay predicted by time relative to EOS (Q_t) and that predicted by patient characteristics.

A-1.4 Required Selection on Unobservables (Altonji et al, 2005)

This appendix section details a procedure similar to that outlined in Altonji et al. (2005). The goal of this exercise is to quantify the amount of selection on unobservables necessary to explain decreases in length of stay for patients assigned at each hour near EOS. The basic intuition is that the possibility that selection on unobservables explains estimated effects can be quantified by the extents to which selection and outcomes can be explained by observables.

A-1.4.1 Conceptual Framework

Consider a condensed form of the outcomes regression Equation (1):

$$\begin{aligned} Y &= \sum_{\overline{m}} \alpha_{\overline{m}} A_{\overline{m}} + \mathbf{\Omega}'\mathbf{\Gamma} \\ &= \sum_{\overline{m}} \alpha_{\overline{m}} A_{\overline{m}} + \mathbf{W}'\mathbf{\Gamma}_W + \xi, \end{aligned} \tag{A-1.11}$$

where I omit subscripts for simplicity. I define $A_{\overline{m}} \equiv \mathbf{1}\left(\lceil \overline{t}(j,t) - t \rceil = \overline{m}\right)$ for whether the time t that patient i was assigned to physician j was in the $\overline{m}^{\text{th}}$ hour from j's EOS. $\alpha_{\overline{m}}$ is the causal effect of a patient being assigned in the $\overline{m}^{\text{th}}$ hour prior to EOS. $\mathbf{\Omega}$ is the full set of other variables, both observed and unobserved, that determine outcome Y, while $\mathbf{W}$ includes only observed patient, time, and provider characteristics (to be distinguished from $\mathbf{X}_{it}$ in Equation (1), which only includes patient characteristics). $\mathbf{\Gamma}$ is the causal effect of $\mathbf{\Omega}$ on Y. $\mathbf{\Gamma}_W$ is the subvector of $\mathbf{\Gamma}$ that corresponds to $\mathbf{W}$ within $\mathbf{\Omega}$, and ξ is an index of the unobserved variables.

Since variables in $\mathbf{W}$ are likely correlated with ξ, rewrite Equation (A-1.11) as

$$Y = \sum_{\overline{m}} \alpha_{\overline{m}} A_{\overline{m}} + \mathbf{W}'\gamma_W + \varepsilon, \tag{A-1.12}$$

where γ_W and ε are constructed so $\text{Cov}\,(\varepsilon, \mathbf{W}) = 0$ *by definition*. Thus γ_W captures both the causal effect of $\mathbf{W}$ on Y, or$\mathbf{\Gamma}_W$, as well as the portion of ξ that may be correlated with $\mathbf{W}$. Note that, for the regression estimate of α_m to be unbiased, the standard OLS assumption is that $\text{Cov}\,(\varepsilon, A_{\overline{m}}) = 0$, or $E\left[\varepsilon \,|A_{\overline{m}} = 1\right] - E\left[\varepsilon \,|A_{\overline{m}} = 0\right] = 0$.

A-1.4.2 Measure of Selection on Unobservables

Altonji et al. (2005) argue for upper bound of selection on unobservables, specified by

$$\frac{E\left[\varepsilon \,|A_{\overline{m}} = 1\right] - E\left[\varepsilon \,|A_{\overline{m}} = 0\right]}{\text{Var}(\varepsilon)} = \frac{E\left[\mathbf{W}'\gamma_W \,|A_{\overline{m}} = 1\right] - E\left[\mathbf{W}'\gamma_W \,|A_{\overline{m}} = 0\right]}{\text{Var}(\mathbf{W}'\gamma_W)}, \tag{A-1.13}$$

which states that the relationship between the index of unobservables in Equation (A-1.12) and the indicator for selection $A_{\overline{m}}$ is equal in magnitude to the relationship between observable predictors of Y and $A_{\overline{m}}$, respectively normalizing for variance.

They argue that this condition represents an upper bound because of observed variables are not randomly collected but rather represent characteristics that are collected precisely because they are more important for outcomes of interest. Furthermore, because many observed variables are in fact collected after the selection event, they include random shocks that cannot have influenced the selection event. This latter argument is related to the fact that patient clinical characteristics are generally unknown by the physician at the time of assignment.

A-1.4.3 Estimation of Potential Bias

In order to estimate the potential bias at the upper bound implied by Equation (A-1.13), consider the following linear selection equation:

$$A_{\overline{m}} = \mathbf{W}'\beta_W^{\overline{m}} + A_{\overline{m}}^*, \tag{A-1.14}$$

where $A_{\overline{m}}^*$ is a residual that is orthogonal to $\mathbf{W}$. Then Equation (A-1.12) can be stated as

$$Y = \sum_{\overline{m}} \alpha_{\overline{m}} A_{\overline{m}}^* + \mathbf{W}' \left(\gamma_W + \sum_{\overline{m}} \alpha_{\overline{m}} \beta_W^{\overline{m}} \right) + \varepsilon.$$

This leads to a statement of the potential bias due to selection on unobservables:

$$\begin{aligned}\operatorname{plim}\hat{\alpha}_{\overline{m}} &\approx \alpha_{\overline{m}} + \frac{\operatorname{Cov}\left(A_{\overline{m}}^{*},\varepsilon\right)}{\operatorname{Var}\left(A_{\overline{m}}^{*}\right)} \\ &= \alpha_{\overline{m}} + \frac{\operatorname{Var}\left(A_{\overline{m}}\right)}{\operatorname{Var}\left(A_{\overline{m}}^{*}\right)}\left(E\left[\varepsilon\left|A_{\overline{m}}=1\right.\right]-E\left[\varepsilon\left|A_{\overline{m}}=0\right.\right]\right),\end{aligned}$$

From Equation (A-1.13), the bias can be stated in terms of $E\left[\mathbf{W}'\gamma_W\left|A_{\overline{m}}=1\right.\right]-E\left[\mathbf{W}'\gamma_W\left|A_{\overline{m}}=0\right.\right]$:

$$\text{Bias} = \frac{\operatorname{Var}\left(A_{\overline{m}}\right)\operatorname{Var}\left(\varepsilon\right)}{\operatorname{Var}\left(A_{\overline{m}}^{*}\right)\operatorname{Var}\left(\mathbf{W}'\gamma_W\right)}\left(E\left[\mathbf{W}'\gamma_W\left|A_{\overline{m}}=1\right.\right]-E\left[\mathbf{W}'\gamma_W\left|A_{\overline{m}}=0\right.\right]\right) \tag{A-1.15}$$

Under the null hypothesis that $\alpha_{\overline{m}} = 0$, γ_W can be consistently estimated by Equation (A-1.11).

I can then arrive at a consistent estimate of bias in Equation (A-1.15) with the following procedure, with results shown in Table A-1.3: For each $\overline{m} \in \{-6,\ldots,-1\}$, I define $A_{\overline{m}}$ over all observations and empirically calculate $\widehat{\operatorname{Var}}\left(A_{\overline{m}}\right)$. I also calculate $\widehat{\operatorname{Var}}\left(A_{\overline{m}}^{*}\right)$ after estimating Equation (A-1.14) for each $\overline{m}$. Similarly, I estimate $\widehat{\operatorname{Var}}\left(\varepsilon\right) = 0.160$ and $\widehat{\operatorname{Var}}\left(\mathbf{W}'\gamma_W\right) = 0.580$ from Equation (A-1.12). Equation (A-1.12) also allows me to form an estimate of selection on observables, $\hat{E}\left[\mathbf{W}'\gamma_W\left|A_{\overline{m}}=1\right.\right]-\hat{E}\left[\mathbf{W}'\gamma_W\left|A_{\overline{m}}=0\right.\right]$, for each $\overline{m}$. Using the condition in Equation (A-1.13) that normalized selection on unobservables is bounded by normalized selection on observables, I then calculate an upper bound of the bias due to selection on unobservables with Equation (A-1.15). As shown in Table A-1.3, the upper bound of the bias in $\hat{\alpha}_1$, the effect of arriving in the last hour of shift on the length of stay, estimated by Equation (1), is -0.00124. Given that $\hat{\alpha}_1 = -0.5873$, this implies that normalized selection on unobservables would have to be 475 times greater than normalized selection on observables. As a comparison, in their example of the impact of Catholic school on educational attainment, Altonji et al. (2005) argue that selection on unobservables is highly unlikely with a ratio 3.55.

A-2 Effects Relative to Shift Beginning

The literature on shift work has almost exclusively focused on cumulative health effects and fatigue (e.g., Brachet et al., 2012; Shetty and Bhattacharya, 2007; Volpp and Rosen, 2007), while I explore the possibility of strategic behavior in this paper. Unlike shifts of 36 hours in the residency work-hours debate, significant fatigue is less likely near the end of a shift of nine hours, the modal shift length in this setting. Nonetheless, I specifically address this issue by exploiting variation in shift length to control for effects, such as fatigue, correlated with time since the beginning of shift. I assume that, conditional on time since beginning of shift, fatigue is independent of time to EOS.

In the full model of Equation (1), I show robust EOS effects controlling for time since the beginning of shift. The effect attributable to time since shift beginning is minor compared to the overall effect for length of stay. Here I illustrate the robustness of EOS effects more directly by simply showing the effect on length of stay for each hour prior to EOS separately for three categories of shift lengths. I study shifts that are nine hours in length, as well as shifts that are seven or eight hours in lengths and shifts that are ten hours in length. Figure A-2.1 plots coefficients α_m from Equation (1) estimated separately for each shift-length category. Panel A plots coefficients according to time relative to EOS and shows coefficients largely similar across shift lengths and within hour prior to EOS. Panel B arranges the coefficients according to time from shift beginning, illustrating the corollary that the EOS effect is largely independent of the time since beginning the shift.

A-3 Time Components of Length of Stay

In Section 4, length of stay decreases while formal utilization increases near EOS. This suggests that formal utilization is a net substitute for time in patient care. In this appendix, I further examine this hypothesis by a closer look at the time components of length of stay. In practice, time is not neatly divided into pure substitute or complement components with formal utilization (call these components τ_1 and τ_2, respectively), but some intuitive distinctions can be made: Time before the first formal order likely belongs to τ_1 (e.g., time spent interviewing the patient or performing serial abdominal examination as opposed to CT scan). Time after the last formal order likely belongs to τ_2, reflecting time needed to follow up on utilization (e.g., waiting for CT scan report). Although time in between the first and last orders could belong to either τ_1 or τ_2, the spacing of these orders often reflects clinical monitoring and reasoning more closely related to τ_1.

Measuring length of stay in three component shares – time between pod arrival and first order, time between first and last (non-discharge) orders, and time between last and discharge orders – I estimate a fractional logit model (Papke and Wooldridge, 1996) using similar regressors as in Equation (1). Figure A-3.1 presents results of marginal effects relative to EOS. Panel A scales time shares by the median predicted length of stay in each hour prior to EOS according to Equation (1); Panel B simply plots the unscaled proportional shares. These proportions remain relatively unchanged except for the last hour prior to EOS, when the proportions for time prior to first order and inter-order time both decrease. These results suggest relative reductions in τ_1, particularly in the last hour prior to EOS, and are consistent with the increase in formal utilization (net substitution) in the last hour shown in Table 3 and Figure 6.

A-4 Stylized Model Proposition Proofs

A-4.1 Proof of Proposition 1

Denote inputs in Section 6.1 that maximize expected utility in Equation (7), conditional on patient assignment ($a = 1$), as $(\tau^*(t), z^*(t))$. Denote corresponding inputs that maximize welfare as

$\left(\tau^{FB}(t), z^{FB}(t)\right)$. *Assume that* $F_{p'}(p^*) < \frac{1}{2}$.

(a) As $t \to \bar{t}$, $\tau^*(t)$ *weakly decreases,* $z^*(t)$ *may weakly increase (if* τ *and* z *are net substitutes) or decrease (if* τ *and* z *are net complements), and* $E[d|\tau^*(t), z^*(t)]$ *weakly increases.*

Proof. The partial derivative of $E[U|\, a = 1]$ with respect to τ is

$$\frac{\partial E[U|\, a = 1]}{\partial \tau} = -\frac{\partial \tilde{c}_\tau}{\partial \tau} + \lambda\left(\frac{\partial}{\partial \tau} E\left[v\left(\theta, d\left(\hat{\theta}\right)\right)\right] - \frac{\partial c}{\partial \tau}\right). \tag{A-4.1}$$

Define $\Delta_v^{\theta=1} \equiv v(1,1) - v(1,0)$ and $\Delta_v^{\theta=0} \equiv v(0,0) - v(0,1)$. Using this notation, note that

$$E\left[v\left(\theta, d\left(\hat{\theta}\right)\right)\right] = \begin{cases} E[v(\theta,0)] + \Delta_v^{\theta=1} p' q, & p' < p^* \\ E[v(\theta,1)] + \Delta_v^{\theta=0}(1-p')\, q, & p' \geq p^* \end{cases},$$

which implies that

$$\frac{\partial}{\partial \tau} E\left[v\left(\theta, d\left(\hat{\theta}\right)\right)\right] = \begin{cases} \Delta_v^{\theta=1} p' \frac{\partial q}{\partial \tau}, & p' < p^* \\ \Delta_v^{\theta=0}(1-p') \frac{\partial q}{\partial \tau}, & p' \geq p^* \end{cases}. \tag{A-4.2}$$

From Equation (A-4.1), the first-order condition with respect to τ is

$$\frac{\partial}{\partial \tau} E\left[v\left(\theta, d\left(\hat{\theta}\right)\right)\right] = \frac{1}{\lambda}\frac{\partial \tilde{c}_\tau}{\partial \tau} + \frac{\partial c}{\partial \tau}.$$

The only part of this relationship that depends on t is $\partial \tilde{c}_\tau / \partial \tau$. As $t \to \bar{t}$, $\partial \tilde{c}_\tau / \partial \tau$ weakly increases, which implies that τ^* weakly decreases as a function of t. Note that this relationship is stronger when λ is smaller (i.e., when the physician cares less about patient outcomes relative to her leisure and income).

The corresponding first-order condition with respect to z is

$$\frac{\partial}{\partial z} E\left[v\left(\theta, d\left(\hat{\theta}\right)\right)\right] = \frac{\partial c}{\partial z},$$

where

$$\frac{\partial}{\partial z} E\left[v\left(\theta, d\left(\hat{\theta}\right)\right)\right] = \begin{cases} \Delta_v^{\theta=1} p' \frac{\partial q}{\partial z}, & p' < p^* \\ \Delta_v^{\theta=0}(1-p') \frac{\partial q}{\partial z}, & p' \geq p^* \end{cases}.$$

If z and τ are complements (i.e., if $\partial^2 q / (\partial z \partial \tau) > 0$), then $\partial q / \partial z$ decreases as τ decreases. This implies that z^* decreases as τ^* decreases. Given that τ^* weakly decreases as a function of t, z^* weakly decreases as a function of t. On the other hand, if z and τ are substitutes (i.e., if $\partial^2 q / (\partial z \partial \tau) < 0$), then $\partial q / \partial z$ increases as τ decreases. This implies that z^* increases as τ^* decreases. Given that τ^* weakly decreases as a function of t, z^* weakly increases as a function of t.

As $t \to \bar{t}$, the cost of producing q weakly increases since $\partial \tilde{c}_\tau / \partial \tau$ weakly increases. Thus $q\left(\tau^*\left(t\right), z^*\left(t\right)\right)$ weakly decreases in t. If $\Pr\left(p' > p^*\right) > \Pr\left(p' < p^*\right)$ (i.e., $F_{p'}\left(p^*\right) > \frac{1}{2}$ as assumed), then $E\left[d\,|\tau^*\left(t\right), z^*\left(t\right)\right]$ weakly increases since the probability that $\hat{\theta} = \varnothing$ (i.e., $1-q$) weakly increases, and discharge is the default when $p' > p^*$. □

(b) For all t, $\tau^*\left(t\right) \leq \tau^{FB}\left(t\right)$, and $E\left[d|\tau^*\left(t\right), z^*\left(t\right)\right] \geq E\left[d|\tau^{FB}\left(t\right), z^{FB}\left(t\right)\right]$.

Proof. Now define the first best, by $\lambda = 1$ (relative to $\lambda > 1$). For any t,

$$\frac{1}{\lambda}\frac{\partial \tilde{c}_\tau}{\partial \tau} \geq \frac{\partial \tilde{c}_\tau}{\partial \tau},$$

which implies that $\tau^{FB} \geq \tau^*$. Similarly, $q^{FB} \geq q^*$, which implies that $E\left[d|\tau^*\left(t\right), z^*\left(t\right)\right] \geq E\left[d|\tau^{FB}\left(t\right), z^{FB}\left(t\right)\right]$ *if* $F_{p'}\left(p^*\right) < \frac{1}{2}$. □

(c) If τ and z are net substitutes, then $z^*\left(t\right) > z^{FB}\left(t\right)$ for all t, and $z^*\left(t\right) - z^{FB}\left(t\right)$ weakly increases in w_t, holding t constant. The reverse is true if τ and z are net complements.

Proof. By similar argument, if $\tau^{FB} \geq \tau^*$, then $z^* \geq z^{FB}$ if τ and z are net substitutes, and $z^* \leq z^{FB}$ if τ and z are net complements. If $\partial^2 q / \left(\partial \tau \partial w_t\right) < 0$, then $\tau^{FB} - \tau^*$ weakly increases with w_t for any t. This implies that $z^* - z^{FB}$ weakly increases with w_t if τ and z are net substitutes, or weakly decreases with w_t if τ and z are net complements. Regardless of whether τ and z are net substitutes or net complements, $\left|z^* - z^{FB}\right|$ weakly increases with w_t for any t. □

A-4.2 Proof of Proposition 2

Consider a^* as the patient assignment in Section 6.1 that maximizes expected utility in Equation (7), a^{FB} as the assignment that maximizes expected welfare when optimal $\left(\tau^{FB}, z^{FB}\right)$ is publicly known and contractible, and a^{SB} as the assignment that maximizes expected welfare when $\left(\tau^{FB}, z^{FB}\right)$ is either publicly unknown or non-contractible. Assignment will follow threshold rules in which assignment occurs if and only if $E\left[O\left(\theta; \mathcal{E}_t\right)\right]$ is greater than a threshold. The respective threshold rules are $\underline{O}^*$, $\underline{O}^{FB}$, and $\underline{O}^{SB}$, where $\underline{O}^* < \underline{O}^{SB} < \underline{O}^{FB}$. $\underline{O}^{FB} - \underline{O}^{SB}$ and $\underline{O}^{SB} - \underline{O}^*$ increase as $t \to \bar{t}$ decreases or as λ decreases.

Proof. The expected utility under $a = 0$ is $E\left[O\left(\theta; \mathcal{E}_t\right)\right]$, and the expected utility under $a = 1$ is

$$E\left[U|\, a=1\right] = y + \max_{\mathbf{z}}\left\{\lambda\left(E\left[v\left(\theta, d\left(\hat{\theta}\right)\right)\middle|\,\tau, z\right] - c\left(\tau, z\right)\right) - \tilde{c}_\tau\left(\tau\right)\right\},$$

where

$$E\left[v\left(\theta, d\left(\hat{\theta}\right)\right)\middle|\,\tau, z\right] = \begin{cases} E\left[v\left(\theta, 0\right)\right] + \Delta_v^{\theta=1} p q\left(\tau, z\right), & p < p^* \\ E\left[v\left(\theta, 1\right)\right] + \Delta_v^{\theta=0}\left(1-p\right) q\left(\tau, z\right), & p \geq p^* \end{cases}.$$

Note that at the time of patient assignment, the physician only knows the (publicly known) probability that $\theta = 1$, p, instead of the probability p' privately observed by the physician after assignment.

Denote $\underline{O}^*$ as the threshold rules such that accepting the patient maximizes expected utility ($a^* = 1$) if and only if $E\left[O\left(\theta; \mathcal{E}_t\right)\right] > \underline{O}^*$. Define $W\left(\tau, z\right) \equiv E\left[v\left(\theta, d\left(\hat{\theta}\right)\right)\middle| \tau, z\right] - c\left(\tau, z\right) - \tilde{c}_\tau\left(\tau\right)$ as the net expected social welfare from assigning the patient to the physician, conditional on patient care inputs (τ, z). It is easy to see that $\underline{O}^* = W\left(\tau^*, z^*\right) - \left(\lambda^{-1} - 1\right)\tilde{c}_\tau\left(\tau^*\right)$, where the second term represents the wedge from the physician overvaluing her leisure relative to other welfare-relevant objects. The corresponding threshold that determines the first-best assignment a^{FB} is $\underline{O}^{FB} = W\left(\tau^{FB}, z^{FB}\right)$, when optimal $\left(\tau^{FB}, z^{FB}\right)$ can be implemented. Finally, consider the second-best assignment policy, in which the patient may be assigned as a policy, $a^{SB} \in \{0, 1\}$, but the physician controls (τ, z). In this policy, $a^{SB} = 1$ if and only $E\left[O\left(\theta; \mathcal{E}_t\right)\right] > \underline{O}^{SB} = W\left(\tau^*, z^*\right)$, because the physician will choose inputs $\left(\tau^*, z^*\right)$ downstream.

Since $\left(\tau^{FB}, z^{FB}\right)$ maximizes W, it must be that $W\left(\tau^{FB}, z^{FB}\right) \geq W\left(\tau^*, z^*\right)$. Furthermore, Proposition 1 shows us how $\left(\tau^*, z^*\right) \neq \left(\tau^{FB}, z^{FB}\right)$ given $\lambda < 1$ and when EOS distortions are binding through $\tilde{c}_\tau > 0$. In other words, when $\tilde{c}_\tau > 0$, $W\left(\tau^{FB}, z^{FB}\right) > W\left(\tau^*, z^*\right)$. This implies that $\underline{O}^{SB} < \underline{O}^{FB}$. Furthermore, if $\tilde{c}_\tau\left(\tau^*\right) > 0$ and $\lambda < 1$, $\underline{O}^* < \underline{O}^{SB}$.

As $t \to \bar{t}$ or as λ decreases, $\left(\lambda^{-1} - 1\right)\tilde{c}_\tau\left(\tau^*\right)$ increases, and $W\left(\tau^{FB}, z^{FB}\right) - W\left(\tau^*, z^*\right)$ increases because $\left(\tau^*, z^*\right)$ is increasingly distorted (Proposition 1). This implies that $\underline{O}^{FB} - \underline{O}^{SB}$ and $\underline{O}^{SB} - \underline{O}^*$ increase as $t \to \bar{t}$ or as λ decreases. $\square$

A-5 Counterfactual Simulations

This appendix details the procedure to simulate outcomes under counterfactual assignment policies, as discussed at a high level in Section 7. To summarize, I first estimate a dynamic discrete choice model. Second, I construct counterfactual assignment policies and use the dynamic discrete choice model to compute counterfactual physician discharge choice probabilities under these policies. I use the counterfactual assignment policies and conditional choice probabilities to simulate patient arrivals and discharges. Third, I use the sequences of simulated assignments and discharges to impute welfare-relevant costs of physician time, patient time, and hospital resources.

A-5.1 Dynamic Programming Model

Consider a doctor-time observation (j, t), in state $S_{j,t}$. To simplify the decision space, I model time in five-minute intervals and assume that the doctor may discharge at most one patient $i \in \mathcal{I}\left(j, t\right)$ in the interval, including $i = \varnothing$ indicating no discharge of any patient. Denote the same doctor in the next period as (j, t'), in the corresponding state $S_{j,t'}$ simply S'. Suppressing notation for j and t for now, the integrated value function is

$$V(S) = E\left[\max_{i\in\mathcal{I}}\left\{u(i,S) + \delta\int_{S'} V(S')\, dF(S'|i,S)\right\}\right]. \tag{A-5.1}$$

$u(i,S)$ includes both a fixed component of utility and a random error term, ε_i, i.i.d. as Type I extreme value:

$$u(i,S) = \overline{u}(i,S) + \varepsilon_i.$$

$F(S'|i,S)$ is the Markov transition probability function, and implicit in this transition function is the assignment policy function, mapping S to probabilities of new patient assignments.

The corresponding choice-specific value function (not including the error term) is

$$\overline{v}(i,S) = \overline{u}(i,S) + \delta\int_{S'} V(S')\, dF(S'|i,S). \tag{A-5.2}$$

The physician chooses $i^* = \arg\max_{i\in I}(\overline{v}(i,S) + \varepsilon_i)$. Given that ε_i is distributed as Type I extreme value, this implies the physician's conditional discharge choice probabilities, $\Pr(i|S)$:

$$\Pr(i|S) = \frac{\exp(\overline{v}(i,S))}{\sum_{i'\in\mathcal{I}}\exp(\overline{v}(i',S))}. \tag{A-5.3}$$

The expected utility flow and subsequent value function from the choice i^* is in fact the integrated value function and can be stated differently using Euler's constant γ:

$$\begin{aligned} V(S) &= \gamma + \log\left[\sum_{i\in I}\exp(\overline{v}(i,S))\right] \\ &= \gamma + \log\left[\sum_{i\in I}\exp\left(\overline{u}(i,S) + \delta E\left[V(S')\,|i,S\right]\right)\right]. \end{aligned} \tag{A-5.4}$$

The setting has a finite horizon and is non-stationary. Because assignment policies do not differ past EOS (i.e., no patients are assigned past EOS), discharge policies past EOS will be the same in any counterfactual assignment policy. I therefore take EOS as the terminal period of the dynamic programming problem to estimate, bypassing the fact that the time a physician leaves work is unobserved. In other words, I directly estimate conditional choice probabilities for states past EOS using a simple logit model, and I use these choice probabilities regardless of any potential counterfactual assignment policy prior to EOS.

A-5.1.1 Estimating the Transition Probability Function

Considering time in discrete five-minute intervals, I estimate the probability of being assigned a given number of patients. In 23,990 shifts in the study period ranging from June 2005 to December 2012, I observe 1,151,888 observations over time t and shift s. Of 370,843 patients arriving during valid times, I further restrict the estimation sample to arrivals and discharges of 350,053 patients whose length of stay is at most twelve hours and who arrived at most twelve hours prior to EOS. The remaining 20,790 patients, whom I denote as $i \in I^{outside}$, are therefore not modeled in either arrivals or discharges, but I count them toward workload defined below. In simulations described below in Appendix A-5.3, I take arrivals and discharges of patients $i \in I^{outside}$ as fixed in every simulation.

I estimate an ordered logit model of the number of patients assigned at t to shift s, $a(s,t) \in \{0,1,2,3\}$. I consider the shift type $\langle \ell, \underline{o}, \overline{o} \rangle_s$, the time of EOS $\overline{t}(s)$, and physician j's census (or workload) $w(j,t-1)$ in the previous period (for j satisfying $s(j,t) = s$), the number of patients assigned in the last hour, the hour of the day, and the pod as the relevant variables in S determining assignments. $w(j,t)$ is defined in Equation (3), which I slightly rephrase here as

$$w(j,t) \equiv \sum_i \mathbf{1}\left(t \geq t^a(i)\right) \mathbf{1}\left(t \leq t^d(i)\right) \mathbf{1}\left(j = J(i)\right), \tag{A-5.5}$$

where $t^a(i)$ is the arrival (assignment) time of i, $t^d(i)$ is the corresponding discharge order time, and $J(i)$ is the physician corresponding to i. This model represents the assignment policy function $\mathcal{A}_0(S)$.

In addition to the number of patients assigned, I also estimate the expected log length of stay of patients assigned, as a function of hour of the day, day of the week, month-year interactions, pod, time relative to EOS, and the physician identity.

A-5.1.2 Estimating Utility Flow Parameters

The state space includes characteristics of all patients currently under the physicians' care as well as time to EOS. In order to accommodate such a complex state space, I use a sieve approach by Barwick and Pathak (2015). This approach also allows for continuous states and states that are either never or rarely encountered in the data. Additionally, I collapse many patient characteristics into a single index, such as expected length of stay, and sum these indices across patients to reduce the dimensionality of the state space.

1. Set a functional form approximation for $\overline{u}(i,S) \approx \mathbf{b}(i,S)\,\theta_u$, where $\mathbf{b}(i,S)$ is a $1 \times K$ vector of splines.

2. Set a functional form approximation for $V(S) \approx \mathbf{h}(S)\,\theta_V$, where $\mathbf{h}(S)$ is a $1 \times L$ vector of splines.

3. A given set of parameters $\hat{\theta}_u$ implies $\hat{\theta}_V$ from Equation (A-5.4). One way of estimating $\hat{\theta}_V$ is minimizing the L_2 norm

$$\hat{\theta}_V = \arg\max_{\theta_V} \left\| \mathbf{h}(S)\,\theta_V - \log\left[\sum_{i \in \mathcal{I}} \exp\left(\mathbf{b}(i,S)\,\hat{\theta}_u + \delta E\left[\mathbf{h}\left(S'|i,S\right)\right]\theta_V\right)\right]\right\| \tag{A-5.6}$$

by nonlinear least squares across observations $\{(j,t)\}$. In practice, I implement Equation (A-5.4) by setting the constraint

$$E\left[\mathbf{h}(S)\,\hat{\theta}_V - \log\left[\sum_{i\in\mathcal{I}}\exp\left(\mathbf{b}(i,S)\,\hat{\theta}_u + \delta E\left[\mathbf{h}\left(S'|i,S\right)\right]\hat{\theta}_V\right)\right]\right] = 0. \tag{A-5.7}$$

Expectations $E\left[\mathbf{h}\left(S'|i,S\right)\right]$ are given by the transition probability function, estimated as described in Section A-5.1.1.

4. Perform constrained maximum likelihood estimation, where the relevant data include the indicator variables $\{d(i,t)\}$, such that $d(i,t) = \mathbf{1}\left(i_{j(i),t} = i^*_{j(i),t}\right)$, and the set of basis functions $\{\mathbf{b}(i,S_{j,t})\}$ and $\left\{E\left[\mathbf{h}\left(S_{j,t'}\right)|i,S_{j,t}\right]\right\}$ for each physician-time (j,t) and choice i. The log likelihood of the data is

$$\log\mathcal{L} = \sum_{j,t}\sum_{i\in I(j,t)} d(i,t)\log\widehat{\Pr}\left(d(i,t)\,|i,S_{j,t}\right), \tag{A-5.8}$$

where $\widehat{\Pr}\left(d(i,t)\,|i,S_{j,t}\right)$ is given by Equation (A-5.3). I maximize Equation (A-5.8) as a function of $\hat{\theta}_u$, subject to constraints in Equation (A-5.7) that imply $\hat{\theta}_V$. In estimation, I fix $\delta = 0.98$, informed by comparing the maximum log likelihood of various candidate values of δ.

A-5.1.3 Identification

For identification of the dynamic model, features of the state space need to be excluded from the utility flow (Magnac and Thesmar, 2002). In this problem, these exclusion restrictions are grounded in what I consider relevant to the physician's utility: Physicians receive flow utility from predicted length of stay and the difference between length of stay and predicted length of stay, since discharging patients with lower predicted length of stay could be easier and since physicians would prefer to discharge patients close to predicted length of stay. A variety of other features of the state space, including time to EOS, only enter the value function. Given these exclusion restrictions, the dynamic model is identified by conditional choice probabilities as shown by Hotz and Miller (1993).

In the conceptual framework (Section 6), I also consider foregone leisure as relevant for physician utility. However, because I do not observe the actual time that a physician goes home, and because patients are never assigned past EOS, I conveniently ignore modeling this utility explicitly.[1] Instead, I include the number of patients remaining on the physician's census at

[1] In an alternative specification of the dynamic programming problem, I model future utility flows from simulated outcomes, as in Hotz et al. (1994), including the times that physicians are *likely* to go home, based on a simple rule that they are likely to go home between when all but one or two patients have been discharged (as assumed in Section A-5.4). This results in utility estimates that suggest that leisure time is quite important: staying an extra hour is two to five times as important as deviating from ideal (predicted) length of stay by 50%

EOS as a state variable in the value function. As argued above, because patients are never assigned past EOS, counterfactual assignment policies and therefore conditional discharge choice probabilities only differ prior to EOS.

Finally, as is well known in discrete choice, utility is normalized for a reference choice. I normalize the choice $i = \varnothing$ to have flow utility of zero, since this choice exists for all observations (j, t). Similarly, the effect on the value function of characteristics of S' that never vary in expectation across i is not directly identified, although the impact on the value function of these characteristics interacted with other characteristics that do vary across i is identified.

A-5.1.4 Specification and Model Fit

I specify a relatively simple baseline model. Utility flow $\overline{u}(i, S)$ is a linear function of an indicator for whether $i \neq \emptyset$, predicted length of stay (based on patient characteristics for i) if $i \neq \varnothing$, and two cubic splines of the current difference between length of stay and predicted length of stay if $i \neq \emptyset$. Importantly, I rule out physicians receiving different flow utilities from discharging patients when they are at different times relative to EOS. The value function $V(S)$ is a linear function of the number of patients remaining on census, the sum of differences between length of stay and predicted length of stay (across patients on census), and the interaction between time to EOS and the number of patients on census. I show parameter estimates for various specifications of $\overline{u}(i, S)$ in Table A-5.1.

Figure 9 shows the model fit according to discharge probabilities using a sample of 792,687 patient-time observations. The figure shows three types of discharge probabilities $\Pr(i|S)$, along characteristics of i (e.g., predicted length of stay, difference between length of stay and predicted length of stay) and variables in S (e.g., time to EOS, number of patients on census). The first discharge probability is calculated by binning the raw data and involves no model. The second is calculated by a flexible multinomial logit model, with no restrictions on how state variables affect choices (i.e., this model is outside of a dynamic programming framework). The third is the probability implied by Equation (A-5.3), with the restrictions in the dynamic model that separate flow utilities from value functions. Figure 9 shows that the dynamic model fits quite well compared to the flexible model, despite the exclusion restrictions on utility and relatively parsimonious state variables. I consider more state spaces for $V(S)$ – additional variables (e.g., the average time of arrival for patients on census, the time since the last patient discharge), nonlinear relationships, and interactions between the variables – but do not find that these complications appreciably improve the fit of the dynamic model.

In Figure A-5.2, I further evaluate model fit by average shift outcomes with respect to time to EOS. In particular, I evaluate fit by the following averages for patients arriving in each 30-minute period with respect to EOS: the number of patients arriving in the period; the length of stay (and its log); and the average workload during a patient's length of stay, or $\overline{w}(i)$ defined in

for a single patient. However, this specification does not allow me to calculate counterfactual discharge policies.

Equation (5). The panels of this figure evaluating length of stay and average workload are based on the same simulation algorithm, described in Appendix A-5.3, used to evaluate counterfactual assignment policies. This figure shows that the fit is reasonably good using the conditional choice probabilities implied by Equation (A-5.3) in the dynamic model, although length of stay (particularly, its log) does not seem to decrease by as much in the last two hours prior to EOS.

A-5.2 Counterfactual Assignment Policies

The counterfactual assignment policies $\mathcal{A}_\Delta(S) \equiv \mathcal{A}_0(\overline{m}(\Delta), S^-)$ are constructed by modifying time to EOS. That is, for a true time to EOS $\overline{m}$, the assignment function considers a modified time to EOS $\overline{m}(\Delta)$ that is a function of the scalar parameter Δ:

$$\overline{m}(\Delta) = \begin{cases} \max(\Delta, \overline{m}) & \Delta > 0 \\ \overline{m}, & \Delta = 0\,. \\ \overline{m}\left(1 - \min\left(\max\left(|\Delta| - \overline{m}, 0\right), 1\right)\right) & \Delta < 0 \end{cases} \tag{A-5.9}$$

In this convenient parameterization, $\overline{m}(\Delta)$ is at most Δ greater than $\overline{m}$ if $\Delta > 0$ and at most Δ smaller than $\overline{m}$ if $\Delta < 0$. In the case where $\Delta < 0$, the function is slightly more complicated so that it is continuous; starting at $\overline{m} = |\Delta|$, $\overline{m}(\Delta) = \overline{m}$ and decreases to $\overline{m}(\Delta) = 0$ by $\overline{m} = |\Delta| - 1$. $\mathcal{A}_\Delta(S)$ correspondingly increases assignments relative to $\mathcal{A}_0(S)$ for $\Delta > 0$ and decreases assignments for $\Delta < 0$. Figure A-5.1 shows both the time modification function $\overline{m}(\Delta)$ as well as the corresponding counterfactual assignment policies $\mathcal{A}_\Delta(S)$ for $\Delta \in \{-4, -2, 2, 4\}$.

Anticipating a counterfactual assignment policy $\mathcal{A}_\Delta$, the physician will adopt counterfactual conditional choice probabilities that can be characterized by the parameters $\hat{\theta}_u$ and $\hat{\theta}_V^\Delta$, where $\hat{\theta}_V^\Delta$ is calculated by Equation (A-5.6) and counterfactual transition probabilities implied by $\mathcal{A}_\Delta$. I constrain the counterfactual value function to take the same value as the actual value function at EOS, i.e., $\mathbf{h}\left(S|t = \bar{t}(s)\right)\hat{\theta}_V^\Delta = \mathbf{h}\left(S|t = \bar{t}(s)\right)\hat{\theta}_V$. The conditional choice probabilities characterize the physician's discharge policy, $\mathcal{D}_\Delta$.

A-5.3 Simulation

For a given assignment policy $\mathcal{A}_\Delta(S)$ and corresponding discharge policy $\mathcal{D}_\Delta(S)$, I simulate patient arrivals and discharges to create a set of patient arrival and discharge observation. Specifically, I follow this procedure for each simulation r:

1. Start t at three hours before the beginning of each shift s. Set $w_{\Delta,r}(j, t-1) = 0$.

2. Determine new assignments at t for each s.

   (a) Simulate $a_{\Delta,r}(s,t)$ new assignments for s at t, using $\mathcal{A}_\Delta$. Denote each of these new assignments with an unused $i \notin I^{outside}$, note that $t^a_{\Delta,r}(i) = t$, and also simulate predicted log length of stay for each i.

(b) Assign patients $i \in I^{outside}$ where $t^{a}_{outside}(i) = t$ to the relevant shifts s.

3. Calculate workload $w_{\Delta,r}(j,t)$ by Equation (A-5.5).

4. If $t \geq \underline{t}(s)$ and $w_{\Delta,r}(j,t) > 0$, determine discharges at t for each s.

   (a) Simulate $d_{\Delta,r}(i,t) \equiv \mathbf{1}\left(t^{d}_{\Delta,r}(i) = t\right)$ for each $i \notin I^{outside}$ where $d_{\Delta,r}(i,t-1) = 0$, using $\mathcal{D}$.

   (b) Discharge patients $i \in I^{outside}$ where $t^{d}_{outside}(i) = t$ from the relevant shifts s.

5. The procedure is complete for s such that $t \geq \bar{t}(s)$ and $w^{\Delta,r}(j,t) = 0$. For the remaining s, revise $t = t + 1$ and return to Step #2.

The resulting collection of visits $I_{\Delta,r} = \bigcup_{j,t} I_{\Delta,r}(j,t)$, where $t^{a}_{\Delta,r,}(i)$ and $t^{d}_{\Delta,r,}(i)$ are observed for each $i \in I_{\Delta,r}$, form the data under Δ in simulation r. Simulated workload-adjusted length of stay for patient i under physician j can be calculated by dividing i's simulated length of stay by simulated average censuses under j during i's length of stay.

A-5.4 Imputing Cost Outcomes

Having simulated arrivals and discharges, I am now in the position to impute overall costs for each counterfactual simulation r of Δ. Overall costs include physician-time, patient-time, and hospital-resource costs. Repeating Equation (10):

$$\text{Costs}_{\Delta,r} = \text{PhysicianTime}_{\Delta,r} + \text{PatientTime}_{\Delta,r} + \text{HospitalResources}_{\Delta,r}. \tag{A-5.10}$$

The first cost, physician-time costs, represents the value of leisure foregone. Physician hours in a given shift s can increase either if a peer must arrive earlier before the index physicians EOS, or if the index physician must stay longer past EOS:

$$\text{PhysicianTime}_{\Delta,r} = \text{Wage} \times \sum_{s} \left(\text{WorkCompletionTime}_{\Delta,r}(s) - \text{PeerArrivalTime}_{\Delta,r}(s)\right).$$

"Slacking off" in the assignment policy, by assigning fewer patients to the physician ending shift, mechanically requires peers to arrive earlier. In the actual data, there are generally two unseen patients at the time of peer arrival (see Figure A-7.3). I therefore model $\text{PeerArrivalTime}_{\Delta,r}(s)$ as when there are two unseen patients near $\bar{t}(s)$, based on the assignment policy and an exogenous pod flow rate of 2.22 patients per hour (see Figure 3). I model $\text{WorkCompletionTime}_{\Delta,r}(s)$ (when the physician on shift s leaves the ED) as the midpoint between when the third-to-last patient is discharged and when the second-to-last patient is discharged. This empirically matches the stated work completion time of generally two to three hours past EOS, although results are insensitive to the precise definition of work completion. Implicit in this rule is that physicians are not more likely to pass off patients with more work at EOS; given that work completion time is really insensitive to being assigned more work at EOS (due to quicker discharges), this is unlikely to be quantitatively important. I multiply physician-hours by a wage of \$120 per hour.

The second cost, patient-time costs, reflects the value of patient time:

$$\text{PatientTime}_{\Delta,r} = \text{TimeValue} \times \sum_i \tau_{\Delta,r}(i),$$

where TimeValue = \$20/hour, or roughly the average hourly wage in the US, and $\tau_{\Delta,r}(i)$ is the simulated length of stay implied by $t^a_{\Delta,r}(i)$ and $t^d_{\Delta,r}(i)$ in discrete time.

The third cost in Equation (A-5.10), hospital-resource costs, represents resource costs, via formal utilization and admissions, incurred by the physician. As shown in Section A-3 and Table 3, workload-adjusted length of stay, formal orders, admissions, and total costs all increase only in the last hour of shift, suggesting that workload-adjusted length of stay is a good measure of time that increases patient-care costs as it is decreased. In each simulation r of each policy Δ, I estimate the EOS effect on workload-adjusted length of stay by coefficients $\hat{\alpha}^{\Delta,r}_{\overline{m}}$ in this regression:

$$\log\left(\tau(i)/w(i)\right)_{\Delta,r} = \alpha^{\Delta,r}_{\overline{m}(i,t)} + \mathbf{g}\left(\underline{m}(i,t)\right)' \gamma^{\Delta,r}_g + \varepsilon^{\Delta,r}_i, \tag{A-5.11}$$

where $\left(\tau(i)/w(i)\right)_{\Delta,r}$ is simulated workload-adjusted length of stay, $t = t^a(i)$ is the simulated time of arrival, and $\mathbf{g}(\cdot)$ is a vector of cubic splines of assignment time relative to shift beginning.

In simulated data with $\Delta = 0$, I estimate $\hat{\alpha}^0_1 \equiv \frac{1}{20}\sum_{r=1}^{20} \hat{\alpha}^{0,r}_1 = -0.240$ and $\hat{\alpha}^0_2 \equiv -0.059$, which implies that workload-adjusted length of stay decreases by 18.1% in the last hour of shift under the observed assignment policy. Note that this difference is slightly smaller (more conservative) than that implied by coefficients $\hat{\alpha}_1 = -0.232$ and $\hat{\alpha}_2 = -0.069$ estimated without simulation using actual data (Table A-7.3 and Figure A-7.4). Given that total costs increase by 20.8% in the last hour prior to EOS, I estimate the elasticity of hospital-resource costs to workload-adjusted length of stay, for decreases in workload-adjusted length of stay that are 5.9% below baseline, as $20.8\%/-18.1\% = -1.15$. I thus calculate hospital-resource costs as

$$\begin{aligned}\text{HospitalResources}_{\Delta,r} &= \sum_s \sum_{\overline{m}} \sum_{t=\underline{t}(s)}^{\overline{t}(s)} \mathbf{1}\left(\left\lceil \overline{t}(s) - t \right\rceil = \overline{m}\right) a_{\Delta,r}(s,t) \times \\ &\quad \exp\left(\text{BaseLogCosts} - 1.15 \cdot \min\left(0, \hat{\alpha}^{\Delta,r}_{\overline{m}} - \hat{\alpha}^0_2\right)\right),\end{aligned} \tag{A-5.12}$$

where BaseLogCosts = [log \$ +] 6.750. Note hospital-resource costs increase with greater assignments (higher Δ) both because per–patient costs increase, and the number of patients that this applies to also increases. I assume that increases in workload-adjusted length of stay above baseline do not reduce costs. Rather, increases seen in the data prior to the last hour of shift could be consistent with "foot-dragging," in which physicians delay discharge but do not otherwise change patient care (Chan, 2016). Finally, as discussed in the main paper, I conservatively assume no negative effects on patient health, even as physicians produce less information for the discharge decision, since I observe none in sample (Table 3).

A-5.4.1 Imputing the Value of Leisure

I can also impute the revealed value of leisure in terms of hospital-resource costs by calculating the ratio between extra hospital-resource costs incurred and leisure time gained as a result of the physician discharge behavior near EOS. The discharge function $\mathcal{D}_0(S)$ increases the discharge hazard as t approaches $\bar{t}(s)$, shortening workload-adjusted length of stay and increasing hospital-resource costs. I examine what discharges would look like if not influenced by EOS behavior by modifying t in the discharge function. That is, I consider a modified discharge function $\underline{\mathcal{D}}(\overline{m}, S^-) \equiv \mathcal{D}_0(\max(4, \overline{m}), S^-)$ that does change as a function of time relative to EOS, at least in the four hours prior to EOS.

I then evaluate differences in hospital-resource costs and work-completion time under both of these discharge functions. The ratio between these two differences reveals physicians' implicit valuation of an hour of leisure in terms of hospital-resource costs:

$$\text{LeisureValue}_{\Delta,r} = -\frac{\text{HospitalResources}_{\Delta,r}\,|\mathcal{A}_\Delta, \mathcal{D}_\Delta - \text{HospitalResources}_{\Delta,r}\,|\mathcal{A}_\Delta, \underline{\mathcal{D}}_\Delta}{\text{WorkCompletionTime}_{\Delta,r}\,|\mathcal{A}_\Delta, \mathcal{D}_\Delta - \text{WorkCompletionTime}_{\Delta,r}\,|\mathcal{A}_\Delta, \underline{\mathcal{D}}_\Delta}.$$

Because $\text{LeisureValue}_{\Delta,r}$ is specific to a counterfactual assignment policy Δ, I link the value of an hour of incremental leisure to the time when the physician would have been able to go home under $\mathcal{A}_\Delta$ and $\mathcal{D}_\Delta$ (i.e., $\text{WorkCompletionTime}_{\Delta,r}\,|\mathcal{A}_\Delta, \mathcal{D}_\Delta$). This time will be earlier for $\Delta < 0$ and later for $\Delta > 0$.

A-6 Heterogeneous Effects by Physician and Peer Types

In this appendix, I consider heterogeneous EOS effects using regressions of the following form:

$$Y_{it} = \alpha^{Type(i)}_{\overline{m}(i,t)} + \gamma_{\underline{m}(i,t)} + \beta \mathbf{X}_i + \eta \mathbf{T}_t + \kappa_{Type(i)} + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_{it}, \tag{A-6.1}$$

where Y_{it} is workload-adjusted log length of stay, defined in Equation (5), and $Type(i) \in \{0, 1\}$ refers to some category that observation i belongs to ($Type(i) = 1$) or the complement set ($Type(i) = 0$).

The first set of categories refer to the physician $j(i)$ at t: whether $j(i)$ is male, whether the $j(i)$ is older than average at t, whether $j(i)$ has greater tenure than average at t, and whether $j(i)$ is faster than average. The last characteristic is estimated from a regression of log length of stay on patient characteristics, time categories, pod identities, and (the object of interest) physician dummies. Results in Figure A-6.1 show roughly similar results, regardless of physician type, except that faster physicians are much less likely to have EOS distortions.

The second set of categories refer to the relationship between physician $j(i)$ and the peer $j^{-}(i)$ of the subsequent shift in the same managerial location, if there is one. This peer will potentially assume the care of remaining patients who are not seen or who need to be transferred, and he or she also observes the index physician ending shift to a greater degree than any other physician. I consider whether $(j(i), j^{-}(i))$ are of the same sex, whether $j(i)$ has less tenure than $j^{-}(i)$, whether $(j(i), j^{-}(i))$ are "familiar" (i.e., they have previously worked more than 60 hours in the same location), and whether $j(i)$ is slower than $j^{-}(i)$. Results in Figure A-6.2 show that physicians are more likely to have EOS distortions when working with a senior peer, when they are not familiar with the peer, and when the peer is faster.

A-7 Additional Results

This appendix presents the following additional results in tables and figures:

- Table A-7.2 describes the process of constructing the sample, including the number of observations in each step.
- Table A-7.1 lists the number of observations for each shift type. Observations are counted in terms of unique shifts, hours, potential patients (who could be assigned to a shift of that shift type at time of arrival), and actual patients (who are assigned to a shift of that shift type).
- Table A-7.3 reports coefficients for EOS effects on workload-adjusted length of stay, as a continuation of Table 4. Results in this table only control for time relative to shift beginning. I use these more parsimonious regressions to operationalize workload-adjusted length of stay as the key substitute for hospital-resource costs in the structural model in Section 7, in which simulating the full set of covariates would be impractical. Results are estimated on both actual and simulated data.
- Figure A-7.1 shows evidence on how long physicians stay past EOS in terms of the share of shifts in which an order written by the attending physician of record (AOR) is yet to be written, out of shifts in which an order written by any attending physician is yet to be written.
- Figure A-7.2, Panel A, shows the AOR order share from Figure A-7.1 at one hour past EOS for shifts ending at each hour of the day. Panel B of the figure shows the corresponding census of patients remaining at EOS.
- Figure A-7.3 shows average patient counts ("censuses") for physicians in shifts with different overlap $\bar{o}$.
- Figure A-7.4 shows coefficients for EOS effects on workload-adjusted length of stay, reported in Table A-7.3, estimated on both actual and simulated data.

References

ALTONJI, J. G., T. E. ELDER, AND C. R. TABER (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151–184.

BARWICK, P. J. AND P. A. PATHAK (2015): "The costs of free entry: an empirical study of real estate agents in Greater Boston," *The RAND Journal of Economics*, 46, 103–145.

BRACHET, T., G. DAVID, AND A. M. DRECHSLER (2012): "The Effect of Shift Structure on Performance," *American Economic Journal: Applied Economics*, 4, 219–246.

CHAN, D. (2016): "Teamwork and Moral Hazard: Evidence from the Emergency Department," *Journal of Political Economy*, 124, 734–770.

CHETTY, R., J. N. FRIEDMAN, AND J. E. ROCKOFF (2014): "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104, 2593–2632.

HOTZ, V. J. AND R. A. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60, 497.

HOTZ, V. J., R. A. MILLER, S. SANDERS, AND J. SMITH (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *The Review of Economic Studies*, 61, 265–289.

JACOB, B. A., L. LEFGREN, AND D. P. SIMS (2010): "The Persistence of Teacher-Induced Learning," *Journal of Human Resources*, 45, 915–943.

MAGNAC, T. AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816.

PAPKE, L. E. AND J. M. WOOLDRIDGE (1996): "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics*, 11, 619–632.

SHETTY, K. D. AND J. BHATTACHARYA (2007): "Changes in hospital mortality associated with residency work-hour regulations," *Annals of Internal Medicine*, 147, 73–80.

VOLPP, K. G. AND A. K. ROSEN (2007): "Mortality among hospitalized Medicare beneficiaries in the first two years following ACGME resident duty hour reform," *The Journal of the American Medical Association*, 298, 975–983.

Table A-1.1: Balance Tests, Other Predicted Outcomes

	(1)	(2)	(3)	(4)	(5)
	A. Assumption 1				
Predicted outcome	Order count	Inpatient admission	Log total cost	30-day mortality	14-day bounce-back
Hours prior to EOS					
F-test p-value	0.748	0.867	0.855	0.579	0.208
Resid. char. distrib.					
Mean	13.88	0.274	6.777	0.019	0.060
10th percentile	12.16	0.186	6.487	0.008	0.053
90th percentile	15.46	0.352	7.035	0.027	0.067
	B. Assumption 2				
Predicted outcome	Order count	Inpatient admission	Log total cost	30-day mortality	14-day bounce-back
Hours prior to EOS					
F-test p-value	0.740	0.869	0.833	0.793	0.962
Resid. char. distrib.					
Mean	13.88	0.274	6.777	0.019	0.060
10th percentile	10.29	0.132	6.256	0.003	0.043
90th percentile	17.43	0.413	7.286	0.031	0.075

Note: This table continues Table 1 in assessing balance with respect to Assumptions 1 and 2. In Panel A, outcomes are predicted by *ex post* clinical characteristics, and these predicted outcomes are regressed, as in Equation (A-1.2), on hour relative to EOS while controlling for *ex ante* clinical characteristics, time categories, pod, and providers. In Panel B, outcomes are predicted by all patient characteristics, and the predicted outcomes are regressed, as in Equation (A-1.3), on hourly propensities for assignment relative to EOS while controlling only for time categories, pod, and providers. Further details are given in Appendix A-1.2. Estimates of coefficients for individual hours relative to EOS are omitted for brevity; the p-value of the F-test that all coefficients are jointly 0 is given instead for each model. As in Table 1, summary statistics of the variation in residualized predicted outcomes are reported at the patient level (Panel A), or the residualized predicted outcomes are first averaged within each hour with summary statistics of these averages reported at the hourly level (Panel B).

Table A-1.2: Assessing Selection Bias Using Arrival Times

	(1) Mean actual, Y_t	(2) Mean predicted, $\hat{Y}_t^{ante}$	(3) Mean predicted, $\hat{Y}_t^{full}$
Mean EOS effect, Q_t	1.029***	0.029	0.024
	(0.060)	(0.025)	(0.025)
Number of visits	409,352	409,352	409,352
Number of shifts	22,501	22,501	22,501
Number of hour cells	63,345	63,355	63,355

Note: This table reports regressions assessing tests for selection bias using only variation in patients arriving at the ED in different hours; its graphical form is presented in Figure A-1.2. Column 1 asks whether residualized length of stay averaged within hour of arrival is predicted by EOS effects averaged within hour, as in Equation (A-1.7). Columns 2 and 3 ask whether predicted length of stay averaged within hour of arrival is correlated with EOS effects averaged within hour, as in Equation (A-1.10). Average EOS effects within hour of arrival, Q_t, is defined by Equation (A-1.4) and calculated as follows: (i) Coefficients on time relative to EOS are calculated from (1) using a leave-shift-out sampling; (ii) these coefficients are averaged across shifts in process at hour t, weighted by visits. Residualized length of stay averaged within hour, Y_t (Column 1), is calculated as follows: (i) Calculate residualized actual log length of stay, by subtracting expected log length of stay based on all covariates listed in the note for Table 2, using only variation within time to EOS; (ii) average within hour. To predicted log length of stay by patient characteristics (Columns 2 and 3), I residualize the characteristics by time categories and use within-EOS-time variation to predict log length of stay. As with Q_t and Y_t, I also average these predictions within hour of arrival. Patient characteristics and time categories are described in the note for Table 2. OLS is performed keeping visits as observations, though each observation within an hour t is identical and standard errors are clustered by t. Standard errors are in parentheses. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. Details are given in Appendix A-1.3.

Table A-1.3: Potential Bias from Selection on Unobservables

	$\widehat{\text{Var}}\,(A_{\overline{m}})$	$\widehat{\text{Var}}\,(A^*_{\overline{m}})$	Selection on observables	Bias upper bound	$\hat{\alpha}_m$	$\hat{\alpha}_m$ as bias multiple
Patient selection into hour prior to EOS ($A_{\overline{m}}$)						
Last hour (A_{-1})	0.00249	0.00062	-0.00111	-0.00124	-0.5873	474.93
Second hour (A^{-2})	0.02658	0.00387	-0.01103	-0.02086	-0.2869	13.75
Third hour (A^{-3})	0.07442	0.00784	0.00223	0.00584	-0.1230	-21.05
Fourth hour (A^{-4})	0.08956	0.01053	-0.00381	-0.00893	-0.0907	10.16
Fifth hour (A^{-5})	0.10191	0.01287	-0.03295	-0.07192	-0.0232	0.32
Sixth hour (A^{-6})	0.10851	0.01391	-0.04192	-0.09014	-0.0103	0.11

Note: This table reports estimates in a procedure based on Altonji et al. (2005) to calculate potential bias from selection on unobservables, as described in Appendix A-1.4. Selection is modeled for whether a patient is assigned in the $\overline{m}^{\text{th}}$ hour prior to EOS ($A_{\overline{m}}$) by Equation (A-1.14), the residual of which is $A^*_{\overline{m}}$. Selection on observables is defined as $\hat{E}\,[\mathbf{W}'\gamma_W\,|A_{\overline{m}}=1] - \hat{E}\,[\mathbf{W}'\gamma_W\,|A_{\overline{m}}=0]$, where γ_W is estimated from Equation (A-1.12). Using the condition from Altonji et al. (2005), in Equation (A-1.13), which states that normalized selection on unobservables is at most equal in magnitude to normalized selection on observables, an upper bound of bias from selection on unobservables is calculated from Equation (A-1.15). I use $\widehat{\text{Var}}\,(\varepsilon) = 0.160$ and $\widehat{\text{Var}}\,(\mathbf{W}'\gamma_W) = 0.580$ in this calculation. $\hat{\alpha}_{\overline{m}}$ is estimated by Equation (1); for convenience, results are repeated from Column 5 of Table 2. Finally $\hat{\alpha}_{\overline{m}}$ is stated as a multiple of the bias upper bound in the last column of this table.

Table A-5.1: Dynamic Model Utility Parameter Estimates

	(1)	(2)	(3)	(4)	(5)
	Discharge choice utility				
Deviation of log LOS from predicted					
Deviation $= -1.5$	-2.406	-2.320	-2.290	-2.072	-1.864
	(0.179)	(0.179)	(0.179)	(0.178)	(0.175)
Deviation $= -1$	-1.629	-1.572	-1.549	-1.406	-1.264
	(0.119)	(0.119)	(0.119)	(0.119)	(0.117)
Deviation $= -0.5$	-1.000	-0.978	-0.946	-0.886	-0.795
	(0.061)	(0.061)	(0.061)	(0.061)	(0.060)
Deviation $= 0$	-0.666	-0.689	-0.616	-0.659	-0.586
	(0.023)	(0.023)	(0.022)	(0.022)	(0.022)
Deviation $= 0.5$	-0.635	-0.714	-0.567	-0.733	-0.645
	(0.069)	(0.069)	(0.069)	(0.069)	(0.068)
Predicted log LOS	-0.823	-0.932	-0.745	-0.978	-1.137
	(0.023)	(0.034)	(0.024)	(0.035)	(0.061)
Log workload-adjusted LOS		0.085		0.193	0.216
		(0.022)		(0.023)	(0.023)
Hours from last discharge			-0.244	-0.289	-0.802
			(0.017)	(0.018)	(0.047)
Other cubic splines	N	N	N	N	Y
Observations	792,687	792,687	792,687	792,687	792,687
Groups	115,674	115,674	115,674	115,674	115,674
Log likelihood	-77,577.04	-77,568.27	-77,463.99	-77,422.82	-77,321.02

Note: This table shows estimated utility flow parameters, with standard errors in parentheses. for incurred by discharging a patient in discrete dynamic programming models. The utility of not discharging any patient is normalized to 0. All models include utility flows for splines of deviation in current log length of stay from predicted log length of stay (or "deviation LOS") and linear predicted log length of stay. To interpret the spline coefficients on deviation LOS, the first five rows report estimates for the utility of discharging a patient at deviation LOS values of $(-1.5, -1, -0.5, 0, 0.5)$, representing approximately equally spaced quantiles in the data. These estimates are linear combinations of the spline coefficient estimates, and standard errors for these linear combinations are calculated using the delta method. Log workload-adjusted length of stay of a discharged patient is included in the utility flow of models 2, 4, and 5; hours from last discharge is included in the utility flow of models 3 to 5. Model 5 is similar to Model 4 but includes cubic splines for predicted log length of stay and hours from last discharge. In all models, the value function at EOS is only a function of the number of patients remaining on census (or "census") and the average patient arrival time for these patients relative to EOS. Value functions at times prior to EOS are otherwise constrained by Equation (A-5.7) from the dynamic programming Bellman equation; in all models, I fit value functions with the census, the sum of deviation LOS (across patients on census), and the census interacted with time to EOS and hours from last discharge. Details of the estimation procedure are given in Appendix A-5.1. All models are fit on a sample of 792,687 patient-interval observations, corresponding to 115,674 physician-interval choice sets (groups). Log likelihoods are given in the last row. For comparison, a fully flexible static logit model (i.e., with no restrictions on what enters the utility flow) had a log likelihood of -77,071.90. Graphical fit of discharge probabilities is given in Figure 9.

Table A-7.1: Shift Type Observation Numbers

Shift type	Shifts	Hours	Potential patients	Actual patients
$\langle 7, 0, 1\rangle$	95	665	1,645	1,160
$\langle 7, 1, 0\rangle$	237	1,659	6,674	2,597
$\langle 7, 1, 1\rangle$	101	707	4,281	1,783
$\langle 8, 0, 1\rangle$	319	2,552	8,453	4,952
$\langle 8, 1, 0\rangle$	174	1,392	7,440	1,981
$\langle 9, 0, 1\rangle$	3,453	30,879	84,292	58,589
$\langle 9, 0, 2\rangle$	325	2,349	6,411	4,541
$\langle 9, 0, 4\rangle$	408	2,898	9,326	4,839
$\langle 9, 0, 6\rangle$	364	3,276	16,186	5,899
$\langle 9, 1, 0\rangle$	3,414	30,528	118,030	59,897
$\langle 9, 1, 1\rangle$	2,909	26,181	116,108	54,221
$\langle 9, 1, 4\rangle$	2,249	19,170	80,279	28,694
$\langle 9, 1, 5\rangle$	60	540	2,554	892
$\langle 9, 1, 6\rangle$	211	1,899	8,157	2,524
$\langle 9, 2, 0\rangle$	464	3,294	12,027	6,317
$\langle 9, 3, 1\rangle$	485	3,277	17,013	6,699
$\langle 9, 3, 3\rangle$	60	540	3,226	1,089
$\langle 9, 4, 0\rangle$	347	2,347	9,996	3,994
$\langle 9, 4, 1\rangle$	212	1,908	8,974	3,370
$\langle 9, 4, 3\rangle$	426	2,752	16,730	5,344
$\langle 9, 4, 4\rangle$	772	5,094	26,094	9,413
$\langle 9, 4, 6\rangle$	2,141	19,269	99,726	29,007
$\langle 9, 5, 3\rangle$	60	540	2,851	1,043
$\langle 9, 6, 0\rangle$	634	5,706	34,943	9,244
$\langle 9, 6, 1\rangle$	1,504	13,536	61,197	21,861
$\langle 9, 6, 4\rangle$	575	5,175	31,088	9,597
$\langle 9, 9, 1\rangle$	353	3,177	15,965	4,598
$\langle 10, 0, 0\rangle$	176	1,760	4,812	2,578
$\langle 10, 0, 1\rangle$	243	2,430	5,783	4,615
$\langle 10, 0, 2\rangle$	137	1,040	2,631	1,901
$\langle 10, 0, 4\rangle$	139	1,050	3,616	2,378
$\langle 10, 1, 0\rangle$	277	2,770	9,092	4,401
$\langle 10, 4, 0\rangle$	139	1,050	4,335	1,834
$\langle 12, 0, 0\rangle$	142	1,704	4,119	2,423
$\langle 12, 4, 9\rangle$	319	3,828	16,490	5,566
Total	23,924	206,942	860,544	369,841

Note: This table lists the number of observations for each shift type, each defined as $\langle \ell, \underline{o}, \overline{o}\rangle$, where ℓ is the shift length in hours, $\underline{o}$ is the overlap in hours with a previous shift, and $\overline{o}$ is the overlap in hours with a subsequent shift in the same location. Observations are counted in terms of unique shifts, hours, potential patients (patients who arrive at the ED during a time when there is a shift of type $\langle \ell, \underline{o}, \overline{o}\rangle$ in progress), and actual patients (patients who are treated by a physician on a shift of type $\langle \ell, \underline{o}, \overline{o}\rangle$).

Table A-7.2: Sample Definition

Sample description or step	Variables added	Observations
1. Raw visit data	Patient demographics, clinical diagnoses, process times (arrival at ED, arrival at bed, discharge order, discharge with destination), treatment pod, 30-day mortality, providers of record (physician, resident or physician assistant, nurse)	442,244
2. Drop visits with patients leaving before being assigned by physician or discharged		426,899
3. Merge with physician order data and bed audit data	Detailed physician orders with timestamps for medication, intravenous fluids, laboratory tests, radiology tests, and nursing orders; timestamps for bed movements	411,198
4. Merge with pod schedules	Shift types, start times, end times, and managerial locations	398,563
5. Identify visits with physician of record in visit data matching with schedules		372,224

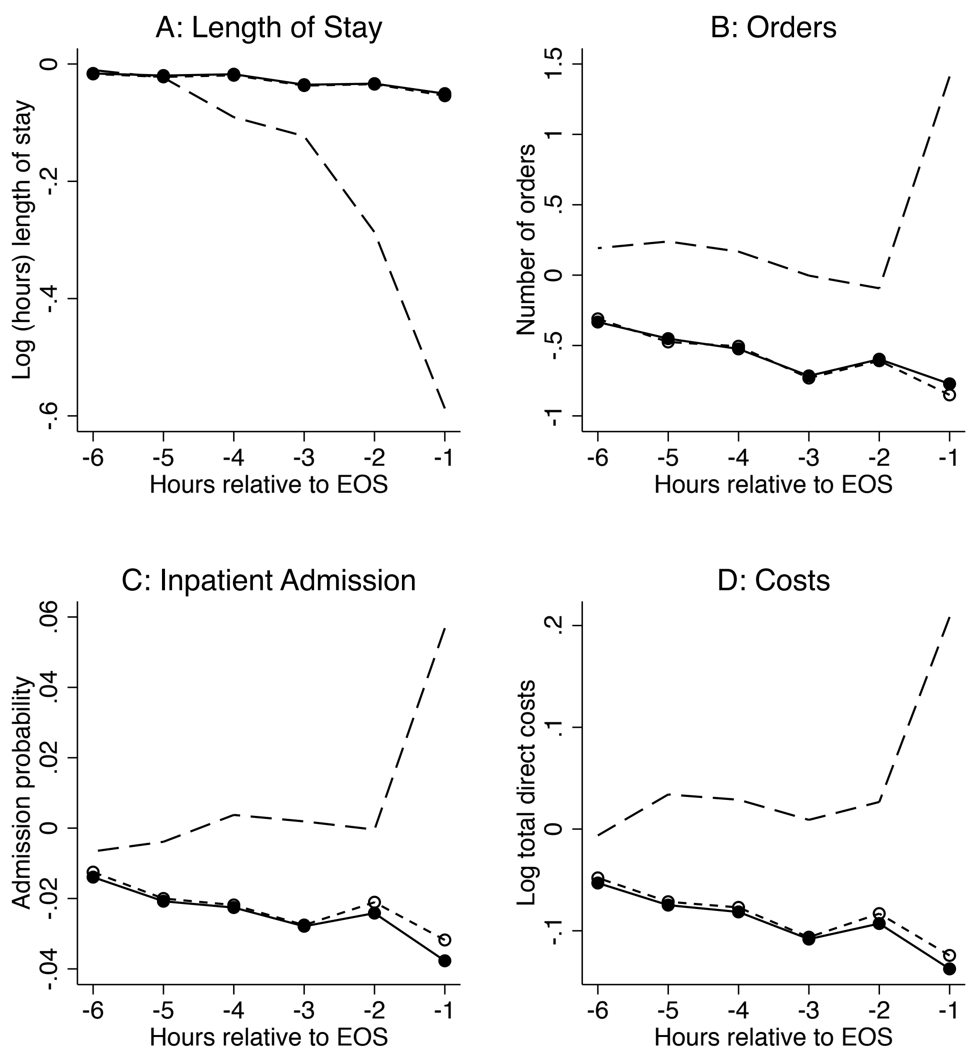
Note: This table describes each step in sample construction. Variables included in each step are listed in the second column, and the number of observations resulting from each step are in the third column.

Table A-7.3: Effect on Workload-adjusted Length of Stay by Shift Overlap

	(1)	(2)	(3)	(4)	(5)	(6)
	All $\bar{o}$	$\bar{o} \leq 1$	$\bar{o} \geq 2$	All $\bar{o}$	$\bar{o} \leq 1$	$\bar{o} \geq 2$
Hour prior to EOS						
Last hour	-0.232***	-0.339***	0.031	-0.200***	-0.240***	-0.053
	(0.037)	(0.05)	(0.061)	(0.037)	(0.049)	(0.063)
Second hour	-0.069***	-0.089***	0.168***	-0.067***	-0.064**	0.101***
	(0.019)	(0.025)	(0.035)	(0.019)	(0.025)	(0.036)
Third hour	-0.016	-0.027	0.176***	-0.013	-0.007	0.122***
	(0.015)	(0.020)	(0.028)	(0.015)	(0.020)	(0.029)
Fourth hour	-0.077***	-0.076***	0.087***	-0.052***	-0.044**	0.075***
	(0.014)	(0.017)	(0.026)	(0.014)	(0.017)	(0.027)
Fifth hour	-0.052***	-0.048***	0.046**	-0.037***	-0.029**	0.031
	(0.012)	(0.014)	(0.022)	(0.012)	(0.014)	(0.022)
Sixth hour	-0.032***	-0.028***	0.013	-0.028***	-0.021**	0.002
	(0.008)	(0.010)	(0.014)	(0.008)	(0.010)	(0.015)
Control for time relative to shift beginning	Y	Y	Y	Y	Y	Y
Patient, provider, and other time controls	N	N	N	N	N	N
Sample	Full, actual	$\bar{o} \leq 1$, actual	$\bar{o} \geq 2$, actual	Full, simulated	$\bar{o} \leq 1$, simulated	$\bar{o} \geq 2$, simulated
Number of observations	334,955	231,576	101,657	334,955	231,710	101,692
Adjusted R-squared	0.010	0.011	0.001	0.009	0.011	0.001
Sample mean outcome	-0.920	-0.987	-0.789	-0.927	-0.973	-0.798

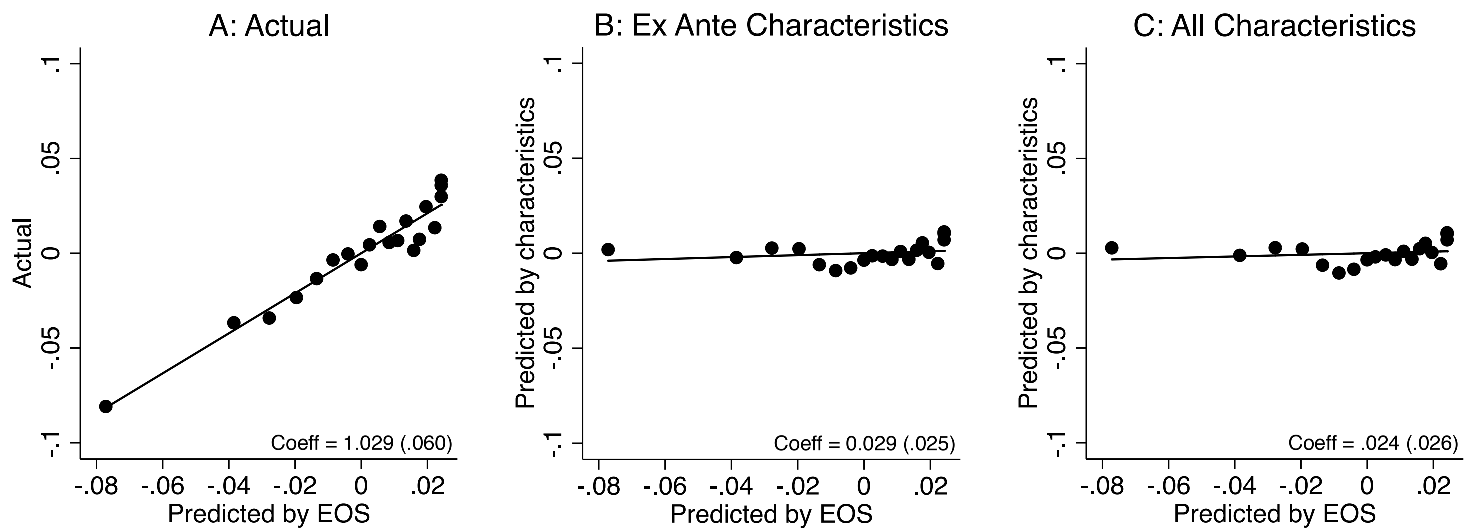
Note: This table is a continuation of Table 4, reporting coefficient estimates and standard errors in parentheses for EOS effects on workload-adjusted length of stay, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Models (1) to (3) are estimated on actual data, while Models (4) to (6) are estimated on simulated data. Models also differ by which shifts, based on overlap $\bar{o}$, are included. All models are estimated with Equation (A-5.11), which controls for time relative to shift beginning but not for other variables, in order to facilitate comparison between actual and simulated data. Workload-adjusted length of stay is calculated by Equation (5). * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. Results are graphically shown in Figure A-7.4.

Figure A-1.1: Patient Selection on Observables Relative to End of Shift



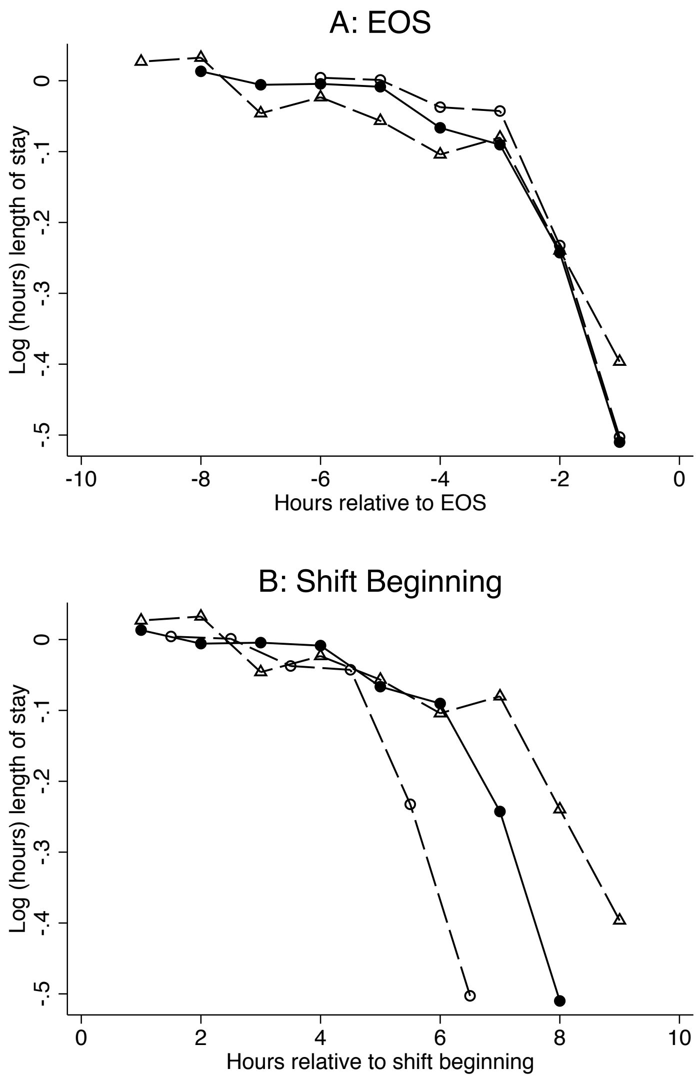
Note: This figure shows selection on observables for each hour prior to end of shift (EOS) on length of stay (Panel A), orders (Panel B), inpatient admissions (Panel C), and costs (Panel D). Each outcome is predicted based on patient characteristics observable prior to acceptance (age, sex, ESI) (closed circles) and on the full set of characteristics usually unobservable until after patient acceptance (e.g., 29 Elixhauser indices, race, language) (short-dashed line, open circles). Coefficients are estimated for predicted outcome using Equation (A-1.1). For reference, adjusted effects on actual outcomes from Figure 6 are shown with the dashed line. The reference category is any time greater than six hours prior to EOS.

Figure A-1.2: Assessing Selection Bias Using Arrival Times



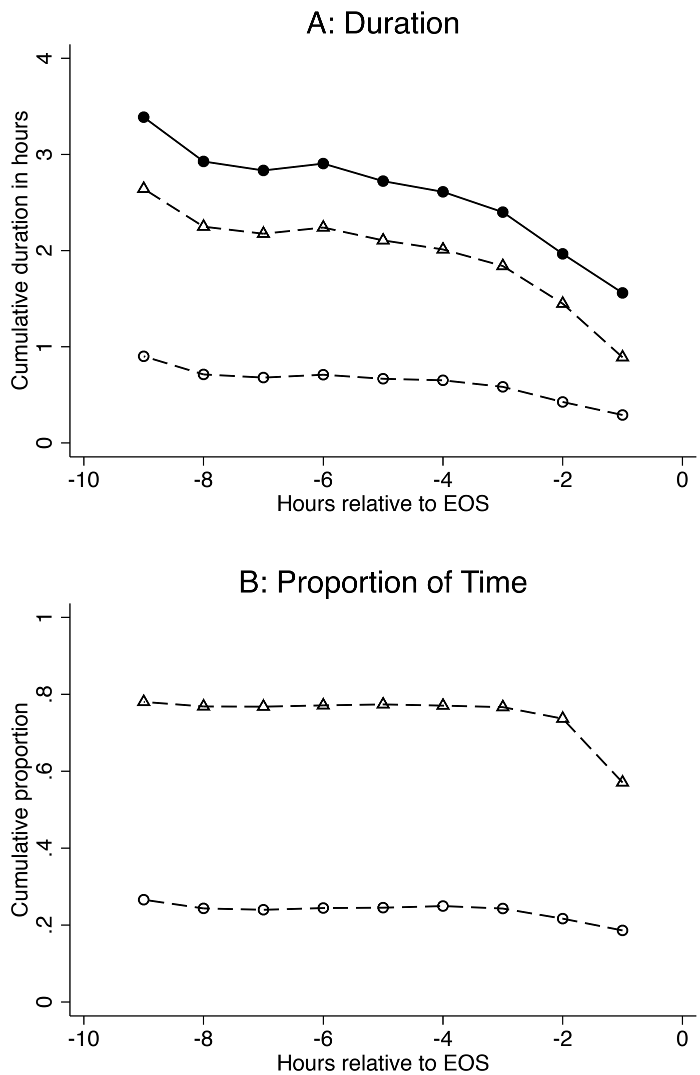
Note: This figure is the graphical representation of Table A-1.2, using only variation in arrival times to assess selection bias. It shows binned scatterplots of actual (residualized) log length of stay (Panel A), log length of stay predicted on "ex ante" characteristics possibly observable to physicians prior to assignment (Panel B), and log length of stay predicted on all characteristics including clinical diagnoses usually observable only after assignment (Panel C). Predicted and actual log lengths of stay are all averaged within hour cell and weighted by visit. The core data for the x-axis on all three panels is the log length of stay predicted by the times to EOS, defined by Equation (A-1.4) as Q_t. Q_t is calculated as follows: First, coefficients on time relative to EOS are calculated from Equation (1) using leave-shift-out sampling. Next, these coefficients are averaged across shifts in process at hour t, weighted by visits. To calculate residualized actual log length of stay (Panel A), I subtract expected log length of stay based on all covariates listed in the note for Table 2, except for time to EOS, using only variation within time to EOS. To calculate predicted log length of stay by patient characteristics (Panels B and C), I residualize the characteristics by time categories and use within-EOS-time variation to predict log length of stay. Patient characteristics and time categories are described in the note for Table 2, respectively. To construct each of the binned scatterplots, I demean values on the x- and y-axis, separate the data into 20 equal-sized groups (by patient visits) ordered by A_t, then plot the mean value within each bin. Solid lines show the best linear fit by OLS on the underlying microdata, clustered by hour (coefficients and standard errors are given as notes in each panel, also given in Table A-1.2). Details are given in Appendix A-1.3.

Figure A-2.1: Effects on Length of Stay by Shift Length



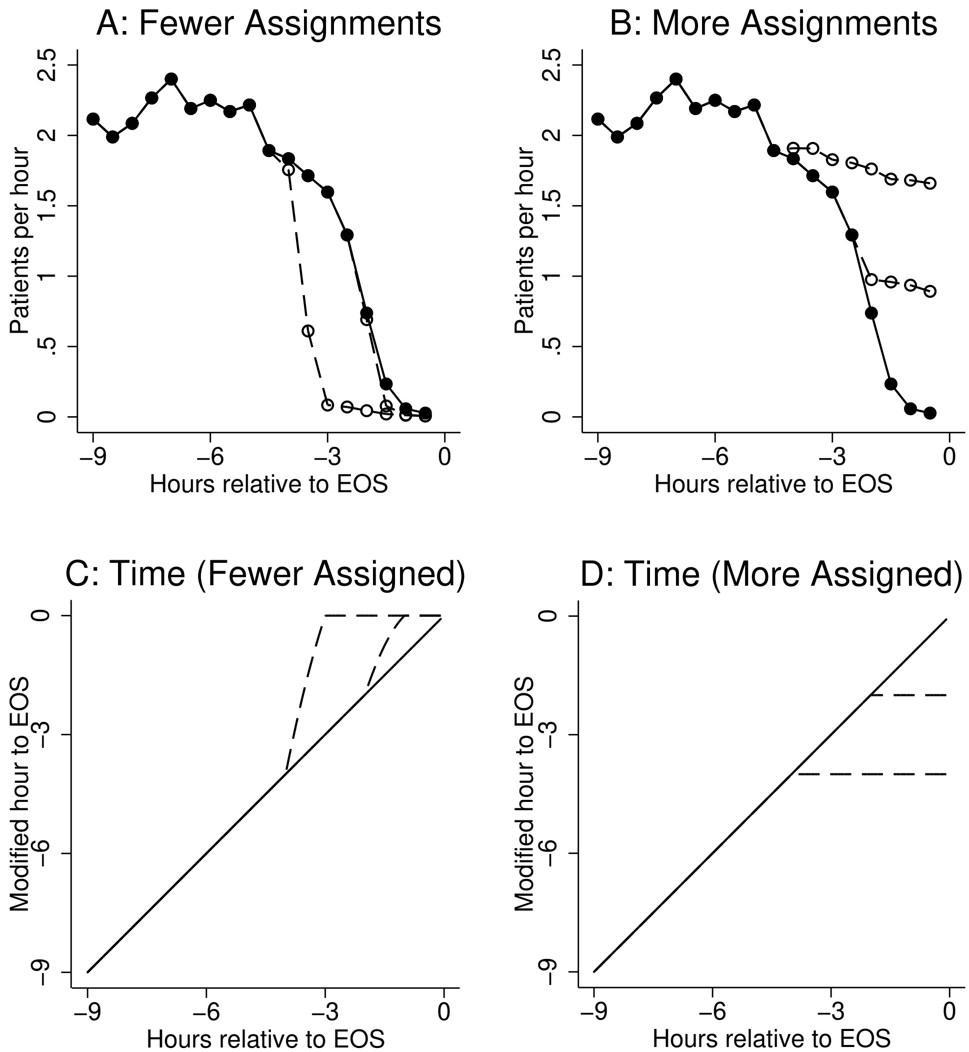
Note: This figure shows coefficients from Equation (1) estimated separately for shifts of seven or eight hours in length (open circles), nine hours in length (closed circles), and ten hours in length (open triangles). Panel A arranges estimates by hours relative to end of shift (EOS). Panel B arranges estimates by hours relative to shift beginning.

Figure A-3.1: Time Components



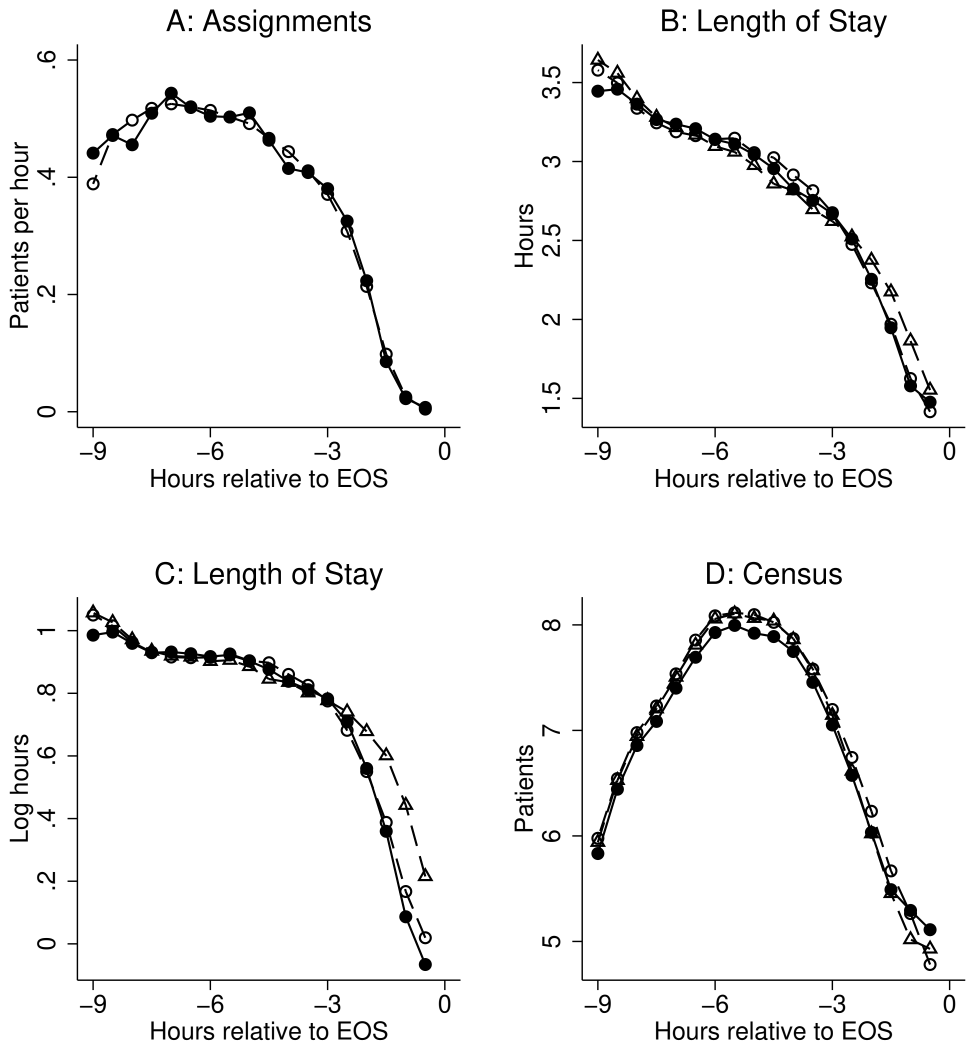
Note: This figure plots time components of length of stay as a function of hours relative to end of shift (EOS): time from pod arrival to first order (open circles), time from first to last (non-discharge) order (open triangles), and time from last order to discharge order (closed circles). Panel B shows marginal effects from a fractional logit model on these shares. Panel A represents these results as time in hours, incorporating results on the EOS effect on length of stay.

Figure A-5.1: Example Counterfactual Assignment Policies



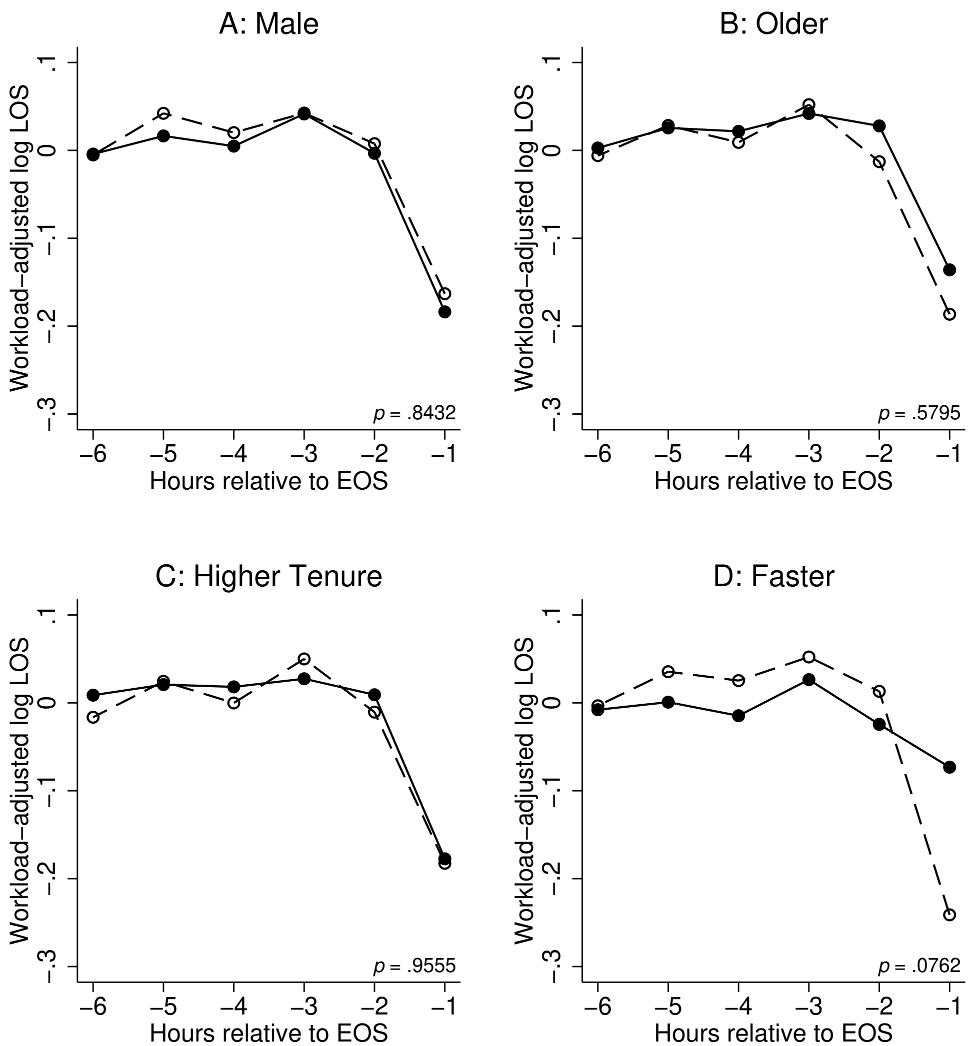
Note: This figure shows example counterfactual assignment policies, parameterized as hours Δ that time relative to EOS can be modified by. With dashed lines, Panel A shows counterfactual policies that reduce assignment near EOS ($\Delta \in \{-4, -2\}$); Panel B shows counterfactual policies this increase assignment near EOS ($\Delta \in \{2, 4\}$). These counterfactual policies are constructed by modifying the way time to EOS is considered, as specified by Equation (A-5.9). $\Delta < 0$ reduces time to EOS starting at $|\Delta|$ hours prior to EOS; $\Delta > 0$ increases time to EOS starting at Δ hours prior to EOS. These modifications in time to EOS are shown in the bottom Panels C and D, corresponding to Panels A and B, respectively. Further details are given in Appendix A-5.2.

Figure A-5.2: Model Fit by Simulated Outcomes on Shift



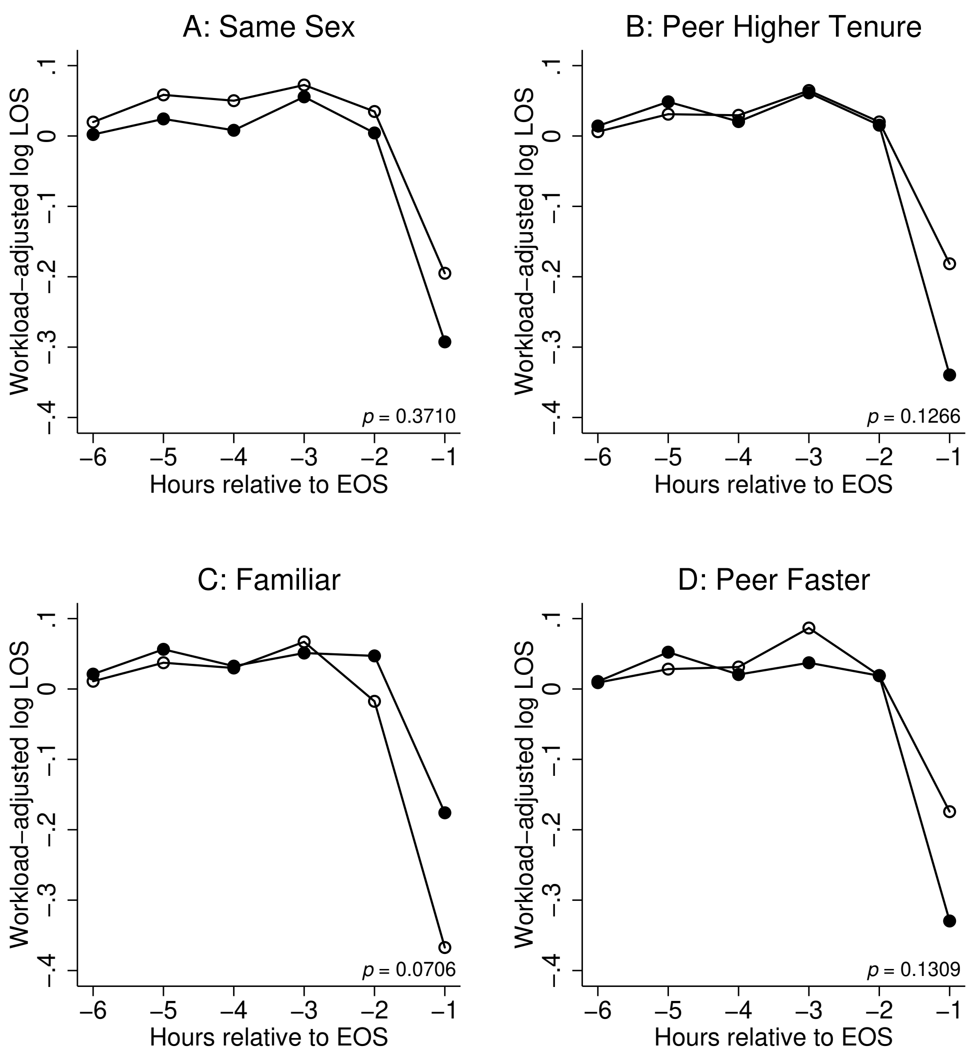
Note: This figure shows the fit of simulated outcomes with actual outcomes observed on 22,434 shifts with non-missing shift types out of the universe of 24,499 shifts. Outcomes are plotted against the x-axis of hours relative to EOS, in 30-minute intervals. Panel A evaluates the fit of actual and simulated patient assignments at each point in time, given the actual census of patients (i.e., discharges are not modeled). Panels B to D evaluate the fit between actual outcomes (solid dots) and those simulated by two discharge models, for any patient assigned in the relevant 30-minute interval: a fully flexible "static" model of conditional discharge probabilities (hollow dots), and a dynamic model of the discharge probabilities with restrictions on what can enter into the utility flow (triangles). The dynamic model, described in greater detail in Appendix A-5.1, corresponds to the simplest model in Table A-5.1, Model 1, although the fit and the estimated parameters do not qualitatively differ across specifications. The simulation algorithm is given in Section A-5.3. Panel D ("Census") reflects actual and simulated average census during the length of stay of a patient i, $\overline{w}(i)$ in Equation (5).

Figure A-6.1: End of Shift Effect by Physician Type



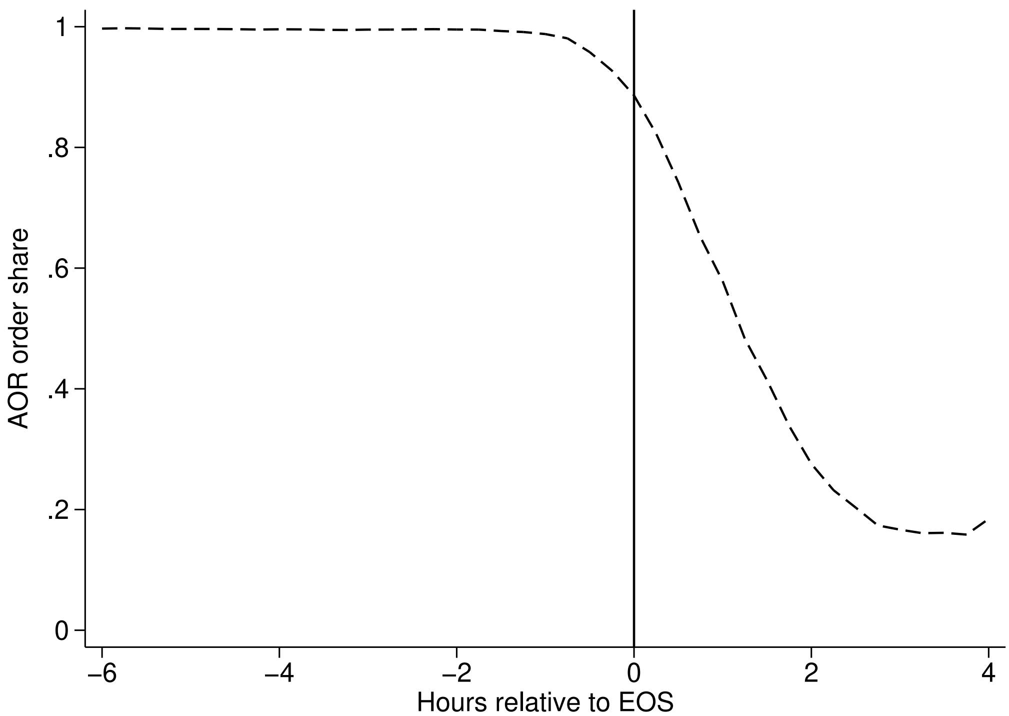
Note: This figure shows estimates of effects of hour relative to end of shift (EOS) for physicians of different types, using Equation (A-6.1). In each panel, estimates for physicians of the stated type are shown in solid dots, while estimates for physicians not of the stated type are shown in hollow dots. Panel A shows male vs. female; Panel B shows physicians who are older than average (about 39 years old) vs. not; Panel C shows physicians who have higher tenure at the ED (about 5.5 years) vs. not; Panel C shows physicians who have are faster than average (i.e., have a lower fixed effect in a regression of length of stay) vs. not. p-values for the significance of the difference between the last hour effect for physicians with and without the characteristic are given as notes in the lower right.

Figure A-6.2: End of Shift Effect by Physician-Peer Relationship



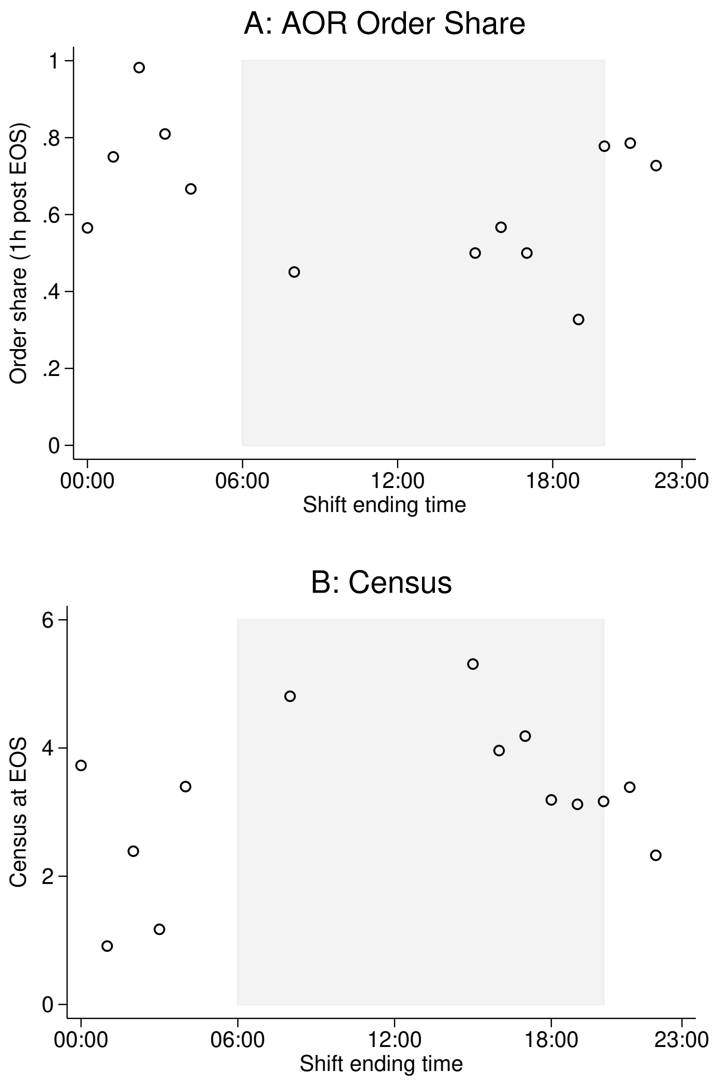
Note: This figure shows estimates of effects of hour relative to end of shift (EOS) for different types of relationships between physicians and the peer in the subsequent shift, if there is one. Estimates are from Equation (A-6.1). In each panel, estimates for physician-peer relationships of the stated type are shown in solid dots, while relationships not of the stated type are shown in hollow dots. Panel A shows physician-peer pairs of the same vs. different sex; Panel B shows physicians who have higher vs. lower tenure than their peer; Panel C shows physicians who have worked more vs. less than 60 hours together; Panel D shows physicians who are faster vs. slower than their peer. p-values for the significance of the difference between the last hour effect for pairs with and without the characteristic are given as notes in the lower right.

Figure A-7.1: Attending Physician Order-writing over Time



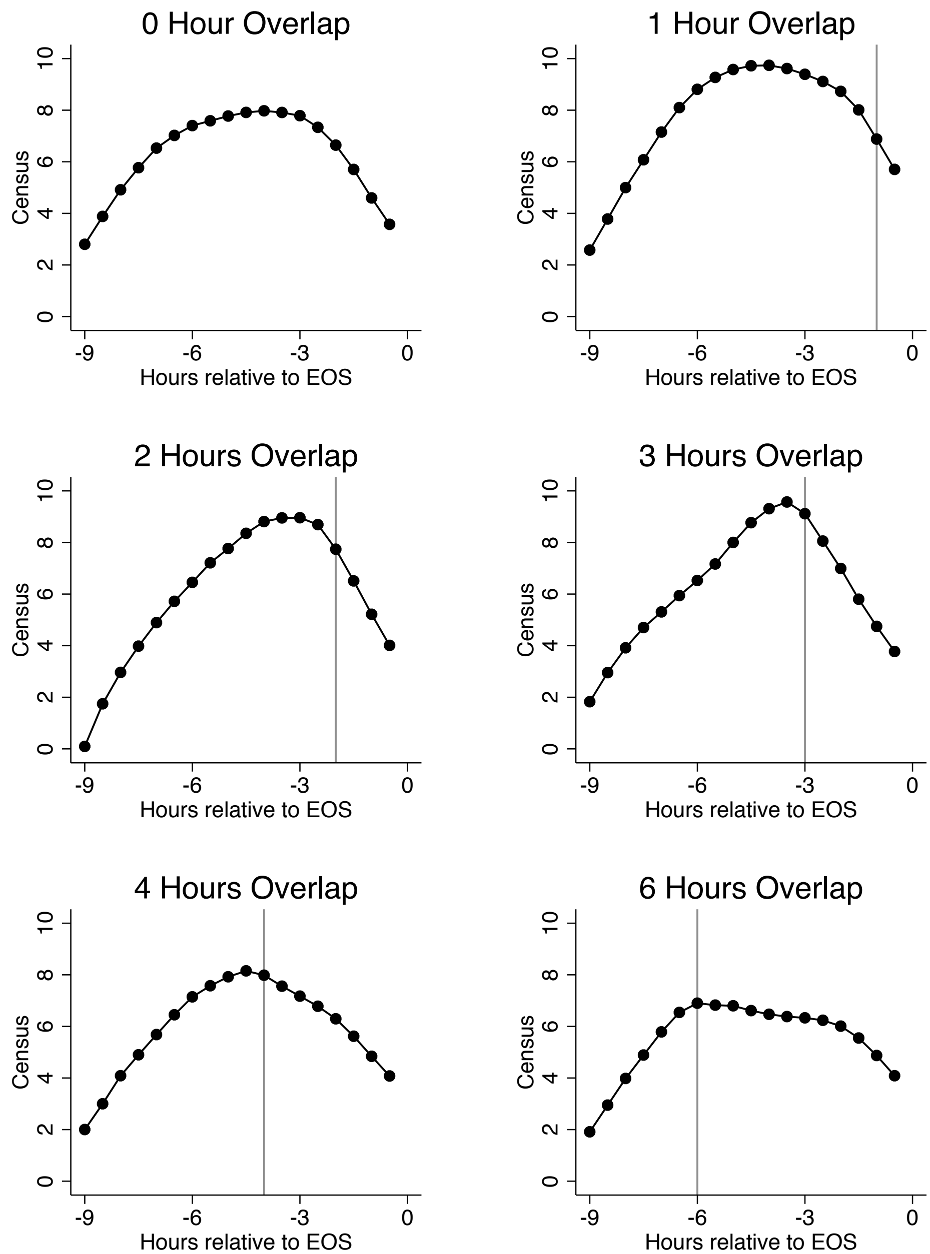
Note: This figure shows the activity of order-writing by the attending physician of record (AOR), conditional on an order written by any attending physician, at various points in time relative to end of shift (EOS). The AOR is the physician on the bill for patient care, corresponding to the physician whose shift is matched to a patient visit. The AOR order share on the y-axis is the number of shifts in which there exists an AOR order at a later time divided by the number of shifts in which there exists any attending physician order at a later time, as a function of time relative to EOS.

Figure A-7.2: Order-writing and Censuses by Shift Ending Time



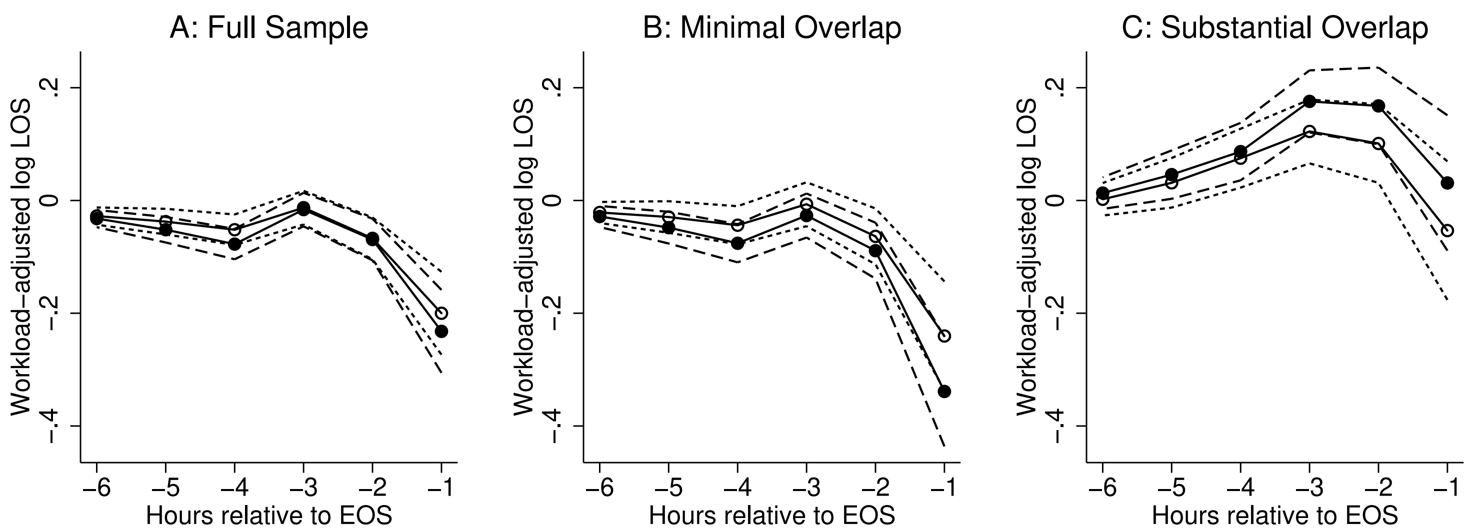
Note: This figure shows the attending of record (AOR) order share, defined in Figure A-7.1, at one hour past end of shift (EOS) (Panel A) and the average number of patients remaining on census at EOS (Panel B), for shifts ending at various times of the day. The shaded gray area indicates "daytime" hours between 6:00 a.m. and 8:00 p.m.

Figure A-7.3: Censuses over Time



Note: This figure plots average censuses over time relative to the end of shift (EOS). Each panel shows results for physicians in shifts with a given EOS overlap time. Subsequent shift starting times are marked with a vertical line.

Figure A-7.4: Workload-adjusted Length of Stay



Note: This figure shows coefficients for regressions, described in Equation (A-5.11), of the log of workload-adjusted length of stay (length of stay divided by average census) on time relative to end of shift (EOS), controlling for time from beginning of shift using both actual data (closed circles, confidence intervals in long-dashed lines) and simulated data (open circles, confidence intervals in short-dashed lines). Panel A shows results using actual or simulated data for all shifts. Results using actual or simulated data either only for shifts where EOS overlap $\bar{o} \leq 1$ or only for shifts with $\bar{o} \geq 2$ are shown in Panels B and C, respectively. Numbers for this figure are shown in Table A-7.3.