

The Efficiency of Slacking Off: Evidence from the Emergency Department*

David C. Chan[†]

December 22, 2016

Abstract

Work schedules play an important role in utilizing labor in organizations. In this study of emergency department physicians in shift work, schedules induce two distortions: First, physicians “slack off” by accepting fewer patients near end of shift (EOS). Second, physicians distort patient care, incurring higher costs as they spend less time on patients assigned near EOS. Examining how these effects change with shift overlap reveals a tradeoff between the two. Within an hour after the normal time of work completion, physicians are willing to spend hospital resources more than six times their market wage to preserve their leisure. Accounting for overall costs, I find that physicians slack off at approximately second-best optimal levels.

JEL Codes: D20, I10, L23, L84, M11, M54

*I am thankful to Jason Abaluck, Panle Barwick, Lanier Benkard, Amitabh Chandra, Daniel Chen, David Cutler, Mike Dickstein, Joe Doyle, Liran Einav, Bob Gibbons, Jeremy Goldhaber-Fiebert, Jon Gruber, Igal Hendel, Rob Huckman, Dan Kessler, Eddie Lazear, Robin Lee, Sara Machado, Grant Miller, Paul Oyer, Maria Polyakova, Stephen Ryan, Ori Shelef, Jonathan Skinner, and Chris Walters, and seminar participants at AEA/ASSA, ASHE-con, Cornell, Columbia, ETH Zurich, iHEA, MIT, NBER Productivity/Innovation/Entrepreneurship, NBER Summer Institute, RAND, Queen’s University, UC Irvine, USC, and University of Toronto for helpful comments and suggestions. I acknowledge early support from the NBER Health and Aging Fellowship, under the National Institute of Aging Grant Number T32-AG000186; the Charles A. King Trust Postdoctoral Fellowship, the Medical Foundation; and the Agency for Healthcare Research and Quality Ruth L. Kirschstein Individual Postdoctoral Fellowship F32-HS021044.

[†]Address: 117 Encina Commons, Room 217; Stanford, CA 94306. Phone: 650-725-9582. Fax: 650-723-1919. Email: david.c.chan@stanford.edu.

1 Introduction

Canonical models of production consider labor as an input but are silent on two important questions about how firms use workers’ time: First, how should worker availability be scheduled? Second, how should work be distributed across workers conditional on availability? This paper analyzes scheduling, a widespread form of coordination in organizations, as a principal-agent problem.¹ By defining boundaries of worker availability, schedules may open a margin for distortionary behavior if the appropriate time to complete work tasks is private information.

In this paper, I theoretically and empirically consider the implications of this problem on work assignment. If workers overvalue their leisure time relative to other consequences of their workplace actions, schedules induce two distortions near end of shift (EOS): First, on an extensive margin, workers “slack off” by accepting fewer tasks than socially optimal. Second, on an intensive margin, workers may rush to complete their work, spending less time than socially optimal on assigned tasks near EOS. Since workers usually have much more discretion on the intensive margin, the second distortion could be significantly costlier and implies that some slacking off is second-best optimal. While the empirical setting of this paper is in health care, the setting shares characteristics with other time-sensitive and information-rich workplaces:² Tasks are uncertain and largely non-contractible; compensation contracts are based on availability (or *minimum quantity* of hours worked); delaying assignment is costly; and worker-task specificity imposes some cost of transferring tasks to other workers once assigned.

Emergency department (ED) shiftwork is well-suited to allow me to estimate the effect of schedules on behavior. Shifts ending at different times allow me to separate effects related to shift work from differences due to the time of day. Shifts of different lengths allow separating

¹A large and active literature in operations management has viewed scheduling workers as mechanical inputs (e.g., Perdikaki et al., 2012; He et al., 2012; Green, 2004, 1984), including recent investigations that describe how worker throughput responds to environmental features such as “system load” (Kc and Terwiesch, 2009). In economics, a team-theoretic literature (e.g., Marschak and Radner, 1972; Radner, 1993; Garicano, 2000) has taken a similar approach. In practice, algorithmic approaches, e.g., using computerized staffing tools, are widely used by firms (Maher, 2007).

²Examples of such workplaces could include large-scale construction, management consulting, and software engineering. A number of online worker scheduling services have emerged, and case studies of client firms across industries can be found, for example, at <https://www.shiftplanning.com/casestudies>. The industry need not be 24-7, although the size of the economy involved in 24-7 activities has grown (Presser, 2003), and the number of workers with non-standard work times has grown to more than two-fifths (Beers, 2000). The key factors involve worker discretion and an assignment decision across workers.

these effects from “fatigue,” which I consider to depend on the time since the beginning of shift. Physicians work in virtually all types of shifts. I show that physicians accept fewer patients near EOS. For patients they are assigned, I also show that physicians shorten the duration of care (“length of stay”) in the ED and increase formal utilization, inpatient admissions, and hospital costs as the time of patient arrival approaches EOS. I find evidence that differential selection of patient types (i.e., selecting healthier patients near EOS) is negligible compared to the size of the effect on length of stay and in the opposite direction of utilization, admissions, and costs. To interpret EOS effects on patient care as causal, I utilize patient characteristics plausibly unobserved at the time of assignment and quasi-random patient arrival times at the ED.

Further, to interpret changes in patient care as distortions, I use another source of variation from shift structure: the overlapping time between when a peer arrives on a new shift and when the index physician reaches EOS. My identifying assumption is that, conditional on the volume of work, the time from the beginning of the shift, and the time from the peer’s arrival, the EOS should have no bearing on efficiency, since it is merely when physicians *may* go home if work is complete. I show that distortions on the intensive margin of patient care are greatest when physicians have the least time to offload work onto a peer before EOS. In fact, there is no increased utilization or admissions when overlap is four or more hours.

This evidence suggests a policy tradeoff between the extensive and intensive margins of distortion. On the extensive margin, workers “slack off” by accepting fewer patients. Although assigning more patients to physicians near EOS would reduce slacking off, it would increase the workload near EOS and the pressure for physicians to rush, inefficiently substituting other inputs for time.³ Using a structural model based on physician discharge behavior relative to EOS, as well as the natural experiment above that implies a marginal rate of technical substitution between physician time and hospital costs, I consider a wider range of counterfactual policies of patient assignment near EOS. I find that observed assignment patterns approximately minimize overall costs of physician time, patient time, and hospital resources. Assigning more patients near EOS such that physicians stay an additional hour induces an additional \$3,300 in hospital spending

³The idea of time per effective work is related to work by Coviello et al. (2014), who discuss the effect of dividing time among tasks, although with a single worker who works indefinitely. The time for completing a project mechanically is lower when fewer projects are active because time is divided among fewer projects.

per shift; physicians also reveal that they are willing to spend more than \$790 in hospital dollars per each hour of leisure saved, which is more than six times greater than the market wage.

This paper contributes to two strands of literature. First, a central economic question is how to induce workers to work efficiently, analyzed through the lens of incomplete contracts and the principal-agent problem (Simon, 1947; Hart and Holmstrom, 1987). Following seminal papers that evaluate a manager’s second-best optimal policy under hidden action or information, (e.g., Shapiro and Stiglitz, 1984; Aghion and Tirole, 1997; Milgrom and Roberts, 1988), I apply this framework to the design of scheduling and assignment, and I find that work assignment should be lower than first-best near EOS. This paper also contributes to an empirical literature on the relationship between workplace design and productivity.⁴ In particular, recent literature suggests that workplaces that grant greater flexibility to workers in how, when, and where work is performed have greater productivity (Ichniowski et al., 1996; Bloom et al., 2014).

Second, this paper sheds empirical light on the balance between extrinsic and intrinsic motivation (e.g., Benabou and Tirole, 2003). While workers no doubt care about their income and leisure, a now-substantial literature in economics recognizes that workers care about the “mission” of their job.⁵ In medical care, where information is continuous, multidimensional, and difficult to communicate, it would be extremely difficult to design incentives to provide the right care for patients if physicians only cared about income and leisure. By construction, salaries and schedules provide an environment in which extrinsic motives are muted relative to intrinsic ones, but the boundaries of schedules present a unique opportunity to study the tradeoff between private and intrinsic mission-oriented goals. In this paper, I find that the reduced-form tradeoff depends on the time during shift and quickly grows large in favor of private goals.

The issues I study in this paper are particularly relevant to health care delivery, which has experienced broad changes in the use of labor over the last few decades. Technological advances have caused a proliferation in the diagnostic and therapeutic decisions that should be made

⁴Relatedly, an interesting set of papers has studied timing distortions in nonlinear contracts such as sales incentive plans and government budgets (e.g., Oyer, 1998; Liebman and Mahoney, 2013; Larkin, 2014).

⁵The general case of intrinsic motivation has been discussed by Tirole (1986) and in later papers (Dewatripont et al., 1999; Akerlof and Kranton, 2005; Besley and Ghatak, 2005; Prendergast, 2007). Physicians balancing profit and patient welfare has been considered by Ellis and McGuire (1986), for example. In contrast, related empirical work has been relatively new, e.g., peer effects due to social incentives (Bandiera et al., 2005, 2009; Mas and Moretti, 2009) and the response to information arguably orthogonal to profits (Kolstad, 2013).

in rapid order from a patient’s presentation.⁶ Further, changes in work and society, including the emergence of dual-earner families, have driven worker preferences for more predictable yet flexible hours (e.g., Goldin, 2014; Presser, 2003). Thus, increasingly, health care is delivered by organizations, and schedules play an important role in assigning uncertain work (e.g., Briscoe, 2006; Casalino et al., 2003). These changes of course have parallels in other industries, which also feature increasingly interrelated and complex production.

The remainder of this paper is organized as follows: Section 2 describes the institutional setting and data. Section 3 discusses a conceptual framework to consider EOS effects. Section 4 investigates patient assignment rates relative to EOS. Section 5 reports EOS effects for patients who are assigned. Section 6 considers the relationship between shift overlap, workload, and patient-care distortion. Section 7 presents simulations of counterfactual policies of patient assignment. Section 8 discusses additional points of interpretation, and Section 9 concludes.

2 Institutional Setting and Data

2.1 Shift Work

I study a large, academic, tertiary-care ED in the US with a high frequency of patient visits. Like in virtually all other EDs around the country, work is organized by shifts. In the study sample from June 2005 to December 2012, shifts range from seven to twelve hours in length (ℓ). Shifts also differ in overlap with a previous shift (\underline{o}) or with a subsequent shift (\bar{o}) in the same location. I observe 23,990 shifts in 35 different shift types summarized by $\langle \ell, \underline{o}, \bar{o} \rangle$ (Table A-6.1).

For physicians working in these shifts, the end of shift (EOS) is simply the time after which they are allowed to go home if they have completed their work. Because I focus on behavior at EOS, I pay special attention to \bar{o} . This overlap is the time prior to EOS during which a physician shares new work with another physician who has begun work in the same location.⁷

“Location” refers to a set of beds in the ED in which a physician may treat patients. This

⁶A related result of technological advances is specialized knowledge, which requires care delivered in teams. Although technological advances have been widespread, see Messerli et al. (2005) for the particularly impressive example of modern cardiovascular care, compared to Dwight Eisenhower’s heart attack treatment in 1955.

⁷I distinguish between shifts that end with the closure of a patient location, or “terminal shifts” with $\bar{o} = 0$, and those continuing patient care with another shift in the same location, or “transitioned shifts” with $\bar{o} > 0$.

managerial definition may differ from broader physical areas, or “pods,” where physicians may see each other but may not share the same beds. That is, a pod may contain more than one managerial location. During my sample period, I observe two to three pods, with a new pod opening in May 2011, that at various times were divided into two to five managerial locations.

In the study period, the ED underwent 15 different shift schedule changes at the location-week level. Within each regime, the pattern of shifts could differ across day of the week (see Figure 1). As is common in scheduled work, shift times were designed around estimated workload needs, and schedule changes reflected changes in the flow of patients to ED. Some shift regime changes were merely minor tweaks in the times of specific shifts, while others involved larger changes. Shifts are scheduled many months in advance, and physicians are expected to work in all types of shifts at all times and locations. Physicians may only request rare specific shifts off, such as holidays and vacation days, and shift trades are rare. During a shift, physicians cannot control the volume of patients arriving to the ED or the patient types that the triage nurse assigns to beds. Throughout the entire study period, physicians were exposed to the same financial incentives: They were paid a clinical salary based on the number of shifts they work with a 10% bonus based on clinical revenue (measured by Relative Value Units, or RVUs, per hour) and modified by research, teaching, and administrative metrics.⁸ Although their salaries are based on numbers of shifts worked, physicians are not compensated for time worked past EOS.⁹

2.2 Patient Care

After arrival at the ED, patients are assigned to a bed by a triage nurse. This assignment determines the managerial location for the patient and therefore the one or more physicians who may assume care for the patient. Once the patient arrives in a bed, a physician may sign up for that patient on the computer order entry system. Physicians are expected to complete work on

⁸The metric of Relative Value Units (RVUs) per hour is a financial incentive that encourages physicians to work faster, because RVUs are mostly increased on the extensive margin by seeing more patients and are rarely increased by doing more for the same patients.

⁹This is the standard financial arrangement for salaried physicians across the US. Specifically, physicians are exempt from overtime pay as per the Fair Labor Standards Act of 1938 (FLSA). A large number of worker categories are exempt from overtime pay, including most positions with a high degree of discretion (see <http://www.dol.gov/elaws/esa/flsa/screen75.asp>).

any patient for whom they have assumed care, in order to reduce information loss with hand-offs (e.g., Apker et al., 2007), except in uncommon cases where the patient is expected to stay much longer in the ED. Because of this, physicians report often staying two to three hours past EOS.¹⁰ For patients arriving near EOS, physicians may opt not to start work and leave the patient for another physician. This option is more acceptable if this physician peer will arrive soon or has already arrived in the same location.

In addition to the attending physician (or simply “physician”), patient care is also provided by resident physicians or physician assistants and by nurses (not to be confused with the triage nurse). These other providers also work in shifts. Generally shifts of different team members do not end at the same time as each other, except when a location closes. More importantly, unlike physicians, care by nurses, residents, and physician assistants is more readily transferred between providers in the same role when they end their respective shifts, perhaps reflecting the lesser importance of their information in decision-making. For example, only physicians have the authority to make patient discharge decisions.

For physicians in the ED, the concept of patient discharge is a matter of discretion. Patient care is usually expected to continue after discharge, in either outpatient or inpatient settings. The key criterion for completion of work – or discharge – is whether the physician believes that sufficient information has been gathered for a discharge decision out of the ED. This decision is often made with incomplete diagnosis and treatment. Rather, the physician may decide to discharge a patient home with outpatient follow-up after “ruling out” serious medical conditions, or the physician may admit the patient for inpatient care if the patient could still possibly have a serious condition that would make discharge home unsafe.¹¹

Physicians may gather the information they need to make the discharge decision in several ways. Formal diagnostic tests are an obvious way to gain more information on a patient’s clinical condition. Treatment can also inform possible diagnoses by patient response, such as response to bronchodilators for suspected asthma. But time – for a careful history-taking, physical exam-

¹⁰In shifts with greater overlap, which have become more common, physicians report staying shorter amounts of time, but still up to one hour past EOS. Quantitative evidence using physician orders and patient discharge times is presented in Figure A-7.1 with a brief discussion in Appendix A-7.

¹¹In this ED, there is yet a third discharge destination to “ED observation,” if the patient meets certain criteria that make discharge either home or to inpatient unclear and justify watching the patient in the ED for a substantial period of time (usually overnight) to watch clinical progress.

ination, serial monitoring, or a well-planned sequence of formal tests and treatment – remains an important input in the production of information. Diagnostic tests and treatments can be complements or substitutes for time: Formal tests (e.g., CT and MRI scans) take time to complete and can thus prolong the length of stay, but testing can also substitute for a careful questioning or serial monitoring to gather information more rapidly.

2.3 Observations and Outcomes

From June 2005 to December 2012, I observe 442,244 raw patient visits to the ED. I combine visit data with detailed timestamped data on physician orders, patient bed locations, and physician schedules to yield a working sample of 372,224 observations. Details of the sample definition process are described in Table A-6.2. In the sample, I observe the identities of 102 physicians, 1,146 residents and physician assistants, and 393 nurses.

Because I focus on behavior near EOS, I present in Figure 2 the key variation across the 23,990 shifts in the time of day for EOS, shift length, and the overlap with another shift at EOS. Table A-6.1 lists the underlying number of observations for each shift type, in terms of hours, potential patients who arrive during a time when a shift of that type is in progress, and actual patients who are seen by a physician working in a shift of that type.

ED length of stay not only captures an important input of time in patient care but also largely determines when a physician can leave work. I measure length of stay from the arrival at the pod to entry of the discharge order. The timing of the discharge order, as opposed to actual discharge, is relatively unaffected by downstream events (e.g., inpatient bed availability, patient home transportation, or post-ED clinical care). I also use timestamped orders as measures of utilization and to create intervals of time within length of stay that are likely to be rough substitutes or complements with formal utilization, which I discuss further in Appendix A-3.

Since the primary product of ED care is the physician’s discharge decision, I focus on the decision to admit a patient as a key outcome measure, which has also received attention as a source of rising system costs (Schuur and Venkatesh, 2012; Forster et al., 2003). I accordingly measure total direct costs, which are the hospital’s internal measures of costs incurred both in

the ED and possibly during a subsequent admission.¹² Finally, I measure thirty-day mortality, occurring in 2% of the sample visits, and return visits to the ED within 14 days (“bounce-backs”), occurring in 7% of the sample (Lerman and Kobernick, 1987). However, these latter outcomes are less strongly influenced by the ED physician and depend on a host of factors outside the ED and hospital system, reducing the precision of their estimated effects.

2.4 Patient Observable Characteristics

When patients arrive at the ED, they are evaluated by a triage nurse and assigned an emergency severity index (ESI), which ranges from 1 to 5, with lower numbers indicating a more severe or urgent case (Tanabe et al., 2004). When the patient is assigned a bed, this information is communicated via a computer interface, together with the patient’s last name, age, sex, and “chief complaint” (a phrase that describes why the patient arrived at the ED).

In addition to elements displayed via the ED computer interface, I observe patient language, race, zip code of residence, and diagnostic information. The last characteristic of diagnoses is only incompletely known by physicians (or anyone) prior to assignment (via the “chief complaint”), especially since physicians do not interact with patients or examine their medical records prior to accepting them. I codify the diagnostic information into 30 Elixhauser indicators based on diagnostic ICD-9 codes for comorbidities (e.g., renal disease, cardiac arrhythmias) that have been validated for predicting clinical outcomes using administrative data (Elixhauser et al., 1998). Diagnostic codes of course are also partly determined by patient care.

3 Conceptual Framework

I introduce a simple model to consider how physician decisions – accepting patients and choosing inputs to care – may be distorted under work schedules. While the model is tailored to ED physicians, the key distortionary elements of the model are the following: (1) Workers have private information about their tasks; (2) workers care less about the social consequences of their actions, relative to their own income and leisure; and (3) *ex post* worker-task specificity

¹²Direct costs are for services that physicians control and are directly related to patient care. Indirect costs include administrative costs (e.g., paying non-clinical staff, rent, depreciation, and overhead). These costs are internal valuations of actual resources. They are not charges and are unrelated to billing or revenue source.

prevents workers from simply passing off tasks at EOS (Briscoe, 2007; Goldin, 2014).

3.1 Model Setup

Consider a physician in a shiftwork arrangement: She has a contract to arrive at shift beginning \underline{t} and stay until EOS \bar{t} or whenever she discharges her last patient, whichever is later, and she will receive a lump-sum payment y for this. Now consider a patient arriving at time $t < \bar{t}$. The relevant welfare parameters of her work environment is captured by \mathcal{E}_t , which includes the start time of the physician's shift, \underline{t} , and her current workload, w_t , and the start time and workload of a potential peer who may take the patient instead, (\underline{t}', w'_t) . The patient's underlying health state, $\theta \in \{0, 1\}$ for whether the patient is healthy ($\theta = 0$) or sick ($\theta = 1$), is unobservable, but $\Pr(\theta = 1) = p$ is publicly known. The physician takes the following actions:

1. Given t , \mathcal{E}_t , and p , the physician decides on $a \in \{0, 1\}$, whether to accept the patient ($a = 1$) or not ($a = 0$).
2. If she accepts the patient, she observes private information \mathcal{I} so that $\Pr(\theta = 1|\mathcal{I}) = p'$, and $|p' - \theta| < |p - \theta|$.¹³ She decides on inputs \mathbf{z} in patient care: time τ and formal tests and treatments z .
3. The physician observes $\hat{\theta} = \theta$ with probability $q(\mathbf{z}) \in (0, 1)$ and no information ($\hat{\theta} = \emptyset$) with probability $1 - q(\mathbf{z})$. She decides $d \in \{0, 1\}$, to admit ($d = 1$) or discharge home the patient ($d = 0$).
4. The patient's health state θ is observed, and the physician receives the following utility:

$$U = \begin{cases} y + \lambda O(\theta; \mathcal{E}_t), & a = 0 \\ y - \tilde{c}_\tau(\tau) + \lambda(v(\theta, d) - c(\mathbf{z})), & a = 1 \end{cases}. \quad (1)$$

Utility is stated in dollar terms, where physician income y does not depend on her actions.¹⁴

¹³I rule out private information before patient acceptance in this model. This is generally consistent with the institutional setting, and I examine this empirically as Assumption 1 below.

¹⁴In scheduled work y mostly depends *ex ante* availability, not *ex post* time past EOS. This model can accommodate some rewards correlated with staying past EOS (e.g., financial incentives for seeing more patients, social recognition); all that it requires is that physicians are relatively uncompensated for leisure.

$O(\theta; \mathcal{E}_t)$ is the value of the “outside option” if $a = 0$, which depends on θ and the work environment \mathcal{E}_t . $v(\theta, d)$ is the value of making discharge decision d for patient with health θ . $c(\mathbf{z})$ is the cost of patient care inputs, from which I separate $\tilde{c}_\tau(\tau)$, the cost of foregone leisure if the physician stays past EOS. $\lambda \in (0, 1)$, and $1 - \lambda$ is the wedge by which the physician undervalues the mission of patient care.

To be clear about the wedge, first consider the social welfare function as equivalent to Equation (1), except without λ (i.e., $\lambda = 1$). As $\lambda \rightarrow 1$, physician utility approaches social welfare, and the agency problem disappears. As $\lambda \rightarrow 0$, utility approaches the standard labor supply model in which workers only care about consumption and leisure. If $\lambda = 0$ (which I rule out), the physician would have no incentive to make the right decisions (despite observing \mathcal{I} and sometimes θ).

3.2 Patient Care

I first examine EOS effects on patient care and the discharge decision, conditional on assignment ($a = 1$). Discharge decisions have important efficiency implications for resource utilization and patient health. Formally, patients with $\theta = 0$ should be discharged home, while those with $\theta = 1$ should be admitted: $v(0, 0) > v(0, 1)$ and $v(1, 1) > v(1, 0)$. Discharging a sick patient home is particularly harmful: $v(1, 1) - v(1, 0) > v(0, 0) - v(0, 1)$. Because of this last fact, if θ remains unobserved, the physician will admit if and only if $p' > p^*$, where $p^* < \frac{1}{2}$.¹⁵ In other words, the realized discharge decision d is a function of $\hat{\theta}$:

$$d(\hat{\theta}) = \begin{cases} \theta, & \hat{\theta} = \theta \\ \mathbf{1}(p' > p^*), & \hat{\theta} = \varnothing \end{cases}.$$

The probability q of observing θ is in turn a function of patient care \mathbf{z} .¹⁶ $q(\mathbf{z})$ is increasing

¹⁵This can be straightforwardly shown by noting that $E[V|d=0, p'=p^*] = E[V|d=1, p'=p^*]$.

¹⁶I abstract away from treatment within the ED that can improve the patient’s health. This can easily be incorporated into the model and would not change qualitative results, except that if z^* is increasing in p , then physicians will be less likely to accept *ex ante* sicker patients.

and concave with respect to both τ and z . τ and z may be net substitutes ($\partial^2 q / (\partial \tau \partial z) < 0$) or net complements ($\partial^2 q / (\partial \tau \partial z) > 0$) in production. Effective time per patient is reduced with higher workload w_t : $\partial^2 q / (\partial \tau \partial w_t) < 0$. This contrasts with formal inputs, for which I make the normalizing assumption $\partial^2 q / (\partial z \partial w_t) = 0$. The intuition behind this is that with more patients, a physician has to divide her time and attention between them, but formal utilization can be ordered with the click of a mouse. By the normalizing assumption, I focus attention on substitutability or complementarity between time and formal utilization. Costs in $c(\mathbf{z})$ and $\tilde{c}_\tau(\tau)$ are positive, continuous, increasing, and convex in their arguments. Define $\tilde{c}_\tau = 0$ for $\tau + t - \bar{t} \leq 0$. For simplicity, assume additive separability of each element of \mathbf{z} in $c(\mathbf{z})$, again to focus on the production of $q(\mathbf{z})$.

Proposition 1. *Denote inputs in Section 3.1 that maximize expected utility in Equation (1), conditional on patient assignment ($a = 1$), as $\mathbf{z}^*(t) = (\tau^*(t), z^*(t))$. Denote corresponding inputs that maximize welfare as $\mathbf{z}^{FB}(t) = (\tau^{FB}(t), z^{FB}(t))$.*

- (a) *As $t \rightarrow \bar{t}$, $\tau^*(t)$ weakly decreases, $z^*(t)$ may weakly increase (if τ and z are net substitutes) or decrease (if τ and z are net complements), and $E[d|\mathbf{z}^*(t)]$ weakly increases if $F_{p'}(p^*) < \frac{1}{2}$.*
- (b) *For all t , $\tau^*(t) < \tau^{FB}(t)$, and $E[d|\mathbf{z}^*(t)] > E[d|\mathbf{z}^{FB}(t)]$.*
- (c) *If τ and z are net substitutes, then $z^*(t) > z^{FB}(t)$ for all t , and $z^*(t) - z^{FB}(t)$ weakly increases in w_t , holding t constant. The reverse is true if τ and z are net complements dominates.*

As the physician nears EOS, she will shorten length of stay τ . The intensity of diagnostic tests and treatments may increase or decrease, depending on whether τ and z are net substitutes or complements, respectively. Finally, she observes θ with lower probability q . This increases admissions, as long as $F_{p'}(p^*) < \frac{1}{2}$, where $F_{p'}(\cdot)$ is the c.d.f. of p' conditional on p and $a^* = 1$ (i.e., as long as $\theta = 1$ with sufficient probability). Note that distortions in discharges are due solely through distortions in \mathbf{z}^* , which affect q : conditional on \mathbf{z} , physicians make the right d decisions. These distortions also increase with workload w_t , as this reduces the effective time per patient to produce q .

3.3 Patient Assignment

I next consider the physician's upstream decision to accept the new patient, $a \in \{0, 1\}$. The physician compares the outside option under $a = 0$, including whether the patient is likely to wait for care, and expected utility under $a = 1$,

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

where

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

Denote \underline{Q}^* as the threshold rules such that accepting the patient maximizes expected utility ($a^* = 1$) if and only if $E[O(\theta; \mathcal{E}_t)] > \underline{Q}^*$. Define $W(\mathbf{z}) \equiv E \left[v \left(\theta, d \left(\hat{\theta} \right) \right) \middle| \mathbf{z} \right] - c(\mathbf{z}) - \tilde{c}_\tau(\tau)$ as the net expected social welfare from assigning the patient to the physician, conditional on patient care inputs \mathbf{z} . It is easy to see that $\underline{Q}^* = W(\mathbf{z}^*) - (\lambda^{-1} - 1) \tilde{c}_\tau(\tau^*)$, 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 $\underline{Q}^{FB} = W(\mathbf{z}^{FB})$, when optimal \mathbf{z}^{FB} 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 \mathbf{z} . In this policy, $a^{SB} = 1$ if and only if $E[O(\theta; \mathcal{E}_t)] > \underline{Q}^{SB} = W(\mathbf{z}^*)$.

Proposition 2. *Consider a^* as the patient assignment decision in Section 3.1 that maximizes expected utility in Equation (1), a^{FB} as the assignment that maximizes expected welfare when optimal \mathbf{z}^{FB} is publicly known and contractible, and a^{SB} as the assignment that maximizes expected welfare when \mathbf{z}^{FB} is either publicly unknown or non-contractible. Assignment will follow threshold rules in which assignment occurs if and only if $E[O(\theta; \mathcal{E}_t)]$ is greater than a threshold. The respective threshold rules are \underline{Q}^* , \underline{Q}^{FB} , and \underline{Q}^{SB} , where $\underline{Q}^* < \underline{Q}^{SB} < \underline{Q}^{FB}$. $\underline{Q}^{FB} - \underline{Q}^{SB}$ and $\underline{Q}^{SB} - \underline{Q}^*$ increase as $t \rightarrow \bar{t}$ decreases or as λ decreases.*

There are first-best reasons for assignment to decrease near EOS. As $t \rightarrow \bar{t}$, the outside option $O(\theta; \mathcal{E}_t)$ increases because a peer is more likely to be arriving soon or already present, and $W(\mathbf{z})$ may also decrease due to fatigue and the possibility of foregone leisure.¹⁷ However, beyond this decrease, assignment decided by the physician, a^* , will be inefficiently low near EOS ($\underline{Q}^* < \underline{Q}^{FB}$). The second-best policy, in which physicians continue to choose \mathbf{z}^* , will assign patients at a threshold \underline{Q}^{SB} in between \underline{Q}^* and \underline{Q}^{FB} . The relative distance between these policy thresholds will depend on the curvature of W (i.e., $\partial^2 W / \partial \tau^2 < 0$ and $\partial^2 W / \partial z^2 < 0$ are very negative) relative to the misvaluation of leisure, $(\lambda^{-1} - 1) \tilde{c}_\tau(\tau^*)$. If W is very curved, then the physician will not distort \mathbf{z}^* by much for a given misvaluation of leisure, but if otherwise, then there is larger scope for patient-care distortions. The greater the patient-care distortions, the closer \underline{Q}^{SB} will be to \underline{Q}^* than to \underline{Q}^{FB} .

3.4 Remarks

The inefficiency in the model is fundamentally informational. First, physicians observe private information p' , so management does not know \mathbf{z}^{FB} or $E[d|\mathbf{z}^{FB}]$. Second, they are imperfect agents, overvaluing consumption and leisure relative to patient care. The canonical way to implement first-best would be to pay physicians an hourly overtime wage, in this case $(1 - \lambda) \tilde{c}'_\tau$ if one knows λ and \tilde{c}'_τ . However, this is often impractical within firms due to uncertainty and complexity.¹⁸ As Weitzman (1974) notes, under uncertainty, when there is more concavity in the benefits of production inputs than convexity in the input costs, setting quantities (e.g., schedules) dominates setting prices (e.g., overtime pay). Further, errors in setting the right prices in complex environments and could be distortionary, especially with moral hazard. In fact, prespecifying schedules and fixed pay removes patient-care distortion *within* the shift.¹⁹

¹⁷Another version of the patient assignment question is patient selection (i.e., how $E[p|a^* = 1]$ changes as $t \rightarrow \bar{t}$). Selection will likely be towards healthier patients: For low p and as $t \rightarrow \bar{t}$, expected utility under $a = 1$ likely diminishes less quickly, and expected utility under $a = 0$ likely increases less quickly. I will examine this empirically in Section 5.1.

¹⁸As in footnote 9, physicians and other professionals with discretion, as opposed to nurses for example, are largely not paid overtime. Complex work not easily defined by time, discretion, and moral hazard issues noted in this paragraph perhaps form an economic microfoundation for this.

¹⁹This is possible as long as physicians can be guaranteed to leave by EOS, which can be mostly implemented by avoiding new work near EOS, rather than discharging patients earlier within the shift.

In this informational environment, work assignment is a natural policy lever, since assignment is easy to observe and influence. Physicians may be assigned too little work to justify the value of their time. However, assigning more work worsens distortions in patient-care decisions, \mathbf{z} , both by assigning patients to physicians under time pressure, and via the dynamic of increasing workload and therefore reducing effective time per patient.²⁰ In Section 6.2, I empirically focus on shift overlap $\bar{o} \equiv \bar{t} - \underline{t}'$ as one mechanism that influences a_t^* through changing $O(\theta; \mathcal{E}_t)$. More broadly, a_t may be implemented by a variety of managerial instruments, such as piece-rate pay, social norms, or formal assignment policies. Therefore, while I model a_t as a physician choice here, in Section 7, I more generally consider it as a sufficient-statistic policy instrument.

4 Patient Assignment

In this section, I describe patient assignment near EOS. As in Proposition 2, it is natural that physicians will be less likely to accept patients as EOS nears, because time for patient care is costlier. Patient assignment to physicians can also be influenced by assignment to locations (particularly in location-times with only one physician). The simple analysis in this section presents unadjusted average rate of patient assignment to a physician nearing EOS across a variety of shift types. In particular, I verify that physicians are assigned fewer patients near EOS, and that greater overlap \bar{o} allows physicians to decline patients earlier relative to EOS.

Figure 3 presents the hourly average rates of new patient assignments in 30-minute bins relative to EOS. Each panel represents shifts with a different \bar{o} and shows assignments for the index physician (patients accepted) and for the location inclusive of the index physician (patients assigned by the triage nurse). Physicians are generally assigned two or three new patients in an hour, and assignment rates are highest near the beginning of shift. For $\bar{o} > 0$, assignment rates show two relationships with time. First, patient flow declines in the hour *prior* to the transitioning peer’s arrival at the location. Second, patient flow declines close to zero in the two to three hours prior to EOS. If there is sufficient \bar{o} , patient flow is relatively constant but

²⁰I address this dynamic empirically but have not explicitly modeled in this conceptual framework. This may be formally considered in an expanded dynamic model with two patients arriving at different times, t and $t + 1$, and respective decisions (a_t, \mathbf{z}_t, d_t) and $(a_{t+1}, \mathbf{z}_{t+1}, d_{t+1})$. Increasing a_t increases w_{t+1} and thus, from Proposition 1, reduces welfare by worsening distortions in \mathbf{z}_{t+1}^* and d_{t+1}^* .

diminished in that duration. For $\bar{o} = 0$, the decline in patient flow begins earlier, at least four hours prior to EOS. Also in Figure 3, patients who are not accepted by the index physician may wait up to an hour to be seen by a peer yet to arrive, but patient flow to transitioning peers generally at least makes up for the decline in flow for the index physician. That is, despite declines in patient assignment to the leaving physician, patients continue to arrive at the pod at similar or greater rates prior to the peer’s transitioning shift.

The earlier arrival of peers allows for earlier reductions in patient assignment relative to EOS. The reductions are in fact *prior* to peer arrival, especially in shifts with shorter transitions, suggesting anticipatory behavior. For terminal shifts with no peer arriving in the same location, the long decline in patient flow rates is implemented by the *triage nurse* assigning fewer patients to the physician nearing EOS. Thus, “slacking off” is achieved between coworkers sharing a location and, in cases without coworkers (when physicians have no choice), by managerial assignment itself.

5 Effect on Patient Care

5.1 Identification and Balance

The panel nature of the data allows for me to control for time categories (e.g., time of the day or day of the week) because shifts start at different times. Furthermore, variation in shift lengths identifies the EOS effect separately from fatigue or other effects that depend on time relative to the beginning of shifts.²¹ Finally, I observe the same physicians and support staff (i.e., physician assistants, nurses, and residents) and can therefore control for their identities as a given physician approaches EOS. However, given that the number of assigned patients clearly declines as physicians near EOS, an important question is how the EOS effect on patient care is identified separately from patient selection correlated with time to EOS.

Figure 4 shows statistics of characteristics for patients assigned to physicians in 30-minute bins relative to EOS. Compared to numbers of patients assigned (Figure 3), and compared to variation *within* bins, patient characteristics are relatively constant across bins. There is a slight

²¹In alternative models, I also control for cubic splines of total number of patients seen prior to the index patient’s arrival. Results (not shown) are essentially identical with these additional controls.

trend toward younger patients being assigned to physicians nearing EOS. Figure 5 similarly shows that the distribution of predicted length of stay, using *ex ante* characteristics of age, sex, ESI, race, and language that are presumably observable at the time of assignment, is quite stable across time relative to EOS. This descriptive evidence suggests that assignment policies relative to EOS are primarily based on patient numbers rather than patient characteristics.

To be clear about identification, I consider two independent assumptions:

Assumption 1 (Excludable Characteristics). *Conditional on ex ante patient characteristics, time of arrival, pod, and providers, patient excluded characteristics are orthogonal to assigned time relative to EOS.*

Assumption 2 (Random Arrival Times). *Conditional on time of arrival, pod, and providers, patient characteristics are orthogonal to ED arrival times with different propensities for assignment to times relative to EOS.*

The intuition behind Assumption 1 is that assignment within the ED operates through patient numbers and (to a much lesser extent) on *ex ante* patient characteristics. Clinical characteristics that determine outcomes are mostly unknown to physicians before assignment and are assumed to be excluded from assignment policies. This assumption is testable to the extent that I observe *ex post* clinical characteristics and can show that, conditional on *ex ante* patient characteristics, time categories, and staff identities, there exists no correlation between clinical characteristics and assignment to times relative to EOS. The intuition behind Assumption 2 is that, although physicians may influence assignment conditional on patient arrival, patients arrive at the ED without any systematic selection toward times when there may or may not be a physician near EOS. The shifts at the time of ED arrival drive the propensity of being assigned to a physician near EOS but are conditionally orthogonal to any characteristics of the arriving patients.²²

I assess the plausibility of each assumption in a regression framework. To assess Assumption 1, I regress presumably excludable patient characteristics based on Elixhauser diagnoses – specif-

²²I condition on pod and providers because arrival time to triage is not always observed in the data and thus time to *ED floor* is my preferred measure of time. This mechanism of random assignment therefore primarily addresses assignment to physician conditional on assignment to pod but also asserts that the triage nurse decision only depends on pod and physician identities. In practice, however, controlling for physician and pod identities does not matter much for balance.

ically (i) predicted length of stay by diagnoses and (ii) count of diagnoses – on hour relative to EOS, controlling for *ex ante* patient characteristics, time categories (hour relative to beginning of shift, hour of the day, day of the week, and month-year interactions), pod, and provider identities. To evaluate Assumption 2, I regress both *ex ante* and *ex post* patient characteristics on hourly propensities for assignment to a physician at a time relative to her EOS, controlling for time categories, pod, and provider identities. I consider (i) predicted length of stay by all patient characteristics, (ii) age, (iii) male sex, (iv) ESI, (v) white race, (vi) black race, (vii) English language, (viii) Spanish language, and (ix) count of Elixhauser indices. Table A-1.1 presents balance results for other predicted outcomes, assessing both Assumptions 1 and 2. These regression specifications closely parallel those used to estimate main EOS effects next.²³

Table 1 reports results consistent with conditional random assignment under either Assumption 1 (the first two columns) or Assumption 1 (the remaining columns). None of the regressions yield jointly significant coefficients on hour relative to EOS. Furthermore, Table 1 shows that variation in any of the patient characteristics, residualized by covariates depending on the assumption, remains quite large. In fact, if Assumption 1 were violated, and if patients with the lowest predicted length of stay based on *ex post* clinical characteristics were perfectly sorted to hours closest to EOS when more than one physician is available, the bias in the last hour of shift would be -0.177 , more than 17 standard errors apart from the corresponding coefficient in the corresponding balance test (Column 1 of Table 1).

5.2 Main EOS Effects

In the full specification, I estimate the following equation:

$$Y_{it} = \alpha_{\overline{m}(i,t)} + \gamma_{\underline{m}(i,t)} + \beta \mathbf{X}_i + \eta \mathbf{T}_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_{it}. \quad (2)$$

Outcome Y_{it} is indexed for patient visit i at time t , and the object of interest is arrival hour

²³Appendix A-1 gives further formal details. Included in this appendix is an approach introduced by Altonji et al. (2005) to quantify selection on unobservables necessary to explain observed treatment effects. The appendix additionally takes an approach closely following Chetty et al. (2014) to quantify “forecast bias.” Finally, the appendix considers selection on other predicted outcomes of cost, orders, and admission.

$\overline{m}(i, t) = \lceil \overline{t}(i) - t \rceil$ prior to EOS, where seven or greater hours prior to EOS is the omitted category. I control for time relative to the shift beginning $\underline{m}(i, t) = \lfloor t - \underline{t}(i) \rfloor$, patient characteristics \mathbf{X}_i , time categories \mathbf{T}_t (for month-year, day of the week, and hour of the day), pod $p(i)$, and physician $j(i)$ and team $k(i)$ (i.e., resident or physician assistant, and nurse) identities.

Table 2 shows results for log length of stay from versions of Equation (2) with varying sets of controls. All models estimate highly significant and negative coefficients for approaching time to EOS, with visits seven or more hours prior to EOS being the reference category. The reduction in length of stay grows larger in magnitude as time approaches EOS. By the last hour prior to EOS, versions of Equation (2) estimate effects on log length of stay ranging from -0.53 to -0.72 . The full model, shown in Column 5 of Table 2 and plotted in Panel A of Figure 6, estimates an effect on log length of stay of -0.59 in the last hour and serves as the baseline model for this paper.

Results are essentially indistinguishable whether or not all patient characteristics or only *ex ante* characteristics are included (Columns 5 and 6 in Table 2), as suggested by Assumption 1. More generally, results are qualitatively unchanged regardless of whether I control for *any* patient characteristics, time categories, pod dummies, provider identities, or time relative to shift beginning. For example, the difference between Columns 4 and 5 columns represents the effect of time relative to shift beginning, which can include fatigue and is separately identified from EOS effects due to variation in shift lengths. This difference, about 0.13 in the last hour prior to EOS, also accounts for only a minor portion of the overall effect.²⁴

Columns 7 and 8 in 2 report alternative estimates of the EOS effect on length of stay under Assumption 2. Under this assumption, estimates are robust to selection across physicians within hour of arrival and are identified by variation in available shifts (and corresponding hours relative to EOS) *across* hours of arrival. The specification under this assumption is

$$\overline{Y}_t^* = \sum_{\overline{m}=6}^1 \alpha_{\overline{m}} P_{\overline{m}}(t) + \varepsilon_t, \quad (3)$$

where $\overline{Y}_t^* \equiv \frac{1}{N_t} \sum_{i \in I(t)} Y_{it}^*$ is the average of residualized length of stay Y_{it}^* for the set $I(t)$ of N_t

²⁴See Appendix A-2 for more direct results on effects relative to shift beginning.

patients arriving at t , and $P_{\overline{m}}(t)$ is the fraction of these patients being assigned to a shift with \overline{m} hours prior to its end.²⁵ In this specification each observation is an arrival hour. Results in Column 7 and 8, which respectively either include or omit *all* patient characteristics, closely match each other and thus support Assumption 2.

Table 3 shows results for other outcome measures, including the order count, inpatient admission, log total cost, 30-day mortality, and 14-day bounce-backs. Estimates for α_m are generally insignificant for hours before the last hour prior to EOS, but are significantly positive in the last hour. Patients arriving and assigned in the last hour prior to EOS have 1.4 additional orders for formal tests and treatment, from a sample mean of 13.5 orders.²⁶ These patients are also 5.7 percentage points more likely to be admitted, which is 21% relatively higher than the sample mean of 27%. Log total costs are 0.21 greater in the last hour prior to EOS. Mortality and bounce-backs do not exhibit a significant effect with respect to EOS, although these outcomes are either rare (mortality) or imprecisely predicted (bounce-backs). I plot coefficients for orders, admissions, and total costs in Panels B to D of Figure 6.

6 Shift Overlap, Workload, and Distortion

I evaluate how workload and patient-care effects vary across shifts with varying overlap near EOS, for two purposes: First, this supports the interpretation that EOS effects reflect inefficiency, under the identifying assumption that the EOS by itself has no first-best implications for patient care, conditional on volume of work, time since beginning work, and time since a peer’s arrival. Formal overlap (i.e., time between peer arrival and EOS) only changes when a physician is *allowed* to leave work. Second, this analysis uses shift structure as a concrete example of patient assignment as a policy lever. Through patient assignment, a planner can influence the efficiency of patient care: Assigning fewer patients to physicians on schedules may underutilize the value of their time, but assigning more patients worsens the EOS distortion in the use of time as an

²⁵As argued by Chetty et al. (2014), Y_{it}^* is calculated using *within* \overline{m} variation. Section A-1.3 presents an approach and corresponding results that closely follow Chetty et al. (2014) in a way that estimates a single measure of “forecast bias” and additionally accounts for within-shift endogeneity by estimating “jack-knifed” predicted $\hat{\alpha}_{\overline{m}}$ for shift using data that does not include that shift. As suggested elsewhere, results in that section fail to reject the null hypothesis of no forecast bias.

²⁶This suggests that formal orders are a net substitute for time. See Appendix A-3 for more direct results supporting this hypothesis.

input to care.

6.1 Patient Censuses over Time

As a descriptive exercise, I first measure workload $w(j, t)$ as the number of patients cared for by physician j (her “census”) at time t :

$$w(j, t) = \sum_{j(i)=j} \mathbf{1}(t \geq t'(i)) \mathbf{1}(t \leq t'(i) + \tau(i)), \quad (4)$$

the count of visits arriving prior to t at $t'(i)$ and staying past t until $t'(i) + \tau(i)$, where $\tau(i)$ is length of stay.

Figure A-7.2 shows unadjusted census averages in 30-minute intervals in different shift types by \bar{o} . Average censuses start at around two patients at the beginning of all shift types, representing unstaffed patients from the previous shift, except for shift types with $\bar{o} = 2$, which happen not to transition from another shift (i.e., $\underline{o} = 0$). Patients remain on the census at EOS. The number of patients remaining on census in the last 30 minutes prior to EOS is consistently close to four, with the exception of shifts with $\bar{o} = 1$, which have censuses of about six.

Because of client-worker specificity, physicians must usually either reduce censuses to close to zero or at least have a well-defined pass-off plan to instruct another physician.²⁷ With smaller \bar{o} , EOS defines earlier times relative to peer arrival after which physicians are *allowed* to go home. Holding constant time from beginning of shift and time from peer arrival, smaller \bar{o} thus induces a greater scope for distortionary care near EOS.

6.2 EOS Effects by Shift Overlap

I empirically consider how patient-care EOS effects may differ by shift overlap. Larger patient-care effects with small \bar{o} , conditional on time from beginning of shift, are consistent with distortionary care. Further, the interaction provides evidence of the intuitive tradeoff between extensive and intensive margins of distortion: If physicians have more time to slack off before

²⁷As described in Section 2.2, physicians report that they therefore generally stay at least an hour after EOS. This is also supported by the timing of physician orders relative to EOS, shown in Figure A-7.1. Also, while physicians have an average of four patients at EOS in shifts with $\bar{o} = 0$, these patients are much harder to transfer than in transitioned shifts.

EOS (because EOS is later relative to peer arrival), patient-care distortions will be smaller under this tradeoff.

I consider three categories \overline{O} of overlap at EOS – terminal shifts ($\overline{o} = 0$), minimally transitioned shifts ($\overline{o} = 1$), and substantially transitioned shifts ($\overline{o} \geq 2$)²⁸ – and estimate

$$Y_{it} = \alpha_{\overline{m}(i,t)}^{\overline{O}} \mathbf{1}(\overline{o}(i) \in \overline{O}) + \gamma_{\underline{m}(i,t)} + \beta \mathbf{X}_i + \eta \mathbf{T}_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_{it}, \quad (5)$$

which is similar to Equation (2) but interacts the hourly EOS effects with the categories \overline{O} , where $\overline{o}(i)$ is a function assigning visit i to overlap \overline{o} of the shift to which the visit is assigned. I normalize coefficients so that, as before, the reference category includes times that are seven hours or more before EOS in each of the overlap categories.

Figure 7 shows interacted EOS effects on length of stay, orders, admission, and total costs. The EOS effect on length of stay is largely similar across shift categories (Panel A). All three shift categories show a substantial decline in length of stay as EOS approaches. However, EOS effects are notably absent in shifts with $\overline{o} \geq 2$ for orders, admission probability, and total costs (Panels B to D). In contrast, shifts with $\overline{o} \leq 1$ show large increases in orders, admissions, and total costs at EOS.

6.3 Effective Time per Patient

The evidence above suggests a link between patient assignment, workload, and patient care: Assigning physicians more patients near EOS increases workload and thus decreases the effective time physicians spend on each patient’s care. To directly assess this concept, I create a new outcome measure of *workload-adjusted length of stay*, which normalizes length of stay by the physician’s average census during the stay. That is, for visit i arriving at t , I divide length of stay, $\tau(i)$, by the average census $\overline{w}(i)$ under the assigned physician, $j(i)$, over the course of the

²⁸While I observe shifts with $\overline{o} \in \{2, 3\}$, they entail very few observations, as listed in Table A-6.1. Results are essentially unchanged whether I omit these observations or consider them as belonging to the minimally transitioned shift category.

i 's length of stay (from t to $\tau(i)$):

$$\tau(i) / \bar{w}(i) = \tau(i) \left[\frac{1}{\tau(i)} \int_{\tilde{t}=t}^{t+\tau(i)} w(j(i), \tilde{t}) d\tilde{t} \right]^{-1}, \quad (6)$$

where census $w(j, t)$ is defined by Equation (4).

I then regress the log of workload-adjusted length of stay using Equation (2).²⁹ The last column of Table 3 shows that workload-adjusted length of stay decreases significantly only in the last hour prior to EOS. Thus, adjusting length of stay for workload reconciles previous results in which length of stay progressively decreases as EOS approaches, but orders, admissions, and costs increase only in the last hour. Further, in Table 4, I show that workload-adjusted length of stay is only decreased in the last hour of shift when $\bar{o} \leq 1$. When $\bar{o} \geq 2$, workload-adjusted length of stay does not decrease near EOS and, if anything, slightly increases prior to the last hour of shift.³⁰ These relationships suggests workload-adjusted length of stay as a relevant measure of time that enters into the physician production function, particularly when considering distortions in orders, admissions, and costs relative to EOS.

7 Counterfactual Assignment Policies

Despite important variation in patient assignment across shifts with different \bar{o} , observed assignment – either between physician peers or by the triage nurse – dramatically diminishes near EOS in all shifts. In this section, I assess the efficiency implications of a fuller range of counterfactual assignment policies, including policies that do not significantly diminish assignment near EOS. I consider the assignment of work as a sufficient statistic for a wider range of managerial policies (e.g., rules, financial incentives) including but not limited to shift overlap.

While I consider assignment policies as exogenous, I allow physicians full discretion in how they respond to assignment. The intuition is that, while work assignment is easily observable and

²⁹This is different than controlling for current census; results in Table 2 are unchanged when flexible splines of current census are included in Equation (2). Instead, workload-adjusted length of stay solely captures *future* actions by the physician, including future censuses. Otherwise including future censuses as covariates in a regression framework would be problematic.

³⁰If anything workload-adjusted length of stay slightly increases prior to EOS when $\bar{o} \geq 2$. Such increases do not appear to be associated with changes in other outcomes of orders, admissions, or costs, which could be consistent with increases in length of stay for strategic purposes, or “foot-dragging,” as discussed in Chan (2016).

therefore a natural managerial policy, the downstream patient care decisions – particularly the use of time – are much more difficult to monitor or manage. To evaluate welfare, I consider overall costs due to physician time, patient time, and hospital resources. While assigning more patients to physicians near EOS reduces the number of physician-hours (i.e., the amount of overlap) required to process patient flow, it increases patient-care costs as physicians endogenously reduce time spent with patients near EOS.

7.1 Simulation Routine

At a high level, there are three components of simulating outcomes under counterfactual assignment policies (details in Appendix A-4). First, I specify counterfactual assignment policies as exogenous and sufficient statistics of policies under a planner’s control. Second, I model physician discharge responses, which conditional on assignment policies, are not subject to a planner’s control or are otherwise non-contractible. Third, I link discharge responses to other welfare-relevant outcomes such as hospital costs. I discuss each of these in turn.

Consider an ED physician in state S , which includes shift characteristics $\langle \ell, \underline{o}, \bar{o} \rangle$ that the physician is currently working in, time categories (e.g., time of day, day of week), time relative to EOS, and the set of patients I (and their characteristics) already under her care. An assignment policy $\mathcal{A}(S)$ is a mapping from S to some distribution of discrete assignments. The actual assignment policy $\mathcal{A}_0(S)$ is nonparametrically identified, and using discrete time intervals of five minutes, I estimate it using an ordered logit model. As a convenient specification based on patterns observed in Figure 3, I consider counterfactual assignment policies $\mathcal{A}_\Delta(S)$ that adjust the way time to EOS is considered by up to a constant Δ . That is, if $S = (\bar{m}, S^-)$, where \bar{m} is time relative to EOS, and S^- is everything else, then $\mathcal{A}_\Delta(S) \equiv \mathcal{A}_0(\bar{m}(\Delta), S^-)$. New patients are assigned similarly, except that time is *as if* the physician were $\bar{m}(\Delta)$ hours instead of \bar{m} hours from EOS, holding S^- fixed. If $\bar{m}(\Delta) < \bar{m}$ (or $\Delta < 0$), then fewer patients will be assigned; if $\bar{m}(\Delta) > \bar{m}$ (or $\Delta > 0$) then more patients will be assigned.

Physician discharge policies $\mathcal{D}(S)$ map S to probabilities of each potential patient discharge $i \in I$. The actual physician discharge policy $\mathcal{D}_0(S)$ is also nonparametrically identified, and I estimate $\Pr(i|S)$ using a multinomial logit model. These probabilities are indeed the conditional

choice probabilities of a dynamic programming problem, as explained in Hotz and Miller (1993). In the actual equilibrium, $\mathcal{D}_0(S)$ incorporates both the utility flow of discharging i appropriately and dynamic considerations of foregone leisure by staying past EOS. If physicians do not anticipate systematic changes in assignment from $\mathcal{A}_0(S)$ to $\mathcal{A}_\Delta(S)$, then $\mathcal{D}_0(S)$ can be used to simulate counterfactual discharge decisions. Although I will consider assignment policies $\mathcal{A}_\Delta(S)$ that substantially change assignments near EOS, expected workloads under $\mathcal{A}_\Delta(S)$ near EOS remain very much within those observed in the data. This suggests that using $\mathcal{D}_0(S)$ could be a valid strategy with the considerable advantage that it does not rely on any other modeling assumptions. Nonetheless, using a recent method by Barwick and Pathak (2015) that permits complex state spaces, I also specify a formal dynamic programming model and use estimated primitives to calculate a counterfactual discharge policy $\tilde{\mathcal{D}}_\Delta(S)$ for each Δ , where the tilde indicates additional restrictions for the dynamic model to be identified (details in Appendix A-5).

Finally, I model the relationship between discharge decisions and outcomes such as hospital costs. Recall that under random assignment (Section 5.1) comparing patients treated by physicians at varying times relative to EOS implies causal effects on length of stay and formal costs (i.e., those incurred by lab tests, procedures, and hospital admission). Under the assumption that formal costs are affected via the EOS mechanism of reducing lengths of stay, results in Sections 5 and 6 imply a marginal rate of technical substitution between time and cost from formal inputs and hospital admission. This is again nonparametrically identified, although based on results in Section 6.3, I choose to use workload-adjusted length of stay as the relevant measure impacting hospital costs.³¹

For each counterfactual policy Δ in the range of $[-4, 4]$ hours, I simulate assignments via \mathcal{A}_Δ and discharges via \mathcal{D}_0 or $\tilde{\mathcal{D}}_\Delta$, from the beginning to the end of each shift in the data. In each simulation $r = 1, \dots, 20$ of each policy Δ , I calculate total counterfactual costs,

$$\text{Costs}_{\Delta,r} = \text{PhysicianTime}_{\Delta,r} + \text{PatientTime}_{\Delta,r} + \text{HospitalResources}_{\Delta,r}, \quad (7)$$

³¹The elasticity estimate is motivated by the fact that both observed total costs (Figure 6) and observed workload-adjusted length of stay (Figure A-7.3) increase only in the last hour prior to EOS, in Section 6.2. In simulated data, I calculate workload-adjusted length of stay decreases by 18.1% in the last hour of shift when $\Delta = 0$, an estimate very close to but more conservative than based on actual data in Table A-6.3. Since total costs increase by 20.8% in the last hour prior to EOS, I calculate the elasticity as $20.8\% / -18.1\% = -1.15$. More detail is given in Appendix A-4.4.

which I take as a measure of welfare, under the conservative assumption that patient health is unaffected despite EOS distortions in time, formal utilization, and admissions.³² $\text{PhysicianTime}_{\Delta,r}$ captures additional wages that the ED must pay in order to meet patient flow and may increase for two reasons. First, if fewer patients are assigned before EOS, a peer must arrive earlier because backlog occurs earlier. On the other hand, if more patients are assigned prior to EOS, the index physician must stay later past EOS, and this foregone leisure is also valuable. I value physician time with a base-case wage of \$120/hour, which is close to actual wages in this ED and national averages of hourly pay, although results are insensitive to wages several multiples higher. I value patient time, in $\text{PatientTimeCosts}_{\Delta}^r$, at \$20/hour, so that shorter lengths of stay are valuable from a patient-perspective, all else equal. Finally, I model $\text{HospitalResources}_{\Delta}^r$ as the EOS effect of increasing hospital costs – from formal inputs and hospital admission – when physicians decrease workload-adjusted length of stay, as discussed above.

7.2 Results

Increasing assignment near EOS results in large increases in resource-utilization costs, which dominates any physician-time savings that may accrue from a later-arriving peer. For example, an assignment policy that results in physicians staying an extra hour past EOS by assigning them about four additional patients also results in an increase of \$3,700 in hospital spending per shift, for a net increase of \$3,300 in total costs, compared to a policy that assigns those four patients to a physician not ending shift. Figure 8 shows average changes in total costs per shift, stated in Equation (7), under counterfactual policies, where policies are shown in terms of changes in the number of patients assigned.³³ Both reducing and increasing patient assignment near EOS increase total costs relative to those under the actual assignment policy. This suggests that the observed pattern of assignment – implemented by shift overlap, triage nurse assignment, and physician responses – is approximately second-best optimal. This qualitative finding is robust

³²In sample, recall that I find no effect on mortality or bounce-backs (Table 3). Out of sample, with greater assignments near EOS, the EOS distortion may also worsen health outcomes, which implies even greater importance to reduce assignments near EOS. However, this should not matter for Δ close to 0 and therefore for statements about optimal assignment if the optimal assignment policy occurs close to the observed assignment regime.

³³Changes in costs with respect to changes in patients assigned is easier to understand than the policy index Δ , since Δ is only *maximum* amount of change in time in the counterfactual assignment policy (i.e., $|\overline{m}(\Delta) - \overline{m}| \leq \Delta$). This can be appreciated in plots of curtailed assignment policies in Figure A-4.1, in which $\Delta = -2$ is still very similar to $\Delta = 0$.

to whether the actual discharge policy \mathcal{D}_0 or the model-based counterfactual policy $\tilde{\mathcal{D}}_\Delta$ is used and is true even under an extreme assumption of a \$600 per hour physician wage.

The intuition behind this robust finding is that, in the actual assignment regime \mathcal{A}_0 , very few patients are assigned in the last hour of shift, and this is precisely when costs start increasing. Reducing assignment near EOS would not substantially change in absolute terms the number of patients with EOS-induced distortions in hospital resources, but of course would increase the physician wage bill by needing to staff more physician hours before shift changes. On the other hand, increasing assignment near EOS is much costlier for two reasons. First, patient care costs are much higher than physician wages. Second, increasing assignment not only exposes more patients to EOS cost distortions but also worsens the distortions on a per patient basis if the relevant time input measure is time per patient. This also suggests that, if a planner were uncertain about when the optimal assignment policy should begin reducing assignments, it would be relatively cheaper to pay for idle physician time as opposed to risking patient care distortions.

Another way to use the simulation is to assess the implicit tradeoff physicians make between foregoing leisure and increasing resource-utilization costs. At each point in time relative to normal completion, I compute the dollar value of extra resource-utilization costs incurred per leisure hour gained, shown in Figure 9 (details in Appendix A-4.5). As indicated earlier, the actual assignment policy results in minimal patient-care distortions, reflecting a low “value” of leisure (below the market wage of \$120 per hour) immediately at EOS. This could reflect norms to stay past EOS, reportedly two to three hours past EOS, which I discuss in the subsequent section. However, the value of leisure quickly rises above market wage at 10 minutes past the normal time of completion. At about one hour past EOS, physicians are willing to expend \$790 in hospital resources in order to avoid an additional hour at work. Under a strict interpretation of Equation (1), this implies that λ is decreasing after EOS and that $\lambda = \$120/\$790 = 0.15$ at about one hour past EOS.³⁴

³⁴I do not estimate a single λ during calibration. Rather, the simulation exercise calibrates the physician discharge decision to many moments in the data (see Figures A-5.1 and A-5.2). Given no (assumed) worsening of patient health, these values of extra resource-utilization costs per leisure hour gained thus reveals λ at each point in time. In Appendix A-6 (Figure A-6.1) I show reduced-form evidence that the value of leisure appears higher in shifts ending at “prime” hours between 4 p.m. and 8 p.m. after which others are likely to be off work or out of school and awake.

8 Discussion

The main focus of this paper is to assess a simple but, to my knowledge, unexplored consequence of work schedules: When work past scheduled availability is undercompensated, workers will avoid new work, and the use of time for work will be distorted, possibly in costly ways. While these effects are illustrated concretely in the setting of health care, there are several general points of interpretation. I discuss some of these briefly here.

Presenteeism and Slacking Off. The terms “presenteeism” and “slacking off” have become common in everyday usage. Some definitions of presenteeism describe workers “stay[ing] beyond the time needed for effective performance on the job” (Simpson, 1998).³⁵ Slacking off has been described as tapering work, particularly in the context of shirking near the end of scheduled work.³⁶ Both of these concepts are closely related to the phenomenon described in this paper. Despite the negative connotation of these terms, I argue that informational frictions (i.e., when work can be assigned but not much else is observable or contractible) imply some slacking off would be optimal in a second-best sense. This may explain why the practice – either tapering assignments or equivalently scheduling time at work to be longer than necessary – is not only prevalent but also tolerated. In the setting of this ED, I find that the observed assignment policy is in fact approximately (second-best) optimal.

Social and Behavioral Mechanisms of Distortion. It is standard to assume that workers care about their own income and leisure more than the productive consequences of their workplace actions. Therefore, a natural interpretation of distortions near EOS is that they arise from strategic behavior, or moral hazard. However, other mechanisms could lead to the same welfare-reducing distortions and would have equivalent implications on how availability should be scheduled and worked assigned. For example, social norms may pressure workers not to stay too long after EOS (e.g., doing so would signal incompetence), so that moving EOS too early without tapering work generates the same inefficient time pressure.³⁷ Workers may take sched-

³⁵Other definitions have described presenteeism as showing up to work while ill.

³⁶See, for example, definitions in the *McGraw-Hill Dictionary of American Idioms* and the *American Heritage Dictionary of Phrasal Verbs*.

³⁷A related distortion caused by social incentives is that ED physicians care more about their peers than inpatient doctors they would be admitting their patients to. This is one (distortionary) source of client-worker specificity. In Appendix A-6 (Figures A-6.3 and A-6.2), I show that EOS distortions are more likely under certain

ules as a contractual “reference point” (Hart and Moore, 2008), and there may even be accepted routines (e.g., sign-out rounds) that reinforce this sense. Workers in practice may underestimate the time it takes to complete work, paying too much attention to when EOS is officially set.

Even rational workers would view uncertainty with respect to the time they leave work as costly, if they have plans that cannot be easily changed in the short term, such as picking up children from school or having a dinner reservation. This implies that the canonical model of a constant value of leisure is misspecified, and that short-term reductions in “leisure” are much costlier than prearranged work, as has been noted in the press with practices such as “flexible” or “just-in-time” scheduling that benefit employers but could be costly for employees (e.g., Greenhouse, 2014). Accounting for this would imply even greater welfare benefits from slacking off to ensure that workers leave close to a prespecified time.

Eliminating Transfer Costs. Without transfer costs between physicians, there would be no EOS distortion. The results from simulating counterfactual assignment policies, e.g., that assigning merely four additional patients would increase net costs for a shift by \$3,300, suggest that potential transfer costs are quite large, at least in terms of hospital resources. Transfer costs could reflect technological difficulties in passing information to subsequent physicians or could simply reflect behavioral norms that discourage transfers. A hypothetical scenario in which transfer costs are eliminated could further improve welfare by removing the need to slack off. However, it appears that welfare improvements in this hypothetical scenario would be relatively limited for two reasons: (i) very few patients are affected by increased hospital costs under the actual assignment policy, so reductions in hospital costs would be minimal; and (ii) the cost of an hour or two of physician time is (relatively) cheap, so reducing slacking off would not save much. Further, there are likely general behavioral foundations for transfer costs that could prove difficult to eliminate.³⁸

physician types and physician-peer relationships, e.g., slower physicians and physicians handing off to a senior peer, a faster peer, or a peer with whom they are unfamiliar.

³⁸Transfer costs have been discussed in the economics literature as “worker-task specificity” (e.g., Goldin, 2014) and in the medical literature manifesting as difficulties in passing information between providers transitioning care (e.g., Apker et al., 2007). More generally, Polanyi (1966) has described tacit knowledge that is difficult to communicate, which could apply to complex patient care, at least in the sense that successful communication is costly both for the sender and the receiver (Dewatripont and Tirole, 2005). Although patient care shared with nurses and residents who overlap with attending physicians changing shifts could reduce transfer costs, the extent of this is also limited by knowledge and communication issues with attending physicians responsible for the discharge decision.

Schedules, Prices, and Spending Budgets. It is reasonable to ask whether this inefficiency can be mitigated by price or budget policies. For example, in the conceptual framework, a wage in the form of overtime pay at $(1 - \lambda)\tilde{c}'_\tau$ per hour would exactly cancel out any distortionary incentive near EOS. Similarly, one might speculate whether a global budget on spending for each physician might restrain the incentive to overutilize formal resources and admit patients near EOS. Under certainty and perfect information, prices (e.g., wages, or costs imposed on physicians for utilization) and quantities (e.g., physician hours) are equivalent. However, under uncertainty, control via quantities can be superior when benefits are more concave than costs are convex, which characterizes production within most organizations at least in the short term (Weitzman, 1974).

This reasoning likely underlies the very existence of schedules. For example, in the ED, given uncertain work flow, it could be very costly to simply post wages *in advance* and hope that the appropriate number of physicians show up for the appropriate number of hours, as opposed to specifying the number of hours each physician should be there. With asymmetric information and moral hazard, price or budget mechanisms could be even worse. For example, under a global budget, physicians have greater incentive to cherry-pick healthier patients to stay within budget. If physicians already have the right incentives to care for patients outside the scheduling distortion (i.e., $v(\theta, d)$ and $c(\mathbf{z})$ are appropriately weighted), then a global budget could distort care uniformly toward underprovision. Instead, simple contracts that allow physicians to treat the neediest patients first are likely to be the least distortionary.

9 Conclusion

I examine ED physicians working in shifts and find evidence consistent with behavioral distortions due to scheduled work: On an extensive margin, physicians are less likely to accept new patients near EOS. On an intensive margin, physicians complete their work earlier as end of shift (EOS) approaches. As the input of time becomes costlier, physicians modify the mix of inputs in patient care, and as they produce less information for discharge decisions, they are more likely to admit patients.

The EOS phenomenon documented in this paper reflects a definitional issue of scheduled work:

Although scheduled availability begins and ends at set times, the true nature of work usually blurs across these constructed boundaries. Further, *ex post* worker-task specificity is often substantial in work that is information-rich. I show a tradeoff between extensive and intensive margins of distortion. In fact, observed patterns of “presenteeism” or “slacking off” appear approximately second-best optimal, as I simulate large welfare losses in assignment policies without tapering near EOS. Key to this result is that physicians are willing to spend increasingly large amounts of hospital dollars for each hour of their leisure time, a finding that sheds light on the tradeoff between intrinsic and extrinsic motivations. This is relevant for a wide set of policy levers that act via assignment. Attempting to prevent workers from sitting idly could be quite costly when used at the wrong time in scheduled work.

References

- AGHION, P. AND J. TIROLE (1997): “Formal and real authority in organizations,” *The Journal of Political Economy*, 105, 1–29.
- AKERLOF, G. A. AND R. E. KRANTON (2005): “Identity and the Economics of Organizations,” *The Journal of Economic Perspectives*, 19, 9–32.
- 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.
- APKER, J., L. A. MALLAK, AND S. C. GIBSON (2007): “Communicating in the Gray Zone: Perceptions about Emergency Physician-hospitalist Handoffs and Patient Safety,” *Academic Emergency Medicine*, 14, 884–894.
- BANDIERA, O., I. BARANKAY, AND I. RASUL (2005): “Social preferences and the response to incentives: Evidence from personnel data,” *The Quarterly Journal of Economics*, 120, 917–962.
- (2009): “Social connections and incentives in the workplace: Evidence from personnel data,” *Econometrica*, 77, 1047–1094.
- 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.
- BEERS, T. (2000): “Flexible schedules and shift work: Replacing the 9-to-5 workday?” Tech. rep., Bureau of Labor Statistics.
- BENABOU, R. AND J. TIROLE (2003): “Intrinsic and Extrinsic Motivation,” *The Review of Economic Studies*, 70, 489–520.

- BESLEY, T. AND M. GHATAK (2005): “Competition and Incentives with Motivated Agents,” *American Economic Review*, 95, 616–636.
- BLOOM, N., J. LIANG, J. ROBERTS, AND Z. J. YING (2014): “Does Working from Home Work? Evidence from a Chinese Experiment,” *The Quarterly Journal of Economics*.
- BRACHET, T., G. DAVID, AND A. M. DRECHSLER (2012): “The Effect of Shift Structure on Performance,” *American Economic Journal: Applied Economics*, 4, 219–246.
- BRISCOE, F. (2006): “Temporal Flexibility and Careers: The Role of Large-Scale Organizations for Physicians,” *Industrial and Labor Relations Review*, 60, 88.
- (2007): “From Iron Cage to Iron Shield? How Bureaucracy Enables Temporal Flexibility for Professional Service Workers,” *Organization Science*, 18, 297–314.
- CASALINO, L. P., K. J. DEVERS, T. K. LAKE, M. REED, AND J. J. STODDARD (2003): “Benefits of and barriers to large medical group practice in the united states,” *Archives of Internal Medicine*, 163, 1958–1964.
- 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.
- COVIELLO, D., A. ICHINO, AND N. PERSICO (2014): “Time Allocation and Task Juggling,” *The American Economic Review*, 104, 609–623.
- DEWATRIPONT, M., I. JEWITT, AND J. TIROLE (1999): “The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies,” *The Review of Economic Studies*, 66, 199–217.
- DEWATRIPONT, M. AND J. TIROLE (2005): “Modes of Communication,” *Journal of Political Economy*, 113, 1217–1238.
- ELIXHAUSER, A., C. STEINER, D. R. HARRIS, AND R. M. COFFEY (1998): “Comorbidity Measures for Use with Administrative Data,” *Medical Care*, 36, 8–27.
- ELLIS, R. P. AND T. G. MCGUIRE (1986): “Provider behavior under prospective reimbursement: Cost sharing and supply,” *Journal of Health Economics*, 5, 129–151.
- FORSTER, A. J., I. STIELL, G. WELLS, A. J. LEE, AND C. VAN WALRAVEN (2003): “The Effect of Hospital Occupancy on Emergency Department Length of Stay and Patient Disposition,” *Academic Emergency Medicine*, 10, 127–133.

- GARICANO, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108, 874–904.
- GOLDIN, C. (2014): "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 104, 1091–1119.
- GREEN, L. (1984): "A Multiple Dispatch Queueing Model of Police Patrol Operations," *Management Science*, 30, 653–664.
- GREEN, L. V. (2004): "Capacity Planning and Management in Hospitals," in *Operations Research and Health Care*, ed. by M. L. Brandeau, F. Sainfort, and W. P. Pierskalla, Springer US, no. 70 in International Series in Operations Research & Management Science, 15–41.
- GREENHOUSE, S. (2014): "Part-Time Schedules, Full-Time Headaches," *The New York Times*.
- HART, O. AND B. HOLMSTROM (1987): "The theory of contracts," in *Advances in Economic Theory*, ed. by T. F. Bewley, Cambridge, UK: Cambridge University Press, 71–156.
- HART, O. AND J. MOORE (2008): "Contracts as Reference Points," *The Quarterly Journal of Economics*, 123, 1–48.
- HE, B., F. DEXTER, A. MACARIO, AND S. ZENIOS (2012): "The Timing of Staffing Decisions in Hospital Operating Rooms: Incorporating Workload Heterogeneity into the Newsvendor Problem," *Manufacturing and Service Operations Management*, 14, 99–114.
- 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.
- ICHNIOWSKI, C., T. A. KOCHAN, D. LEVINE, C. OLSON, AND G. STRAUSS (1996): "What Works at Work: Overview and Assessment," *Industrial Relations: A Journal of Economy and Society*, 35, 299–333.
- JACOB, B. A., L. LEFGREN, AND D. P. SIMS (2010): "The Persistence of Teacher-Induced Learning," *Journal of Human Resources*, 45, 915–943.
- KC, D. S. AND C. TERWIESCH (2009): "Impact of workload on service time and patient safety: an econometric analysis of hospital operations," *Management Science*, 55, 1486–1498.
- KEANE, M. P. AND K. I. WOLPIN (1994): "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *The Review of Economics and Statistics*, 76, 648.

- KOLSTAD, J. T. (2013): “Information and Quality When Motivation Is Intrinsic: Evidence from Surgeon Report Cards,” *American Economic Review*, 103, 2875–2910.
- LARKIN, I. (2014): “The cost of high-powered incentives: Employee gaming in enterprise software sales,” *Journal of Labor Economics*, 32, 199–227.
- LERMAN, B. AND M. S. KOBERNICK (1987): “Return Visits to the Emergency Department,” *The Journal of Emergency Medicine*, 5, 359–362.
- LIEBMAN, J. B. AND N. MAHONEY (2013): “Do Expiring Budgets Lead to Wasteful Year-End Spending? Evidence from Federal Procurement,” Working Paper 19481, National Bureau of Economic Research.
- MAGNAC, T. AND D. THESMAR (2002): “Identifying Dynamic Discrete Decision Processes,” *Econometrica*, 70, 801–816.
- MAHER, K. (2007): “Wal-mart seeks flexibility in worker shifts,” *The Wall Street Journal*.
- MARSCHAK, J. AND R. RADNER (1972): *Economic Theory of Teams*, New Haven, CT: Yale University Press.
- MAS, A. AND E. MORETTI (2009): “Peers at Work,” *The American Economic Review*, 99, 112–145.
- MESSERLI, F. H., A. W. MESSERLI, AND T. F. LÄESCHER (2005): “Eisenhower’s Billion-Dollar Heart Attack – 50 Years Later,” *New England Journal of Medicine*, 353, 1205–1207.
- MILGROM, P. AND J. ROBERTS (1988): “An Economic Approach to Influence Activities in Organizations,” *The American Journal of Sociology*, 94, 154–179.
- OYER, P. (1998): “Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality,” *The Quarterly Journal of Economics*, 113, 149–185.
- 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.
- PERDIKAKI, O., S. KESAVAN, AND J. M. SWAMINATHAN (2012): “Effect of Traffic on Sales and Conversion Rates of Retail Stores,” *Manufacturing and Service Operations Management*, 14, 145–162.
- POLANYI, M. (1966): *The Tacit Dimension*, New York: Doubleday Press.
- PRENDERGAST, C. (2007): “The Motivation and Bias of Bureaucrats,” *The American Economic Review*, 97, 180–196.

- PRESSER, H. B. (2003): *Working in a 24/7 Economy: Challenges for American Families.*, Russell Sage Foundation.
- RADNER, R. (1993): "The Organization of Decentralized Information Processing," *Econometrica*, 61, 1109–46.
- SCHUUR, J. D. AND A. K. VENKATESH (2012): "The Growing Role of Emergency Departments in Hospital Admissions," *The New England Journal of Medicine*, 367, 391–393.
- SHAPIRO, C. AND J. E. STIGLITZ (1984): "Equilibrium Unemployment as a Worker Discipline Device," *The American Economic Review*, 74, 433–444.
- SHETTY, K. D. AND J. BHATTACHARYA (2007): "Changes in hospital mortality associated with residency work-hour regulations," *Annals of Internal Medicine*, 147, 73–80.
- SIMON, H. (1947): *Administrative Behavior*, New York: Macmillan.
- SIMPSON, R. (1998): "Presenteeism, Power and Organizational Change: Long Hours as a Career Barrier and the Impact on the Working Lives of Women Managers," *British Journal of Management*, 9, 37–50.
- TANABE, P., R. GIMBEL, P. R. YARNOLD, D. N. KYRIACOU, AND J. G. ADAMS (2004): "Reliability and Validity of Scores on the Emergency Severity Index Version 3," *Academic Emergency Medicine*, 11, 59–65.
- TIROLE, J. (1986): "Hierarchies and bureaucracies: On the role of collusion in organizations," *Journal of Law, Economics and Organization*, 2, 181–214.
- 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.
- WEITZMAN, M. L. (1974): "Prices vs. Quantities," *The Review of Economic Studies*, 41, 477–491.

Table 1: Balance Tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Assumption 1</i>				<i>Assumption 2</i>			
	Predicted LOS	Diagnostic count	Predicted LOS	Age	ESI	White	Spanish- speaking	Diagnostic count
Hour prior to EOS								
Last hour	0.003 (0.011)	0.008 (0.070)	-0.008 (0.066)	0.851 (3.297)	-0.004 (0.137)	0.157* (0.092)	0.024 (0.051)	0.140 (0.194)
Second hour	-0.002 (0.006)	-0.028 (0.037)	-0.006 (0.021)	-0.305 (1.088)	-0.001 (0.044)	0.022 (0.030)	-0.006 (0.019)	-0.046 (0.064)
Third hour	0.002 (0.005)	0.001 (0.031)	0.016 (0.015)	0.499 (0.784)	-0.032 (0.031)	0.019 (0.021)	0.004 (0.013)	0.034 (0.046)
Fourth hour	0.004 (0.004)	0.014 (0.026)	-0.006 (0.014)	-0.332 (0.699)	0.000 (0.027)	0.009 (0.019)	0.007 (0.012)	0.014 (0.041)
Fifth hour	0.004 (0.003)	0.011 (0.021)	0.007 (0.013)	0.282 (0.645)	-0.027 (0.025)	-0.022 (0.018)	0.010 (0.011)	0.025 (0.040)
Sixth hour	0.000 (0.003)	-0.010 (0.017)	0.007 (0.012)	-0.075 (0.621)	-0.021 (0.025)	0.000 (0.017)	0.008 (0.010)	0.037 (0.038)
<i>F</i> -test <i>p</i> -value	0.598	0.699	0.852	0.963	0.878	0.339	0.945	0.872
Observations	371,421	371,421	371,421	371,421	371,421	371,421	371,421	371,421
Resid. char. distrib.								
Mean	1.078	1.012	1.079	47.65	2.826	0.495	0.101	1.013
10th percentile	0.985	0.381	0.969	41.69	2.596	0.329	0.017	0.681
90th percentile	1.167	1.623	1.188	53.72	3.058	0.664	0.190	1.356

Note: This table assesses balance with respect to two identifying assumptions. Columns 1 and 2 assess Assumption 1, based on Equation (A-1.1), controlling for *ex ante* patient characteristics time categories, pod, and providers. The remaining columns assess Assumption 2, based on Equation (A-1.2), controlling for time categories, pod, and providers. Further details are given in Appendix A-1.1. Coefficient estimates and standard errors in parentheses are given for each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. The *p*-value of the *F*-test that all coefficients are jointly 0 is also given for each model. The last three rows report summary statistics of the variation in residualized patient characteristics, either across all patients (Columns 1 and 2) or across patient averages at the hourly level (remaining columns). * denotes significance at 10% level. Black race and English-speaking were also assessed for Assumption 2 and similarly yielded null results but are omitted from the table for brevity. Table A-1.1 reports results for other predicted outcomes.

Table 2: End of Shift Effect on Log Length of Stay

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	OLS	OLS	OLS	OLS	Between	Between
Hour prior to EOS								
Last hour	-0.607*** (0.028)	-0.547*** (0.025)	-0.529*** (0.025)	-0.716*** (0.039)	-0.587*** (0.050)	-0.584*** (0.050)	-0.612*** (0.078)	-0.626*** (0.078)
Second hour	-0.316*** (0.008)	-0.282*** (0.008)	-0.330*** (0.008)	-0.461*** (0.012)	-0.287*** (0.026)	-0.287*** (0.026)	-0.325*** (0.033)	-0.344*** (0.035)
Third hour	-0.139*** (0.005)	-0.129*** (0.005)	-0.161*** (0.006)	-0.260*** (0.009)	-0.123*** (0.022)	-0.122*** (0.022)	-0.139*** (0.025)	-0.157*** (0.027)
Fourth hour	-0.112*** (0.005)	-0.092*** (0.004)	-0.111*** (0.005)	-0.173*** (0.008)	-0.091*** (0.018)	-0.089*** (0.018)	-0.098*** (0.021)	-0.107*** (0.022)
Fifth hour	-0.070*** (0.004)	-0.055*** (0.004)	-0.078*** (0.005)	-0.120*** (0.007)	-0.023 (0.015)	-0.022 (0.015)	-0.037*** (0.017)	-0.047*** (0.019)
Sixth hour	-0.065*** (0.004)	-0.048*** (0.004)	-0.057*** (0.005)	-0.090*** (0.007)	-0.010 (0.012)	-0.010 (0.012)	-0.015 (0.014)	-0.025* (0.015)
Patient characteristics	None	All	All	All	All	Ex ante	All	None
Time and pod	N	N	Y	Y	Y	Y	Y	Y
Physician-resident-nurse identities	N	N	N	Y	Y	Y	Y	Y
Time relative to shift beginning	N	N	N	N	Y	Y	Y	Y
Observations	371,107	371,107	371,107	371,107	371,107	371,107	68,044	68,044
Adjusted <i>R</i> -squared	0.008	0.189	0.211	0.400	0.410	0.404	0.061	0.064
Sample outcome	1.050	1.050	1.050	1.050	1.050	1.050	1.074	1.074

Note: This table reports regressions of log length of stay on arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Columns 1 through 6 report results of Equation (2), identified by Assumption 1; Columns 7 and 8 report results of Equation (3), identified by Assumption 2. Patient characteristics include *ex ante* (to assignment to physician) characteristics of demographics, emergency severity index (ESI), time spent in triage; and *ex post* Elixhauser indicators for clinical diagnoses. Time dummies include indicators for hour of day, day of week, and month-year interactions. Standard errors are in parentheses and, for “between” models in Columns 7 and 8, are bootstrapped; * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. The adjusted *R*-squared is given for each overall model in Columns 1 through 6; Columns 7 and 8 report only the adjusted *R*-squared of the second-stage model in (3) that uses residualized outcomes.

Table 3: End of Shift Effect on Other Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	Order count	Inpatient admission	Log total cost	30-day mortality	14-day bounce- back	Workload- adjusted log LOS
Hour prior to EOS						
Last hour	1.411** (0.562)	0.057** (0.024)	0.208** (0.080)	-0.003 (0.008)	-0.028 (0.018)	-0.144*** (0.051)
Second hour	-0.093 (0.302)	0.000 (0.013)	0.027 (0.043)	-0.001 (0.004)	-0.011 (0.010)	0.015 (0.027)
Third hour	-0.003 (0.249)	0.002 (0.011)	0.009 (0.036)	-0.005 (0.004)	-0.005 (0.008)	0.090*** (0.022)
Fourth hour	0.167 (0.207)	0.004 (0.009)	0.029 (0.030)	-0.001 (0.003)	-0.002 (0.007)	0.036* (0.018)
Fifth hour	0.239 (0.171)	-0.004 (0.007)	0.034 (0.024)	-0.002 (0.003)	0.001 (0.005)	0.037** (0.015)
Sixth hour	0.192 (0.137)	-0.007 (0.006)	-0.006 (0.019)	0.001 (0.002)	0.001 (0.004)	0.001 (0.012)
Number of observations	371,421	371,421	366,219	371,421	371,421	371,148
Adjusted <i>R</i> -squared	0.531	0.459	0.472	0.295	-0.044	0.476
Sample mean outcome	13.518	0.269	6.750	0.018	0.060	-0.904

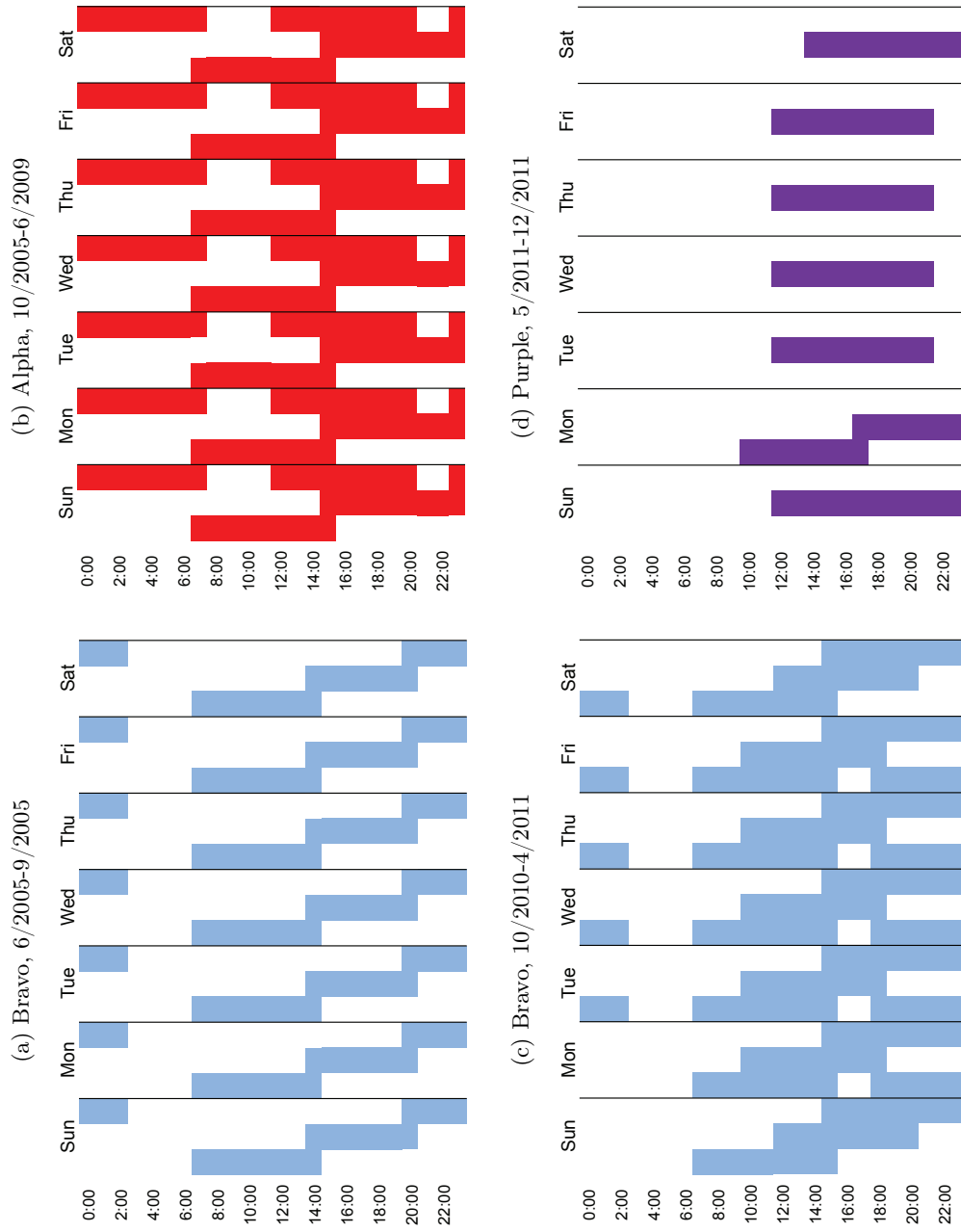
Note: This table reports coefficient estimates of Equation (2) regressing other outcome variables, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Workload-adjusted length of stay (LOS) is calculated by Equation (6), and workload-adjusted log LOS is the log of this value. Controls are the same as in Column 5 in Table 2. Standard errors are in parentheses. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level.

Table 4: Effect on Workload-adjusted Log Length of Stay by Shift Overlap

	(1)		(2)		(3)	
	$\bar{o} \leq 1$	$\bar{o} \geq 2$	$\bar{o} \leq 1$	$\bar{o} \geq 2$	$\bar{o} \leq 1$	$\bar{o} \geq 2$
Hour prior to EOS						
Last hour	-0.167** (0.068)	-0.034 (0.110)	-0.229*** (0.069)	-0.003 (0.143)		
Second hour	0.015 (0.038)	0.126* (0.067)	0.014 (0.039)	0.140 (0.085)		
Third hour	0.05 (0.031)	0.137** (0.056)	0.037 (0.033)	0.099 (0.069)		
Fourth hour	0.007 (0.025)	0.088* (0.051)	-0.002 (0.026)	0.056 (0.058)		
Fifth hour	0.013 (0.021)	0.085* (0.044)	0.009 (0.022)	0.052 (0.047)		
Sixth hour	-0.017 (0.015)	0.044 (0.038)	-0.022 (0.016)	0.029 (0.031)		
Control for time relative to shift beginning	Y		Y		Y	
Patient, provider, and other time controls	Y		Y		Y	
Sample	Full, actual		$\bar{o} \leq 1$, actual		$\bar{o} \geq 2$, actual	
Number of observations	333,233		231,576		101,657	
Adjusted R -squared	0.456		0.491		0.502	
Sample mean outcome	-0.926		-0.987		-0.789	

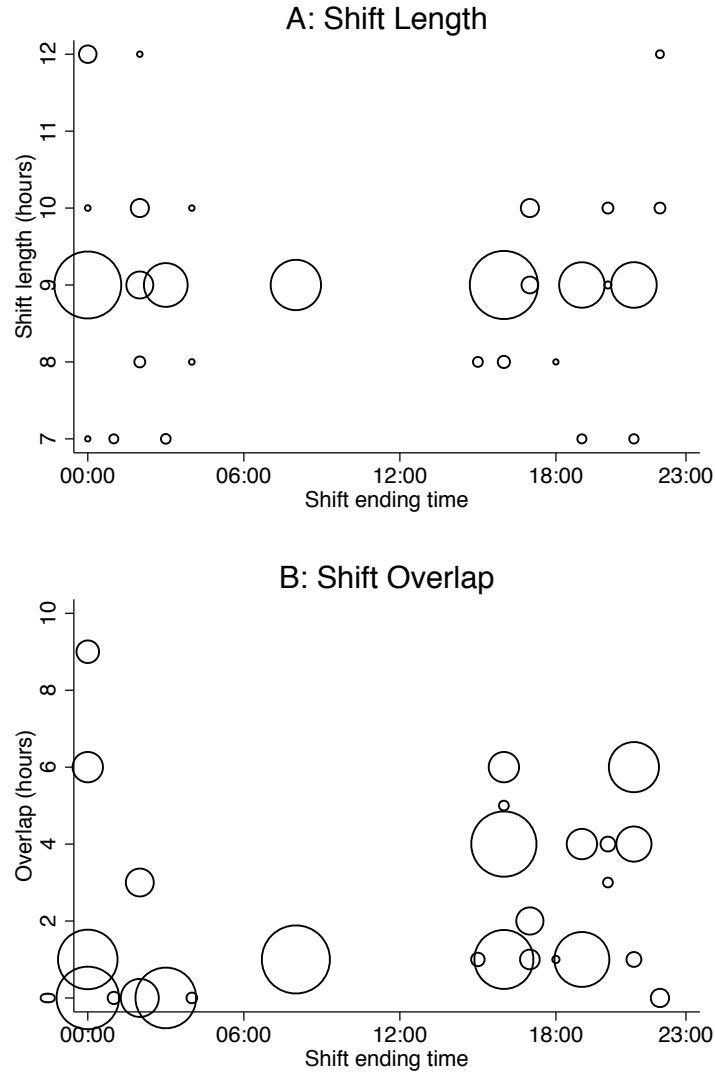
Note: This table reports coefficient estimates and standard errors in parentheses for EOS effects on workload-adjusted log length of stay, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Model 1 is estimated by Equation (5), while models 2 and 3 are estimated separately by Equation (2) on subsamples of the data according to \bar{o} . All three models are estimated with a full set of controls, as in Column 5 in Table 2. Workload-adjusted length of stay (LOS) is calculated by Equation (6), and workload-adjusted log LOS is the log of this value. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. This table is continued by Table A-6.3.

Figure 1: Example Weekly Pod Schedules



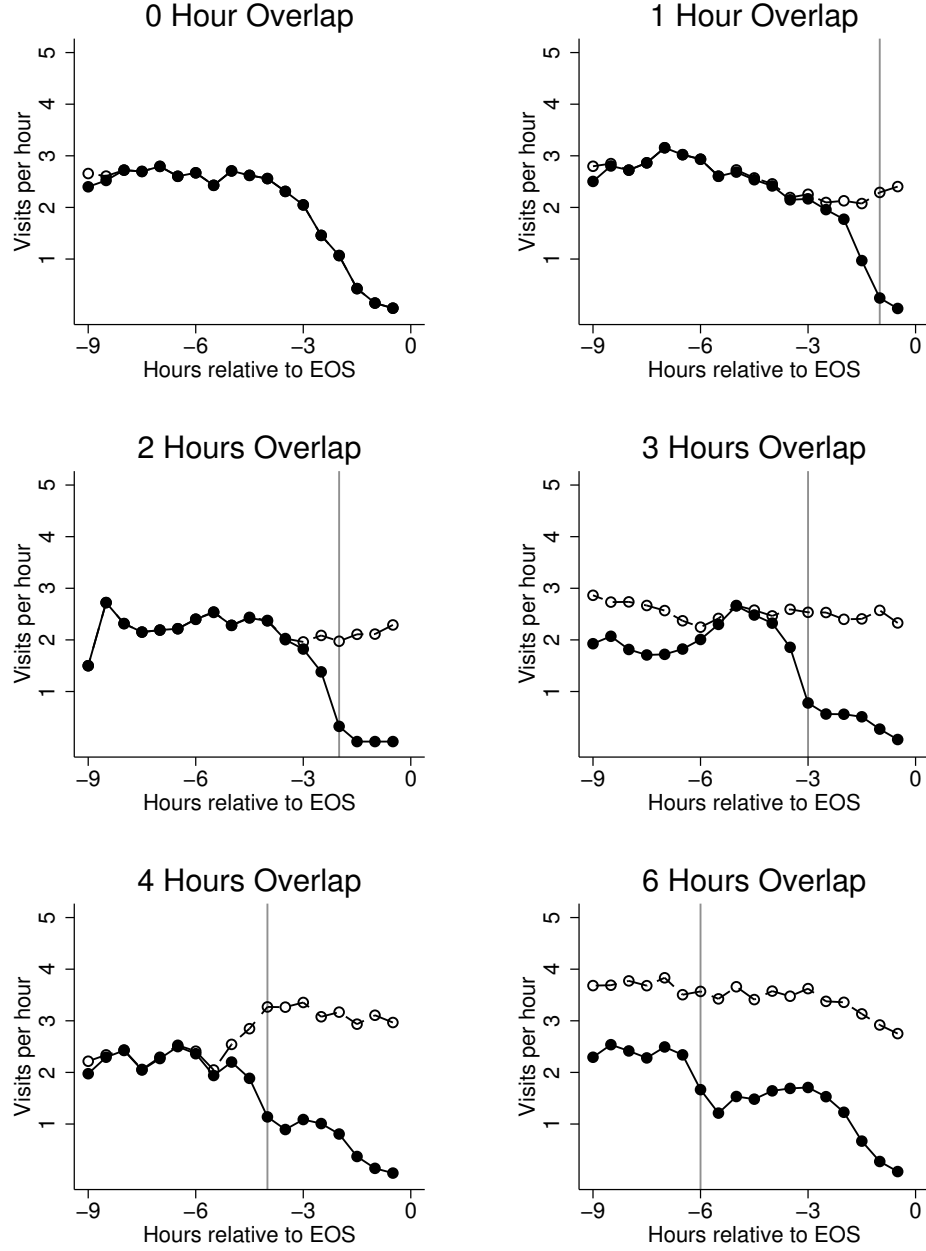
Note: Filled areas in vertical lines represent hours scheduled for a shift for a single physician. Hours when there is more than one physician present are represented by horizontally adjacent filled areas.

Figure 2: Shift Variation



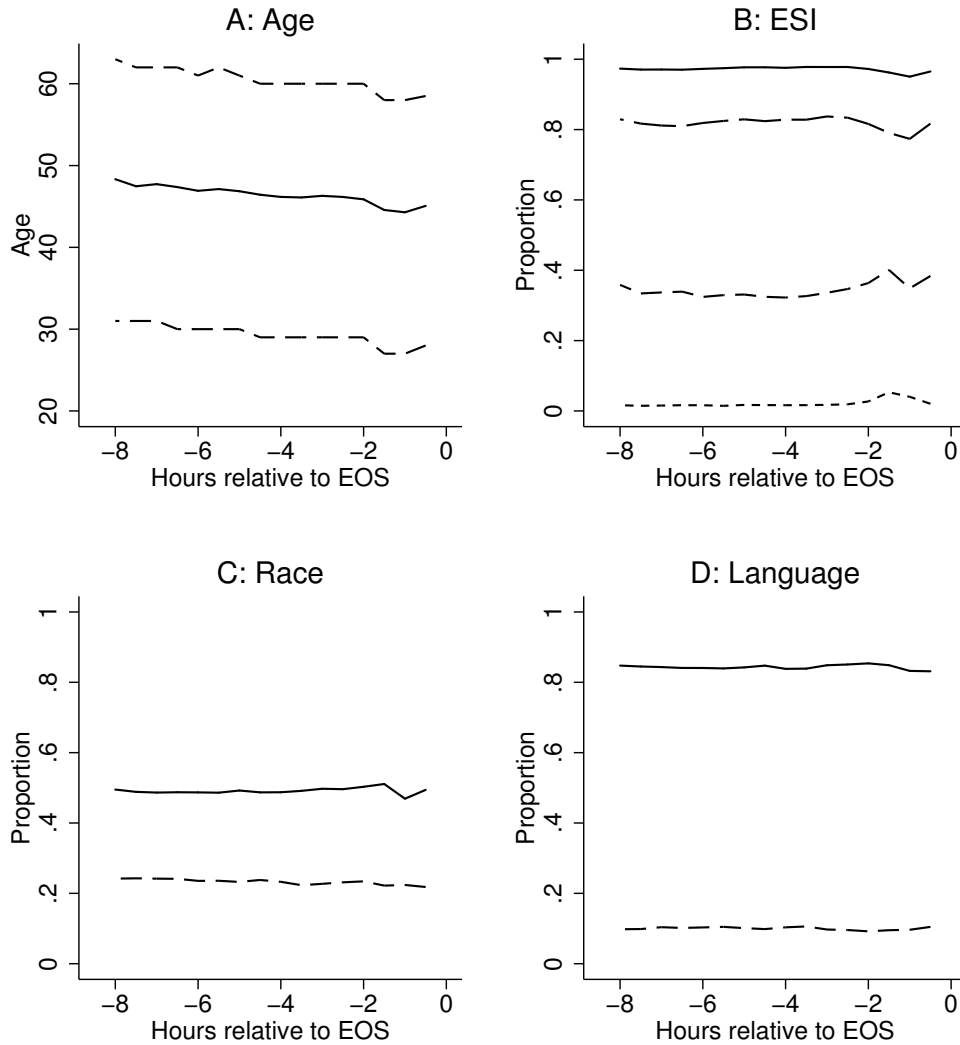
Note: This figure illustrates the variation in observations across shift types. Panel A plots shifts by shift ending time and shift length. Panel B plots shifts by shift ending time and the length of overlapping transition at the end of shift. The sizes of circles in both panels represent the number of shifts out of 23,990 total in each two-dimensional category.

Figure 3: Flow of Patient Visits over Time



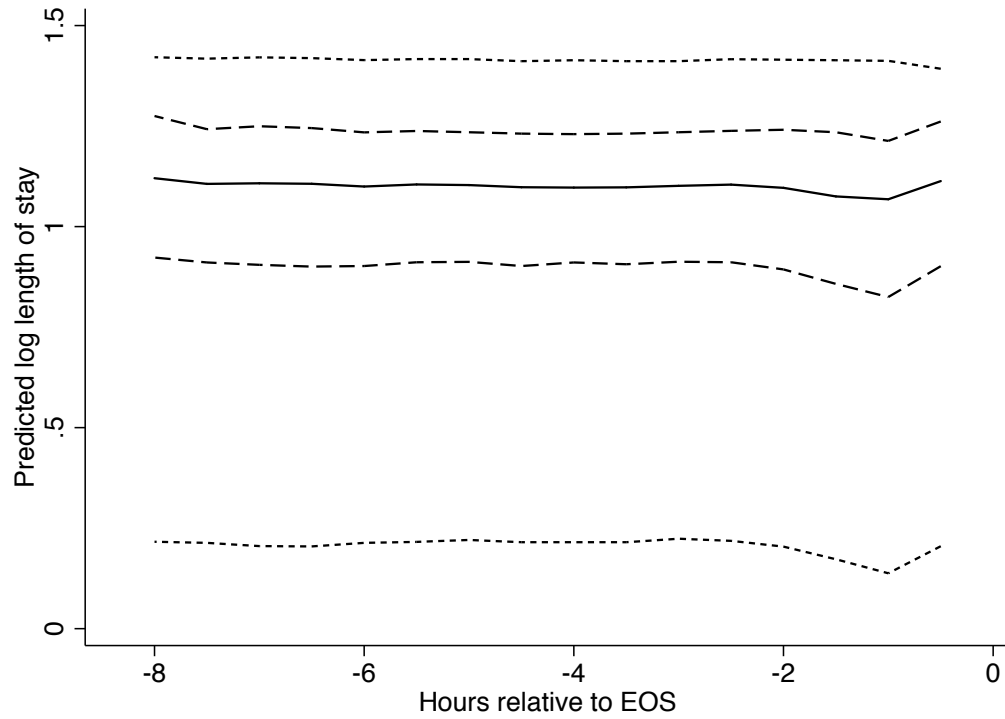
Note: This figure shows unadjusted average hourly rates of patient visits for each 30-minute interval relative to end of shift (EOS). Each panel shows results for shifts with a given EOS overlap time. Patient visits for the index physician are shown in closed circles; patient visits for the location are shown in open circles. Subsequent shift starting times are marked with a vertical line.

Figure 4: Patient Characteristics by Time Relative to EOS



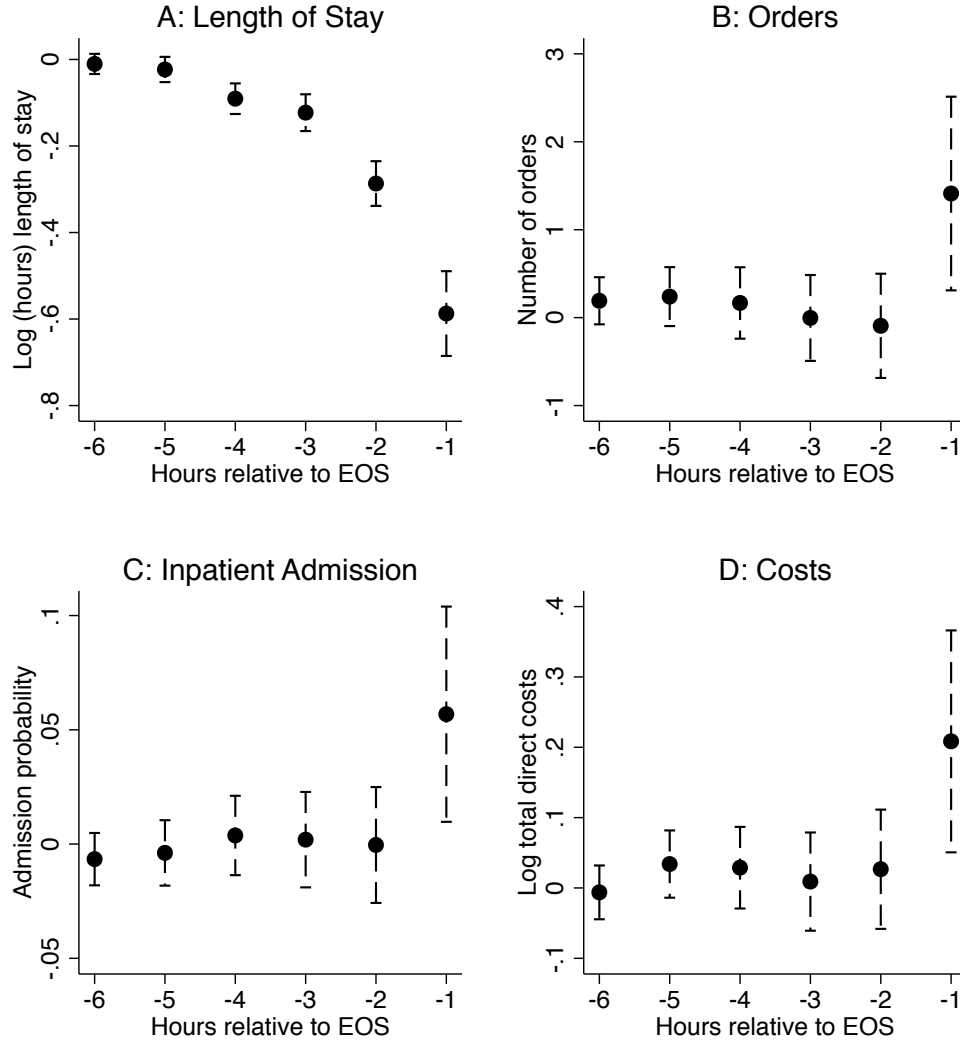
Note: This figure shows statistics of the distribution of characteristics for patients assigned to physicians in each 30-minute bin relative to EOS. Panel A shows mean age (solid line) and 25th and 75th percentiles of age (dashed lines). Panel B shows cumulative proportions of patients with an emergency severity index (ESI) from 1, ≤ 2 , ≤ 3 , and ≤ 4 , respectively. ESI is an integer from 1 (most severe) to 5 (least severe), evaluated by the triage nurse and determined by algorithm (Tanabe et al., 2004). Panel C shows proportions of patient white (solid) and black (dashed) race. Panel D shows proportion of English (solid) and Spanish (dashed) language.

Figure 5: Predicted Length of Stay by Time Relative to EOS



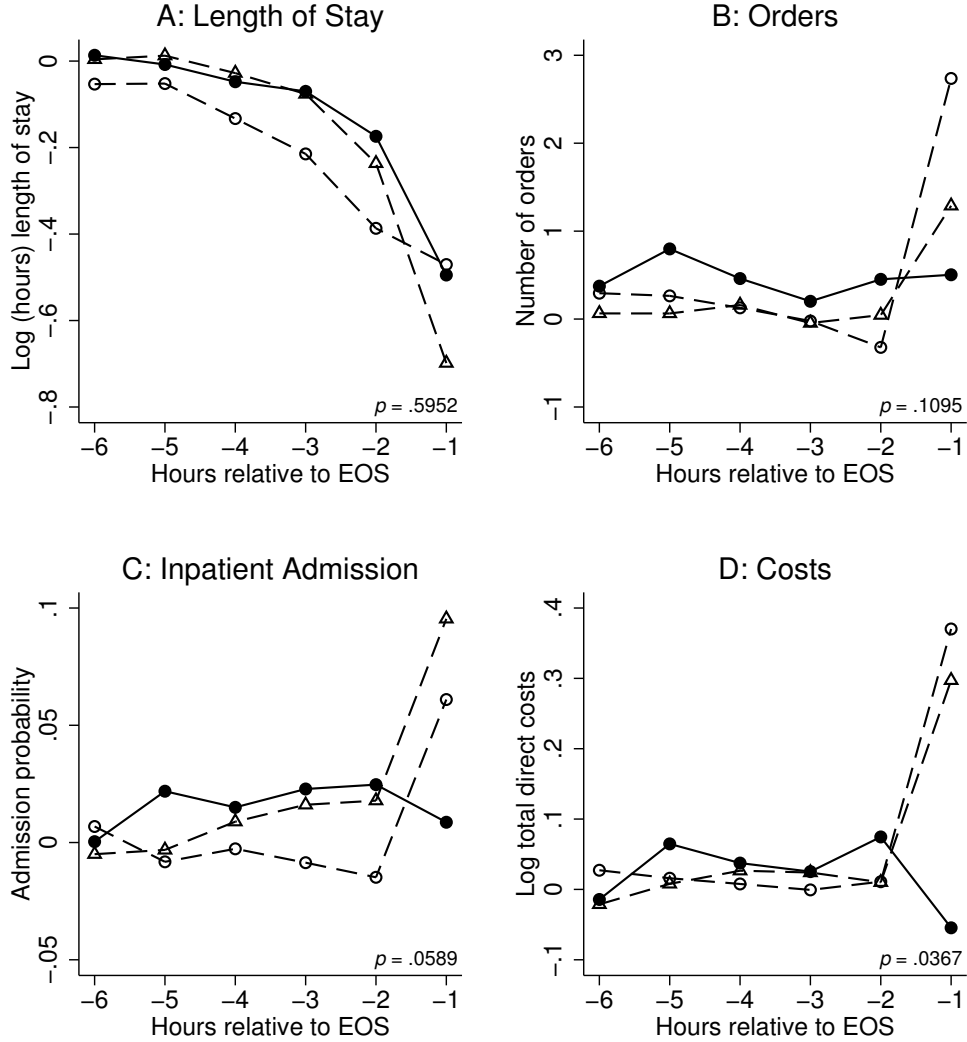
Note: This figure shows quantiles of predicted log length of stay for patients assigned to physicians in each 30-minute bin relative to EOS. Log length of stay is predicted by cubic splines of age, sex, indicators for ESI, indicators for language, and indicators for race. The solid line shows medians; dashed lines show 25th and 75th percentiles; and short-dashed lines show 5th and 95th percentiles.

Figure 6: End of Shift Effects



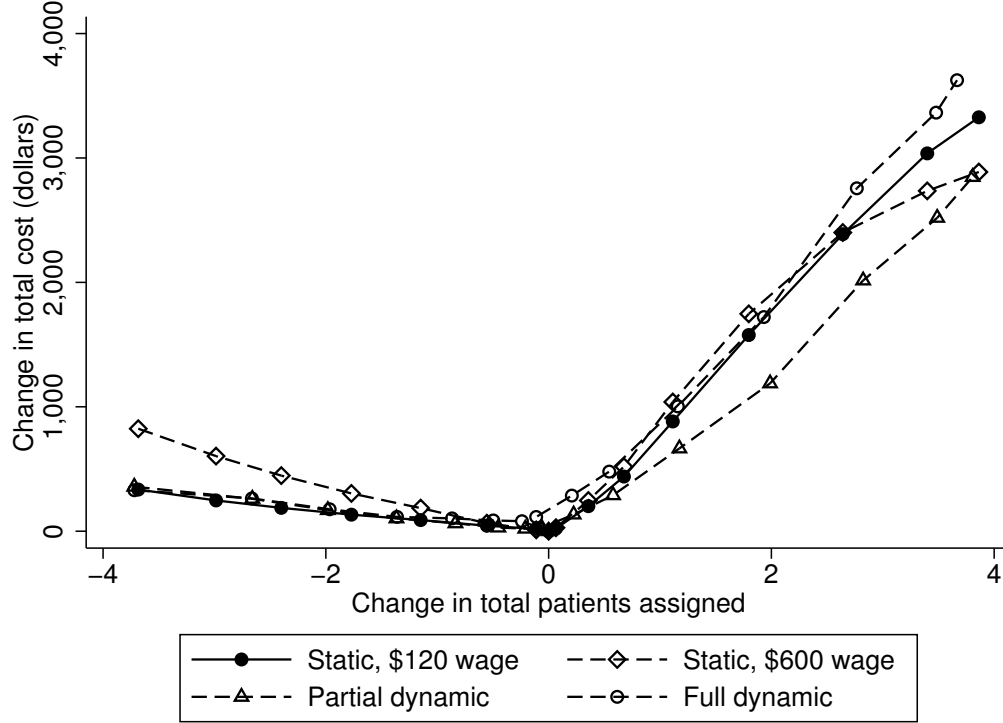
Note: This figure plots average effects for each hour prior to end of shift (EOS) on length of stay (Panel A), orders (Panel B), inpatient admissions (Panel C), and costs (Panel D). Each outcome is estimated separately using Equation (5), and results are the same as in Column 5 in Table 2 for length of stay and Table 3 for the other outcomes. The reference category is any time greater than six hours prior to EOS. Bracketed dashed lines represent 95% confidence intervals for each estimate.

Figure 7: End of Shift Effects by Shift Overlap



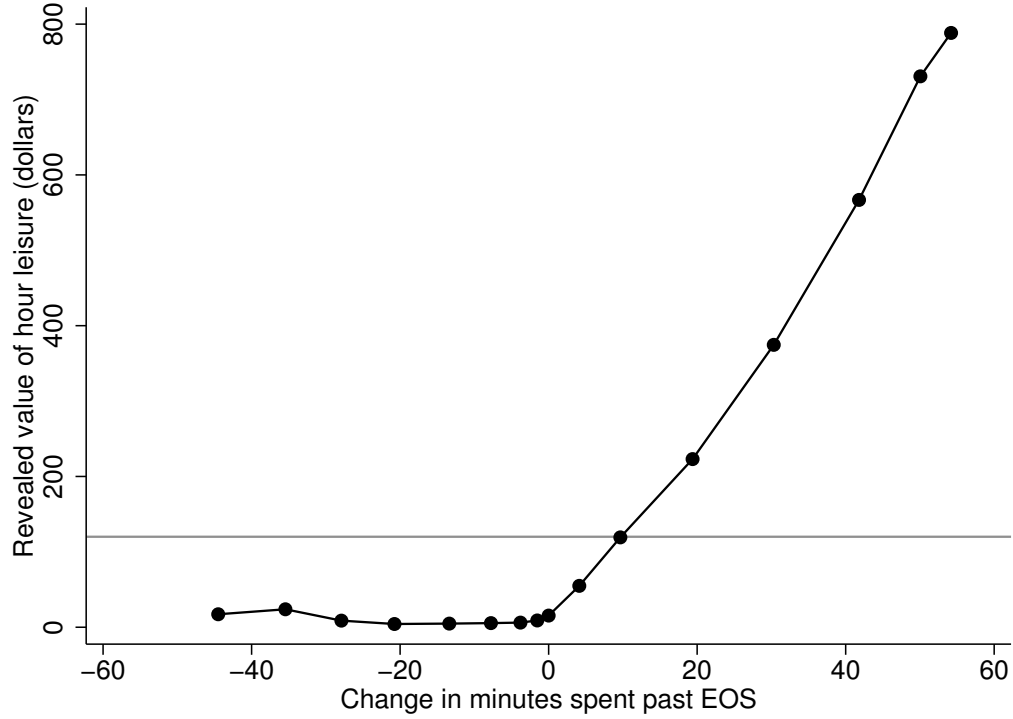
Note: This figure shows end of shift (EOS) effects by EOS overlap times on length of stay (Panel A), orders (Panel B), inpatient admissions (Panel C), and costs (Panel D). Each outcome is estimated separately using Equation (5). Estimates for terminal shifts ($\bar{o} = 0$) are shown in open triangles; estimates for minimally transitioned shifts ($\bar{o} = 1$) are shown in open circles; and estimates for substantially transitioned shifts ($\bar{o} \geq 2$) are shown in closed circles. The noted p -values in each panel are for a test that the coefficient of last hour of shift interacted with $\bar{o} \leq 1$ is different than the coefficient of last hour of shift interacted with $\bar{o} \geq 2$ (these coefficients are slightly different than those shown in the figure in that there are only two categories of shift overlap instead of three).

Figure 8: Simulated Change in Total Cost per Shift under Counterfactual Assignment



Note: This figure plots simulated changes in total cost for each counterfactual assignment regime, as described in Section 7 and Appendix A-4. Assignment regimes may either increase or decrease assignment, as illustrated in Figure A-4.1 The x -axis is the change in total patients assigned per shift as a result of an assignment regime (“0” represents the actual assignment regime). Daily costs include physician-time, patient-time, and hospital-resource costs. Four models of physician discharge policies are considered to simulate outcomes: (i) a “static” model \mathcal{D}_0 based on observed physician discharges without any restrictions, in which physician time costs are \$120 an hour (the baseline model); (ii) the same model but with a \$600 wage; (iii) a “partial dynamic” model $\tilde{\mathcal{D}}_0$ with dynamic programming restrictions that only certain state variables can enter the utility flow, where the value function remains the unchanged under counterfactual assignment policies; and (iv) a “full dynamic” model $\tilde{\mathcal{D}}_\Delta$ using the same flow utility parameters as (iii) but recomputing the value function for each counterfactual assignment policy. The wage in the dynamic programming models is \$120 per hour; further details of the dynamic programming are in Appendix A-5 and discussion of results in Appendix A-5.6. Each marker is an average over 20 simulations in an assignment policy \mathcal{A}_Δ , and each simulation contains 280,000 to 460,000 observations, depending on whether \mathcal{A}_Δ decreases or increases the number of patients assigned relative to \mathcal{A}_0 .

Figure 9: Value of Leisure in Dollars of Patient Care



Note: This figure plots the imputed value of leisure, revealed by increases in resource-utilization costs that shorten the time for completion of work under simulated counterfactual assignment policies. For each counterfactual assignment policy, data are simulated using the baseline discharge policy \mathcal{D}_0 and then again for a counterfactual discharge policy $\underline{\mathcal{D}}$ that is insensitive to time relative to EOS. The difference between these two discharge policies yields a tradeoff of resource-utilization costs for shortened work completion time. The ratio between these two represents the value of leisure, denominated in hospital-resource dollars, and is plotted on the y -axis. The x -axis is the change work completion time using the baseline discharge policy. The horizontal line drawn at \$120 is the effective hourly wage for an hour of a physician's (scheduled) time. Details are described in Appendix A-4.5.

Appendix

A-1 Identification under Patient Selection

A-1.1 Assessing Identifying Assumptions

In Section 5.1 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. Remaining characteristics, including those that are correlated with *ex post* clinical characteristics generally observed only after patient assignment, do not influence patient assignment. I assess this assumption by considering an *ex post* characteristic X^{post} and estimating the following regression:

$$X_{it}^{post} = \alpha_{\bar{m}(i,t)} + \gamma_{\bar{m}(i,t)} + \beta \mathbf{X}_i^{ante} + \eta \mathbf{T}_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_t. \quad (\text{A-1.1})$$

Under Assumption 1, excludable characteristics (and any linear combination of them, such as an outcome prediction based on them) are unrelated to time relative to EOS, conditional on *ex ante* characteristics, time categories, pod identities, and provider identities. Therefore, under this assumption, the set of coefficients $\{\alpha_{\bar{m}}\}$ in Equation (A-1.1) should not be jointly significant. Results from regressions of this type are shown in Table 1, Columns 1 and 2, and in Table A-1.1, Panel A.

Assumption 2 is that there is no patient selection to hours of different propensities for assignment to different times relative to EOS, conditional on time categories, pod identities, and provider identities. Assessing this assumption involves regressing *any* patient characteristic or linear combination of characteristics:

$$X_{it} = \sum_{\bar{m}=6}^1 \alpha_{\bar{m}} P_{\bar{m}}(t) + \eta \mathbf{T}_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_t. \quad (\text{A-1.2})$$

$P_{\bar{m}}(t)$ represents the fraction of patients arriving at time t who are assigned to \bar{m} hours prior to EOS, where the omitted category is 7 or greater hours. Results from these regressions are shown 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 time t . Conditioning on pod and provider identities is of course part of my baseline specification for estimating EOS effects in Equation (2) and thus does not pose an issue for identification.

A-1.2 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-1.2.1 Conceptual Framework

Consider a condensed form of the outcomes regression Equation (2):

$$\begin{aligned} Y &= \sum_{\bar{m}} \alpha_{\bar{m}} A_{\bar{m}} + \mathbf{\Omega}' \mathbf{\Gamma} \\ &= \sum_{\bar{m}} \alpha_{\bar{m}} A_{\bar{m}} + \mathbf{W}' \mathbf{\Gamma}_W + \xi, \end{aligned} \quad (\text{A-1.3})$$

where I omit subscripts for simplicity. I define $A_{\bar{m}} \equiv \mathbf{1}(\lceil \bar{t}(j, t) - t \rceil = \bar{m})$ for whether the time t that patient i was assigned to physician j was in the \bar{m}^{th} hour from j 's EOS. $\alpha_{\bar{m}}$ is the causal effect of a patient being assigned in the \bar{m}^{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 (2), 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-1.3) as

$$Y = \sum_{\bar{m}} \alpha_{\bar{m}} A_{\bar{m}} + \mathbf{W}' \gamma_W + \varepsilon, \quad (\text{A-1.4})$$

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 $\text{Cov}(\varepsilon, A_{\bar{m}}) = 0$, or $E[\varepsilon | A_{\bar{m}} = 1] - E[\varepsilon | A_{\bar{m}} = 0] = 0$.

A-1.2.2 Measure of Selection on Unobservables

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

$$\frac{E[\varepsilon | A_{\bar{m}} = 1] - E[\varepsilon | A_{\bar{m}} = 0]}{\text{Var}(\varepsilon)} = \frac{E[\mathbf{W}' \gamma_W | A_{\bar{m}} = 1] - E[\mathbf{W}' \gamma_W | A_{\bar{m}} = 0]}{\text{Var}(\mathbf{W}' \gamma_W)}, \quad (\text{A-1.5})$$

which states that the relationship between the index of unobservables in Equation (A-1.4) and the indicator for selection $A_{\bar{m}}$ is equal in magnitude to the relationship between observable predictors of Y and $A_{\bar{m}}$, 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-1.2.3 Estimation of Potential Bias

In order to estimate the potential bias at the upper bound implied by Equation (A-1.5), consider the following linear selection equation:

$$A_{\bar{m}} = \mathbf{W}'\beta_{\bar{m}} + A_{\bar{m}}^*, \quad (\text{A-1.6})$$

where $A_{\bar{m}}^*$ is a residual that is orthogonal to \mathbf{W} . Then Equation (A-1.4) can be stated as

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

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

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

From Equation (A-1.5), the bias can be stated in terms of $E[\mathbf{W}'\gamma_W | A_{\bar{m}} = 1] - E[\mathbf{W}'\gamma_W | A_{\bar{m}} = 0]$:

$$\text{Bias} = \frac{\text{Var}(A_{\bar{m}}) \text{Var}(\varepsilon)}{\text{Var}(A_{\bar{m}}^*) \text{Var}(\mathbf{W}'\gamma_W)} (E[\mathbf{W}'\gamma_W | A_{\bar{m}} = 1] - E[\mathbf{W}'\gamma_W | A_{\bar{m}} = 0]) \quad (\text{A-1.7})$$

Under the null hypothesis that $\alpha_{\bar{m}} = 0$, γ_W can be consistently estimated by Equation (A-1.3).

I can then arrive at a consistent estimate of bias in Equation (A-1.7) with the following procedure, with results shown in Table A-1.2: For each $\bar{m} \in \{-6, \dots, -1\}$, I define $A_{\bar{m}}$ over all observations and empirically calculate $\widehat{\text{Var}}(A_{\bar{m}})$. I also calculate $\widehat{\text{Var}}(A_{\bar{m}}^*)$ after estimating Equation (A-1.6) for each \bar{m} . Similarly, I estimate $\widehat{\text{Var}}(\varepsilon) = 0.160$ and $\widehat{\text{Var}}(\mathbf{W}'\gamma_W) = 0.580$ from Equation (A-1.4). Equation (A-1.4) also allows me to form an estimate of selection on observables, $\hat{E}[\mathbf{W}'\gamma_W | A_{\bar{m}} = 1] - \hat{E}[\mathbf{W}'\gamma_W | A_{\bar{m}} = 0]$, for each \bar{m} . Using the condition in Equation (A-1.5) 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-1.7). As shown in Table A-1.2, 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 (2), 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-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. As explained above, instead of using variation in the identity of the assigned physician (or more precisely, the time within that physician’s shift), 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 (2) on all observations except those corresponding to s (Jacob et al., 2010). I denote these estimates as $\{\hat{\alpha}_{\bar{m}}^{-s}\}$. This method thus excludes any idiosyncratic selection within a shift that would otherwise introduce bias into $\{\hat{\alpha}_{\bar{m}}\}$. 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

$$Q_t \equiv \frac{\sum_i \mathbf{1}(t_i^a = t) \sum_{m=1}^6 \hat{\alpha}_{\bar{m}}^{-s} \mathbf{1}(\lceil \bar{t}(J(i, t), t) - t \rceil = \bar{m}) \mathbf{1}(S(J(i, t), t) = s)}{\sum_i \mathbf{1}(t_i^a = t)}, \quad (\text{A-1.8})$$

where $J(i, t)$ is a physician assignment function for patient i at time t , and $S(j, t)$ 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} :

$$Y_t \equiv \frac{\sum_i \mathbf{1}(t_i^a = t) \tilde{Y}_{it}}{\sum_i \mathbf{1}(t_i^a = t)} \quad (\text{A-1.9})$$

where

$$\tilde{Y}_{it} \equiv Y_{it} - \left[\hat{\gamma}_{\underline{m}(i, t)} + \hat{\beta} \mathbf{X}_i + \hat{\eta} \mathbf{T}_t + \hat{\zeta}_{p(i)} + \hat{\nu}_{j(i)} \right]. \quad (\text{A-1.10})$$

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 (2):

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

where I use physician fixed effect ν_k instead of physician-team fixed effects ν_{jk} 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-1.8) with the regression

$$Y_t = a + bQ_t + \chi_t. \quad (\text{A-1.11})$$

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

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

under the assumption that

$$\text{Cov}(Q_t, \chi_t) = 0. \quad (\text{A-1.12})$$

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

Column 1 of Table A-1.3 reports estimates of b from Equation (A-1.11). 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-1.12), I cannot reject the null of no unobservable selection across physicians within time. Panel A of Figure A-1.1 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-1.12) (and Assumption (2)). Similar to the analysis in Appendix A-1.1, 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, $[t - \underline{t}(j, t)]$. Using residualized characteristics in each set, I construct respective predictions \hat{Y}^{ante} and \hat{Y}^{full} . Similar to Equation (A-1.9), I average these predictions over all patients

arriving at a given hour:

$$\hat{Y}_t^{set} \equiv \frac{\sum_i \mathbf{1}(t_i^a = t) \hat{Y}_{it}^{ante}}{\sum_i \mathbf{1}(t_i^a = t)}, \quad (\text{A-1.13})$$

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

The regression

$$\hat{Y}_t^{set} = a + b^{set} Q_t + \chi_t \quad (\text{A-1.14})$$

quantifies the degree of selection across hours, as predicted by characteristics \mathbf{X}^{set} : Under (A-1.12), $b^{set} = \text{Cov}\left(E[\hat{\alpha}_{\bar{m}}^{-s}|t], \hat{Y}_t^{set}\right) / \text{Var}\left(E[\hat{\alpha}_{ms}|t]\right)$ for $\bar{m} = \lceil \bar{t}(j, t) - t \rceil$ and set $s = S(j, t)$. Although the assumption in Equation (A-1.12) 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-1.3 report of estimates b^{ante} and b^{full} , respectively, from Equation (A-1.14). 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-1.1 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 Quantifying Selection in Terms of Other 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 (2):

$$\hat{Y}_{it}^{set} = \alpha_{\bar{m}(i,t)} + \gamma_{\underline{m}(i,t)} + \eta \mathbf{T}_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_{it}, \quad (\text{A-1.15})$$

leaving out variables in \mathbf{X}_{it} as regressors. I interpret the coefficients $\alpha_{\bar{m}}$ as the amount patient selection in the \bar{m}^{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-1.2 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 (2) 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 or less resource-intensive 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-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 (2), 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 (2) 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 5, 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 (2). 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 (2); 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 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, the first step is to construct counterfactual assignment policies to simulate assignments. The second step is to estimate physician discharge policies. Based on either actual discharge policies or counterfactual discharge policies obtained by a dynamic programming model, physicians' responses to assignment are then simulated. The third step is to take the resulting distribution of simulated length of stay as physicians near EOS and impute welfare-relevant costs. This appendix describes each of these steps, with the exception of the formal dynamic programming approach, which I leave to the following Appendix A-5.

A-4.1 Assignment Policies

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 . The number of patients assigned in each observation is

$$a(s, t) \equiv \sum_i \mathbf{1}(t^a(i) = t) \mathbf{1}(S(J(i), t^a(i)) = s),$$

where $t^a(i)$ is the arrival time of visit i , $S(j, t)$ is a shift assignment function for physician j and time t , and $J(i)$ is a physician assignment function.

Of $\sum_{s,t} a_{st} = 370,843$ 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. As I describe in Section A-4.3, I take arrivals and discharges of patients $i \in I^{outside}$ as fixed in every simulation.

For a_{st} , I estimate an ordered logit model, in which $a_{st} \in \{0, 1, 2, 3\}$. This represents the assignment function $\mathcal{A}_0(S)$, which depends on the state S . I consider the shift type $\langle \ell, \underline{o}, \bar{o} \rangle_s$, the time of EOS $\bar{t}(s)$, 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 (4), which I slightly rephrase here as

$$w(j, t) \equiv \sum_i \mathbf{1}(t \geq t^a(i)) \mathbf{1}(t \leq t^d(i)) \mathbf{1}(j = J(i)), \quad (\text{A-4.1})$$

where $t^d(i)$ is the corresponding discharge order time for visit i .

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

$$\bar{m}(\Delta) = \begin{cases} \max(\Delta, \bar{m}) & \Delta > 0 \\ \bar{m}, & \Delta = 0 \\ \bar{m}(1 - \min(\max(|\Delta| - \bar{m}, 0), 1)) & \Delta < 0 \end{cases} \quad (\text{A-4.2})$$

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

A-4.2 Discharge Policies

I estimate a multinomial logit hazard model for the probability that visit $i \in I(j, t)$ is discharged at $t^d(i) = t$, or equivalently the indicator $d(i, t) = 1$:

$$\Pr(d(i, t) | S) \equiv \mathcal{D}(S).$$

I consider time relative to EOS, patient i 's expected length of stay (based on characteristics, time categories, pod, and physician identities), the log difference between i 's length of stay and expected length of stay, the time (relative to EOS) of i 's assignment, and the number of patients on census $w(j, t)$. This function based on observed data constitutes the actual discharge policy, $\mathcal{D}_0(S)$. As in the assignment policy, the discharge policy is calculated using observations not in $I^{outside}$. In order to calculate counterfactual discharge policies in which physicians *correctly anticipate* counterfactual assignment policies, I implement a dynamic programming approach, described further in Appendix A-5.

A-4.3 Simulating Observations

For a given assignment policy $\mathcal{A}(S)$ and discharge policy $\mathcal{D}(S)$, I simulate patient arrivals and discharges to create a set of observations under a potentially counterfactual assignment policy. Specifically, I follow this procedure for each simulation r of policy Δ :

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}_Δ . Denote each of these new assignments with an unused $i \notin I^{outside}$, note that $t_{\Delta,r}^a(i) = t$, and also simulate predicted log length of stay for each i .
 - (b) Assign patients $i \in I^{outside}$ where $t_{outside}^a(i) = t$ to the relevant shifts s .
3. Calculate workload $w_{\Delta,r}(j, t)$ by Equation (A-4.1).
4. If $t \geq \bar{t}(s)$ and $w_{\Delta,r}(j, t) > 0$, determine discharges at t for each s .
 - (a) Simulate $d_{\Delta,r}(i, t) \equiv \mathbf{1}(t_{\Delta,r}^d(i) = t)$ 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 $t_{outside}^d(i) = t$ 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 $t_{\Delta,r}^a(i)$ and $t_{\Delta,r}^d(i)$ 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-4.4 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 (7):

$$\text{Costs}_{\Delta,r} = \text{PhysicianTime}_{\Delta,r} + \text{PatientTime}_{\Delta,r} + \text{HospitalResources}_{\Delta,r}. \quad (\text{A-4.3})$$

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:

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

“Slacking off” in the assignment policy occurs earlier relative to EOS 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-7.2). 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 time between when all but one or two patients have been 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 base-case wage of \$120 per hour but also consider the extreme case of \$600 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 $\text{TimeValue} = \$20/\text{hour}$, or roughly the average hourly wage in the US, and $\tau_{\Delta,r}(i)$ is the simulated length of stay implied by $t_{\Delta,r}^a(i)$ and $t_{\Delta,r}^d(i)$ in discrete time.

The third cost in Equation (A-4.3), 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 $\hat{\alpha}_{\frac{\Delta}{m}}^{\Delta,r}$ in this

regression:

$$\log(\tau(i)/w(i))_{\Delta,r} = \alpha_{\bar{m}(i,t)}^{\Delta,r} + \mathbf{g}(\bar{m}(i,t))' \gamma_g^{\Delta,r} + \varepsilon_i^{\Delta,r}, \quad (\text{A-4.4})$$

where $(\tau(i)/w(i))_{\Delta,r}$ 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 $\hat{\alpha}_1^0 \equiv \frac{1}{20} \sum_{r=1}^{20} \hat{\alpha}_1^{0,r} = -0.240$ and $\hat{\alpha}_2^0 \equiv -0.059$, 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-6.3 and Figure A-7.3). 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

$$\begin{aligned} \text{HospitalResources}_{\Delta,r} = & \sum_s \sum_{\bar{m}} \sum_{t=\bar{t}(s)}^{\bar{t}(s)} \mathbf{1}(\lceil \bar{t}(s) - t \rceil = \bar{m}) a_{\Delta,r}(s, t) \times \\ & \exp\left(\text{BaseLogCosts} - 1.15 \cdot \min\left(0, \hat{\alpha}_{\bar{m}}^{\Delta,r} - \hat{\alpha}_2^0\right)\right), \end{aligned} \quad (\text{A-4.5})$$

where $\text{BaseLogCosts} = [\log \$ +] 6.750$. 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. 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-4.5 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. Similar to modifying time in the counterfactual assignment policy, Equation (A-4.2), I examine what discharges would look like if not influenced by EOS behavior by modifying t in the discharge function. That is, I consider $\underline{\mathcal{D}}(\bar{m}, S^-) \equiv \mathcal{D}_0(\max(4, \bar{m}), S^-)$ as a modified discharge function, where discharge probabilities do not change even as the physician nears 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:

$$\text{LeisureValue}_{\Delta,r} = -\frac{\text{HospitalResources}_{\Delta,r}|\mathcal{D}_0 - \text{HospitalResources}_{\Delta,r}|\underline{\mathcal{D}}}{\text{WorkCompletionTime}_{\Delta,r}|\mathcal{D}_0 - \text{WorkCompletionTime}_{\Delta,r}|\underline{\mathcal{D}}}.$$

A-5 Dynamic Programming Approach

In this appendix, I detail a formal dynamic programming approach to calculate counterfactual discharge policies $\tilde{\mathcal{D}}_{\Delta}$, where the tilde indicates additional restrictions to identify forward-looking behavior. The intuition for a dynamic programming approach is straightforward: If forward-looking physicians *correctly anticipate* the counterfactual assignment policies \mathcal{A}_{Δ} , their discharge policies should also change in the counterfactual.

A-5.1 Approach to Challenges

There are three challenges to be addressed in this setting. First, the decision space is more complicated than a standard dynamic programming problem, because any *combination* of patients could be discharged at any given time. For example, with only 10 patients on census, a physician would have $2^{10} = 1,024$ possible choices per time period. I address this by specifying time in five-minute intervals, which allows me to restrict discharges to at most one patient per period.

Second, the state space is complex and includes characteristics of all patients currently under the physicians' care as well as time to EOS. As described below, I collapse many patient characteristics into a single index and sum across patients. Using an approach introduced by Barwick and Pathak (2015), I consider a state space with three continuous dimensions, in addition to more complicated state spaces, and show a reasonable fit with the data. While the majority of states in this space are either never or rarely encountered in practice, simply expanding the number of potential states does not add any computational burden under this approach, as opposed to standard approaches such as backward recursion or high-dimensional interpolation (e.g., Keane and Wolpin, 1994).

Third, the dynamic programming problem in this setting is finite horizon, but the last time period (i.e., time at which the physician on shift goes home) is unobserved. Rather than explicitly modeling the time at which the physician goes home, which gives rise to both estimation and identification problems, I rely on the fact that patients are never assigned patients past EOS. In this way, all counterfactual assignment policies by construction identical past EOS, and therefore all discharge policies should be identical past EOS. I thus only explicitly consider the dynamic programming problem until EOS, after which I assume that physicians will always revert to the actual discharge policy \mathcal{D}_0 , estimated without any dynamic model. Further, I impose a constraint that the value function of any counterfactual dynamic discharge policy $\tilde{\mathcal{D}}_{\Delta}$ should take the same

value at EOS as that of the non-counterfactual dynamic discharge policy, \widetilde{D}_0 .

A-5.2 Defining Value Functions and Choice Probabilities

Consider a doctor-time observation (j, t) , in state $S_{j,t}$. The doctor may discharge at most one patient $i \in I(j, t)$, where with slight abuse of notation denote $I(j, t)$ as a set that includes patients on census as well as $i = \emptyset$ 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' . The setting has a finite horizon and is non-stationary.

Suppressing notation for j and t for now, the integrated value function is

$$V(S) = \sum_{i \in I} \Pr(d(i) | S) \left[u(i, S) + \delta \int_{S'} V(S') dF(S' | i, S) \right]. \quad (\text{A-5.1})$$

$d(i)$ is an indicator for mutually exclusive discharge decisions (including $i = \emptyset$), such that $\sum_{i \in I} \Pr(d(i) | S) = 1$. $u(i, S)$ includes both a fixed component of utility and an expectation of a random error term,

$$u(i, S) = \bar{u}(i, S) + E[\varepsilon_i | i, S].$$

S' in the next period is continuous, even conditional on m and S , because it depends on the arrival of patients of continuous severity. ε_i is distributed as Type I extreme value. I have also suppressed reference to Δ , but this is implicit in $F(S' | i, S)$, $\Pr(d(i) | S)$, and therefore the value function $V(S)$.

I can also construct choice-specific value functions (not including the error term):

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

The physician chooses i^* (i.e., $d(i^*) = 1$) such that $i^* = \arg \max_{i \in I} (\bar{v}(i, S) + \varepsilon_i)$, which imply the conditional choice probabilities $\Pr(d(i) | S)$:

$$\Pr(d(i) | S) = \frac{\exp(\bar{v}(i | S))}{\sum_{i' \in I} \exp(\bar{v}(i' | S))}. \quad (\text{A-5.3})$$

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 :

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

A-5.3 Algorithm

The algorithm has an outer component that maximizes the log likelihood of observing (i^*, j, t) in the data and an inner component based on the relationship in Equation (A-5.4). Using function form approximations of $V(S)$ and $\bar{u}(i, S)$ that are based on splines.

1. Set a functional form approximation for $\bar{u}(i, S) \approx \mathbf{b}(i, S) \beta$, where $\mathbf{b}(i, S)$ is a $1 \times K$ vector of splines.
2. Set a functional form approximation for $W(S) \approx \mathbf{h}(S) \theta$, where $\mathbf{h}(S)$ is a $1 \times L$ vector of splines.
3. A given set of parameters $\hat{\beta}$ implies $\hat{\theta}$ from Equation (A-5.4). One way of estimating $\hat{\theta}$ is minimizing the L_2 norm

$$\hat{\theta} = \arg \max_{\theta} \left\| \mathbf{h}(S) \theta - \log \left[\sum_{i \in I} \exp \left(\mathbf{b}(i, S) \hat{\beta} + \delta E[\mathbf{h}(S'|i, S)] \theta \right) \right] \right\| \quad (\text{A-5.5})$$

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

$$E \left[\mathbf{h}(S) \hat{\theta} - \log \left[\sum_{i \in I} \exp \left(\mathbf{b}(i, S) \hat{\beta} + \delta E[\mathbf{h}(S'|i, S)] \hat{\theta} \right) \right] \right] = 0. \quad (\text{A-5.6})$$

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

$$\log \mathcal{L} = \sum_{j, t} \sum_{i \in I(j, t)} d(i, t) \log \widehat{\Pr}(d(i, t) | i, S_{j, t}), \quad (\text{A-5.7})$$

where $\widehat{\Pr}(d(i, t) | i, S_{j, t})$ is given by Equation (A-5.3).

5. Parameters $\hat{\beta}$ and $\hat{\theta}$ imply the discharge policy $\tilde{\mathcal{D}}_0$, under the restrictions of the dynamic model. The counterfactual discharge policy $\tilde{\mathcal{D}}_{\Delta}$ – i.e., the policy anticipating a counterfactual assignment policy \mathcal{A}_{Δ} – can be characterized by the parameters $\hat{\beta}$ and $\hat{\theta}_{\Delta}$, where $\hat{\theta}_{\Delta}$ describes a counterfactual value function $V_{\Delta}(S)$ and is calculated by step #3 alone using counterfactual transition probabilities implied by \mathcal{A}_{Δ} . I constrain the counterfactual value function to take the same value as the original value function at EOS, i.e., $V_{\Delta}(S | t = \bar{t}(s)) = V(S | t = \bar{t}(s))$.

A-5.4 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 3), 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.³⁹ Instead, I include the number of patients remaining on the physician’s census at EOS as a state variable in the value function. That is, I consider EOS as the “terminal state” at least in the counterfactual discharge policy $\tilde{\mathcal{D}}_\Delta$, after which the physician adopts \mathcal{D}_0 estimated outside of the dynamic model.

Finally, as is well known in discrete choice, utility is normalized for a reference choice. I normalize the choice $i = \emptyset$ 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-5.5 Specification and Model Fit

I specify a relatively simple baseline model. Utility flow $\bar{u}(i, S)$ is a linear function of an indicator for whether $i \neq \emptyset$, predicted length of stay (based on patient characteristics for i) if $i \neq \emptyset$, 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.

Figure A-5.1 shows the model fit according to discharge probabilities using a sample of ap-

³⁹In 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-4.4). 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% for a single patient. However, this specification does not allow me to calculate counterfactual discharge policies.

proximately 800,000 doctor-patient-time observations. The figure shows three types of discharge probabilities $\Pr(d(i, t) | i, S)$, 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 from the “static” multinomial logit model implied by \mathcal{D}_0 , without any exclusion restrictions to identify dynamics. The third is the probability implied by the dynamic model $\tilde{\mathcal{D}}_0$ from Equation (A-5.3), with the exclusion restrictions specified above to separate flow utilities from value functions. Figure A-5.1 shows that the dynamic model fits quite well compared to the static 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 the baseline dynamic model, the overall likelihood function in Equation (A-5.7) is maximized with a discount factor $\hat{\delta} = 0.98$.

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 $\bar{w}(i)$ defined in Equation (6). The panels of this figure evaluating length of stay and average workload are based on the same simulation algorithm, described in Appendix A-4.3, used to evaluate counterfactual assignment policies. This figure shows that simulated outcomes fit extremely well under the discharge policy \mathcal{D}_0 estimated by the unrestricted “static” multinomial logit model. The fit is also reasonably good using the discharge policy $\tilde{\mathcal{D}}_0$ 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-5.6 Discussion of Results

As stated in the main text, Figure 8 primarily shows that the relationship between counterfactual costs and counterfactual assignment policies is robust to the discharge policy used in the simulation. Hospital resources substantially increase when more patients are assigned near EOS. Physician time costs are not substantial, but since there are no significant hospital resource savings when decreasing patients assigned near EOS (e.g., already very few patients are assigned in the last hour of shift), total costs also increase in counterfactual regimes decreasing assignment near EOS.

Here I discuss nuances in the results under different discharge policies. As mentioned above, although the dynamic programming model $\tilde{\mathcal{D}}_0$ fits the data quite well, especially given its exclusion restrictions, it is inferior to the unrestricted “static” model, which makes no assumptions about utility flow or value functions. The particular issue with its fit is that it implies slightly

slower discharges for patients arriving in the last two hours of shift. Since decreases in length of stay near EOS increase patient-care costs, it is not surprising that counterfactual total costs do not increase by as much under the “partial dynamic” counterfactual simulations using \mathcal{A}_Δ and $\tilde{\mathcal{D}}_0$ as they do under the “static” simulations using \mathcal{A}_Δ and \mathcal{D}_0 .

Relative to $\tilde{\mathcal{D}}_0$, counterfactual discharge policies $\tilde{\mathcal{D}}_\Delta$ implied by the dynamic model primitives (i.e., utility flow parameters $\hat{\beta}$) and \mathcal{A}_Δ increase costs as Δ becomes more positive. This appears to be due to physicians discharging patients more slowly (relative to $\tilde{\mathcal{D}}_0$) at times that are moderately distant from EOS but then discharging patients increasingly quickly as they approach EOS. This behavior is consistent with dynamic “foot-dragging” in which forward-looking physicians who expect more patients arriving may try to foreclose new assignments by keeping patients longer in the medium-term and appearing busier (Chan, 2016). The net result is that workload-adjusted length of stay decreases by even more, since more patients are kept on census but length of stay eventually reaches the same dramatic lows that it does under $\tilde{\mathcal{D}}_0$. Under a model that ties hospital resources to workload-adjusted length of stay, this forward-looking behavioral adjustment thus increases hospital resources by even more.

Nonetheless, outcomes under $\tilde{\mathcal{D}}_0$ and $\tilde{\mathcal{D}}_\Delta$ still remain qualitatively the same. Four related reasons likely underlie the similarity. First, workloads implied by \mathcal{A}_Δ are well within the distribution of workloads that appear under actual assignment \mathcal{A}_0 . Second, counterfactual assignment policies only differ prior to EOS, but recall that starting at EOS, no patients are ever assigned, and therefore discharge policies must be identical starting at EOS. Third, assignments in the ED are quite random, and changes in assignment policy are far from deterministic. While there are patterns in how patients are assigned *in expectation*, receiving an extra patient especially near EOS would represent a substantial update in a physician’s outlook. Finally, perhaps reflecting this, it is noteworthy that the discount factor δ of 0.98 still implies that physicians only consider $0.98^{60/5} = 78\%$ of expected future events an hour in advance.

A-6 Other Heterogeneous Effects

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

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

where Y_{it} is workload-adjusted log length of stay, defined in Equation (6), 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 category is whether the shift ends on a “prime” hour between 4 p.m. and 8 p.m., which correspond to times (two hours later) when others are likely to be off work or out of school and awake. I consider two variations of this definition, one that does not include weekends and another that does. Results in Figure A-6.1 show that the decline in workload-adjusted length of

stay is steeper for shifts ending on prime hours.

The second set of categories refer to the physician $j(i)$ at t . In particular, I consider 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-6.2 show roughly similar results, regardless of physician type, except that faster physicians are much less likely to have EOS distortions.

The third 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-6.3 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 empirical results, with some discussion of some of these results:

- Table A-6.2 describes the process of constructing the sample, including the number of observations in each step.
- Table A-6.1 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-6.3 reports coefficients for EOS effects on workload-adjusted length of stay, as a continuation of Table 4. Results in this table are estimated with the full set of controls but only control for time relative to shift beginning. 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 when three types of orders are written: the last order by the attending physician of record (AOR), the first (non-resident) physician order after the AOR’s orders, and the last discharge order.
- Figure A-7.2 shows average patient counts (“censuses”) for physicians in shifts with different overlap \bar{o} .

- Figure A-7.3 shows coefficients for EOS effects on workload-adjusted length of stay, reported in Table A-6.3, estimated on both actual and simulated data.

Table A-6.3 presents evidence qualitatively consistent with results in Table 4, except that I do not control for patient characteristics, time indicators other than time relative to shift beginning, and provider identities. I consider these more parsimonious regressions in Table A-6.3 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 rich set of covariates would either be impractical. Specifically, as workload-adjusted length of stay decreases, hospital-resource costs increase via increased formal utilization and admission likelihood. Tables 4 and A-6.3 both show that workload-adjusted length of stay *increases* above baseline in the hours before the last of shift (while utilization is unchanged). This could be consistent with “foot-dragging” in which physicians delay discharge but do not otherwise change patient care Chan (2016); in the structural model, I therefore assume that increases in workload-adjusted length of stay do not change hospital-resource costs.

Figure A-7.1 shows the cumulative densities of orders relevant to physicians staying past EOS. I do not observe when physicians actually go home, but I can create some relevant bounds based on three types of orders: the last order by the attending physician of record (AOR), the first (non-resident) physician order after the AOR’s orders, and the last discharge order. Cumulative densities for physician orders are calculated using visits in which the relevant physician order is observed at least once. Last AOR order is a lower bound, since most visits do not have AOR orders (i.e., most visits have orders solely written by resident physicians), and AORs may stay past their last order (e.g., to dictating patient notes). Orders by the next non-resident physician are definitive proof that another physician has started to assume the care of some of the index physicians’ original patients, although the index physician may still be working on other patients and on work that does not require orders. The last discharge order is likely an overestimate of the time spent past EOS, since some patients may be passed off or have well-defined discharge plans before the discharge order.

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-1.1), 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-1.2), on hourly propensities for assignment relative to EOS while controlling only for time categories, pod, and providers. Further details are given in Appendix A-1.1. 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: Potential Bias from Selection on Unobservables

	$\widehat{\text{Var}}(A_{\bar{m}})$	$\widehat{\text{Var}}(A_{\bar{m}}^*)$	Selection on observables	Bias upper bound	$\hat{\alpha}_m$	$\hat{\alpha}_m$ as bias multiple
Patient selection into hour prior to EOS ($A_{\bar{m}}$)						
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-1.2. Selection is modeled for whether a patient is assigned in the \bar{m}^{th} hour prior to EOS ($A_{\bar{m}}$) by Equation (A-1.6), the residual of which is $A_{\bar{m}}^*$. Selection on observables is defined as $\bar{E}[\mathbf{W}'_{\gamma W} | A_{\bar{m}} = 1] - \bar{E}[\mathbf{W}'_{\gamma W} | A_{\bar{m}} = 0]$, where γ_W is estimated from Equation (A-1.4). Using the condition from Altonji et al. (2005), in Equation (A-1.5), 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-1.7). I use $\widehat{\text{Var}}(\varepsilon) = 0.160$ and $\widehat{\text{Var}}(\mathbf{W}'_{\gamma W}) = 0.580$ in this calculation. $\hat{\alpha}_{\bar{m}}$ is estimated by Equation (2); for convenience, results are repeated from Column 5 of Table 2. Finally $\hat{\alpha}_{\bar{m}}$ is stated as a multiple of the bias upper bound in the last column of this table.

Table A-1.3: Assessing Selection Bias Using Arrival Times

	(1)	(2)	(3)
	Mean actual, Y_t	Mean predicted, \hat{Y}_t^{ante}	Mean predicted, \hat{Y}_t^{full}
Mean EOS effect, Q_t	1.029*** (0.060)	0.029 (0.025)	0.024 (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-1.1. 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-1.11). 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-1.14). Average EOS effects within hour of arrival, Q_t , is defined by Equation (A-1.8) and calculated as follows: (i) Coefficients on time relative to EOS are calculated from (2) 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-6.1: 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 $\langle \ell, \underline{o}, \bar{o} \rangle$, where ℓ is the shift length in hours, \underline{o} is the overlap in hours with a previous shift, and \bar{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 $\langle \ell, \underline{o}, \bar{o} \rangle$ in progress), and actual patients (patients who are treated by a physician on a shift of type $\langle \ell, \underline{o}, \bar{o} \rangle$).

Table A-6.2: 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

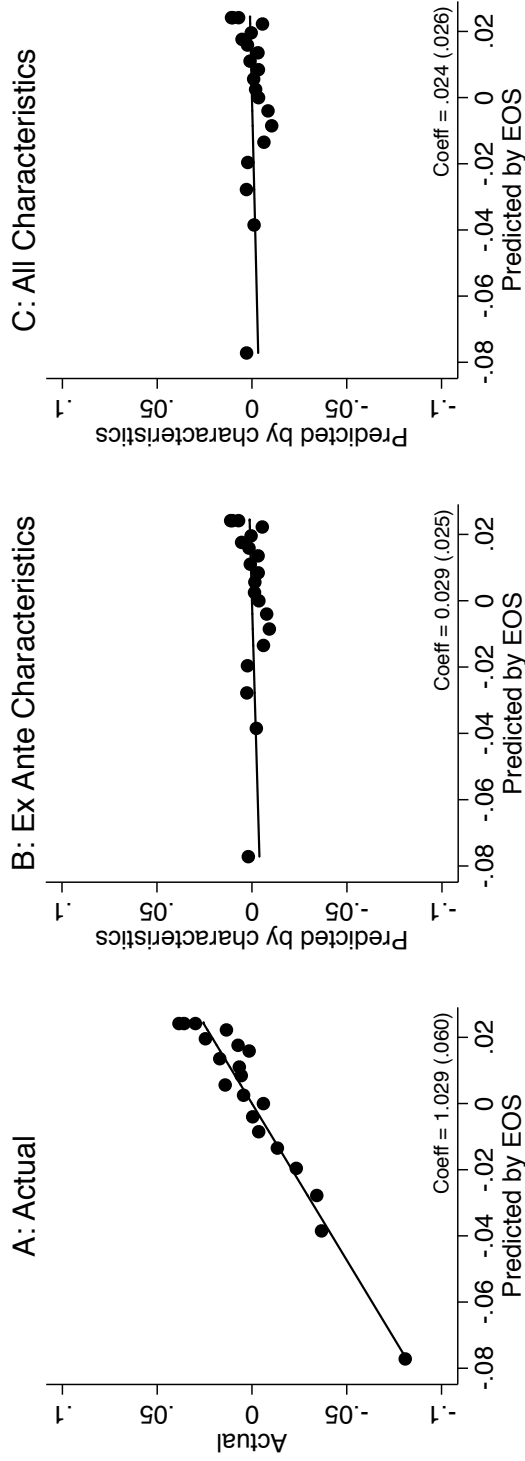
Note: This table describes each step in sample construction. Variables included in each step are listed in the second column, and the number of observations resulting from each step are in the third column.

Table A-6.3: 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*** (0.037)	-0.339*** (0.05)	0.031 (0.061)	-0.200*** (0.037)	-0.240*** (0.049)	-0.053 (0.063)
Second hour	-0.069*** (0.019)	-0.089*** (0.025)	0.168*** (0.035)	-0.067*** (0.019)	-0.064** (0.025)	0.101*** (0.036)
Third hour	-0.016 (0.015)	-0.027 (0.020)	0.176*** (0.028)	-0.013 (0.015)	-0.007 (0.020)	0.122*** (0.029)
Fourth hour	-0.077*** (0.014)	-0.076*** (0.017)	0.087*** (0.026)	-0.052*** (0.014)	-0.044** (0.017)	0.075*** (0.027)
Fifth hour	-0.052*** (0.012)	-0.048*** (0.014)	0.046** (0.022)	-0.037*** (0.012)	-0.029** (0.014)	0.031 (0.022)
Sixth hour	-0.032*** (0.008)	-0.028*** (0.010)	0.013 (0.014)	-0.028*** (0.008)	-0.021** (0.010)	0.002 (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	334,783	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

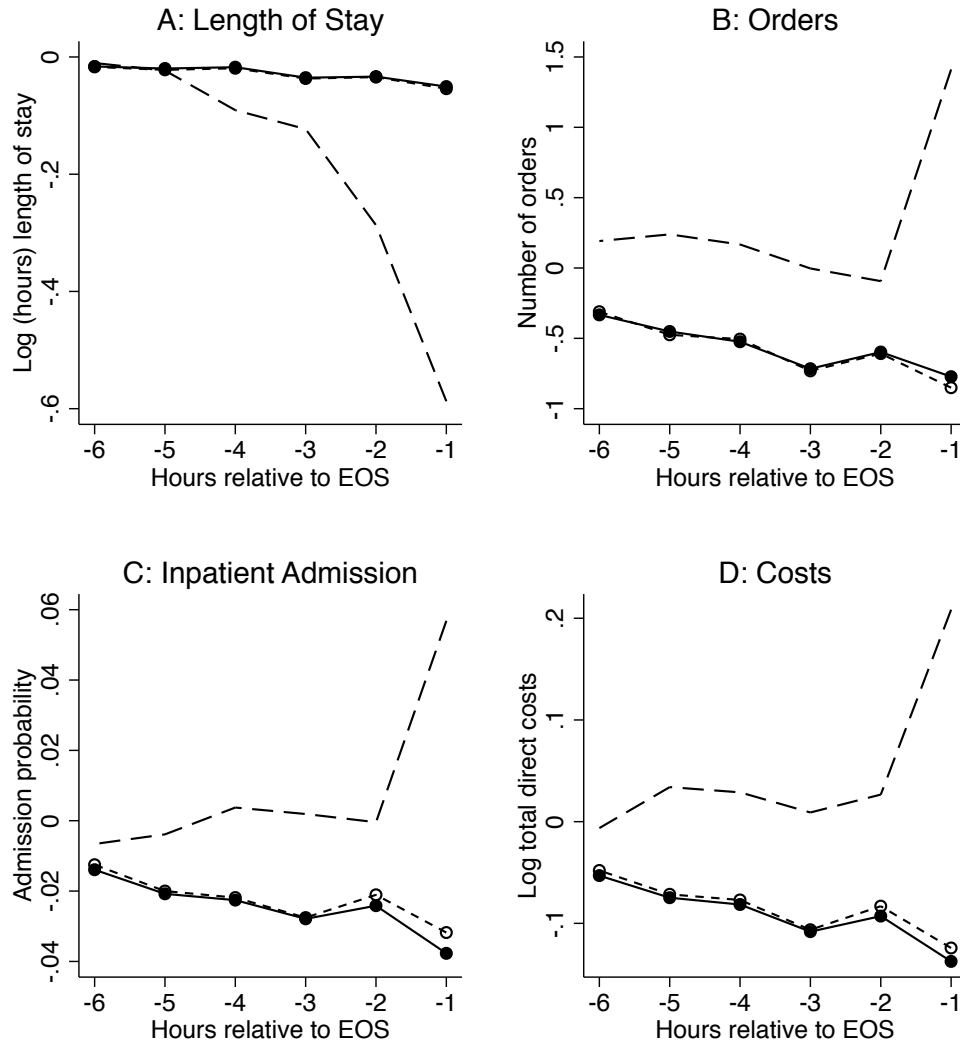
Note: This table is a continuation of Table 4, reporting coefficient estimates and standard errors in parentheses for EOS effects on workload-adjusted length of stay, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Models (1) to (3) 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-4.4), 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 (6). * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. Results are graphically shown in Figure A-7.3.

Figure A-1.1: Assessing Selection Bias Using Arrival Times



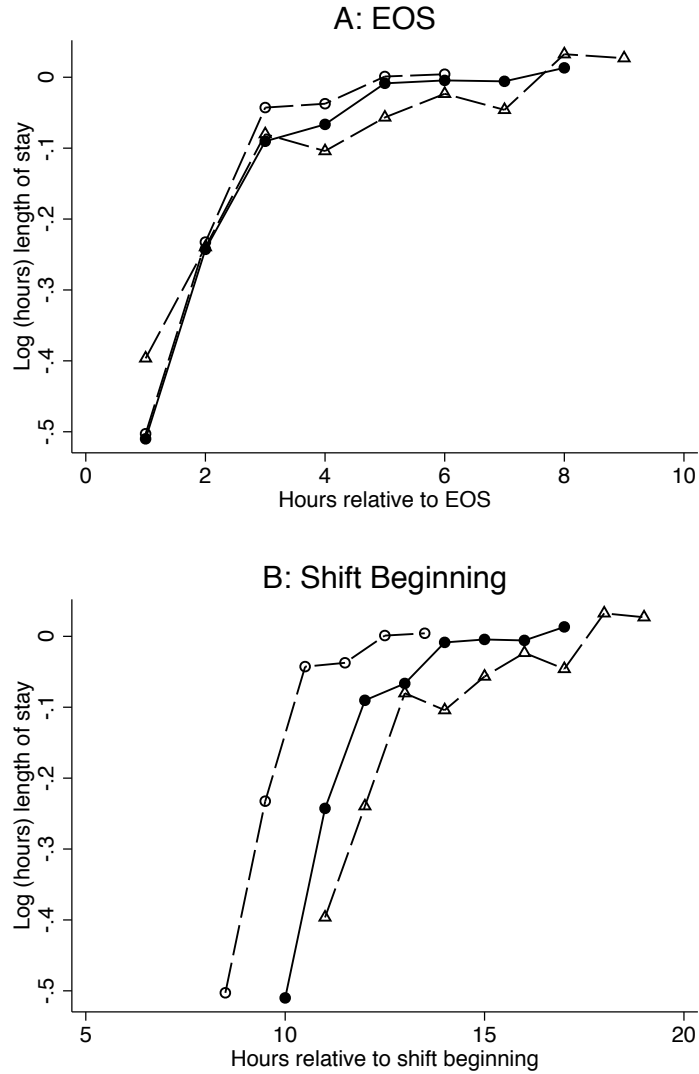
Note: This figure is the graphical representation of Table A-1.3, using only variation in arrival times to assess selection bias. It shows binned scatterplots of actual (residualized) log length of stay (Panel A), log length of stay predicted on “ex ante” characteristics possibly observable to physicians 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-1.8) as Q_t . Q_t is calculated as follows: First, coefficients on time relative to EOS are calculated from (2) 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-1.3). Details are given in Appendix A-1.3.

Figure A-1.2: Patient Selection on Observables Relative to End of Shift



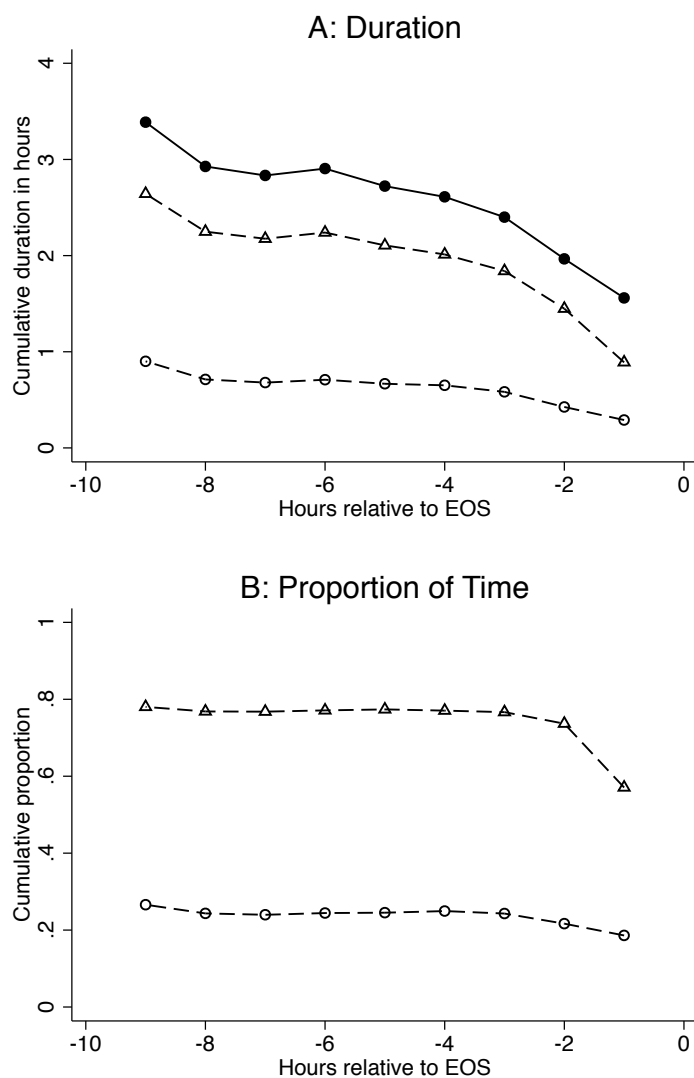
Note: This figure shows selection on observables for each hour prior to end of shift (EOS) on length of stay (Panel A), orders (Panel B), inpatient admissions (Panel C), and costs (Panel D). Each outcome is predicted based on patient characteristics observable prior to acceptance (age, sex, ESI) (closed 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-1.15). 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-2.1: Effects on Length of Stay by Shift Length



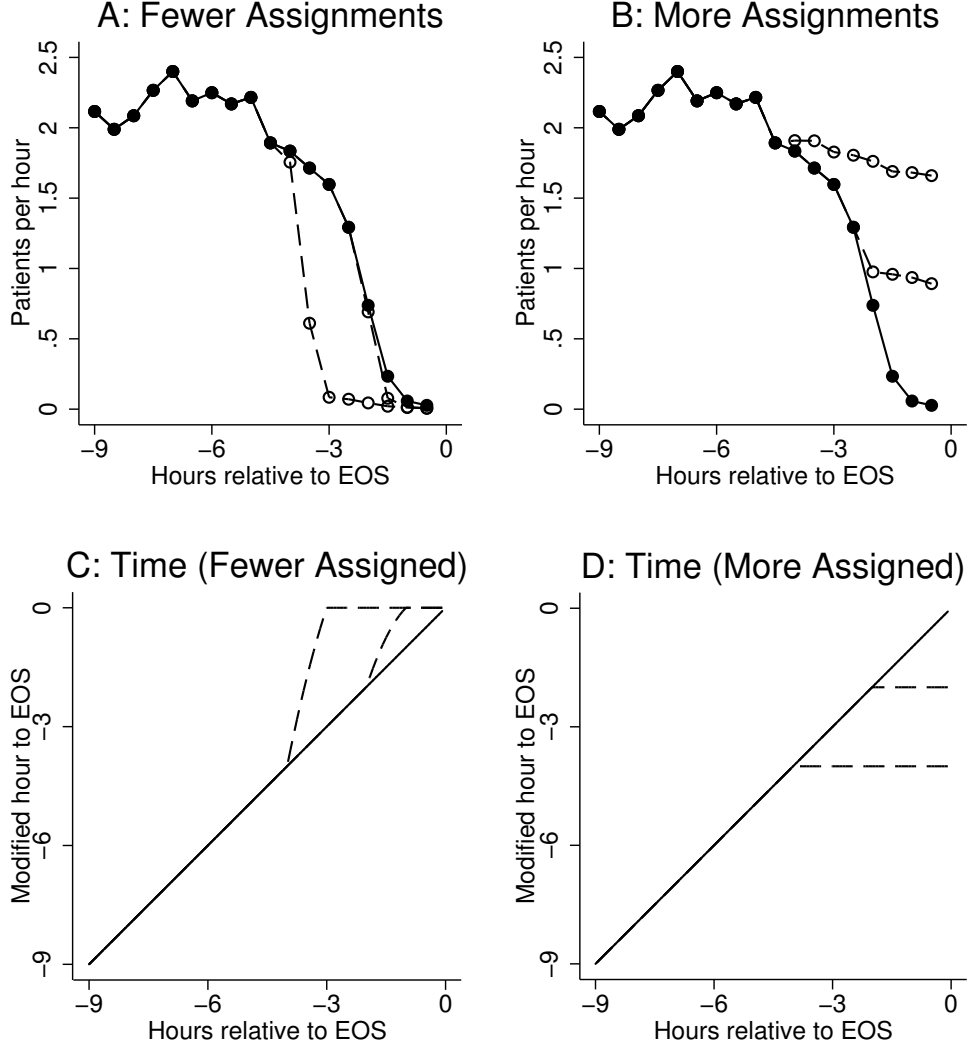
Note: This figure shows coefficients from Equation (2) estimated separately for shifts of seven or eight hours in length (open circles), nine hours in length (closed circles), and ten hours in length (open triangles). Panel A arranges estimates by hours relative to end of shift (EOS). Panel B arranges estimates by hours relative to shift beginning.

Figure A-3.1: Time Components



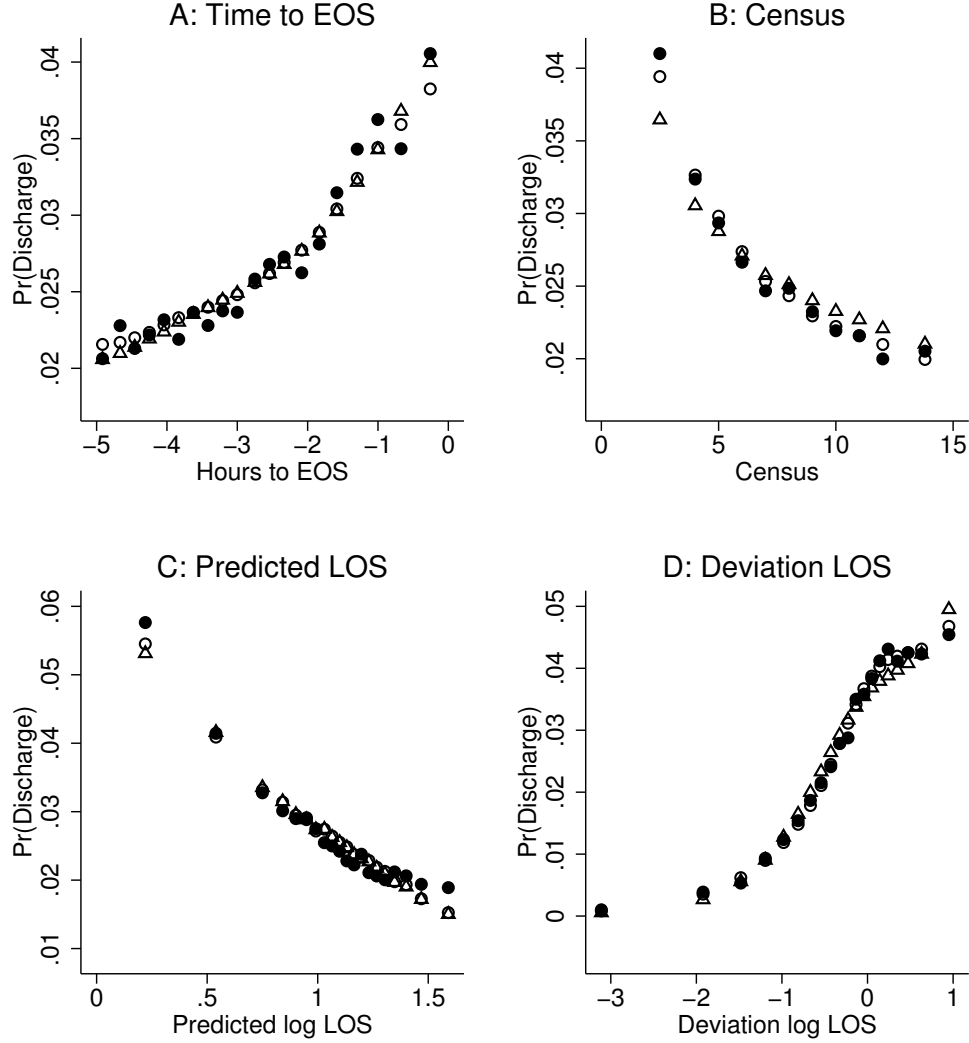
Note: This figure plots time components of length of stay as a function of hours relative to end of shift (EOS): time from pod arrival to first order (open circles), time from first to last (non-discharge) order (open triangles), and time from last order to discharge order (closed circles). Panel B shows 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-4.1: Example Counterfactual Assignment Policies



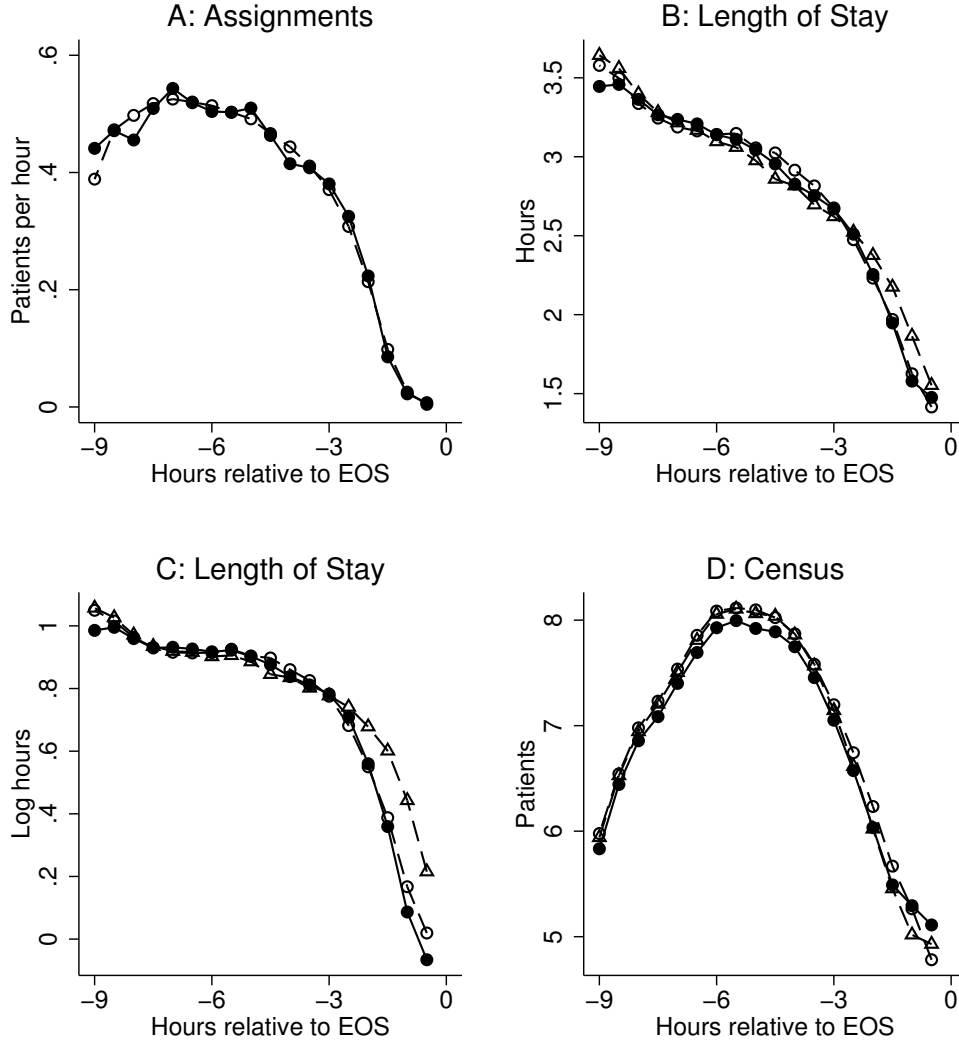
Note: This figure shows example counterfactual assignment policies, parameterized as hours Δ that time relative to EOS can be modified by. With dashed lines, Panel A shows counterfactual policies that reduce assignment near EOS ($\Delta \in \{-4, -2\}$); Panel B shows counterfactual policies that increase 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-4.2). $\Delta < 0$ 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.

Figure A-5.1: Model Fit by Discharge Probabilities



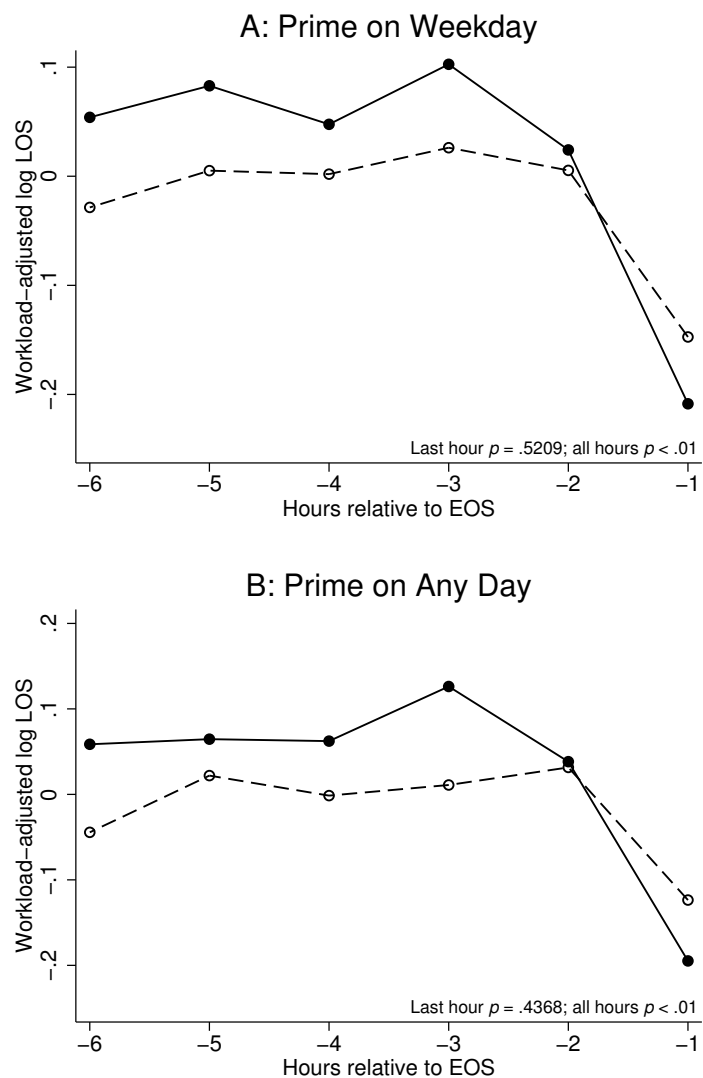
Note: This figure shows the fit of the discharge policies with actual discharge decisions in a sample of approximately 800,000 doctor-patient-time observations. Each panel evaluates the fit between the actual data (solid dots), a multinomial logit “static” model of the discharge policy \mathcal{D}_0 (hollow dots), and a dynamic model yielding discharge policy $\tilde{\mathcal{D}}_0$ and is based on restrictions on what can enter into the utility flow (triangles). The y -axis in all panels is the probability or average of whether a patient i still undischarged as of time $t - 1$ could be discharged at t by doctor j on shift with certain state characteristics. The x -axis is some characteristic of (i, j, t) : hours that (j, t) is away from EOS in Panel A; the number of patients on census, or $\|I(j, t)\|$, in Panel B; predicted length of stay according to i ’s characteristics and time categories of t in Panel C; the deviation in current log length of stay from the predicted log length of stay in Panel D. Each marker potentially represents 1/40th of the data if the x -axis is sufficiently continuous. The dynamic programming approach is described in greater detail in Appendix A-5.

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



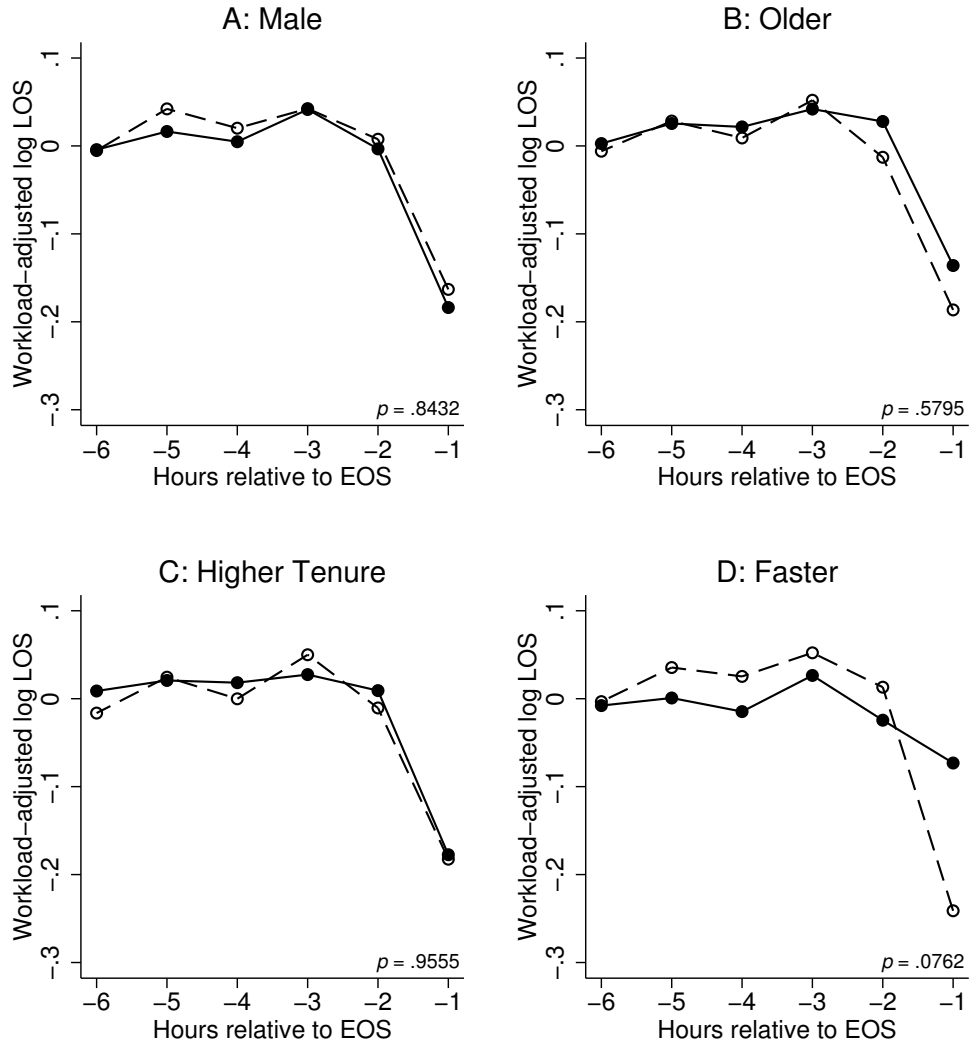
Note: This figure shows the fit of simulated outcomes with actual outcomes observed on 22,434 shifts with non-missing shift types out of the universe of 24,499 shifts. Outcomes are plotted against the x -axis of hours relative to EOS, in 30-minute intervals. Panel A evaluates the fit of actual and simulated patient assignments under \mathcal{A}_0 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 “static” model of the discharge policy \mathcal{D}_0 (hollow dots), and a dynamic model yielding discharge policy $\bar{\mathcal{D}}_0$ and is based on restrictions on what can enter into the utility flow (triangles). Simulated discharge-based outcomes are simulated using the algorithm in A-4.3, except that assignments are taken as given and not simulated (i.e., step 2a is skipped, and all patients will be assigned via step 2b). Panel D (“Census”) reflects actual and simulated average census during the length of stay of a patient i , $\bar{w}(i)$ in Equation (6). The dynamic programming approach is described in greater detail in Appendix A-5.

Figure A-6.1: End of Shift Effect by Whether Last Hour is Prime



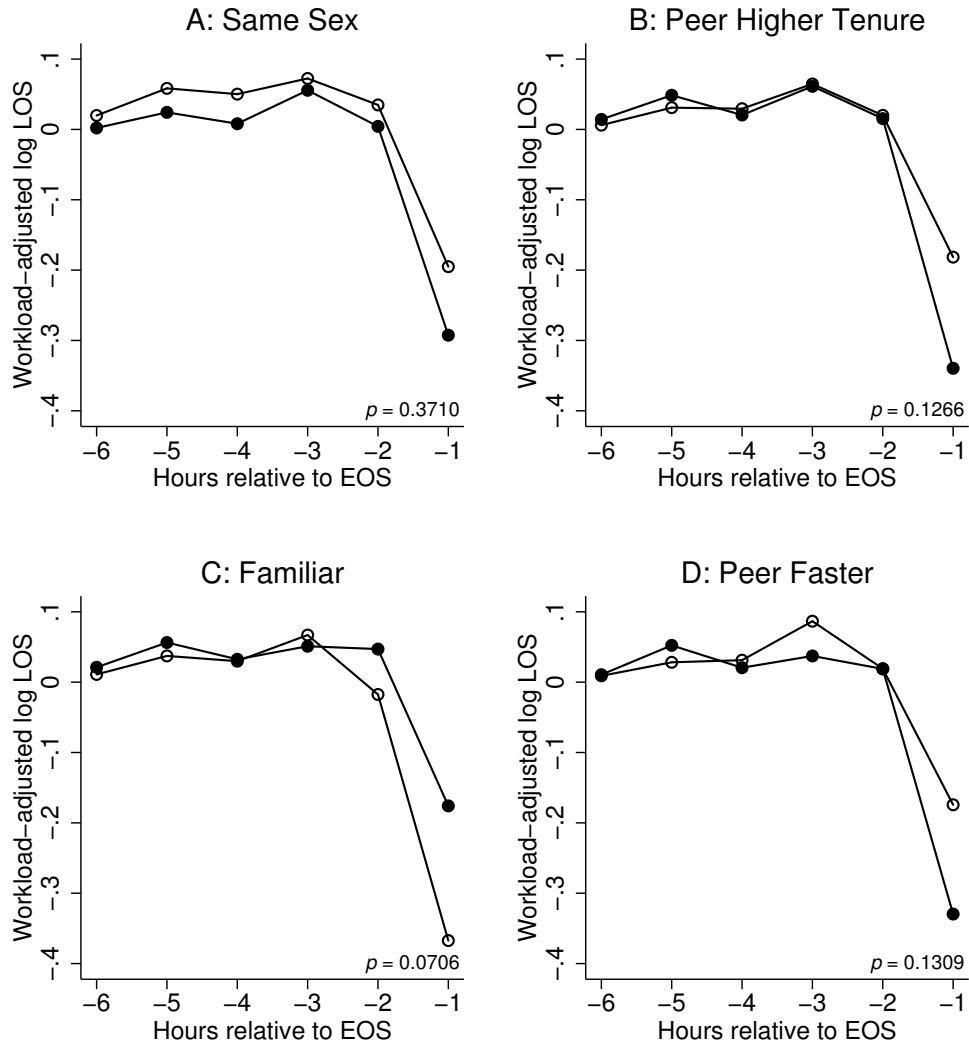
Note: This figure shows estimates of effects of hour relative to end of shift (EOS), using Equation (A-6.1), conditioning on whether the shift ends on a “prime hour” (4 p.m. to 8 p.m.) (solid dots) or not (hollow dots). Panel A categorizes any shift on the weekend as not prime, while Panel B categorizes weekend shifts ending on an hour from 4 p.m. to 8 p.m. as prime. p-values for the null hypothesis that the difference between (prime and non-prime shift) coefficients for the last hour of shift is 0 and for the null hypothesis that the difference between coefficients for all hours is 0 are given in notes for each panel.

Figure A-6.2: End of Shift Effect by Physician Type



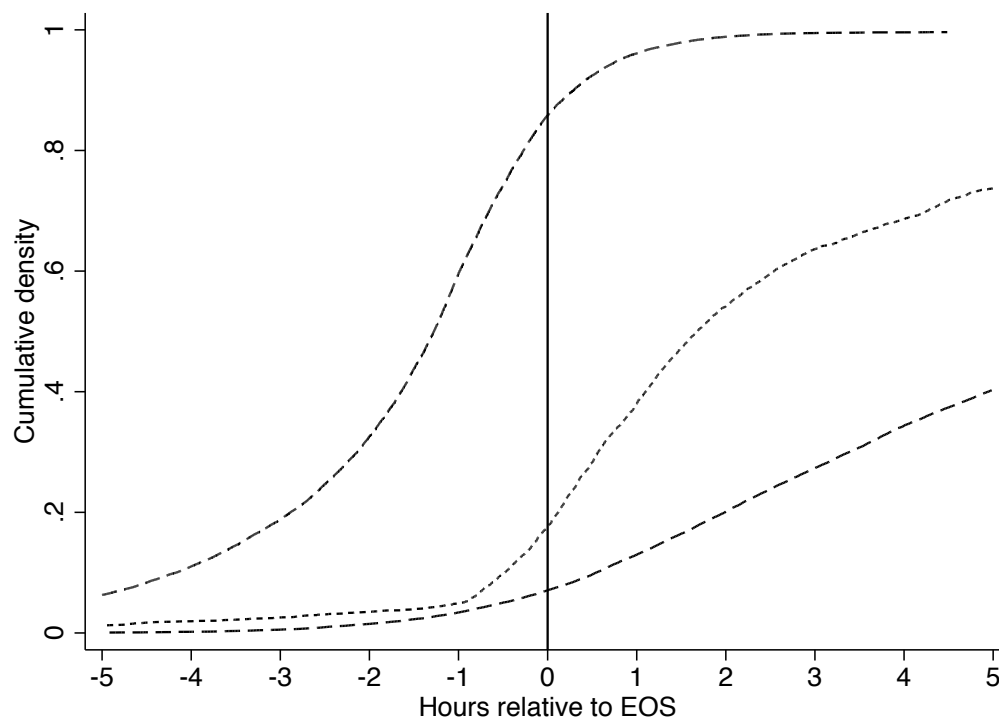
Note: This figure shows estimates of effects of hour relative to end of shift (EOS) for physicians of different types, using Equation (A-6.1). In each panel, estimates for physicians of the stated type are shown in solid dots, while estimates for physicians not of the stated type are shown in hollow 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 D 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-6.3: End of Shift Effect by Physician-Peer Relationship



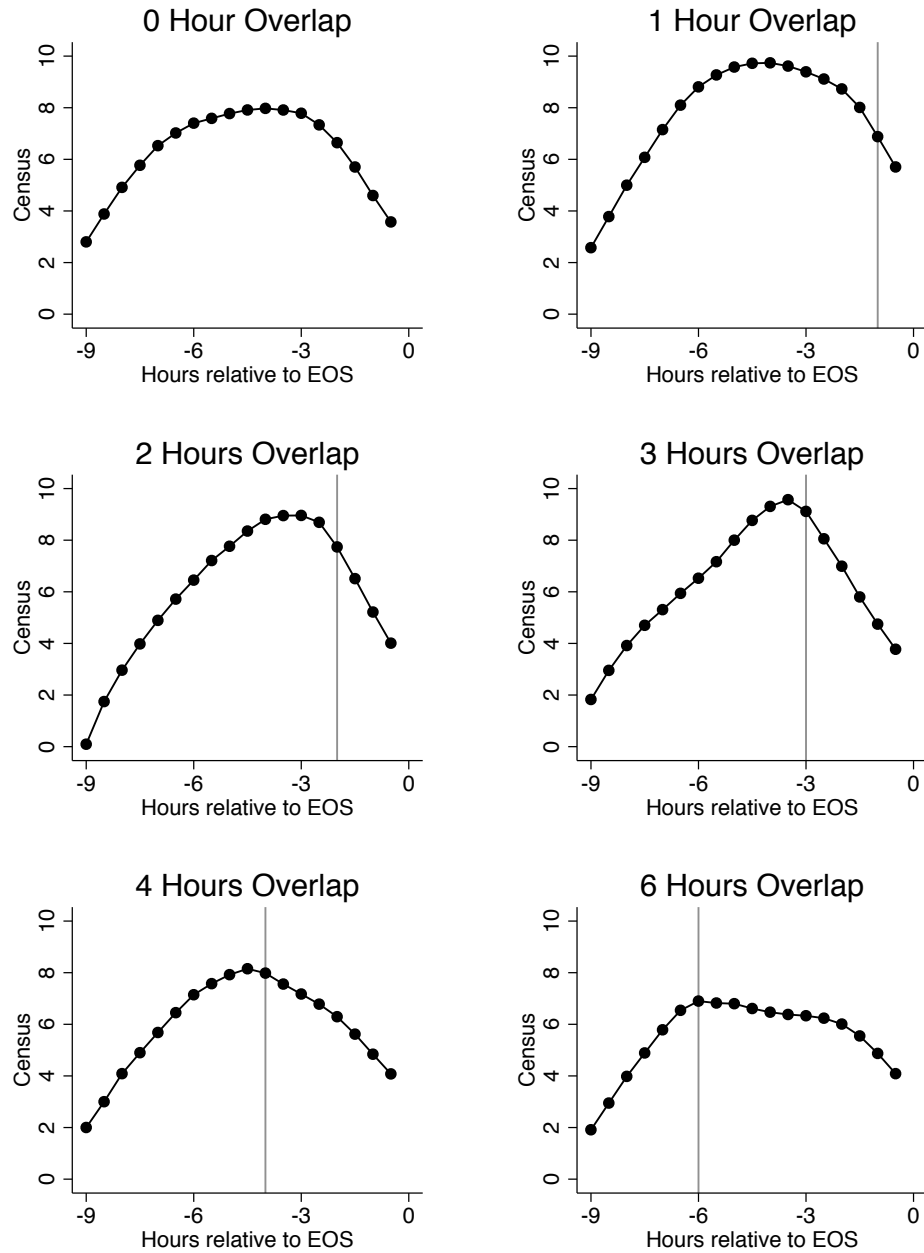
Note: This figure shows estimates of effects of hour relative to end of shift (EOS) for different types of relationships between physicians and the peer in the subsequent shift, if there is one. Estimates are from Equation (A-6.1). In each panel, estimates for physician-peer relationships of the stated 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: Evidence from Orders of Staying Past EOS



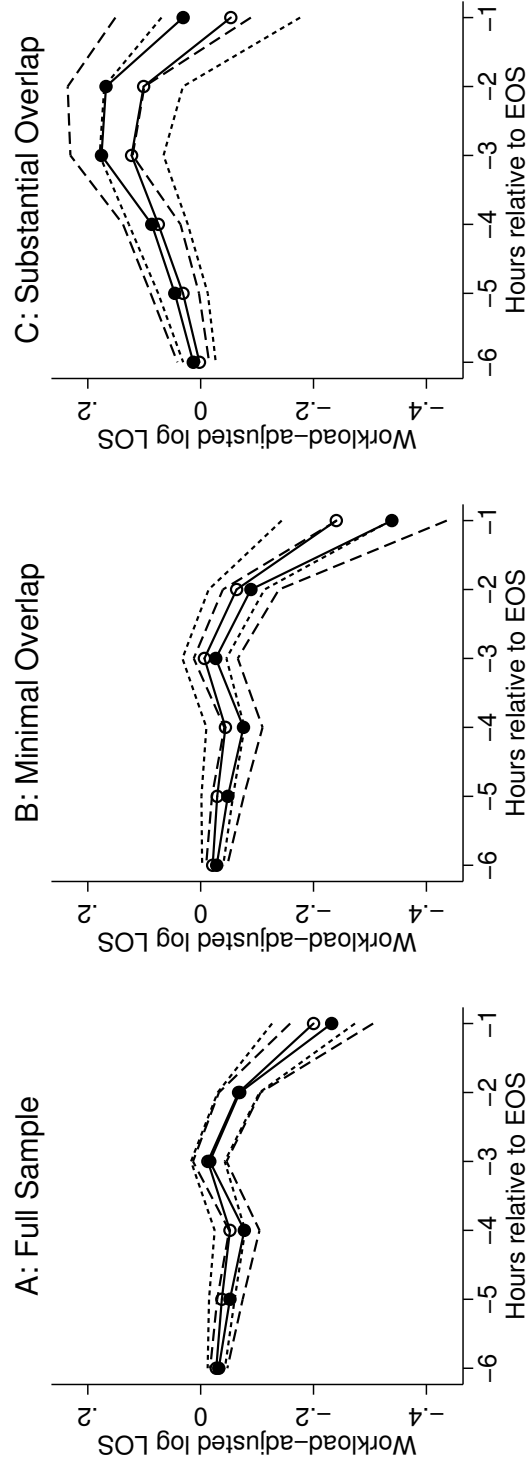
Note: This figure shows the following cumulative density plots, in order from left to right: last order by the attending physician of record (the physician on the bill for patient care, or AOR, corresponding to the physician whose shift is matched to a patient visit), the first (non-resident) physician order after AOR orders, and the last discharge order. The cumulative density of the last order by the AOR at EOS is approximately 86% (i.e., 14% of AORs write orders after their EOS). The median time for the first physician order after AOR orders is 1.7 hours. Only 35% of patient visits have an order by the AOR. Of these visits, only 6% of visits have a subsequent order by an attending physician. Cumulative densities for physician orders are calculated using visits in which the relevant physician order is observed at least once. See notes for this figure in Appendix A-7 for further comments.

Figure A-7.2: 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-7.3: Workload-adjusted Length of Stay



Note: This figure shows coefficients for regressions, described in Equation (A-4.4), 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-6.3.