NBER WORKING PAPER SERIES

# THE EFFICIENCY OF SLACKING OFF:
# EVIDENCE FROM THE EMERGENCY DEPARTMENT

David C. Chan, Jr.

Working Paper 21002
http://www.nber.org/papers/w21002

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2015

I am thankful to Amitabh Chandra, Daniel Chen, David Cutler, Mike Dickstein, Joe Doyle, Liran Einav, Bob Gibbons, Jeremy Goldhaber-Fiebert, Jon Gruber, Rob Huckman, Dan Kessler, Eddie Lazear, Sara Machado, Grant Miller, Paul Oyer, Maria Polyakova, Orie Shelef, Jonathan Skinner, and Chris Walters, and seminar participants for helpful comments and suggestions. I acknowledge support from the NBER Health and Aging Fellowship, under the National Institute of Aging Grant Number T32-AG000186; the Charles A. King Trust Postdoctoral Fellowship, the Medical Foundation; and the Agency for Healthcare Research and Quality Ruth L. Kirschstein Individual Postdoctoral Fellowship F32-HS021044. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by David C. Chan, Jr.. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Efficiency of Slacking Off: Evidence from the Emergency Department
David C. Chan, Jr.
NBER Working Paper No. 21002
March 2015
JEL No. D20,I10,L23,M50

## ABSTRACT

Work schedules play an important role in time-sensitive production utilizing workers interchangeably. Studying emergency department physicians in shift work, I find two types of strategic behavior induced by schedules. First, on an extensive margin, physicians "slack off" by accepting fewer patients near end of shift (EOS). Second, on an intensive margin, physicians distort patient care, incurring higher costs as they spend less time on patients accepted near EOS. I demonstrate a tradeoff between these two strategic behaviors, by examining how they change with shift overlap. Accounting for both costs of physician time and patient care, I find that physicians slack off at approximately second-best optimal levels.

David C. Chan, Jr.
Center for Health Policy and
Center for Primary Care and Outcomes Research
117 Encina Commons
Stanford, CA 94305
and NBER
david.c.chan@stanford.edu

# 1 Introduction

*The clock, which made industrial work possible, fundamentally changed the nature of work and, subsequently, the rest of life, making time scarce.*

Robinson and Godbey (2010)

*There is no more wasteful entity in medicine than a rushed doctor.*

Jauhar (2014)

Two broad changes have altered the intertemporal use of labor in health care production. First, technological advances have caused a proliferation in the diagnostic and therapeutic decisions that should be made in rapid order from a patient's presentation.[1] Second, 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 uncertain work is assigned by schedules (e.g., Briscoe, 2006; Casalino et al., 2003). While an expanding operations literature has focused on mechanically efficient scheduling,[2] this paper observes that the very existence of schedules induces a margin for strategic behavior: Schedules define boundaries of worker availability, but the nature of work often is neither sharply nor predictably delimited in time. This may be exploited by workers who have a private value for their time.

This possibility has implications for workplace design, which I consider in emergency department (ED) shift work. As in other settings, shifts in the ED are meant to specify a *minimum quantity* of hours worked, in which the end of shift (EOS) represents the time when physicians *may* stop accepting new work and go home if their work is complete. If physicians incur a private cost for time spent past EOS (e.g., if they are not compensated adequately, or at all,

[1] A 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.

[2] A large and active literature in operations management has investigated how staffing can be made more efficient by reducing labor costs and increasing profitability (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). Many service companies are increasingly using computerized staffing tools (Maher, 2007).

for this marginal time), schedules induce two forms of strategic behavior near EOS: First, on an extensive margin, physicians will accept fewer patients than socially optimal. Second, on an intensive margin, physicians may rush to complete their work, distorting their production to spend less time on patients they do accept near EOS.

Exploiting variation in the arrival of patients relative to physician schedules, I empirically ask whether patients are treated differently by physicians nearing their shift ends. 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 these effects from "fatigue," which I consider to depend on the time since the beginning of shift. I show that physicians accept fewer patients near EOS. For patients they do accept, I also show that physicians shorten the duration of care ("length of stay") in the ED and increase formal utilization, inpatient admissions, and overall costs as the time of arrival approaches EOS.

To interpret changes in patient care as strategic behavior, 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 not otherwise matter for socially efficient reasons, 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 strategic behavior. On the extensive margin, workers "slack off" by accepting fewer patients, akin to "presenteeism" in which workers are present but not working or stay at work apparently longer than useful. While slacking off represents a waste of physicians' time, it reduces workload when time becomes more costly and, on the intensive margin, mitigates physicians inefficiently substituting other inputs for time.[3] The assignment of work is largely observable and therefore manageable, but distortion on the intensive margin could be much more costly. Considering a

[3]The idea of time per effective work is related to work by Coviello et al. (2014), who discuss of 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.

wide range counterfactual regimes of patient assignment near EOS in a structural model based on the connection between workload-adjusted length of stay and patient care costs, I find that observed assignment patterns approximately minimize costs of physician time and patient care. In other words, given information asymmetry between workers and managers in patient care, slacking off is approximately second-best optimal at least in this setting.[4]

This paper is related to several strands of literature. First, a central economic question is how to induce workers to work efficiently. A mostly theoretical literature has broadly examined levers of financial incentives (Lazear, 2000), social incentives (Kandel and Lazear, 1992), performance measurement (Baker, 1992), and monitoring (Holmstrom, 1982).[5] This paper empirically demonstrates a general tradeoff betweeen the assignment of work and the subsequent performance of work, implying a second-best policy that allows workers to avoid work. This finding is thus related to the design of second-best incomplete contracts, for schedules as contracts that specify *minimum quantity* (or scheduled amount of time) of labor (Weitzman, 1974).

Second, this paper relates to a literature on improving productivity through workplace design (Ichniowski et al., 1997; Lazear, 1995), including recent work by Bloom et al. (2014) investigating the effect of working at home on productivity. The canonical neoclassical model considers worker choices between labor and leisure, but its mapping to current work environments seems at best coarse. In practice, availability for work is scheduled, with frequent, often daily boundaries between work and time off. This paper shows the productivity implications of potential worker agency, with respect to time, for the key workplace design issues of scheduling availability and distributing work. When managerial levers such as piece-rate pay emphasize performing "more" work, overall efficiency may be reduced when workers are preparing to leave work.

Third, my findings contribute to research on the health and social consequences of shift work (e.g., Gordon et al., 1986; Gold et al., 1992), as well as an increasing realization of the difficulties imposed by non-standard schedules on families (Kelly et al., 2011; Kantor, 2014;

[4]Another informational friction taken as given is the ability of workers to hand off work to each other. This friction of *client-worker specificity* is fundamental and has been discussed as the root of temporal inflexibility in many jobs (Briscoe, 2007; Goldin, 2014). In this case, the specificity occurs *ex post* patient assignment.

[5]A growing empirical component has studied social incentives (Mas and Moretti, 2009; Bandiera et al., 2005, 2009; Jackson and Schneider, 2011) and monitoring (Nagin et al., 2002; Duflo et al., 2013). A related empirical literature has examined nonlinear financial incentives with windows of performance measurement (Oyer, 1998; Larkin, 2014).

Greenhouse, 2014), as two-fifths of Americans now work mostly at non-standard times (Presser, 2003). Within health care, there has been a vigorous policy debate about physician work hours from a patient-safety standpoint (e.g., Shetty and Bhattacharya, 2007; Volpp and Rosen, 2007; Nasca et al., 2010). I draw attention to strategic behavior in scheduled work more generally. Considering the worker value of certainty and convenience may inform important production issues in health care and beyond, in which rushing to get home may incur wasteful costs.

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 physician acceptance of new patients. Section 5 reports EOS effects for patients who are accepted and considers evidence for patient selection and physician fatigue. Section 6 considers the relationship between shift overlap, workload, and patient-care distortion. Section 7 presents simulations of counterfactual regimes of patient assignment and shows that the observed assignment regime is approximately second-best optimal. Section 8 concludes.

# 2 Institutional Setting and Data

## 2.1 Shift Work

I study a large, academic, tertiary-care ED 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 ($\ell$). Shifts also differ in overlap with a previous shift ($\underline{o}$) or with a subsequent shift ($\overline{o}$) in the same location. I observe 23,990 shifts in 35 different shift types summarized by $\langle \ell, \underline{o}, \overline{o} \rangle$. Table A-2 lists the number of observations for each shift type in terms of shifts, hours, and patients.

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 $\overline{o}$. 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.[6]

[6] I distinguish between shifts that end with the closure of a patient location, or "terminal shifts" with $\overline{o} = 0$,

"Location" refers to a set of beds in the ED in which a physician may treat patients. This 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. 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. In particular, the regime change in May 2011 included the introduction of a new pod to increase the number of available beds to meet increasing ED volume. All regime changes, however, can be summarized as a set of shifts, each described by a shift type $\langle \ell, \underline{o}, \overline{o} \rangle$, a starting day and time, location, and range of months that the shift was in effect (see Figure A-1; Table A-1 details these shift descriptions).

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% productivity bonus based on clinical productivity (measured by Relative Value Units, or RVUs, per hour) and modified by research, teaching, and administrative metrics.[7]

## 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

and those continuing patient care with another shift in the same location, or "transitioned shifts" with $\overline{o} > 0$.

[7]The 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.

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 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. 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 readily transferred between providers in the same role when they end their respective shifts, perhaps reflecting the lesser importance of their information and discretion. In contrast, 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 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.[8]

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 and physical, serial

[8]In 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.

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-1. In the sample, I observe the identities of 102 physicians, 1,146 residents and physician assistants, and 393 nurses.

Table A-2 summarizes the 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. Because I focus on behavior near EOS, I also present in Figure 1 key variation in the time of day for EOS and the shift length, which separates EOS behavior from the time of the day and from time relative to the shift beginning, as well as variation in the time of day and the overlap with another shift at EOS.

I measure ED length of stay as a important input in patient care. Length of stay for patients arriving near EOS is exactly the measure that determines when a physician can leave work. I measure length of stay from the arrival at the pod to entry of the discharge order, which is 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.

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 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, including costs incurred both by formal utilization in the ED and during a subsequent admission.[9] 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). I observe all this information displayed to physicians prior to patient acceptance.

In addition, I observe patient characteristics that are usually known (if ever) by physicians only after patient acceptance – insurance status, language, race, zip code of residence, and rich diagnostic information – since physicians do not interact with patients or examine their charts 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.

### 2.5 Descriptive Evidence

Figure 2 shows a plot of the distribution of visits over arrival time prior to EOS and length of stay. Panel A shows the raw patient visit count in each fifteen-minute bin of arrival time interacted with each fifteen-minute bin of length of stay. Some findings are apparent from these visit plots. First, few patients are seen within the last two hours prior to EOS.[10] Second, lengths of stay are shorter for patients who arrive and are accepted by a physician closer to EOS than for

[9]Direct 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).

[10]Although relatively few patients are also seen arriving greater than nine hours prior to EOS, this fact reflects that relatively few shifts are greater than nine hours in length.

patients farther from EOS. There also appears to be an additional density of visits just prior to the 45-degree line mapping when length of stay roughly equals the time prior to EOS, implying that patients are more likely to be discharged just prior to EOS than at times before or after.

In order to examine more closely the discharge of patients conditional on acceptance, I plot in Panel B of Figure 2 the density of length of stay conditional on arrival time (and acceptance) prior to EOS. This plot shows a greater density of early discharges with arrival times closer to EOS. As in Panel A, for visits with arrival times between two to seven hours prior to EOS, there appears to be a linear mass of discharges along the 45-degree line in which discharges are roughly just prior to EOS.

# 3 Conceptual Framework

I introduce a simple model in this section to consider how schedules may distort three types of decisions made by ED physicians. First, they decide whether they will care for a patient arriving in their location. Second, if they assume care of the patient, they decide on the inputs of care: time and formal diagnostic tests and treatment. Third, based on information from these inpus, they produce a discharge decision, either for inpatient admission or to home. Because ED physicians may go home after EOS only if she has completed patient care, time becomes more privately costly as a physician nears EOS, with implications for all three decisions.

## 3.1 Model Setup

Consider a physician at the time $t$ of patient arrival, in work environment $\mathcal{E}_t \equiv (\mathcal{W}_t, \mathcal{W}'_t)$. $\mathcal{W}_t \equiv (\underline{t}, w_t)$ captures the start time of the physician's shift, $\underline{t}$, and her current workload, $w_t$. $\mathcal{W}'_t \equiv (\underline{t}', w'_t)$ describes similar information for a potential peer in the subsequent shift in the same managerial location. The patient's underlying health state is $\theta \in \{0, 1\}$, which is unobservable on arrival, but the patient is known on arrival to have $\theta = 1$ with probability $p$. Given the patient's arrival, 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 decides on inputs $\mathbf{z}$ in patient care: observation time $\tau_1$, tests and treatments $z$, and subsequent time $\tau_2$ to follow up on these tests and treatments.

3. After duration $\tau \equiv \tau_1 + \tau_2$, $\theta$ is observed with probability $q(\mathbf{z})$, and the physician decides on $d \in \{0,1\}$ ,to admit $(d=1)$ or discharge home the patient $(d=0)$.

4. The patient's health state $\theta$ is observed, and the physician receives the following utility:

$$u(t, \mathcal{E}_t; \theta; a, \mathbf{z}, d) = \begin{cases} O(\mathcal{E}_t; \theta), & a = 0 \\ V(\theta, d) - c_\tau\left(t, \tau, \bar{t}\right) - c_z z, & a = 1 \end{cases}. \tag{1}$$

   $O(\mathcal{E}_t; \theta)$ is the value of the "outside option" if $a = 0$, which captures concerns about the patient and a peer, who may or may not yet be present. $V(\theta, d)$ is the utility due to discharge decision $d$ for patient with health state $\theta$. $c_\tau\left(t, \tau, \bar{t}\right)$ is time cost of care, which is at least partly private and dependent on $\bar{t}$, the time of EOS; $c_z z$ is the test-treatment cost of care. I discuss $O(\mathcal{E}_t; \theta)$, $V(\theta, d)$, and $c_\tau\left(t, \tau, \bar{t}\right)$ more below.

### 3.2 Producing a Discharge Decision

I first examine EOS effects on the inputs to patient care and the discharge decision, assuming that the physician has chosen to accept the new patient $(a = 1)$. 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, or equivalently, physicians are risk-averse: $V(1,1) - V(1,0) > V(0,0) - V(0,1)$. Because of this last fact, if $\theta$ remains unobserved, the physician will admit if and only if $p > p^* < \frac{1}{2}$:

$$\begin{aligned} E[V|d=0, p=p^*] &= E[V|d=1, p=p^*] \\ p^* V(1,0) + (1-p^*) V(0,0) &= p^* V(1,1) + (1-p^*) V(0,1) \\ \frac{1-p^*}{p^*} &= \frac{V(1,1) - V(1,0)}{V(0,0) - V(0,1)} > 1. \end{aligned}$$

Patient care increases the probability $q$ of observing $\theta$ and therefore appropriate discharges.[11] This probability is increased by formal diagnostic tests and treatment, $z$, and by clinical observation and reasoning over time, $\tau$. $\tau$ can be thought of as two components, $\tau_1$ and $\tau_2$, which are substitutes and complements with formal testing and treatment, respectively. Testing, treatment, and time all entail costs to the physician, but I assume that time also becomes privately more costly as the physician nears EOS, since time spent on patient care past EOS prolongs the time the physician must stay at work.

I impose the following production relationships: $q$ is increasing and concave with respect to inputs to patient care $\mathbf{z}$ ($\tau_1$, $\tau_2$, and $z$). $\tau_1$ and $z$ are substitutes in production ($\partial^2 q/\left(\partial\tau_1\partial z\right) < 0$), while $\tau_2$ and $z$ are complements ($\partial^2 q/\left(\partial\tau_2\partial z\right) > 0$). Effective time per patient is reduced with higher workload $w_t$: $\partial^2 q/\left(\partial\tau_1\partial w_t\right) < 0$ and $\partial^2 q/\left(\partial\tau_2\partial w_t\right) < 0$. This contrasts with formal inputs, for which I make the normalizing assumption $\partial^2 q/\left(\partial z\partial w_t\right) = 0$.

Finally, I assume that the cost of time $\tau = \tau_1 + \tau_2$ depends on whether it extends beyond EOS: $c_\tau\left(t, \tau, \bar{t}\right) = c_\tau^0\left(\tau\right) + \tilde{c}_\tau\left(\tau + t - \bar{t}\right)$, where $c_\tau^0\left(\cdot\right)$ and $\tilde{c}_\tau\left(\cdot\right)$ (respectively the social and additional private costs of time) are positive, increasing, and convex.[12] To be precise about the implications of a private wedge in the cost of time due to EOS, consider a welfare function $W$ identical to Equation (1) except that it does not include private costs of time:

$$W\left(t, \mathcal{E}_t; \theta; a, \mathbf{z}, d\right) = \begin{cases} O\left(\mathcal{E}_t; \theta\right), & a = 0 \\ V\left(\theta, d\right) - c_\tau^0\left(\tau\right) - c_z z, & a = 1 \end{cases}. \tag{2}$$

**Proposition 1.** *Denote decisions in Section 3.1 that maximize expected utility in Equation (1), conditional on patient acceptance ($a = 1$), as $\tau^*$, $z^*$, and $d^*$. Denote corresponding decisions that maximize welfare in Equation (2) as $\tau^{**}$, $z^{**}$, and $d^{**}$.*

*(a) As $t \to \bar{t}$, $\tau^*$ decreases, $z^*$ may decrease (if $\tau_1$ dominates) or increase (if $\tau_2$ dominates),*

[11] I 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.

[12] To incorporate fatigue, $c_\tau^0$ may also depend on $t - \underline{t}$, but the crucial assumption is that only private time costs depend on $\bar{t}$. Private time costs could also be positive for $\tau + t - \bar{t} < 0$ if the physician needs to reduce her workload or attend to other duties (e.g., dictating charts) before going home.

*and $E[d^*]$ increases as long as $1 - F(p^*) > \frac{1}{2}$.*

*(b) For all $t$, $\tau^* < \tau^{**}$, and $E[d^*] < E[d^{**}]$.*

*(c) If the effect from $\tau_1$ dominates that from $\tau_2$, then $z^* < z^{**}$, and $z^{**} - z^*$ increases in $w_t$, holding $t$ constant and for all $t$. The reverse is true if $\tau_2$ dominates.*

As the physician nears EOS, she will shorten length of stay $\tau = \tau_1 + \tau_2$. The intensity of diagnostic tests and treatments may decrease or increase, depending on whether decreases in $\tau_1$ or $\tau_2$ dominate. Finally, she observes $\theta$ with lower probability $q$. This increases admissions, as long as $E[\theta] < 1 - F(p^*)$, where $F(\cdot)$ is the c.d.f. of $p$. A sufficient condition for this is $1 - F(p^*) > \frac{1}{2}$, which includes all symmetric distributions over $p \in [0,1]$. These distortions are greater with greater workload $w_t$, because $w_t$ further increases the cost of time by reducing the effective time per patient to produce $q$.

### 3.3 Accepting a Patient

I next consider the physician's upstream decision to assume the care of a new patient arriving near EOS. A patient arriving near EOS may be seen by the physician ending her shift or another physician beginning in the same managerial location. If the other physician has not yet begun his shift, then the patient would have to wait in order to be seen by him.

Formally, the physician decides whether or not to accept the patient, $a \in \{0,1\}$, in environment $\mathcal{E}_t$. This decision is a comparison between expected utility under $a = 1$,

$$E[u(t, \mathcal{E}_t; \theta; 1, \mathbf{z}^*, d^*)] = \max_{\mathbf{z}} \left\{ E\left[\max_d V(\theta, d)\right] - c_\tau\left(t, \tau, \bar{t}\right) - c_z z \right\},$$

where

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

and expected utility under $a = 0$, $E[O(\mathcal{E}_t; \theta)]$.

**Proposition 2.** *Denote $a^*$ as the patient acceptance decision in Section 3.1 that maximizes expected utility in Equation (1). Denote $a^{**}$ as the corresponding acceptance decision that maximizes expected welfare in Equation (2). $a^* \leq a^{**}$, and $a^{**} - a^*$ weakly increases as $t \to \bar{t}$.*

By construction, $a^* = 1$ if $E\left[u\left(t, \mathcal{E}_t; \theta; 1, \mathbf{z}^*, d^*\right)\right] > E\left[O\left(\mathcal{E}_t; \theta\right)\right]$. As $t \to \bar{t}$, the outside option $O\left(\mathcal{E}_t; \theta\right)$, under $a = 0$, captures peer considerations and increases with $t$ before the peer has arrived ($t < \underline{t}'$), since this diminishes the time that an unattended patient would have to wait for care. It also is intuitive that the patient increasingly should be assigned to the peer, as $t \to \bar{t}$, for first-best reasons such as fatigue. Moreover, higher private time costs further decrease expected utility under $a = 1$ as $t \to \bar{t}$, via higher overall input costs and lower-quality discharges (lower $q$) under reoptimization. Thus, $a^{**} - a^*$ weakly increases as $t \to \bar{t}$.

### 3.4 Efficiency Considerations

A fundamental root of this inefficiency is that information the physician has about the patient is difficult to observe by a manager and difficult to communicate between physicians. If a manager could observe the same information as the physician, then management could impose $a^{**}$, $\mathbf{z}^{**}$, and $d^{**}$. Further, information and the rationale behind decisions are difficult to communicate, even in the absence of misaligned incentives, reflective of "worker-client specificity" that is not unique to health care (Briscoe, 2006, 2007; Goldin, 2014).

Considering the informational environment as fixed, I focus on work assignment as the primary policy lever. Two efficiency considerations of this policy are notable. First, imposing $a^{**}$ alone may actually decrease welfare relative to allowing $a^*$, i.e., it may be that $W\left(t, \mathcal{E}_t; \theta; a^{**}, \mathbf{z}^*, d^*\right) < W\left(t, \mathcal{E}_t; \theta; a^*, \mathbf{z}^*, d^*\right)$. Second, when assignment can be controlled, workload acts as the key connection between "extensive" distortions in $a$ and in "intensive" distortions in $\mathbf{z}$ and $d$. In an expanded dynamic model, consider two patients arriving at different times, $t$ and $t+1$, and respective decisions $\left(a_t, \mathbf{z}_t, d_t\right)$ and $\left(a_{t+1}, \mathbf{z}_{t+1}, d_{t+1}\right)$. Increasing $a_t$ may increase $w_{t+1}$ and thus, from Proposition 1, reduce welfare by worsening distortions in $\mathbf{z}^*_{t+1}$ and $d^*_{t+1}$.

In Section 6.2, I empirically focus on overlap $\overline{o} \equiv \overline{t} - \underline{t}'$ as a mechanism to decrease $a_t^*$ through raising $O(\mathcal{E}_t; \theta)$. More broadly, $a_t^*$ may be influenced by a variety of managerial instruments, and in Section 7, I consider a generic set of patient-assignment policies, over which welfare may be (second-best) maximized subject to informational problems.[13]

# 4 Effect on Patient Acceptance

In this section, I describe the effect of approaching EOS on the physician's decision to accept patients. 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 more costly. The simple analysis in this section presents the unadjusted probability that a patient will be accepted by a physician nearing EOS across a variety of shift types. In particular, I will verify that greater overlap $\overline{o}$ allows physicians to decline patients earlier relative to EOS.

Figure 3 presents the hourly average rates of new patient visits, with each panel representing shifts with a different $\overline{o}$, for the index physician (patients accepted), for the location inclusive of the index physician (patients assigned by the triage nurse), and for the entire ED (patients arriving at the ED). Regardless of the shift type, physicians generally accept between two to three new patients per hour at most, and rates of acceptance are highest near the beginning of shift. Thereafter, in transitioned shifts with $\overline{o} > 0$, the average rates of patient flow show two consistent relationships with time. First, patient flow declines precipitously 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 $\overline{o}$, patient flow is relatively constant but diminished in that duration. In terminal shifts, where $\overline{o} = 0$, 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

[13] Another policy might be to fix the *price* of work by overtime, although this generally will not result in $c_\tau\left(t, \tau, \overline{t}\right) = c_\tau^0(\tau)$, given a second-best world, and may be inferior to quantity policies (Weitzman, 1974). Other mechanisms could aim to assign patient types that confer less discretion to physicians near EOS or to reduce of information asymmetry, including facilitating handoffs between physicians. However, it is likely that informational frictions are pervasive across environments and would be more difficult to address (Apker et al., 2007; Briscoe, 2006).

at least makes up for the decline in flow for the index physician. That is, despite declines in patient acceptance, patients continue to arrive at the pod at similar or greater rates prior to the peer's transitioning shift. Finally, Figure 3 plots the flow of patients to the entire ED, showing background patient flow to other pods that seems unrelated to flows to the index physician. Naturally, overall ED flow appears more stable when averaged across greater shift observations and variation across times of the day (see Figure 1, e.g., $\overline{o} = 1$ and $\overline{o} = 6$).

These relationships are remarkably consistent, over different $\overline{o}$, despite being presented as unadjusted averages. It is intuitive that physicians would decrease their acceptance of new patients as they approach EOS, since the cost of seeing new patients increases with proximity to EOS. The cost is both in the time cost to the physician ending her shift and also in terms of the resulting distortion in patient care.

The earlier arrival of peers allows for earlier reductions in patient acceptance relative to EOS. This includes reductions *prior* to peer arrival, especially in shifts with shorter transitions, suggesting anticipatory behavior. For terminal shifts with no peer arriving in the same location, remarkably, 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, by managerial assignment itself.

# 5 Effect on Patient Care

## 5.1 Main EOS Effects

My main analysis addresses the following: What is the effect of a patient's arrival near a physician's EOS on that patient's care by that physician? Although I address patient selection more directly later, I first control for a rich set of patient characteristics. I use variation within the same health care providers working at different times and locations to control for fixed provider unobservables. Using shift variation within locations and within times, I control for unobservables (e.g., patient characteristics and ED resources) that vary by location and time categories, such as time of the day or day of the week. I finally use variation in shift lengths to control for fatigue, which I consider due to time relative to the beginning of shifts.

In the full specification, I estimate the following equation:

$$\begin{aligned} Y_{ijkpt} &= \sum_{m=-6}^{-1} \alpha_m \mathbf{1}\left(\left\lfloor t - \bar{t}(j,t) \right\rfloor = m\right) + \sum_m \gamma_m \mathbf{1}\left(\left\lfloor t - \underline{t}(j,t) \right\rfloor = m\right) + \\ & \quad \mathbf{X}'_{it}\beta + \mathbf{T}'_t\eta + \zeta_p + \nu_{jk} + \varepsilon_{ijkpt}, \end{aligned} \tag{3}$$

where outcome $Y_{ijkpt}$ is indexed for patient $i$, physician $j$ (in shift from $\underline{t}(j,t)$ to $\bar{t}(j,t)$), assisting team $k$ (including the resident or physician assistant, and the nurse), pod $p$, and arrival time $t$. The coefficients of interest in Equation (3) are $\{\alpha_m\}$, indicating the effect of arrival of patient $i$ at $m$ hours (rounded down to the nearest negative integer) prior to physician $j$'s EOS. I control for time relative to the shift beginning $(t - \underline{t}(j,t))$, patient characteristics $\mathbf{X}_{it}$, time categories $\mathbf{T}_t$ (for month-year, day of the week, and hour of the day), pod identities $\zeta_p$, and physician-team identities $\nu_{jk}$.

Table 1 shows results for log length of stay, estimating coefficients $\{\alpha_m\}$ for time prior to EOS, from versions of Equation (3) 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 (3) estimate effects on log length of stay ranging from $-0.53$ to $-0.72$. The full model, shown in the last column of Table 1 and plotted in Panel A of Figure 4, estimates an effect on log length of stay of $-0.59$ in the last hour and serves as the baseline model for this paper.

Although I address selection and fatigue more directly later, results in Table 1 shed light on both of these. The difference in estimates between the first and second columns represents the effect of including a rich set of patient characteristics, which is about 0.06 on log length of stay in the last hour prior to EOS. The difference between the fourth and last 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.

Table 2 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 $\alpha_m$ are generally insignificant for hours before the last hour prior to EOS, but are significantly positive in the last hour. Patients arriving and accepted 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.[14] 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 4.

### 5.2 A Closer Look at Patient Care

I undertake two analyses to further examine the use of time in patient care. First, I consider *effective time per patient* to capture the fact physicians with more patients must devote less time per patient, holding length of stay constant. This concept highlights the 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.

For each patient $i$ accepted by physician $j$ at time $t$, I construct a measure of workload-adjusted length of stay, as length of stay ($\tau_{ijt}$) divided by the average number of patients cared for by $j$ during the $i$'s length of stay ($\overline{w}_{ijt}$):

$$\tau_{ijt}/\overline{w}_{ijt} = \tau_{ijt}\left[\frac{1}{\tau_{ijt}}\int_{\tilde{t}\in[t,t+\tau_{ijt}]} w_{j\tilde{t}}d\tilde{t}\right]^{-1}, \text{ where} \tag{4}$$

$$w_{j\tilde{t}} = \sum_{i,t}\mathbf{1}\left(\tilde{t}\geq t\right)\mathbf{1}\left(\tilde{t}\leq t+\tau_{ijt}\right)\mathbf{1}\left(j=\mathcal{J}\left(i,t\right)\right). \tag{5}$$

Workload $w_{j\tilde{t}}$ for $j$ at time $\tilde{t}$ is the number of patients under $j$'s care at $\tilde{t}$ (i.e. $j$'s patient "census" at $\tilde{t}$, counting patients who were accepted prior to $\tilde{t}$ and discharged after $\tilde{t}$), where $\mathcal{J}(i,t)$ is a function assigning patient $i$ at $t$ to a physician.

I then regress the log of workload-adjusted length of stay using Equation (3). As shown in the last column of Table 2, time relative to EOS has little effect on workload-adjusted length of

[14] I find that the percentage of orders for diagnostic tests, including laboratory and radiology tests, are slightly higher in the last hour prior to EOS, although this is not statistically significant.

stay until the last hour prior to EOS, when this measure decreases significantly. 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. At least in sample, distortions in patient care, including the use of time, appear to only become significant in the last hour prior to EOS.[15]

Second, I examine the components of time, $\tau_1$ and $\tau_2$ in Section 3, that can be respectively thought of as substitutes and complements of formal utilization. Roughly speaking, $\tau_1$ is time spent on clinical monitoring and reasoning that substitutes for brute-force utilization (e.g., serial abdominal examination as opposed to CT scan). On the other hand, $\tau_2$ is time needed to follow up on utilization (e.g., waiting for CT scan report). In practice, $\tau_1$ and $\tau_2$ are not neatly divided, but some intuitive distinctions can be made: I assume that time spent prior to the first formal order belongs to $\tau_1$ and that time spent after the last formal order belongs to $\tau_2$. Although time in between the first and last orders could belong to either $\tau_1$ or $\tau_2$, the spacing of these orders often reflects clinical monitoring and reasoning more closely related to $\tau_1$.

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 (3). Figure A-2 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 (3); 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 $\tau_1$, 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 2 and Figure 4.

## 5.3 Patient Selection

As described in Section 2.4, physicians observe a limited set of information when accepting patients, but I observe a richer set of patient characteristics, including diagnoses and other

[15] In Section 7, I consider counterfactual assignment regimes with greater assignment near EOS than observed in the data, which result in greater and earlier distortions in patient care.

characteristics that generally are unobserved by physician at patient acceptance.[16] In this analysis, I evaluate patient selection based on two sets of characteristics: those that are observed by a physician or triage nurse prior to acceptance, $\mathbf{X}_{it}^{prior}$, and others that include rich diagnosis codes, insurance status, race, and language that are at best incompletely observed until after patient acceptance, $\mathbf{X}_{it}^{full}$.

Separately for each set, I first generate predicted outcomes by

$$Y_{ijkpt} = \left(\mathbf{X}_{it}^{set}\right)' \beta^{set} + \varepsilon_{ijkpt}, \tag{6}$$

where $set \in \{prior, full\}$. Next, I estimate the following regression describing the relationship between the predicted outcomes for selected patients, $\hat{Y}_{ijkpt}^{set} = \hat{\beta}^{set}\mathbf{X}_{it}^{set}$, using $\hat{\beta}^{set}$ estimated from Equation (6), and the time of selection relative to EOS:

$$\begin{aligned} \hat{Y}_{ijkpt}^{set} \quad = \quad & \sum_{m=-6}^{-1} \alpha_m^{set} \mathbf{1}\left(\lfloor t - \bar{t}(j,t) \rfloor = m\right) + \sum_{m} \gamma_m \mathbf{1}\left(\lfloor t - \underline{t}(j,t) \rfloor = m\right) + \\ & \mathbf{T}_t' \eta + \zeta_p + \nu_{jk} + \varepsilon_{ijkpt}, \end{aligned} \tag{7}$$

leaving out variables in $\mathbf{X}_{it}$ as regressors. I interpret each coefficient $\alpha_m^{set}$ as the EOS effect on patient selection as described by a predicted outcome measure. Comparing results between the two sets of patient characteristics roughly assesses the degree of selection on unobservables.

Figure 5 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 (3) for each respective outcome. Coefficients for selection estimated using the two sets of characteristics are remarkably similar, suggesting negligible selection on unobservables. 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 frequencies of admissions, and incur 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

[16] Of course, some of this information may be known at patient acceptance, for example by informal communication by nursing, e.g., "Doctor, this patient has [formally undocumented issues] that need attention right now."

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.

In Appendix A-1, I undertake a more formal analysis, based on Altonji et al. (2005), to compute the degree of selection on patient unobservables relative to selection on observables required to explain my length of stay results. This approach considers, at each hour prior to EOS, the explanatory power of observables in determining whether patients are selected and the explanatory power of observables in determining length of stay. I find that selection on patient unobservables must be 475 times greater than selection on observables in order to explain the entire effect on length of stay for patients arriving in the last hour prior to EOS.

### 5.4 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 (3), 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.[17] 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-3 plots coefficients $\alpha_m$ from Equation (3) estimated separately for each shift-length category. Panel A plots coefficients according to time relative to EOS and shows coefficients largely similar across

[17] For other outcomes (orders, admission, and costs), the size of the EOS effect actually increases when controlling for time since shift beginning.

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.

### 5.5 Discharge Hazard over Time

Recall that Figure 2 revealed a density mass slightly before EOS for patients arriving at various times prior to EOS. It would be unlikely for an increase in the conditional probability (i.e., hazard) of discharge at EOS but not before or after to represent something other than strategic behavior in patient care. Patient selection at acceptance, for example, should apply to all discharge hazard rates before and after EOS. Fatigue should also continuously increase discharge hazard rates.

In order to formally apply this intuition to the pattern of patient arrival and discharge times in Figure 2, I estimate a discrete-time logit hazard model for discharge. I model the hazard as a logistic function, $\Pr\left(d_{it}=1\left|d_{i,t-1}=0\right.\right)=\left(1+\exp\left(-h\right)\right)^{-1}$, where

$$\begin{aligned} h\left(t,\tau,i,j,k,p\right) &= \sum_{m=-9}^{3}\alpha_m \mathbf{1}\left(t-\bar{t}\left(j,t\right)=m\right)+\sum_{m}\gamma_m \mathbf{1}\left(t-\underline{t}\left(j,t\right)=m\right)+ \\ &\quad \gamma_\tau \hat{E}\left[Y_{ijkpt}|X_i,j,k,p,\tilde{\mathbf{T}}_t\right]+\eta_t+\zeta_\tau. \end{aligned} \tag{8}$$

$t$ is discrete in hourly intervals, and $\tau$ is the hour of the stay (time relative to arrival). The coefficients of interest are the set of $\alpha_m$ for each time relative to EOS. In this hazard model framework, I include fixed effects for each hour of the day ($\eta_t$) and fixed effects for each hour of the stay ($\zeta_\tau$). I also control for time relative to shift beginning ($t-\underline{t}_j$) and predicted log length of stay ($\hat{E}\left[Y_{ijkpt}|X_i,j,k,p,\tilde{\mathbf{T}}_t\right]$) conditional on patient characteristics, provider identities, pod identity, and other time categories $\tilde{\mathbf{T}}_t$(month-year and day of the week).

Figure 6 shows marginal effects, estimated from Equation (8), for each hour $m$ relative to EOS. The hourly discharge hazard ranges from 22% to 31% for the span of time between nine hours prior to EOS and four hours past EOS. The peak of 31% corresponds to probability of discharges within the last hour prior to EOS, and the discharge hazard decreases on both sides of the peak with increasing distance.

These results are consistent with strategic behavior in targeting discharge times right around EOS. As EOS approaches, physicians are increasingly likely to discharge remaining patients. However, after EOS has past, the discharge hazard begins to decrease, possibly reflecting the fact that remaining selected patients have conditions requiring further evaluation and treatment in the ED. After the fourth hour past EOS, exceedingly few patients remain and are likely to be transferred to another physician's care.

# 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 strategic behavior, 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. Conditional on these, 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 policy lever for the more general concept of patient assignment, which influences the efficiency of patient care through workload: Assigning more patients reduces effective time per patient, worsening the EOS distortion in patient care. In Section 7, This empirical variation will also allow for a more general and structural consideration of patient assignment, workload, and efficiency, relevant to a wide range of managerial levers (e.g., rules, financial incentives, social incentives) that primarily act on assignment.

## 6.1 Patient Censuses over Time

I measure workload as the number of patients cared for by a physician (her "census") at a given time, as in Equation (5). Figure A-4 shows unadjusted census averages at each 30-minute interval, for the nine hours prior to EOS, in different shift types by $\overline{o}$. On 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 $\overline{o} = 2$, which happen not to transition from another shift. The census peak coincides with the beginning of the transition time in shifts with $\overline{o} \geq 4$ but occurs progressively earlier as $\overline{o}$ decreases. 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.

In order to leave work, physicians must generally reduce censuses to zero. With smaller $\bar{o}$, they have less time and scope to reduce their censuses before EOS.[18] Put more accurately, with smaller $\bar{o}$, physicians are *allowed* to go home (at EOS) at an earlier point relative to peer arrival. Holding constant time from beginning of shift and time from peer arrival, smaller $\bar{o}$ thus induces a greater incentive for strategic behavior on the intensive margin.

## 6.2 Evidence of Tradeoff

I explore the interaction between patient-care effects and shift overlap. Larger patient-care effects with small $\bar{o}$, conditional on time from beginning of shift, are consistent with strategic behavior. Further, the interaction provides evidence of the intuitive tradeoff between extensive and intensive margins of strategic behavior: If physicians have more time to slack off before EOS, workload near EOS will be lower, and patient-care distortions will be smaller.

I consider three categories of overlap at EOS – terminal shifts ($\bar{o} = 0$), minimally transitioned shifts ($\bar{o} = 1$), and substantially transitioned shifts ($\bar{o} \geq 2$)[19] – and estimate

$$\begin{aligned} Y_{ijkpt} \quad = \quad & \sum_{m=-6}^{-1} \sum_{\overline{O}} \alpha_{ms} \mathbf{1} \left( \left\lfloor t_i - \bar{t}\,(j,t) \right\rfloor = m \right) \mathbf{1} \left( \bar{o}\,(j,t) \in \overline{O} \right) + \\ & \sum_{m} \gamma_m \mathbf{1} \left( \left\lfloor t - \underline{t}\,(j,t) \right\rfloor = m \right) + \mathbf{X}_{it}' \beta + \mathbf{T}_t' \eta + \zeta_p + \nu_{jk} + \varepsilon_{ijkpt}, \end{aligned} \tag{9}$$

similar to Equation (3) but interacting the hourly EOS effects by overlap $\bar{o}\,(j,t)$ in categories $\overline{O}$. I normalize coefficients so that, as before, the reference category includes times seven hours or greater prior to EOS in each of the overlap categories.

Figure 7 shows EOS effects, across the three categories of shift types, for length of stay, orders, admission, and total costs. The EOS effect on length of stay is largely similar among shift categories (Panel A). All three shift categories show a substantial decline in length of stay

[18] When $\bar{o} = 0$, workload is mostly decompressed by fewer assignments by the triage nurse, as shown in Figure 3, but note that it si also more difficult to hand off patients in these terminal shifts.

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

as EOS approaches. However, EOS effects are notably absent in shifts with $\bar{o} \geq 2$ for orders, admission probability, and total costs (Panels B to D). In contrast, shifts with $\bar{o} \leq 1$ show large increases in orders, admissions, and total costs at EOS.

Patient acceptance differs across these shift categories over several hours prior to EOS, while differences in EOS effects become apparent in the last hour prior to EOS. This is consistent with the idea that earlier patient assignment affects future workload and future distortion. In Table A-4, I consider effects on workload-adjusted length of stay, computed by Equation (4), and show that workload-adjusted length of stay decreases substantially in the last hour prior to EOS when $\bar{o} \leq 1$. In contrast, 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.[20]

# 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 consider the assignment of work as a sufficient statistic for a wide range of managerial policies (e.g., rules, financial incentives) including but not limited to shift structure and – using model of patient assignment, discharges, workload, and cost – assess the efficiency implications of a fuller range of counterfactual assignment policies.

The thought experiment is that, while a manager can control work assignment via policies, downstream effects on patient care cannot be directly monitored or managed. Patient assignment and discharges determine workload. Discharges and workload determine the effective time spent per patient. Based on Section 5.2, I consider workload-adjusted length of stay as the relevant measure of time in patient care that substitutes for formal utilization and raises patient-care costs (from formal utilization and admissions) when decreased. I then consider overall costs due to both physician time and patient care. Assigning more patients to physicians mechanically improves physician-time costs by reducing the number of physician-hours to process patient flow, but by increasing workload near EOS, it also worsens patient-care costs.

[20] Such potential increases in workload-adjusted length of stay above baseline do not appear to be associated with increases or decreases in other outcomes of orders, admissions, or costs. This could be consistent with *increases* in length of stay for strategic purposes, as discussed in Chan (2014).

I estimate discrete-time functions for patient assignment and discharge that crucially depend on time to EOS. Patient discharge follows a hazard model $\mathcal{D}\left(t, \sigma_s, w_{jt}, \hat{\tau}_{ist}\right)$ that depends on time $t$, shift characteristics $\sigma_s$ for shift $s$ (i.e., shift type $\langle \ell, \underline{o}, \overline{o} \rangle_s$ and time of EOS $\overline{t}\left(s\right)$), physician $j$'s workload $w_{jt}$ at $t$, and patient $i$'s predicted length of stay $\hat{\tau}_{ist}$ (details are given in Appendix A-2, including model fit, shown in Figure A-5). Assignment follows a zero-inflated Poisson model $\mathcal{A}\left(t, \sigma_s, w_{j,t-1}\right)$ that similarly depends on time $t$, shift characteristics $\sigma_s$, and workload $w_{j,t-1}$ from the previous period. The patient-assignment function allows a convenient specification of counterfactual policies in which assignment is modified only by how time to EOS is considered. For example, a counterfactual assignment policy may assign more patients one hour prior to EOS by assigning as if the time were three hours prior to EOS. Although patient assignment may be modified, physicians continue to discharge patients according to actual time to EOS.

Specifically, consider the assignment function $\mathcal{A}\left(t, \sigma_s, w_{j,t-1}\right)$, which assigns $N_{st}^r$ patients in shift $s$ during discrete time $t$ in simulation $r$, and define counterfactual assignment policies $\mathcal{A}_\Delta\left(t, \sigma_s, w_{j,t-1}\right) \equiv \mathcal{A}\left(\breve{t}\left(t, s, \Delta\right), \sigma_s, w_{j,t-1}\right)$ by an index $\Delta$ representing a time shift in assignment patterns near EOS: Starting at $|\Delta|$ hours before EOS, assignments are "curtailed" by using $\breve{t} > t$ if $\Delta < 0$, while they are "extended" by using $\breve{t} < t$ if $\Delta > 0$. Formally,

$$\breve{t}\left(t, s, \Delta\right) = \begin{cases} t + \kappa \max\left(1 + t - \overline{t}\left(s\right), \Delta\right), & \Delta < 0 \\ t, & \Delta = 0 \\ t + \max\left(0, \min\left(\Delta + t - \overline{t}\left(s\right), \Delta\right)\right), & \Delta > 0 \end{cases}, \tag{10}$$

where $\kappa = \max\left(0, \min\left(1, t - \overline{t}\left(s\right) - \Delta\right)\right) \in [0, 1]$ is a scale to ensure that $\breve{t}$ is continuous in $t$ when $\Delta < 0$. Figure 8 shows example counterfactual policies, for $\Delta \in \{-4, -2, 2, 4\}$.

For each counterfactual policy $\Delta \in [-4, 4]$, I simulate assignments and discharges using functions $\mathcal{A}_\Delta\left(t, \sigma_s, w_{j,t-1}\right)$ and $\mathcal{D}\left(t, \sigma_s, w_{jt}, \tilde{\tau}_{ist}\right)$, respectively, where $\tilde{\tau}_{ist}$ is a prediction of $\hat{\tau}_{ist}$. Using observations generated by each simulation $r = 1, \ldots, 100$ of each policy $\Delta$, I calculate counterfactual costs from both physician time and patient care:

$$\text{Costs}_\Delta^r = \text{PhysicianTimeCosts}_\Delta^r + \text{PatientCareCosts}_\Delta^r. \tag{11}$$

I consider physician-time costs as wasted "presenteeism" wage. Assuming a "no-waste" processing rate of 2.22 patients accepted per hour,[21] I calculate this as

$$\text{PhysicianTimeCosts}_{\Delta}^{r} = \sum_{s} \sum_{t=\underline{t}(s)}^{\bar{t}(s)} \text{Wage} \times \max\left(0, 1 - \frac{N_{st}^{\Delta,r}}{2.22\phi}\right), \tag{12}$$

where $\phi = 0.25$ converts hourly rates to arrivals within fifteen-minute intervals. This assumes that the ED may hire more physicians to generate throughput and thus also captures social time costs of patient care, including the cost of patients waiting for empty beds. I use Wage = \$100/hour (close to actual wages) in my base case but also consider Wage = \$500/hour.

I consider patient-care costs as increased costs due to strategic behavior, conditional on assignment. Costs per patient increase as physicians substitute time with formal utilization and are more likely to admit patients. Specifically, in simulation $r$ of assignment policy $\mathcal{A}_{\Delta}$, I measure decreases in workload-adjusted length of stay for that policy as coefficients $\alpha_m^{\Delta,r}$ in this regression:

$$\log \tilde{Y}_{ij} = \alpha + \sum_{m=-6}^{-1} \alpha_m^{\Delta,r} \mathbf{1}\left(\left\lfloor t_i - \bar{t}\left(s\left(j, t_i\right)\right)\right\rfloor = m\right) + \mathbf{g}\left(t_i - \underline{t}\left(s\left(j, t_i\right)\right)\right)' \gamma_g + \varepsilon_{ij}, \tag{13}$$

where $\tilde{Y}_{ij}$ is workload-adjusted length of stay, $t_i$ is the assignment time of patient $i$, and $\mathbf{g}(\cdot)$ creates is a vector of cubic splines of time from beginning of shift.[22] I calculate patient-care costs as

$$\begin{aligned} \text{PatientCareCosts}_{\Delta}^{r} \quad = \quad & \sum_{m} \exp\left(\text{BaseLogCosts} - 1.15 \cdot \min\left(0, \hat{\alpha}_m^{\Delta,r} - \hat{\alpha}_{-2}^{0}\right)\right) \times \\ & \sum_{s} \sum_{t=\underline{t}(s)}^{\bar{t}(s)} \mathbf{1}\left(\left\lfloor t - \bar{t}\left(s\right)\right\rfloor = m\right) N_{st}^{\Delta,r} \end{aligned} \tag{14}$$

[21] This number is the maximum average processing rate observed in an hour relative to EOS, and therefore this assumption is conservative in the sense that it yields the highest possible physician time cost. Processing rates are fairly stable at times sufficiently prior to EOS and peer arrival (see Figure 3).

[22] For notational simplicity, I assume one visit per patient and I drop superscripts $\Delta$ and $r$ on $\tilde{Y}_{ij}$, $t_i$, $\alpha$, $\gamma_g$, and $\varepsilon_{ij}$ (see Equation (A-9) for the fully notated version). I omit other controls, e.g., those in Equation (3), because I regress simulated data, in which many of the controls are meaningless (e.g., nurse-resident-physician identities, month-year interactions, and individual patient characteristics). Results using actual data are qualitatively robust to controlling for all of these and are shown in Table A-4. This is consistent with results in Section 5.3, showing negligible effects due to patient selection.

where BaseLogCosts $=$ [log \$ +] 6.750, and $\hat{\alpha}^{0}_{-2} \equiv \frac{1}{100}\sum_{r=1}^{100}\hat{\alpha}^{0,r}_{-2} = -0.059$. Based on actual data, I use $-1.15$ as the elasticity of patient-care costs to workload-adjusted time, which applies to decreases in workload-length of stay greater than 5.9% as a conservative benchmark.[23] Equation (14) reflects two sources of increasing patient-care costs as more patients are assigned near EOS: Per-patient costs increase (the first line), and this increase applies to more patients (the second line). I conservatively assume that patient health is not harmed even as physicians produce less information for the discharge decision, given no effects on mortality and bounce-backs in sample (Table 2).

Figure 9 shows average changes in daily costs under counterfactual policies $\Delta \in [-4, 4]$. I find that both curtailing and extending patient assignment increase overall costs relative to those under the actual assignment policy. While allowing physicians to slack off incurs the cost of wasted physician time, increasing patient-care costs dominate when extending patient assignment beyond the actual assignment policy. This is true even under the extreme assumption of a \$500 per hour physician wage. These results suggest that, given physician discretion in patient care, attempting to reduce slacking off could be very costly, and observed patterns of assignment may indeed be approximately second-best optimal.

# 8 Conclusion

I examine the behavior of ED physicians working in shifts and find evidence consistent with strategic behavior as a consequence of scheduled work: First, on an extensive margin, physicians are less likely to accept new patients near EOS. Second, on an intensive margin, physicians complete their work earlier as end of shift (EOS) approaches. As the input of time becomes more costly, 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. This increases patient-care costs by 21% in the last hour prior to EOS.

[23] The elasticity estimate is motivated by the fact that both observed total costs (Figure 4) and observed workload-adjusted length of stay (Figure A-6) increase only in the last hour prior to EOS. In simulated data, I find that 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-5). 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-2.3.

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. Due to informational problems, there remain substantial worker discretion and "client-worker specificity" in many job settings. Workers who are nominally on duty are thus subject to behavioral incentives that may distort their acceptance and performance of work. I show a tradeoff between these extensive and intensive margins of strategic behavior. In fact, observed patterns of "presenteeism" or "slacking off" may indeed be approximately second-best optimal. This implies that a wide set of policy levers (e.g., rules and financial incentives) that largely act on the extensive margin to prevent workers from sitting idly could be quite costly when used at the wrong time in scheduled work.

## References


**Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber**, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, February 2005, *113* (1), 151–184.

**Apker, Julie, Larry A. Mallak, and Scott C. Gibson**, "Communicating in the Gray Zone: Perceptions about Emergency Physician-hospitalist Handoffs and Patient Safety," *Academic Emergency Medicine*, October 2007, *14* (10), 884–894.

**Baker, George P.**, "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, June 1992, *100* (3), 598–614.

**Bandiera, Oriana, Iwan Barankay, and Imran Rasul**, "Social preferences and the response to incentives: Evidence from personnel data," *The Quarterly Journal of Economics*, 2005, *120* (3), 917–962.

**_ , _ , and _** , "Social connections and incentives in the workplace: Evidence from personnel data," *Econometrica*, 2009, *77* (4), 1047–1094.

**Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying**, "Does working from home work? Evidence from a Chinese experiment," *Quarterly Journal of Economics*, 2014, *Forthcoming.*

**Brachet, Tanguy, Guy David, and Andrea M Drechsler**, "The Effect of Shift Structure on Performance," *American Economic Journal: Applied Economics*, April 2012, *4* (2), 219–246.

**Briscoe, Forrest**, "Temporal Flexibility and Careers: The Role of Large-Scale Organizations for Physicians," *Industrial and Labor Relations Review*, 2006, *60*, 88.

_ , "From Iron Cage to Iron Shield? How Bureaucracy Enables Temporal Flexibility for Professional Service Workers," *Organization Science*, April 2007, *18* (2), 297–314.

**Casalino, Lawrence P., Kelly J. Devers, Timothy K. Lake, Marie Reed, and Jeffrey J. Stoddard**, "Benefits of and barriers to large medical group practice in the united states," *Archives of Internal Medicine*, September 2003, *163* (16), 1958–1964.

**Chan, David**, "Teamwork and Moral Hazard: Evidence from the Emergency Department," *Journal of Political Economy*, 2014, *Forthcoming.*

**Coviello, Decio, Andrea Ichino, and Nicola Persico**, "Time Allocation and Task Juggling," *The American Economic Review*, 2014, *104* (2), 609–623.

**Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan**, "Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India," *The Quarterly Journal of Economics*, November 2013, *128* (4), 1499–1545.

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

**Forster, Alan J., Ian Stiell, George Wells, Alexander J. Lee, and Carl Van Walraven**, "The Effect of Hospital Occupancy on Emergency Department Length of Stay and Patient Disposition," *Academic Emergency Medicine*, 2003, *10* (2), 127–133.

**Gold, Diane R., Suzanne Rogacz, Naomi Bock, Tor D. Tosteson, Timothy M. Baum, Frank E. Speizer, and Charles A. Czeisler**, "Rotating shift work, sleep, and accidents related to sleepiness in hospital nurses.," *American Journal of Public Health*, 1992, *82* (7), 1011–1014.

**Goldin, Claudia**, "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, April 2014, *104* (4), 1091–1119.

**Gordon, Nancy P., Paul D. Cleary, Claire E. Parker, and Charles A. Czeisler**, "The prevalence and health impact of shiftwork.," *American Journal of Public Health*, 1986, *76* (10), 1225–1228.

**Green, Linda**, "A Multiple Dispatch Queueing Model of Police Patrol Operations," *Management Science*, June 1984, *30* (6), 653–664.

**Green, Linda V.**, "Capacity Planning and Management in Hospitals," in Margaret L. Brandeau, FranÃ§ois Sainfort, and William P. Pierskalla, eds., *Operations Research and Health Care*, number 70. In 'International Series in Operations Research & Management Science.', Springer US, January 2004, pp. 15–41.

**Greenhouse, Steven**, "Part-Time Schedules, Full-Time Headaches," *The New York Times*, July 2014.

**He, Biyu, Franklin Dexter, Alex Macario, and Stefanos Zenios**, "The Timing of Staffing Decisions in Hospital Operating Rooms: Incorporating Workload Heterogeneity into the Newsvendor Problem," *Manufacturing and Service Operations Management*, January 2012, *14* (1), 99–114.

**Holmstrom, Bengt**, "Moral Hazard in Teams," *The Bell Journal of Economics*, 1982, pp. 324–340.

**Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi**, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 1997, *87* (3), 291–313.

**Jackson, C. Kirabo and Henry Schneider**, "Do Social Connections Reduce Moral Hazard? Evidence from the New York City Taxi Industry," *American Economic Journal: Applied Economics*, 2011.

**Jauhar, Sandeep**, "Busy Doctors, Wasteful Spending," *The New York Times*, July 2014.

**Kandel, Eugene and Edward P. Lazear**, "Peer Pressure and Partnerships," *The Journal of Political Economy*, 1992, pp. 801–817.

**Kantor, Jodi**, "Working Anything but 9 to 5: Scheduling Technology Leaves Low-Income Parents With Hours of Chaos," *The New York Times*, August 2014.

**Kc, Diwas S. and Christian Terwiesch**, "Impact of workload on service time and patient safety: an econometric analysis of hospital operations," *Management Science*, September 2009, *55* (9), 1486–1498.

**Kelly, Erin L., Phyllis Moen, and Eric Tranby**, "Changing Workplaces to Reduce Work-Family Conflict Schedule Control in a White-Collar Organization," *American Sociological Review*, April 2011, *76* (2), 265–290.

**Lambert, Diane**, "Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing," *Technometrics*, February 1992, *34* (1), 1–14.

**Larkin, Ian**, "The cost of high-powered incentives: Employee gaming in enterprise software sales," *Journal of Labor Economics*, April 2014, *32* (2), 199–227.

**Lazear, Edward P.**, *Personnel Economics*, MIT Press, 1995.

_ , "Performance pay and productivity," *The American Economic Review*, December 2000, *90* (5), 1346–1361.

**Lerman, Benjamin and Michael S. Kobernick**, "Return Visits to the Emergency Department," *The Journal of Emergency Medicine*, September 1987, *5* (5), 359–362.

**Maher, Kris**, "Wal-mart seeks flexibility in worker shifts," *The Wall Street Journal*, January 2007.

**Mas, Alexandre and Enrico Moretti**, "Peers at Work," *The American Economic Review*, 2009, *99* (1), 112–145.

**Messerli, Franz H., Adrian W. Messerli, and Thomas F. LÃŒscher**, "Eisenhower's Billion-Dollar Heart Attack – 50 Years Later," *New England Journal of Medicine*, September 2005, *353* (12), 1205–1207.

**Nagin, Daniel S., James B. Rebitzer, Seth Sanders, and Lowell J. Taylor**, "Monitoring, motivation, and management: the determinants of opportunistic behavior in a field experiment," *American Economic Review*, September 2002, *92* (4), 850–873.

**Nasca, Thomas J., Susan H. Day, and E. Stephen Amis**, "The new recommendations on duty hours from the ACGME Task Torce," *New England Journal of Medicine*, 2010, *363* (2), e3.

**Oyer, Paul**, "Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality," *The Quarterly Journal of Economics*, 1998, *113* (1), 149–185.

**Papke, Leslie E. and Jeffrey M. Wooldridge**, "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics*, 1996, *11* (6), 619–632.

**Perdikaki, Olga, Saravanan Kesavan, and Jayashankar M. Swaminathan**, "Effect of Traffic on Sales and Conversion Rates of Retail Stores," *Manufacturing and Service Operations Management*, January 2012, *14* (1), 145–162.

**Presser, Harriet B.**, *Working in a 24/7 Economy: Challenges for American Families.*, Russell Sage Foundation, 2003.

**Robinson, John and Geoffrey Godbey**, *Time for Life: The Surprising Ways Americans Use Their Time*, University Park, Pennsylvania: Penn State Press, November 2010.

**Schuur, Jeremiah D. and Arjun K. Venkatesh**, "The Growing Role of Emergency Departments in Hospital Admissions," *The New England Journal of Medicine*, 2012, *367*, 391–393.

**Shetty, Kanaka D. and Jayanta Bhattacharya**, "Changes in hospital mortality associated with residency work-hour regulations," *Annals of Internal Medicine*, July 2007, *147* (2), 73–80.

**Tanabe, Paula, Rick Gimbel, Paul R. Yarnold, Demetrios N. Kyriacou, and James G. Adams**, "Reliability and Validity of Scores on the Emergency Severity Index Version 3," *Academic Emergency Medicine*, 2004, *11* (1), 59–65.

**Volpp, K. G. and A. K. Rosen**, "Mortality among hospitalized Medicare beneficiaries in the first two years following ACGME resident duty hour reform," *The Journal of the American Medical Association*, September 2007, *298* (9), 975–983.

**Weitzman, Martin L.**, "Prices vs. Quantities," *The Review of Economic Studies*, October 1974, *41* (4), 477–491.

Table 1: End of Shift Effect on Log Length of Stay

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Log length of stay | | | | |
| Hour prior to EOS | | | | | |
| Last hour | -0.607*** | -0.547*** | -0.529*** | -0.716*** | -0.587*** |
| | (0.028) | (0.025) | (0.025) | (0.039) | (0.050) |
| Second hour | -0.316*** | -0.282*** | -0.330*** | -0.461*** | -0.287*** |
| | (0.008) | (0.008) | (0.008) | (0.012) | (0.026) |
| Third hour | -0.139*** | -0.129*** | -0.161*** | -0.260*** | -0.123*** |
| | (0.005) | (0.005) | (0.006) | (0.009) | (0.022) |
| Fourth hour | -0.112*** | -0.092*** | -0.111*** | -0.173*** | -0.091*** |
| | (0.005) | (0.004) | (0.005) | (0.008) | (0.018) |
| Fifth hour | -0.070*** | -0.055*** | -0.078*** | -0.120*** | -0.023 |
| | (0.004) | (0.004) | (0.005) | (0.007) | (0.015) |
| Sixth hour | -0.065*** | -0.048*** | -0.057*** | -0.090*** | -0.010 |
| | (0.004) | (0.004) | (0.005) | (0.007) | (0.012) |
| Patient characteristics | N | Y | Y | Y | Y |
| Time and pod dummies | N | N | Y | Y | Y |
| Physician-resident-nurse identities | N | N | N | Y | Y |
| Time relative to shift beginning | N | N | N | N | Y |
| Number of observations | 371,107 | 371,107 | 371,107 | 371,107 | 371,107 |
| Adjusted $R$-squared | 0.008 | 0.189 | 0.211 | 0.400 | 0.410 |
| Sample mean log length of stay (log hours) | 1.050 | 1.050 | 1.050 | 1.050 | 1.050 |

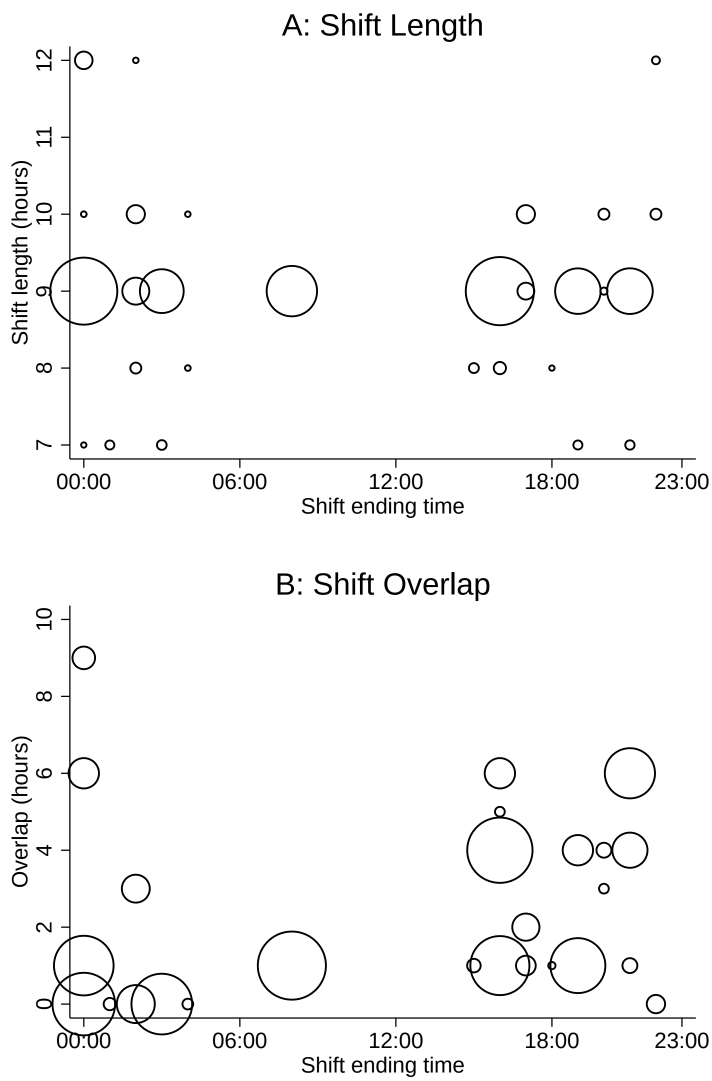
**Note:** This table reports coefficient estimates and standard errors in parentheses for versions of Equation (3) regressing log length of stay, with increasing controls, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Patient characteristics include demographics, emergency severity index (ESI), time spent in triage, and rich indicators for clinical diagnoses (e.g., Elixhauser indices). Time dummies include indicators for hour of day, day of week, and month-year interactions. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level.

Table 2: 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 LOS |
| Hour prior to EOS | | | | | | |
| Last hour | 1.411** | 0.057** | 0.208** | -0.003 | -0.028 | -0.144*** |
| | (0.562) | (0.024) | (0.080) | (0.008) | (0.018) | (0.051) |
| Second hour | -0.093 | 0.000 | 0.027 | -0.001 | -0.011 | 0.015 |
| | (0.302) | (0.013) | (0.043) | (0.004) | (0.010) | (0.027) |
| Third hour | -0.003 | 0.002 | 0.009 | -0.005 | -0.005 | 0.090*** |
| | (0.249) | (0.011) | (0.036) | (0.004) | (0.008) | (0.022) |
| Fourth hour | 0.167 | 0.004 | 0.029 | -0.001 | -0.002 | 0.036* |
| | (0.207) | (0.009) | (0.030) | (0.003) | (0.007) | (0.018) |
| Fifth hour | 0.239 | -0.004 | 0.034 | -0.002 | 0.001 | 0.037** |
| | (0.171) | (0.007) | (0.024) | (0.003) | (0.005) | (0.015) |
| Sixth hour | 0.192 | -0.007 | -0.006 | 0.001 | 0.001 | 0.001 |
| | (0.137) | (0.006) | (0.019) | (0.002) | (0.004) | (0.012) |
| Number of observations | 371,421 | 371,421 | 366,219 | 371,421 | 371,421 | 371, 148 |
| Adjusted $R$-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 |

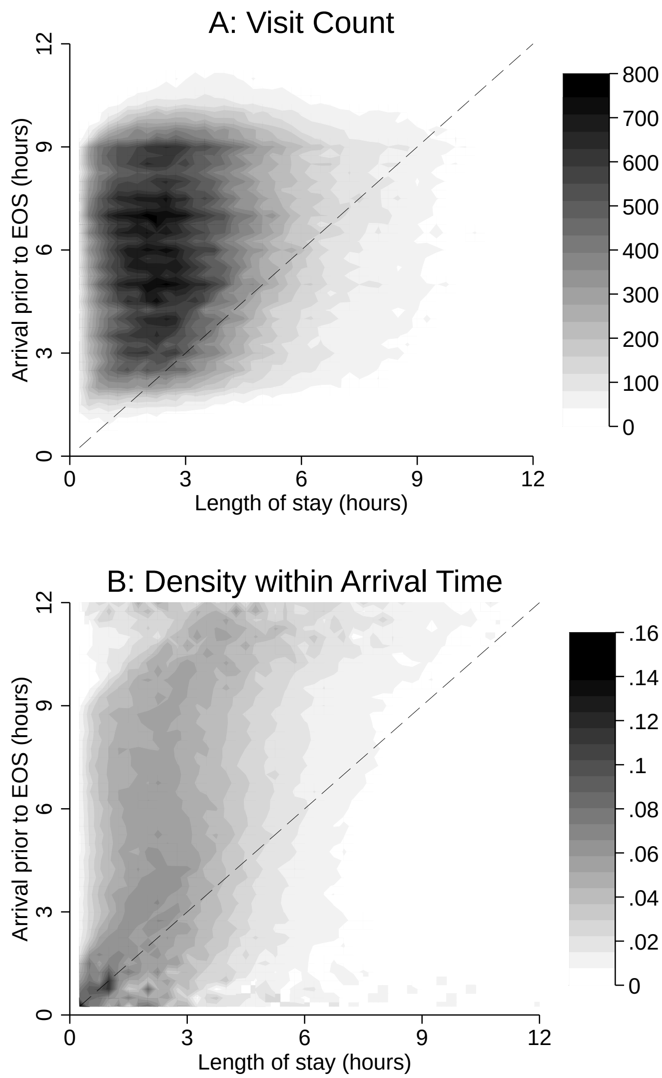
**Note:** This table reports coefficient estimates and standard errors in parentheses for Equation (3) with a full set of controls regressing other outcome variables, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Workload-adjusted length of stay (LOS) is calculated by Equation (4). Controls are as described for Table 1. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level.

Figure 1: Shift Variation



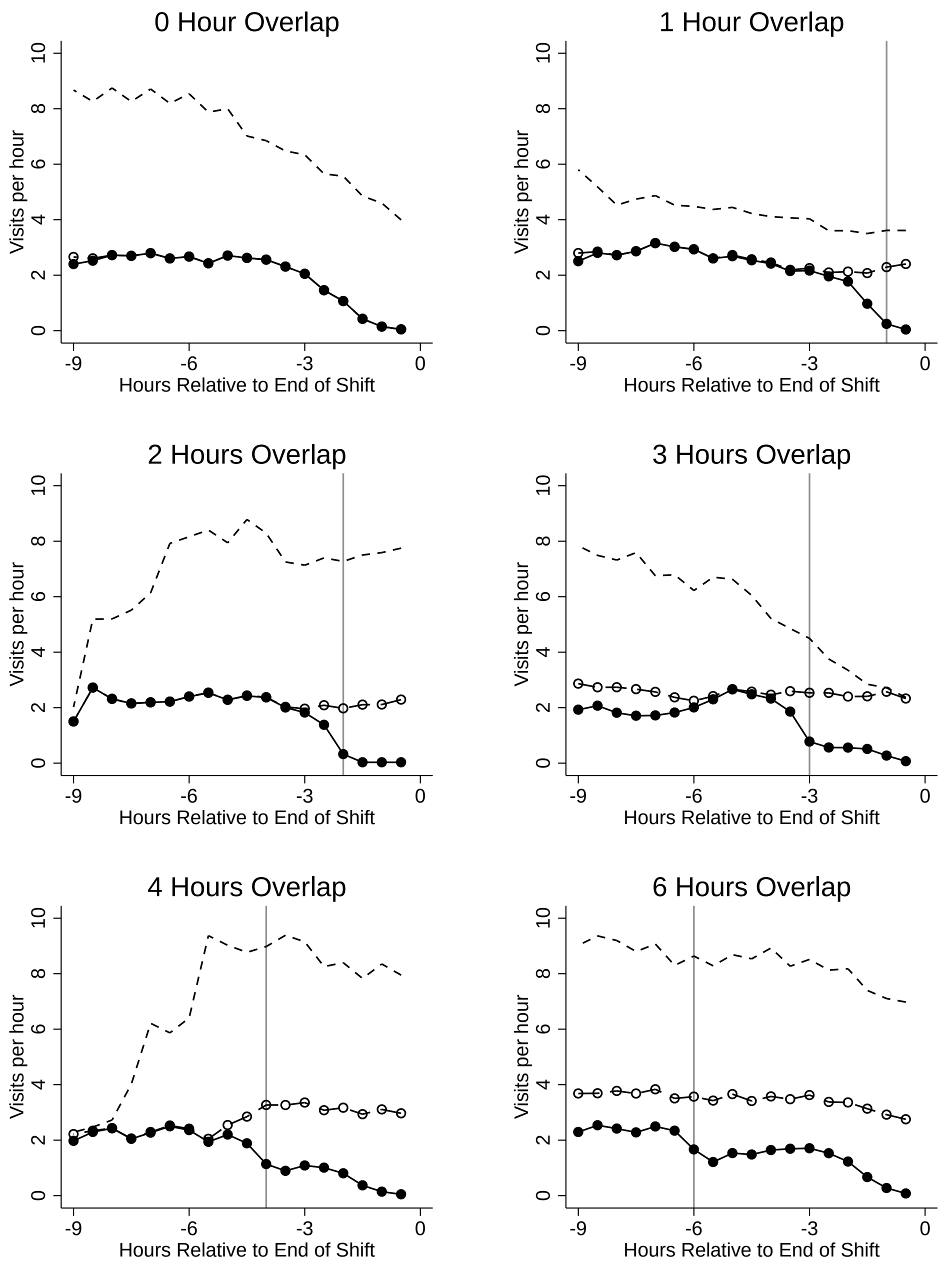

**Note:** This figure illustrates the variation in observations across shift types. Panel A plots shifts by shift ending time and shift length. Panel B plots shifts by shift ending time and the length of overlapping transition at the end of shift.

Figure 2: Density of Visits on Arrival Time and Length of Stay



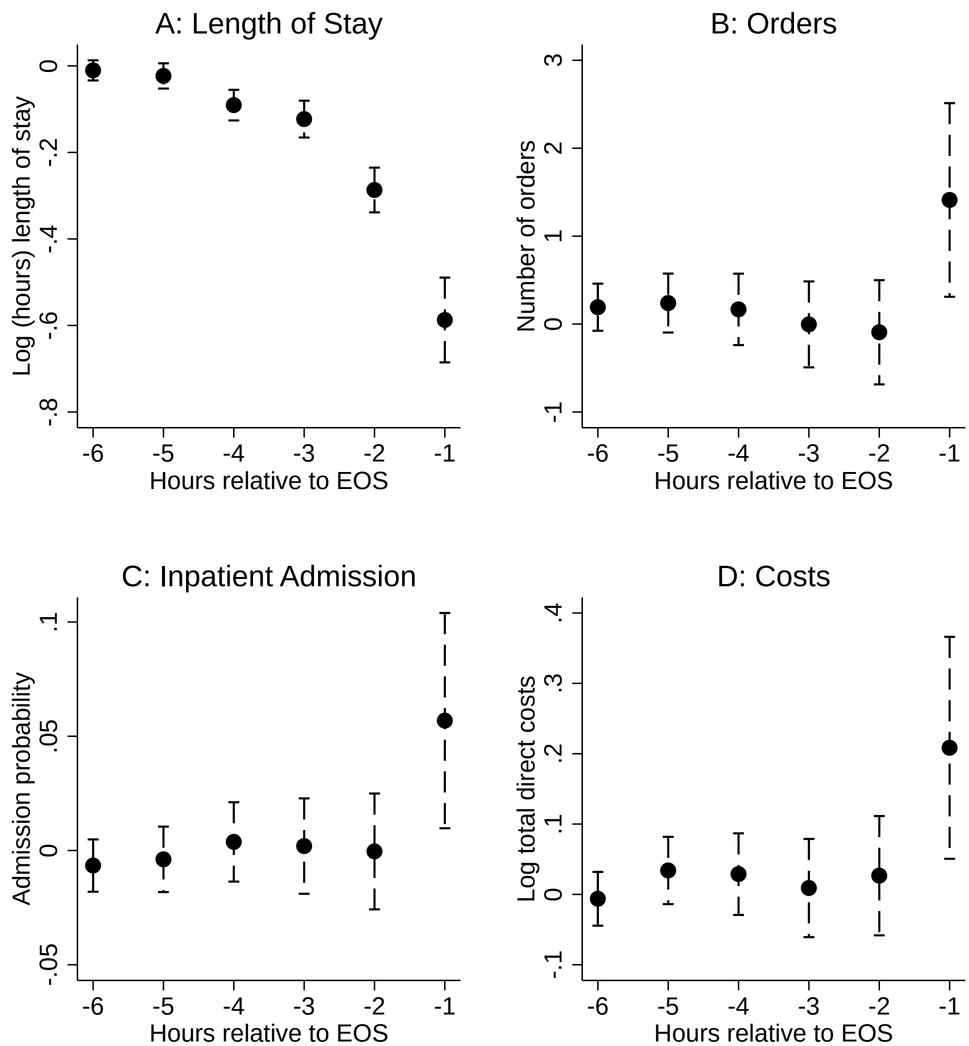

**Note:** This figure plots the distribution of visits over arrival times relative to EOS and length of stay. Panel A plots visit counts within fifteen-minute intervals of arrival time and length of stay. Panel B plots the density of visits, conditional on arrival time.

Figure 3: Flow of Patient Visits over Time



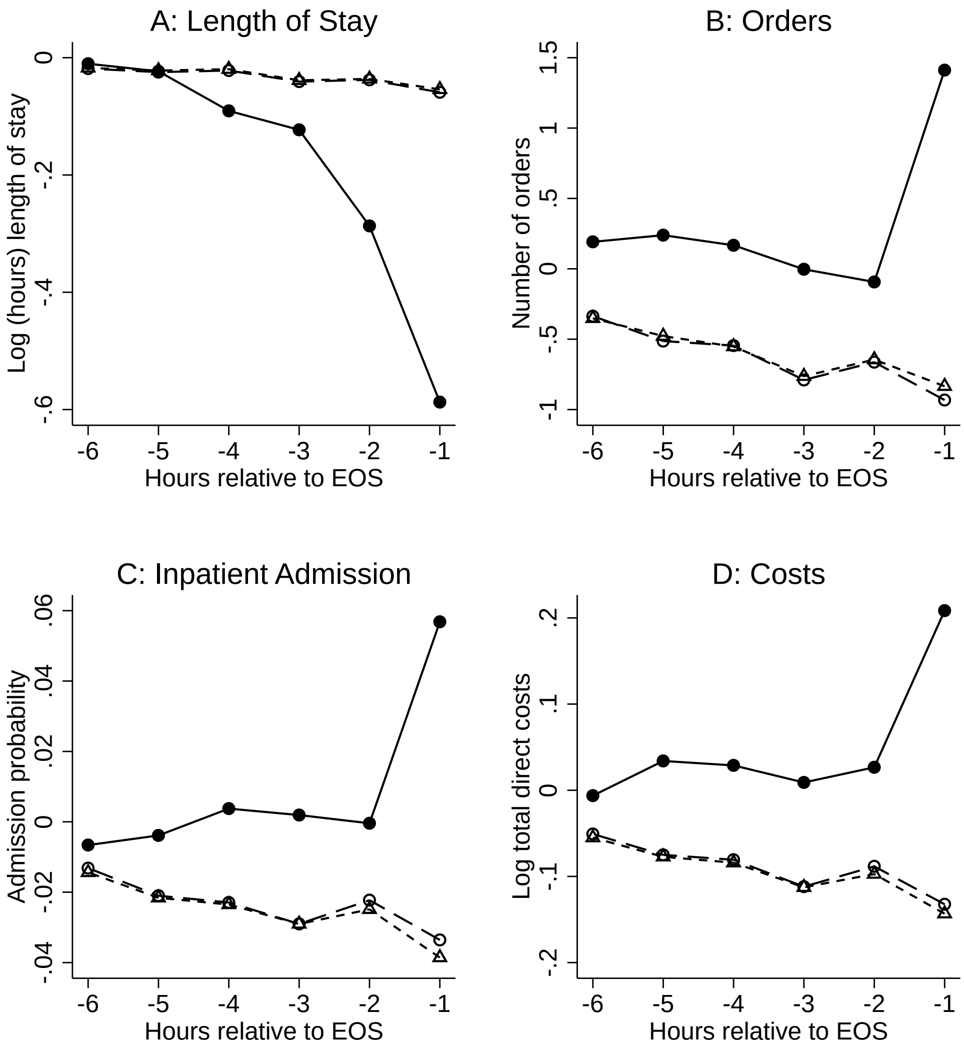

**Note:** This figure shows unadjusted average hourly rates of patient visits for each 30-minute interval relative to end of shift (EOS). Each panel shows results for shifts with a given EOS overlap time. Patient visits for the index physician are shown in closed circles; patient visits for the location are shown in open circles; and patient visits for the entire ED are shown with a dashed line with no markers. Subsequent shift starting times are marked with a vertical line.

Figure 4: End of Shift Effects



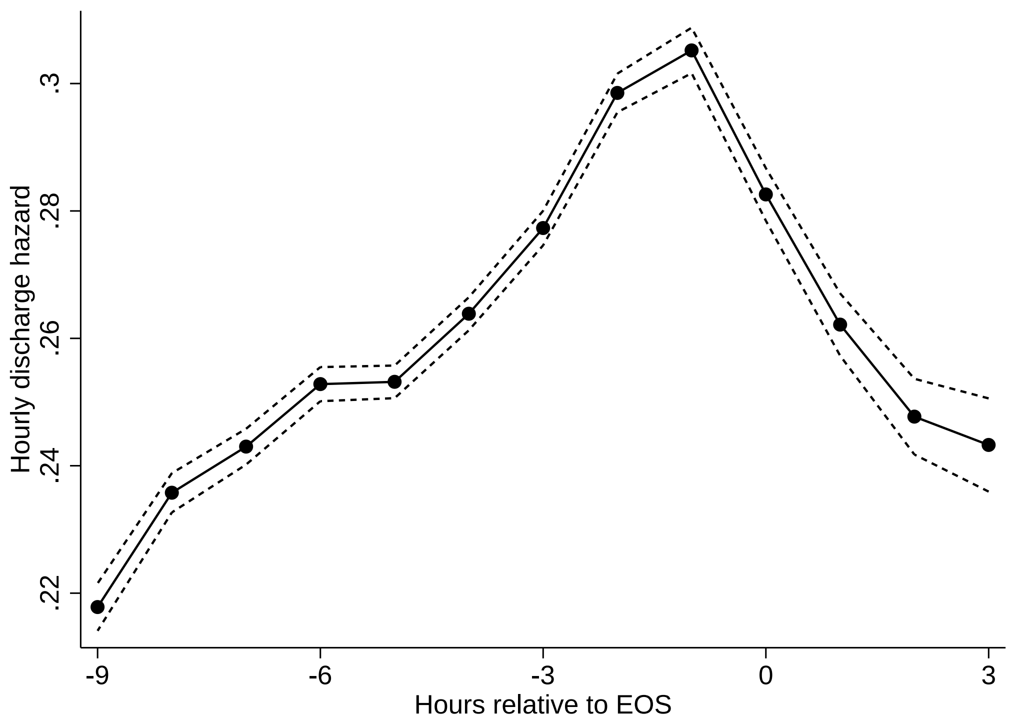

**Note:** This figure plots average effects for each hour prior to end of shift (EOS) on length of stay (Panel A), orders (Panel B), inpatient admissions (Panel C), and costs (Panel D). Each outcome is estimated separately using Equation (5). The reference category is any time greater than six hours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

Figure 5: Patient Selection Relative to End of Shift



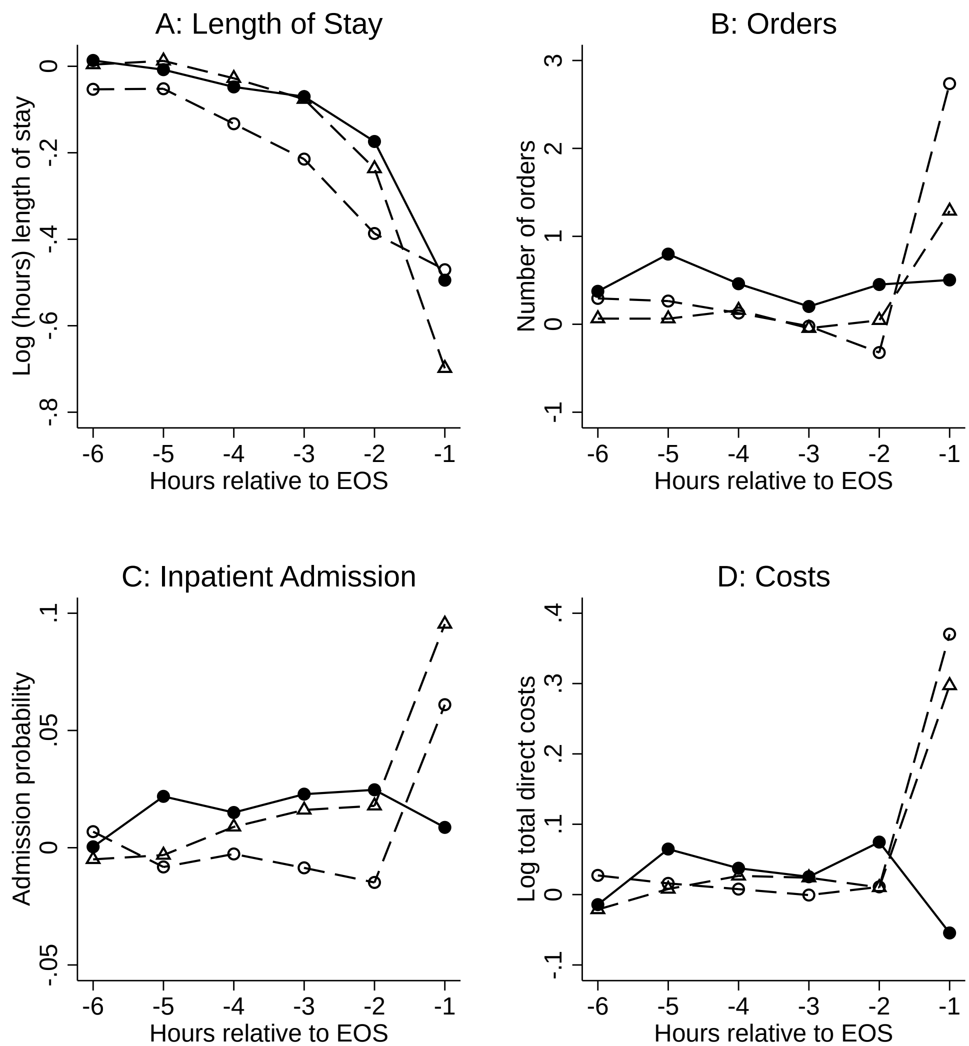

**Note:** This figure shows selection on observables for each hour prior to end of shift (EOS) on length of stay (Panel A), orders (Panel B), inpatient admissions (Panel C), and costs (Panel D). Each outcome is predicted based on patient characteristics only observable prior to treatment (open circles) and on the full set of characteristics, including endogenous diagnoses (open triangles). Coefficients are estimated for predicted outcome using Equation (7). For reference, adjusted effects on actual outcomes from Figure 4 are shown in closed circles. The reference category is any time greater than six hours prior to EOS.

Figure 6: Discharge Hazard over Time



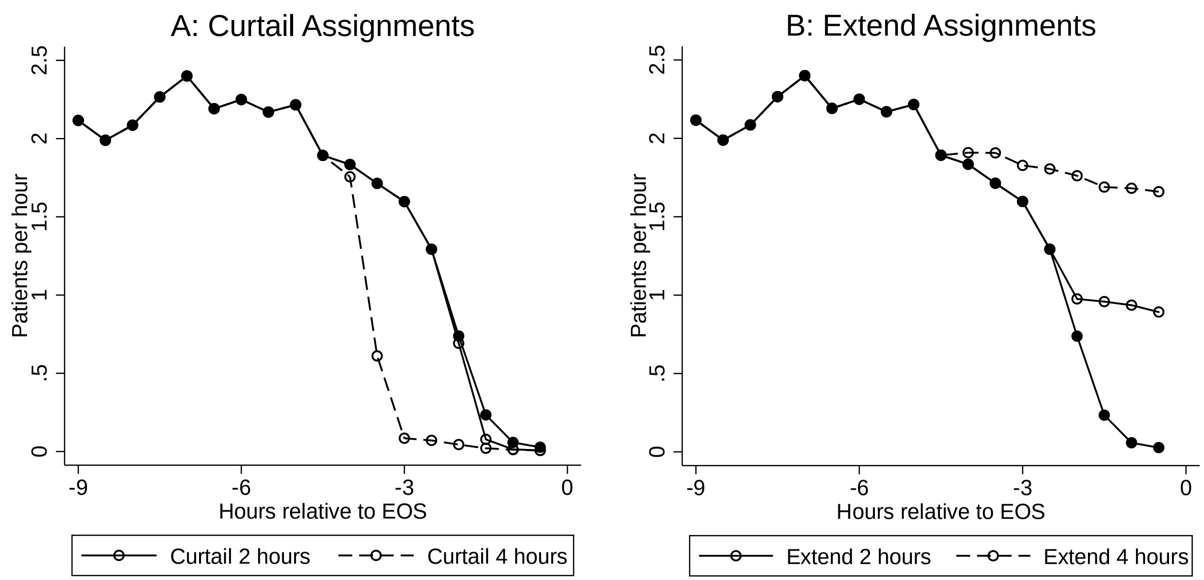

**Note:** This figure plots marginal effects for a discrete-time logit hazard model of discharge for each hour relative to end of shift (EOS), as given in Equation (8). Marginal effects are shown in closed circles; 95% confidence intervals are shown in dashed lines.

Figure 7: End of Shift Effects by Shift Overlap



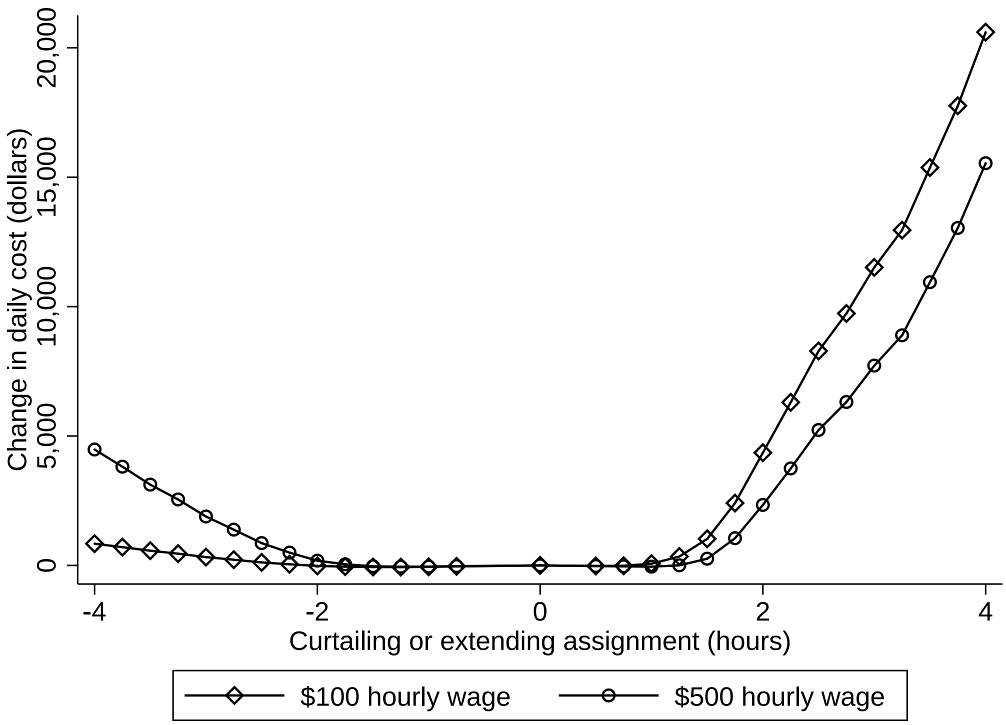

**Note:** This figure shows heterogeneous end of shift (EOS) effects by EOS overlap times on length of stay (Panel A), orders (Panel B), inpatient admissions (Panel C), and costs (Panel D). Each outcome is estimated separately using Equation (9). Estimates for terminal shifts ($\overline{o} = 0$) are shown in open triangles; estimates for minimally transitioned shifts ($\overline{o} = 1$) are shown in open circles; and estimates for substantially transitioned shifts ($\overline{o} \geq 2$) are shown in closed circles.

Figure 8: Example Counterfactual Assignment Regimes




**Note:** This figure shows example counterfactual assignment regimes, parameterized as hours $\Delta$ curtailing or extending assignment, as specified by Equation (10). Panel A shows counterfactual regimes in which assignment is curtailed earlier than actual assignment patterns. Panel B shows counterfactual regimes in which assignment is extended beyond actual assignment patterns. Two and four hours denote the times at which curtailment or extension begins.

Figure 9: Change in Daily Cost over Counterfactual Regimes




**Note:** This figure plots changes in daily cost relative to actual daily cost averaged over 100 simulations for each counterfactual assignment regime. Assignment regimes may either curtail or extend assignment, as illustrated in Figure 8. Hours of curtailment or extension are represented as negative or positive numbers, respectively, and "0" represents the actual assignment regime. Daily costs include both costs spent on physician time and patient care. Changes in daily cost are plotted under the assumption of a $100 hourly wage (diamonds) and $500 hourly wage (circles).

# Appendix

## A-1 Selection on Unobservables

This appendix provides details of how I implement a procedure similar to that outlined in Altonji et al. (2005) to quantify the amount of selection on unobservables necessary to explain decreases in length of stay for patients accepted 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-1.1 Conceptual Framework

Consider a condensed form of the outcomes regression Equation (3):

$$\begin{aligned} Y &= \sum_m \alpha_m A^m + \mathbf{\Omega}'\mathbf{\Gamma} \\ &= \sum_m \alpha_m A^m + \mathbf{W}'\mathbf{\Gamma}_W + \xi, \end{aligned} \tag{A-1}$$

where I omit subscripts for simplicity. $A^m \equiv \mathbf{1}\left(\lfloor t - \bar{t}(j,t) \rfloor = m\right)$ is an abbreviation for the familiar indicator for whether the time $t$ that patient $i$ was assigned to physician $j$ was in the $m^{\text{th}}$ hour from $j$'s EOS. $\alpha_m$ is the causal effect of a patient being assigned in the $m^{\text{th}}$ hour prior to EOS. $\mathbf{W}$ 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 (3), which only includes patient characteristics). $\mathbf{\Gamma}$ is the causal effect of $\mathbf{W}$ on $Y$. $\mathbf{\Gamma}_X$ is the subvector of $\mathbf{\Gamma}$ that corresponds to $\mathbf{W}$ within $\mathbf{\Omega}$, and $\xi$ is an index of the unobserved variables.

Since variables in $\tilde{\mathbf{X}}$ are likely correlated with $\xi$, rewrite Equation (A-1) as

$$Y = \sum_m \alpha_m A^m + \mathbf{W}'\gamma_W + \varepsilon, \tag{A-2}$$

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

### A-1.2 Measure of Selection on Unobservables

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

$$\frac{E[\varepsilon | A^m = 1] - E[\varepsilon | A^m = 0]}{\text{Var}(\varepsilon)} = \frac{E[\mathbf{W}'\gamma_W | A^m = 1] - E[\mathbf{W}'\gamma_W | A^m = 0]}{\text{Var}(\mathbf{W}'\gamma_W)}, \tag{A-3}$$

which states that the relationship between the index of unobservables in Equation (A-2) and the indicator for selection $A^m$ is equal in magnitude to the relationship between unobservable predictors of $Y$ and $A^m$, respectively normalizing for variance.

They argue that this condition represents an upper bound because of observed variables are not randomly collected but rather were collected because they represent important characteristics for outcomes of interest. Furthermore, and more importantly, 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 I observe a rich set of patient characteristics that are either determined by the physician after accepting the patient or are rarely observable by the physician at the time of acceptance.

### A-1.3 Estimation of Potential Bias

In order to estimate the potential bias at the upper bound implied by Equation (A-3), consider the following linear selection equation:

$$A^m = \mathbf{W}'\beta_W + \tilde{A}^m, \tag{A-4}$$

where $\tilde{A}^m$ is a residual that is orthogonal to $\mathbf{W}$. Then Equation (A-2) can be stated as

$$Y = \sum_m \alpha_m \tilde{A}^m + \mathbf{W}'\left(\gamma_W + \alpha\beta_W\right) + \varepsilon.$$

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

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

From Equation (A-3), the bias can be stated in terms of $E\left[\mathbf{W}'\gamma_W\left|A^m = 1\right.\right] - E\left[\mathbf{W}'\gamma_W\left|A^m = 0\right.\right]$:

$$\text{Bias} = \frac{\operatorname{Var}\left(A^m\right)\operatorname{Var}\left(\varepsilon\right)}{\operatorname{Var}\left(\tilde{A}^m\right)\operatorname{Var}\left(\mathbf{W}'\gamma_W\right)}\left(E\left[\mathbf{W}'\gamma_W\left|A^m = 1\right.\right] - E\left[\mathbf{W}'\gamma_W\left|A^m = 0\right.\right]\right) \tag{A-5}$$

Under the null hypothesis that $\alpha_m = 0$, $\gamma_W$ can be consistently estimated by Equation (A-1).

I can then arrive at a consistent estimate of bias in Equation (A-5) with the following procedure, with results shown in Table A-3: For each $m \in \{-6, \ldots, -1\}$, I define $A^m \equiv \mathbf{1}\left(\left\lfloor t - \bar{t}\left(j,t\right)\right\rfloor = m\right)$ over all observations and empirically calculate $\widehat{\operatorname{Var}}\left(A^m\right)$ for each $m$. I also calculate $\widehat{\operatorname{Var}}\left(\tilde{A}^m\right)$ after estimating Equation (A-4) for each $m$. Similarly, I estimate $\widehat{\operatorname{Var}}\left(\varepsilon\right) = 0.160$ and $\widehat{\operatorname{Var}}\left(\mathbf{W}'\gamma_W\right) = 0.580$ from Equation (A-2). Equation (A-2) also allows me to form an estimate of selection on observables, $\hat{E}\left[\mathbf{W}'\gamma_W \left| A^m = 1\right.\right] - \hat{E}\left[\mathbf{W}'\gamma_W \left| A^m = 0\right.\right]$. Using the Altonji et al. (2005) condition in Equation (A-3) 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-5). As shown in Table A-3, the upper bound of the bias in $\hat{\alpha}_{-1}$, the effect of arriving in the last hour of shift on the length of stay, estimated by Equation (3), 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 an ratio 3.55.

## A-2 Structural Model Implementation

This appendix details the procedure to simulate observations under counterfactual assignment policies and impute overall costs under these policies, as discussed more briefly in Section 7. To summarize, I first estimate arrival and discharge functions in discrete time. Actual arrivals and discharges imply patient censuses, or the number of patients under the care of a physician at each point in discrete time, i.e., those who have arrived and have not yet been discharged. I create counterfactual assignment policies by modifying time in the arrival function, and I simulate of patient observations (including arrival times, discharge times, and patient censuses) using this modified arrival function and the (unmodified) discharge function. Finally, I impute of overall costs by regressing simulated workload-adjusted length of stay on time relative to EOS and translating decreases in workload-adjusted length of stay into increases in patient-care costs. Overall costs are the sum of costs due to patient care and physician time.

### A-2.1 Estimating Arrival and Discharge Functions

In each fifteen-minute time interval of each of the 23,990 shifts during the study period ranging from June 2005 to December 2012, I calculate the number of patients assigned to the physician on shift $s$ during time interval $t$. Restricting to $t \in \left[\underline{t}(s) - 3 \text{ hours}, \overline{t}(s)\right]$ yields 1,151,888 observations over $t$ and $s$ with patient arrival (or "assignment") numbers

$$N_{st} \equiv \sum_i \mathbf{1}\left(t_i^a = t\right) \mathbf{1}\left(s\left(\mathcal{J}(i), t_i^a\right) = s\right),$$

where $t_i^a$ is the arrival time of patient $i$, $s(j,t)$ is a shift assignment function for physician $j$ and time $t$, and $\mathcal{J}(i)$ is a physician assignment function, assuming for notational convenience that each patient has only one visit.

Of $\sum_{s,t} N_{st} = 370,843$ 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 $i \in S^{\text{outside}}$, are therefore not modeled in either arrivals or discharges, but I count them toward workload defined below. As I describe in Section A-2.2, I take arrivals and discharges of patients $i \in S^{\text{outside}}$ as fixed in every simulation.

For $N_{st}$, I estimate a zero-inflated Poisson model. I call this an arrival or assignment function $\mathcal{A}(t, \sigma_s, w_{j,t-1})$, which depends on $t$, shift characteristics $\sigma_s$ of $s$ (i.e., shift type $\langle \ell, \underline{o}, \overline{o} \rangle_s$ and time of EOS $\bar{t}(s)$), and physician $j$'s census (or workload) $w_{j,t-1}$ in the previous period (for $j$ satisfying $s(j,t) = s$). $w_{jt}$ is defined in Equation (5), which I slightly rephrase here as

$$w_{jt} \equiv \sum_i \mathbf{1}(t_i^a \geq t) \, \mathbf{1}\left(t \leq t_i^d\right) \mathbf{1}(j = \mathcal{J}(i)), \tag{A-6}$$

where $t_i^d$ is the corresponding discharge order time for patient $i$, again taking advantage of the notational assumption that $i$ refers to a unique patient visit.

As introduced by Lambert (1992), the zero-inflated Poisson model allows for two "regimes," one which always yields $N_{st} = 0$, and the second which follows a Poisson process leading to $N_{st} \geq 0$. In particular,

$$\begin{aligned} \Pr(N_{st} = 0) &= \Pr(\text{Regime 1}) + \Pr(N_{st} = 0 | \text{Regime 2}) \Pr(\text{Regime 2}); \\ \Pr(N_{st} = n) &= \Pr(N_{st} = n | \text{Regime 2}) \Pr(\text{Regime 2}), \; n = 1, 2, \ldots, \end{aligned}$$

where $\Pr(\text{Regime 1}) + \Pr(\text{Regime 2}) = 1$. $\Pr(\text{Regime 1} | t, \sigma_s)$ is specified as the logistic function of a linear combination of shift-type indicators interacted with splines of $\bar{t}_s - t$. $\Pr(N_{st} = n | \text{Regime 2})$ is specified as a Poisson model:

$$\Pr(N_{st} = n | \text{Regime 2}) = \frac{\exp(-\lambda_{st}) \lambda_{st}^n}{n!}, \; n = 0, 1, 2, \ldots,$$

where $\log \lambda_{st} = \log E[N_{st} | t, \sigma_s, w_{j,t-1}]$ is a linear combination of splines of time of the day, splines of month-year interactions, indicators for day of the week, splines of $\bar{t}_s - t$, splines of $w_{j,t-1}$, interactions between splines of $\bar{t}_s - t$ and splines of $w_{j,t-1}$, and interactions between shift-type indicators interacted with splines of $\bar{t}_s - t$.

For discharge times $t_i^d$, I estimate a logit hazard model, similar to but more richly specified than that estimated by Equation (8) in Section 5.5, as

$$\Pr\left(t_i^d = t \middle| t_i^d \geq t - 1\right) \equiv \mathcal{D}(t, \sigma_s, w_{jt}, \hat{\tau}_{ist}) = \frac{1}{1 + \exp(-h(t, \sigma_s, w_{jt}, \hat{\tau}_{ist}))},$$

where $h(t, \sigma_s, w_{jt}, \hat{\tau}_{ist})$ is estimated, separately for shifts of different length $\ell$, as a linear combination of indicators for hour of the day of $t_i^a$, splines of $t - t_i^a$, splines of $\bar{t}(s) - t$, splines of $\bar{t}(s) - t_i^a$, splines of $w_{jt}$, predicted log length of stay ($\hat{\tau}_{ist}$), interactions between splines of $t - t_i^a$ and splines of $\bar{t}(s) - t_i^a$, interactions between splines of $\bar{t}(s) - t$ and splines of $w_{jt}$, and interactions between $\hat{\tau}_{ist}$ and splines of $\bar{t}(s) - t_i^a$.

Predicted log length of stay ($\hat{\tau}_{ist}$) is a linear combination of indicators for day of the week, month-year interactions, patient age and squared age, sex, ESI indicators, Elixhauser indicators, race indicators, language indicators, pod indicators, and physician-nurse-resident joint indicators. As a separate model, I estimate a model $\tilde{\tau}_{ist}$ of $\hat{\tau}_{ist}$ based on day of the week, month-year interactions, pod indicators, shift-type indicators interacted with splines of $\bar{t}(s) - t$, and physician indicators. The reason for the first prediction ($\hat{\tau}_{ist}$ of $\tau_{ijkpt}$) is to condense a large number of characteristics about the patient visit into a single linear score for ease of estimation. The reason for the second prediction ($\tilde{\tau}_{ist}$ of $\hat{\tau}_{ist}$) is to allow simulation of predicted log length of stay without simulating many of the characteristics that enter into $\hat{\tau}_{ist}$, such as nurse and resident identities, which I discuss in the next section.

### A-2.2 Simulating Patient Observations

After estimating arrival and discharge functions from actual data, I simulate patient arrivals and discharges under counterfactual assignment policies. I create the counterfactual policies, $\mathcal{A}_\Delta(t, \sigma_s, w_{j,t-1}) \equiv \mathcal{A}\left(\breve{t}(t, s, \Delta), \sigma_s, w_{j,t-1}\right)$, by modifying patient assignment indexed by a time shift $\Delta$, so that patients are assigned as if time were $\breve{t}(t, s, \Delta)$ rather than $t$. Starting at $|\Delta|$ hours before EOS, assignments are "curtailed" assignments with $\breve{t} > t$ if $\Delta < 0$ or "extended" with $\breve{t} < t$ if $\Delta > 0$ (Equation (10) formally defines $t(t, s, \Delta)$, and Figure 8 shows example counterfactual policies). While assignments are explicitly modified in these policies, the underlying discharge function is fixed, and discharge behavior only changes as $w_{jt}$ increases or decreases near EOS, with $\Delta > 0$ or $\Delta < 0$, respectively.

Specifically, I follow this procedure for each simulation $r$ of counterfactual policy $\Delta$:

1. Start $t$ at three hours before the beginning of each shift $s$. Set $w_{j,t-1}^{\Delta,r} = 0$.
2. Determine new assignments at $t$ for each $s$.
   (a) Simulate $N_{st}^{\Delta,r}$ new assignments for $s$ at $t$, using $\mathcal{A}_\Delta\left(t, \sigma_s, w_{j,t-1}^{\Delta,r}\right)$. Denote each of these new assignments with an unused $i \notin S^{\text{outside}}$, note that $t_{\Delta,r,i}^{a} = t$, and simulate predicted log length of stay $\tilde{\tau}_{ist}^{\Delta,r}$.
   (b) Assign patients $i \in S^{\text{outside}}$ where $t_i^{\text{outside},a} = t$ to the relevant shifts $s$.
3. Calculate workload $w_{jt}^{\Delta,r}$ by Equation (A-6).
4. If $t \geq \underline{t}(s)$ and $w_{jt}^{\Delta,r} > 0$, determine discharges at $t$ for each $s$.
   (a) Simulate $d_{it}^{\Delta,r} \equiv \mathbf{1}\left(t_{\Delta,r,i}^{d} = t\right)$ for each $i \notin S^{\text{outside}}$ where $d_{i,t-1}^{\Delta,r} = 0$, using $\mathcal{D}\left(t, \sigma_s, w_{jt}, \tilde{\tau}_{ist}^{\Delta,r}\right)$.
   (b) Discharge patients $i \in S^{\text{outside}}$ where $t_i^{\text{outside},d} = t$ from the relevant shifts $s$.
5. The procedure is complete for $s$ such that $t \geq \bar{t}(s)$ and $w_{jt}^{\Delta,r} = 0$. For the remaining $s$, revise $t = t + 1$ and return to Step #2.

The resulting collection of $t^a_{\Delta,r,i}$, $t^d_{\Delta,r,i}$, and $w^{\Delta,r}_{jt}$ form the underlying simulated data. 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. Slightly adapting Equation (4) to discrete time,

$$\begin{aligned} \tilde{Y}^{\Delta,r}_{ij} &\equiv \tau^{\Delta,r}_i / \overline{w}^{\Delta,r}_{ij}, \\ &= 0.25 \cdot \max\left(t^d_{\Delta,r,i} - t^a_{\Delta,r,i}, 0.3\right) \left[ \frac{1}{t^d_{\Delta,r,i} - t^a_{\Delta,r,i} + 1} \sum_{\tilde{t}=t^a_{\Delta,r,i}}^{t^d_{\Delta,r,i}} w^{\Delta,r}_{j\tilde{t}} \right]^{-1}. \end{aligned} \tag{A-7}$$

The term $0.25 \cdot \max\left(t^d_{\Delta,r,i} - t^a_{\Delta,r,i}, 0.3\right)$ reflects that, in actual data, $\hat{E}[\tau_i] \approx 0.075$ hours if $t^d_i = t^a_i$, but otherwise $\hat{E}[\tau_i] \approx 0.25\left(t^d_i - t^a_i\right)$ hours if $t^d_i > t^a_i$ (recall that $t$ is in fifteen-minute intervals).

### A-2.3 Imputing Costs

Having simulated arrival and discharge data, I am now in the position to impute overall costs for each counterfactual simulation $r$ of $\Delta$. Overall costs include both physician-time costs and patient-care costs. Repeating Equation (11):

$$\text{Costs}^r_\Delta = \text{PhysicianTimeCosts}^r_\Delta + \text{PatientCareCosts}^r_\Delta. \tag{A-8}$$

The first cost, physician-time costs, represents wasted physician-hours as physicians "slack off" near EOS. Patients they do not accept must be seen by another physician, and the ED must allocate physician-hours accordingly, either choosing to hire more physicians to meet the demand or to allow patients to wait while patient assignments slow down. Thus, assuming efficiently priced wages, the physician wage represents a true opportunity cost for the ED or social planner. I calculate physician-time costs as

$$\text{PhysicianTimeCosts}^r_\Delta = \sum_s \sum_{t=\underline{t}(s)}^{\bar{t}(s)} \text{Wage} \times \max\left(0, 1 - \frac{N^{\Delta,r}_{st}}{2.22 \times 0.25}\right),$$

or simply the wasted "presenteeism" wage incurred when physicians are assigned anything less than the maximum throughput rate of 2.22 patients per hour, again multiplied by 0.25 to correct for the fact that I measure time in fifteen-minutes intervals. Since 2.22 is the maximum observed throughput rate in the data, the assumption that anything less than this is conservative, in the sense that it overestimates physician-time costs. I use Wage = \$100/hour (close to actual wages) in my base case but also consider the extreme case of Wage = \$500/hour.

The second cost in Equation (A-8), patient-care costs, represents costs incurred by the physician, conditional on assignment and time to EOS. Physicians near EOS substitute time spent in patient care with formal utilization and are more likely to admit patients. As shown in Section 5.2 and Table 2, 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 $\Delta$, I estimate the EOS effect on workload-adjusted length of stay by coefficients $\hat{\alpha}_m^{\Delta,r}$ in this regression:

$$\begin{aligned}\log \tilde{Y}_{ij}^{\Delta,r} \quad = \quad & \alpha^{\Delta,r} + \sum_{m=-6}^{-1} \alpha_m^{\Delta,r} \mathbf{1}\left(\left\lfloor t_{\Delta,r,i}^a - \bar{t}\left(s\left(j, t_{\Delta,r,i}^a\right)\right)\right\rfloor = m\right) + \\ & \mathbf{g}\left(t_{\Delta,r,i}^a - \underline{t}\left(s\left(j, t_i\right)\right)\right)' \gamma_g^{\Delta,r} + \varepsilon_{ij}^{\Delta,r},\end{aligned} \tag{A-9}$$

where $\tilde{Y}_{ij}^{\Delta,r}$ is simulated workload-adjusted length of stay from Equation (A-7), and $\mathbf{g}(\cdot)$ creates a vector of cubic splines of assignment time relative to shift beginning.

In simulated data with $\Delta = 0$, I estimate $\hat{\alpha}_{-1}^0 \equiv \frac{1}{100}\sum_{r=1}^{100} \hat{\alpha}_{-1}^{0,r} = -0.240$ and $\hat{\alpha}_{-2}^0 \equiv \frac{1}{100}\sum_{r=1}^{100} \hat{\alpha}_{-2}^{0,r} = -0.059$, 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-5). Given that total costs increase by 20.8% in the last hour prior to EOS, I estimate the elasticity of patient-care 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 patient-care costs as

$$\begin{aligned}\text{PatientCareCosts}_\Delta^r \quad = \quad & \sum_m \exp\left(\text{BaseLogCosts} - 1.15 \cdot \min\left(0, \hat{\alpha}_m^{\Delta,r} - \hat{\alpha}_{-2}^0\right)\right) \times \\ & \sum_s \sum_{t=\underline{t}(s)}^{\bar{t}(s)} \mathbf{1}\left(\left\lfloor t - \bar{t}(s)\right\rfloor = m\right) N_{st}^{\Delta,r}\end{aligned}$$

where BaseLogCosts = [log \$ +] 6.750. Note patient-care costs increase with greater assignments (higher $\Delta$) 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 2).

## A-3 Additional Results

In this appendix, I present the following additional empirical results, as well as a brief discussion of some of these results:

- Table A-1 describes the process of constructing the sample, including the number of observations in each step.

- Table A-2 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-4 reports coefficients for EOS effects on workload-adjusted length of stay. Workload-adjusted length of stay is calculated by Equation (4). Coefficients are estimated on actual data using regressions that adjust for the full set of covariates, as described in Table 1, for example. This table is continued by Table A-5.

- Table A-5 reports coefficients for EOS effects on workload-adjusted length of stay, as a continuation of Table A-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-1 shows example weekly pod schedules.

- Figure A-2 shows EOS effects on time components of length of stay, according to time before first order, time between first and last (non-discharge) order, and time between last order and discharge.

- Figure A-3 shows EOS effects on length of stay for shifts of different lengths, presented both as time relative to EOS and time relative to shift beginning.

- Figure A-4 shows average patient counts ("censuses") for physicians in shifts with different overlap $\bar{o}$.

- Figure A-5 shows the fit between actual and simulated data, where the simulated assignments, length of stay, and censuses are calculated by discrete-time functions of patient assignment and discharge.

- Figure A-6 shows coefficients for EOS effects on workload-adjusted length of stay, reported in Table A-5, estimated on both actual and simulated data.

Tables A-4 and A-5 both report coefficients for EOS effects on workload-adjusted length of stay. As shown in Tables 1 and 2, length of stay continuously decreases as time approaches EOS, but orders, admissions, and costs only increase in the last hour prior to EOS. In Section 5.2, I show that adjusting length of stay for average workload reconciles these different patterns, in that workload-adjusted length of stay only decreases from baseline in the last hour of shift, and thus represents a more accurate measure of the input of time in patient care. Table A-4 further documents that workload-adjusted length of stay decreases with little overlap ($\bar{o} \leq 1$) but shows no decrease relative to baseline with more substantial overlap ($\bar{o} \geq 2$), which is consistent with the fact that orders, admissions, and costs do not change near EOS for $\bar{o} \geq 2$ (Figure 7).

In Table A-5, I present evidence qualitatively consistent with results in Table A-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-5 to operationalize workload-adjusted length of stay as the key substitute for patient-care 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, patient-care costs increase via increased formal utilization and admission likelihood. Tables A-4 and A-5 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, 2014); in the structural model, I therefore assume that increases in workload-adjusted length of stay do not change patient-care costs.

Figure A-2 examines EOS effects on time components defined by patient arrival, physician first order, physician last (non-discharge) order, and patient discharge. As explained in Section 5.2, the purpose of this exercise is to consider components of length of stay that are conceptually substitutes ($\tau_1$) vs. complements ($\tau_2$) with respect to formal utilization. Time spent before the first order is likely to be part of $\tau_1$, whereas time after the last order is likely be part of $\tau_2$. Time between the first and last orders is a mix of the two but is likely to more closely represent $\tau_1$. Results in Figure A-2 are consistent with a decreasing proportion of time in $\tau_1$ and an increasing proportion of time in $\tau_2$, which is consistent with net substitution (increasing formal utilization) in the last hour of shift.

Figure A-3 depicts EOS effects on length of stay in shifts of different lengths, where coefficients are arranged both by time relative to EOS (Panel A) and by time relative to shift beginning (Panel B). As explained in Section 5.4, this illustrates that changes in length of stay with time relative to shift beginning are negligible compared to EOS effects. Of course, essentially all regressions (e.g., Equation (3)) in the paper control for time relative to shift beginning.

Figure A-5 shows the fit between actual data and simulated data from the structural model described in Sections 7 and A-2. Assignments (Panel A) are simulated according to a zero-inflated Poisson model, and length of stay (Panel B) results from discharges simulated according to a logit hazard model, both of which consider time in fifteen-minute discrete intervals. Censuses (Panel C) – or workload as measured by numbers of patients under a physician's care (between assignment and discharge) – are simulated by both the assignment and discharge models. Workload-adjusted length of stay, calculated by Equation (4), is derived from discharges and censuses. Figure A-6 shows the fit in regression coefficients (also reported in Table A-5) of workload-adjusted length of stay between actual and simulated data.

Table A-1: 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-2: Shift Type Observation Numbers

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

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

Table A-3: Potential Bias from Selection on Unobservables

| | $\widehat{\text{Var}}\left(A^m\right)$ | $\widehat{\text{Var}}\left(\tilde{A}^m\right)$ | Selection on observables | Bias upper bound | $\hat{\alpha}_m$ | $\hat{\alpha}_m$ as bias multiple |
|---|---|---|---|---|---|---|
| Patient selection into hour prior to EOS $(A^m)$ | | | | | | |
| 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. Selection is modeled for whether a patient is assigned in the $m^{\text{th}}$ hour prior to EOS $(A^m)$ by Equation (A-4), the residual of which is $\tilde{A}^m$. Selection on observables is defined as $\hat{E}\left[\mathbf{W}'\gamma_W \,|A^m = 1\right] - \hat{E}\left[\mathbf{W}'\gamma_W \,|A^m = 0\right]$, where $\gamma_W$ is estimated from Equation (A-2). Using the condition from Altonji et al. (2005), in Equation (A-3), 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-5). I use $\widehat{\text{Var}}\left(\varepsilon\right) = 0.160$ and $\widehat{\text{Var}}\left(\mathbf{W}'\gamma_W\right) = 0.580$ in this calculation. $\hat{\alpha}_m$ is estimated by Equation (3); for convenience, results are repeated from the last column of Table 1. Finally $\hat{\alpha}_m$ is stated as a multiple of the bias upper bound in the last column of this table.

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

| | (1) | | (2) | (3) |
|---|---|---|---|---|
| | $\bar{o} \leq 1$ | $\bar{o} \geq 2$ | $\bar{o} \leq 1$ | $\bar{o} \geq 2$ |
| Hour prior to EOS | | | | |
| Last hour | -0.167** | -0.034 | -0.229*** | -0.003 |
| | (0.068) | (0.110) | (0.069) | (0.143) |
| Second hour | 0.015 | 0.126* | 0.014 | 0.140 |
| | (0.038) | (0.067) | (0.039) | (0.085) |
| Third hour | 0.05 | 0.137** | 0.037 | 0.099 |
| | (0.031) | (0.056) | (0.033) | (0.069) |
| Fourth hour | 0.007 | 0.088* | -0.002 | 0.056 |
| | (0.025) | (0.051) | (0.026) | (0.058) |
| Fifth hour | 0.013 | 0.085* | 0.009 | 0.052 |
| | (0.021) | (0.044) | (0.022) | (0.047) |
| Sixth hour | -0.017 | 0.044 | -0.022 | 0.029 |
| | (0.015) | (0.038) | (0.016) | (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 |

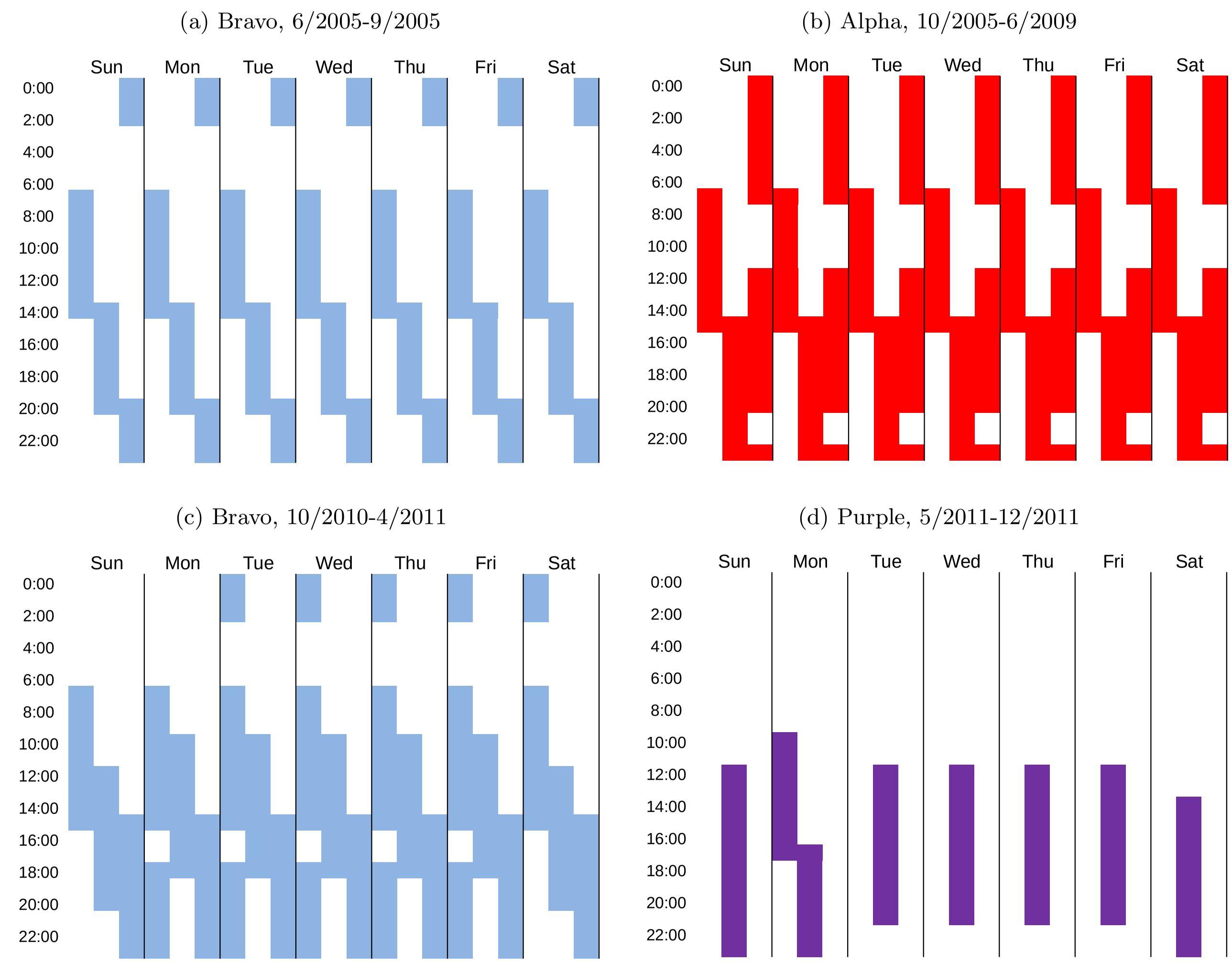
**Note:** This table reports coefficient estimates and standard errors in parentheses for EOS effects on workload-adjusted length of stay, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Model (1) is estimated by Equation (9), while models (2) and (3) are estimated separately by Equation (3) on subsamples of the data according to $\bar{o}$. All three models are estimted with a full set of controls, as described for Table 1. Workload-adjusted length of stay is calculated by Equation (4). * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. This table is continued by Table A-5.

Table A-5: Effect on Workload-adjusted Length of Stay by Shift Overlap

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | All $\bar{o}$ | $\bar{o} \leq 1$ | $\bar{o} \geq 2$ | All $\bar{o}$ | $\bar{o} \leq 1$ | $\bar{o} \geq 2$ |
| Hour prior to EOS | | | | | | |
| Last hour | -0.232*** | -0.339*** | 0.031 | -0.200*** | -0.240*** | -0.053 |
| | (0.037) | (0.05) | (0.061) | (0.037) | (0.049) | (0.063) |
| Second hour | -0.069*** | -0.089*** | 0.168*** | -0.067*** | -0.064** | 0.101*** |
| | (0.019) | (0.025) | (0.035) | (0.019) | (0.025) | (0.036) |
| Third hour | -0.016 | -0.027 | 0.176*** | -0.013 | -0.007 | 0.122*** |
| | (0.015) | (0.020) | (0.028) | (0.015) | (0.020) | (0.029) |
| Fourth hour | -0.077*** | -0.076*** | 0.087*** | -0.052*** | -0.044** | 0.075*** |
| | (0.014) | (0.017) | (0.026) | (0.014) | (0.017) | (0.027) |
| Fifth hour | -0.052*** | -0.048*** | 0.046** | -0.037*** | -0.029** | 0.031 |
| | (0.012) | (0.014) | (0.022) | (0.012) | (0.014) | (0.022) |
| Sixth hour | -0.032*** | -0.028*** | 0.013 | -0.028*** | -0.021** | 0.002 |
| | (0.008) | (0.010) | (0.014) | (0.008) | (0.010) | (0.015) |
| Control for time relative to shift beginning | Y | Y | Y | Y | Y | Y |
| Patient, provider, and other time controls | N | N | N | N | N | N |
| Sample | Full, actual | $\bar{o} \leq 1$, actual | $\bar{o} \geq 2$, actual | Full, simulated | $\bar{o} \leq 1$, simulated | $\bar{o} \geq 2$, simulated |
| Number of observations | 334,955 | 231,576 | 101,657 | 334,783 | 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 A-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 (13), 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 (4). * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. Results are graphically shown in Figure A-6.

Figure A-1: Example Weekly Pod Schedules




(a) Bravo, 6/2005-9/2005

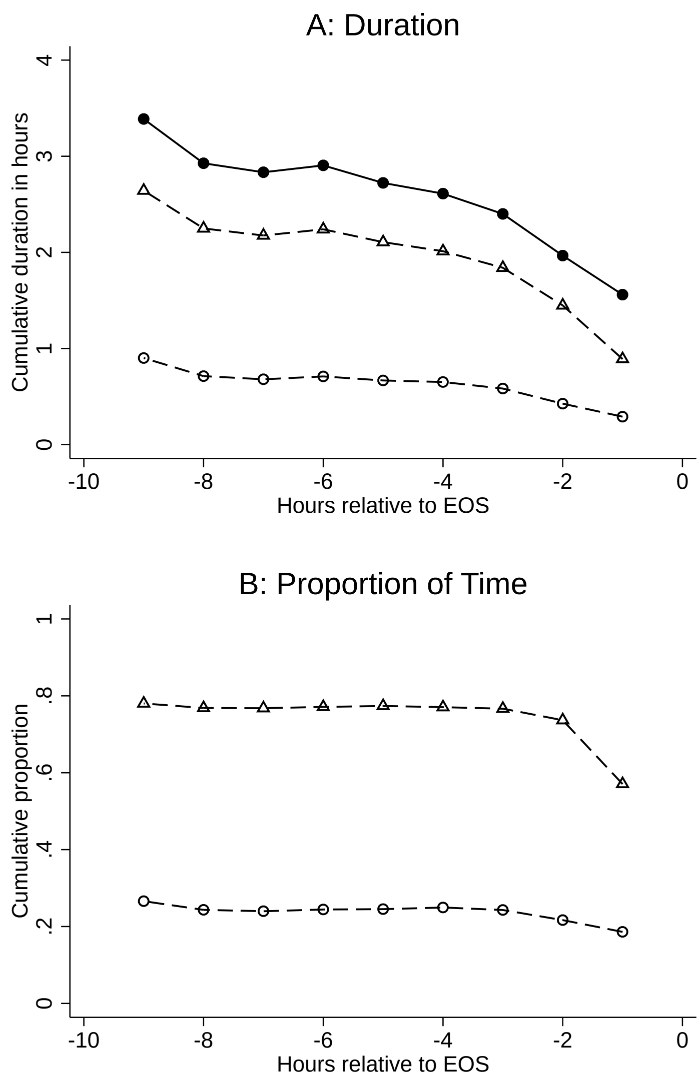
(b) Alpha, 10/2005-6/2009

(c) Bravo, 10/2010-4/2011

(d) Purple, 5/2011-12/2011

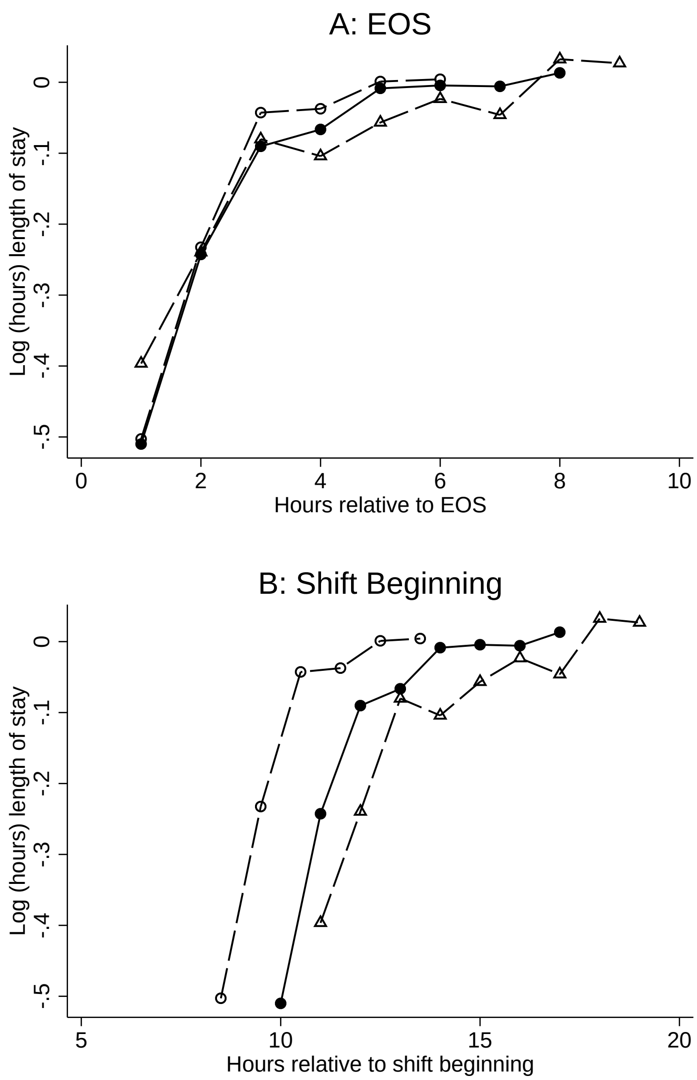
**Note:** Filled areas in vertical lines represent hours scheduled for a shift for a single physician. Hours when there is more than one physician present are represented by horizontally adjacent filled areas.

Figure A-2: Time Components



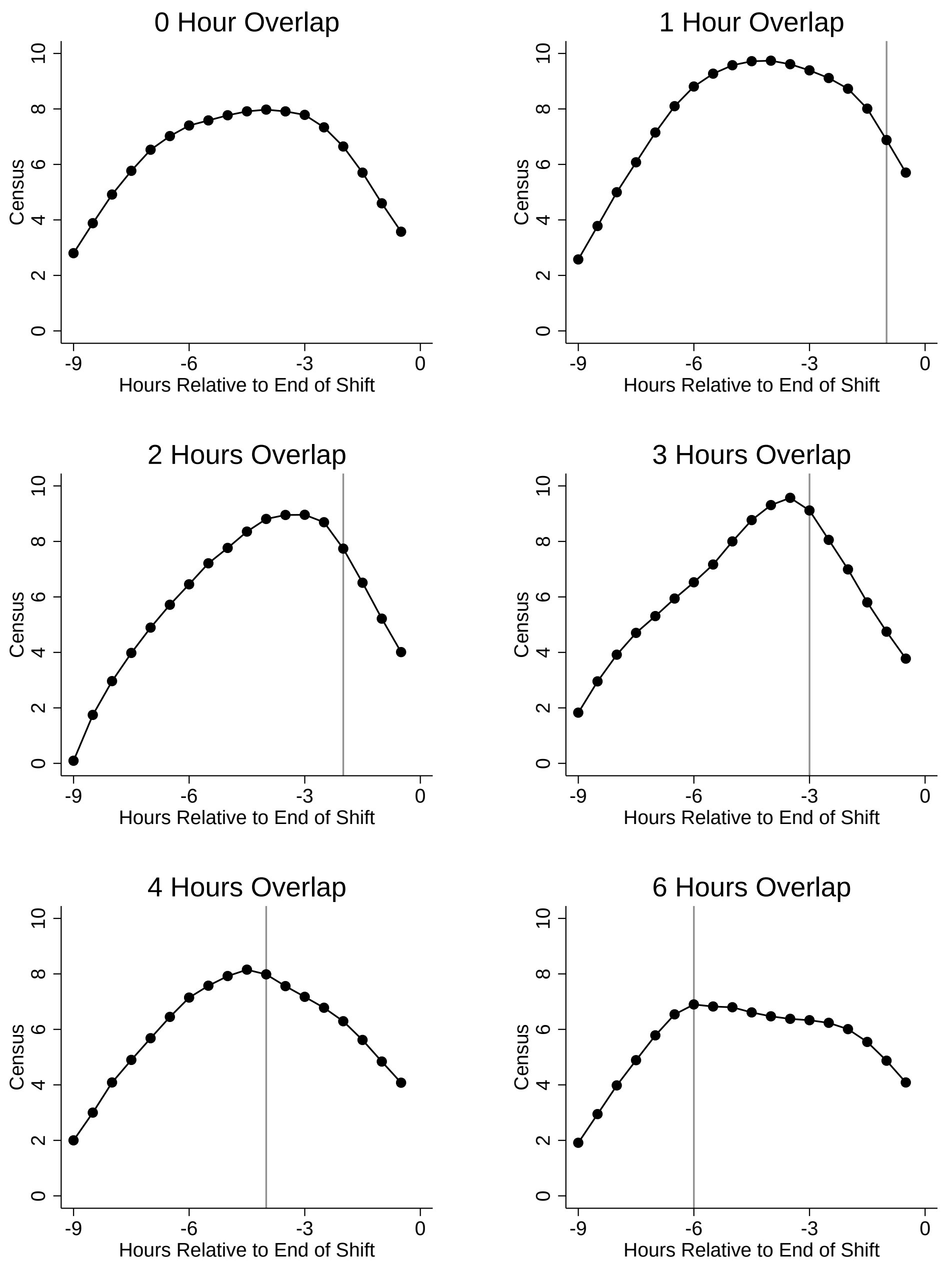

**Note:** This figure plots time components of length of stay as a function of hours relative to end of shift (EOS): time from pod arrival to first order (open circles), time from first to last (non-discharge) order (open triangles), and time from last order to discharge order (closed circles). Panel 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-3: Effects on Length of Stay by Shift Length



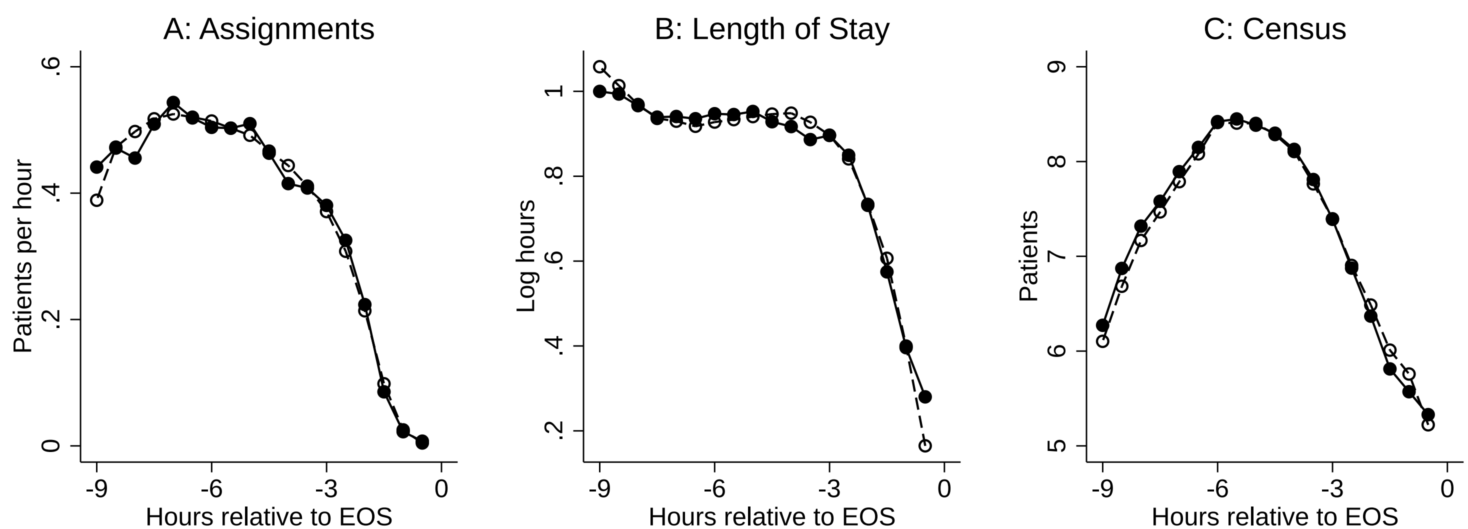

**Note:** This figure shows coefficients from Equation (3) estimated separately for shifts of seven or eight hours in length (open circles), nine hours in length (closed circles), and ten hours in length (open triangles). Panel A arranges estimates by hours relative to end of shift (EOS). Panel B arranges estimates by hours relative to shift beginning.

Figure A-4: Censuses over Time



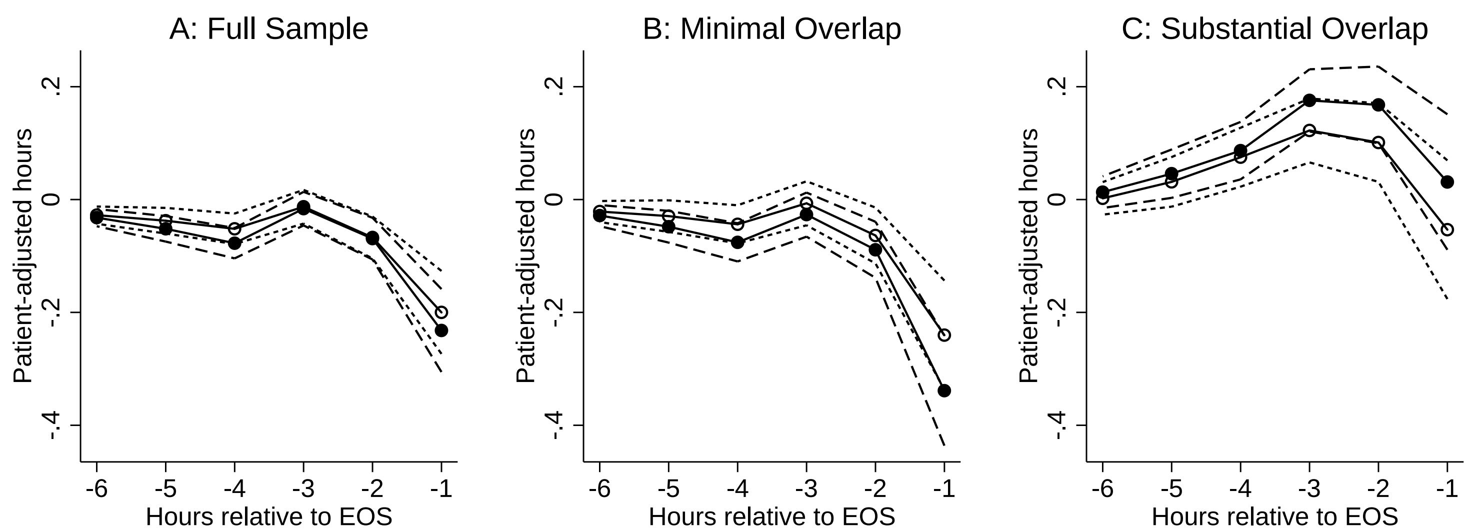

**Note:** This figure plots average censuses over time relative to the end of shift (EOS). Each panel shows results for physicians in shifts with a given EOS overlap time. Subsequent shift starting times are marked with a vertical line.

Figure A-5: Model Fit




**Note:** This figure shows the fit of the structural model described in Section 7 in terms of patient assignments (Panel A), length of stay (Panel B), and census (Panel C), averaged in each 30-minute interval relative to end of shift (EOS). Average actual data are shown in solid circles; average simulated data are shown in hollow circles.

Figure A-6: Workload-adjusted Length of Stay




**Note:** This figure shows coefficients for regressions, described in Equation (13), 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-5.