Clocking Out: Shift Work in the Emergency Department

David C. Chan*

July 21, 2014

Abstract

Work schedules are an increasingly prevalent mechanism of assigning work as production becomes more time-sensitive and utilizes labor more interchangeably. I examine strategic behavior that results from work schedules in the emergency department (ED). I find two types of moral hazard, consistent with privately increasing time costs past end of shift (EOS). First, on an extensive margin, physicians accept fewer patients near EOS. Second, on an intensive margin, physicians both spend less time on patients who arrive near EOS and incur higher costs, ordering more diagnostic tests and admitting patients more often for inpatient care, suggesting a production function that substitutes formal spending for physicians’ time. The intensive margin is worsened in shifts, those with limited overlap with an oncoming physician’s shift, that hamper physicians’ ability to reduce their workload near EOS. This suggests that it may be second-best optimal to allow workers to avoid some work near EOS when the pressure for moral hazard is greatest.

*I am thankful to Amitabh Chandra, Daniel Chen, David Cutler, Joe Doyle, Liran Einav, Bob Gibbons, Jon Gruber, Rob Huckman, Dan Kessler, Grant Miller, Paul Oyer, Jonathan Skinner, and Chris Walters for helpful comments and suggestions. I gratefully 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.
1 Introduction

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


Assigning uncertain future work is a ubiquitous problem across workplaces. Increasingly, as production becomes more time-sensitive and utilizes labor more interchangeably, this problem is addressed in two steps: First, schedules specify which workers are available at a given time. Second, work is distributed based on which workers are available. While an expanding operations literature has studied the importance of scheduling on efficiency, this paper makes the observation that schedules may distort worker behavior for a simple reason: Schedules define boundaries of worker availability, but the nature of work often is neither sharply nor predictably delimited in time. If workers have a net private value for their time, then discretion in the assignment and subsequent performance of work will lead to moral hazard.

I consider this problem of workplace design in the case of emergency department (ED) shift work. Shifts in the ED are meant to represent a minimum quantity of scheduled hours, in the sense that physicians working in shifts may go home after end of shift (EOS) if their work is complete, although they generally require more time to complete the care of their existing patients. If physicians incur a private cost for time spent past EOS (e.g., if they are not compensated adequately or at all for this time), this scheduling creates two forms of moral hazard: First, on an extensive margin, physicians will accept fewer patients than they otherwise would near EOS. 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.

I empirically examine these issues by exploiting variation in the arrival of patients to the ED

---

1 In health care, as in many other industries, these changes have been massive, largely due to changing technology and workforce preferences. The number of process measures to be completed in the first few hours of many disease presentations have grown exponentially, because we now have technologies like coronary stents. Medical knowledge also makes care more standardized. Finally, with dual-earning families, physicians no longer live in hospitals but prefer flexibility and interchangeable roles (Goldin, 2014).

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).
and variation in physician schedules, while controlling for time of the day, day of the week, and location within the ED. Because I observe shifts that end at different times, I separate effects related to shift work from differences due to the time of day. Because I observe shifts of different lengths, I also separate these effects from “fatigue,” which I consider to depend on the time since the beginning of shift, rather than time to EOS. I first show that physicians are substantially less likely to accept patients near their EOS but that this depends on whether patients can be cared for by another physician within a reasonable period of time. Over a wide variety of shift arrangements, I find that physicians reduce their acceptance of patients in the hour prior to the arrival of another physician in an overlapping shift.

For patients they do accept, I find that physicians progressively shorten the duration of care (“length of stay”) as the time of arrival approaches EOS. Patients arriving in the last hour prior to EOS have a length of stay about 40% of the duration that patients arriving six hours prior to EOS have. As length of stay decreases, physicians distort other inputs to patient care, increasing the number of formal orders for tests and treatment for patients arriving in the last hour prior to EOS. This suggests discretion in the production process of patient care, and that formal utilization and time are inputs that can substitute for each other. Patients in the last hour are 21% more likely to be admitted to the hospital, with a corresponding increase in the cost of care. Based on observables, selection appears minimal and predicts lower utilization and fewer admissions, the opposite of my findings for these outcomes.

In order to interpret changes in patient care as a moral hazard, I use another source of variation from shift structure: the overlapping time between when a peer arrives on shift and when the index physician reaches EOS, when she is allowed to go home if her work is complete. 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 first-best reasons, since it is merely when physicians may go home if appropriate. I show that distortions on the extensive 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 distortion in patient care (i.e., no more utilization or admissions) by physicians in shifts with overlap of four or more hours.

---

3For example, a first-best reason to discharge patients earlier and even spend more money doing it is that physicians have fewer patients near EOS, and therefore their time per patient is more valuable.
This paper illustrates two strategic behaviors that workers may take when responding to the
general incentive to leave work near the end of scheduled availability. First, workers may accept
less new work. Second, for cases they have already accepted, workers may distort production to
entail less time but possibly more public resources. Although the first moral hazard is an obvious
one, akin to a definition of “presenteeism” in which workers are present but not working, it
mitigates the second moral hazard of distortion in care. This mitigation occurs through reducing
workload when time becomes more costly.\(^4\) The second moral hazard is often more difficult to
observe and potentially more costly. This tradeoff implies that it is second-best optimal to allow
physicians to “slack off,” avoiding work as they near EOS, particularly because measuring and
inducing the complex set of decisions that comprise patient care is difficult.

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 generic monitoring (Holmstrom, 1982).\(^5\) Because moral hazard
is usually considered in the single dimension of whether workers provide “effort,” there is no
tradeoff between dimensions of moral hazard. This paper empirically demonstrates a general
tradeoff between the choice of work and the subsequent performance of work, under which
second-best optimality allows workers to avoid work. In this sense, this paper is also related
more broadly to the design of second-best incomplete contracts, for schedules as contracts that
specify minimum \textit{quantity} (or scheduled amount of time) of labor (Weitzman, 1974).

Second, this paper is related to a literature on improving productivity through workplace
design (Ichniowski et al., 1997; Lazear, 1995), including recent work by Bloom et al. (2013)
investigating the effect of working at home on productivity. An older literature has studied
the tradeoff between labor and leisure, but its mapping to work environments seems at best
coarse. In practice, availability for work is scheduled within employment, with frequent, often

\(^4\)Coviello et al. (2013) discuss of the effect of dividing time among tasks, 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.

\(^5\)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; Dufo et al., 2013). A related empirical
literature has examined nonlinear financial incentives with windows of performance measurement (Oyer, 1998;
Larkin, 2013).
daily boundaries between scheduled work and time off. For many workers, these times are fairly rigid in the short term, while the nature of work is not. This paper shows the productivity implications of potential worker agency, due to the cost of 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 a large body of research on the long-term health and social consequences of shift work (e.g., Gordon et al., 1986; Gold et al., 1992). Much of this literature has discussed the mechanical effects of long shifts on fatigue. 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 an effect of shift work, and scheduling more generally, due to strategic behavior that extends to shifts with regular hours and short to medium lengths. These findings also speak to a literature concerned with what physicians value (McGuire, 2000). This literature largely considers patient health and financial incentives, while these results suggest that physicians (presumably like other workers) value convenience and certainty of hours. Enriching what is considered in the physician utility function may shed light on important issues in the production of health care, such as the overuse of diagnostic tests (e.g., Baker, 2001), which have at least been anecdotally discussed as a negative consequence of making physicians work too hard (Jauhar, 2014).

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 across shift types. Section 4 investigates the acceptance of new patients by physicians working in a variety of shift structures. 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 patient acceptance, distortion in care, and shift structure. Section 7 discusses efficiency implications, and Section 8 concludes.

Some work has also documented that obstetricians tend to perform more cesarean sections during certain times of the day (Brown, 1996; Spetz et al., 2001), although this work is not linked to shift work or other features or workplace design.
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, physicians in this ED work in shifts. In the study sample from June 2005 to December 2012, I observe shifts with lengths that range from seven to twelve hours. In addition to length, shifts may also differ in any overlap with a previous shift or with a subsequent shift in the same location. I observe about 24,000 shifts that I categorize into 35 different shift types – described as \((l, o_a, o_b)\) by these three parameters of shift length \(l\), prior-shift overlap \(o_a\), and post-shift overlap \(o_b\) – over the sample period. Table 1 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, the most important parameter is the post-shift overlap \(o_b\). This overlap is the time in which a physician nearing EOS may be relieved from new work by another physician who has begun work in the same location. By “location,” I refer 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 in other industries with shift work, shift times were designed around estimated workload needs, and schedule changes reflected changes in the flow of patients to ED, which increased over time. Little attention was paid to the transition times between shifts in a location, with the exception of a change in January 2012, in which more overlap was intentionally introduced to allow for

\[7\]

\(7\)I distinguish between shifts that end with the closure of a patient location, or “terminal shifts” with \(o_b = 0\), and those continuing patient care with another shift in the same location, or “transitioned shifts” with \(o_b > 0\).
smoother transitions of work. 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 in the ED. All regime changes, however, can be summarized as a set of shifts, each described by a shift type \((l, o_a, o_b)\), a starting day and time, location, and range of months that the shift was in effect (see Figure 1; Table A-1 details these shift descriptions).

Shifts are scheduled many months in advance, and physicians are expected to be available 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 exceedingly 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 have always been 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.\(^8\)

### 2.2 Patient Care

After arrival at the ED, patients are assigned to a bed by a triage nurse. This assignment determines the managerial location for the patient and therefore the one or more physicians who may assume care for the patient. Once the patient arrives in a bed, a physician may sign up for that patient on the computer order entry system. Physicians are expected to complete work on any patient for whom they have assumed care, except in uncommon cases where the patient is expected to stay much longer in the ED. For patients arriving near their EOS, physicians may therefore opt not to start work and leave the patient for another physician. This is made easier if another physician 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

---

\(^8\)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.
The other members of the care team 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, while physicians are expected to complete work on their patients, for nurses, residents, and physician assistants, care is readily transferred to another provider in the same role when they end their shifts.

For physicians in the ED, the concept of completing work on patients is a matter of discretion. Care for any given patient complaint is usually expected to continue either as outpatient or as an inpatient. The key criterion for completion of work is whether the physician believes that sufficient information has been gathered for a discharge decision out of the ED. This decision of course could be reached with complete diagnosis and treatment of a condition in the ED. But often much less is accomplished: Rather, the physician may decide to discharge a patient home with outpatient follow-up after ruling out serious medical conditions, or the physician may argue for inpatient admission if the patient has not been treated adequately or could still possibly have a serious condition that would make home unsafe.9

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 rule in or rule out diagnoses, such as bronchodilators for suspected asthma. But time – for a careful history and physical, serial monitoring, or a well-planned sequence of formal tests and treatment – remains an important input in the production of information. Diagnostic tests and treatments can be complements or substitutes for time. Formal tests (e.g., CT and MRI scans) take time to complete and can thus prolong the length of stay. But exhaustive testing can also substitute for a careful questioning or an otherwise less-intensive sequence of investigation to understand the trajectory of a condition.

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

---

9In 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.
physician schedules to arrive at a working sample of 372,224 observations. Details of the sample definition process are described in Table 1. In the sample, I observe the identities of 102 physicians, 1,146 residents and physician assistants, and 393 nurses.

Table 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 2 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 duration of overlap with another shift at EOS.

I measure ED length of stay as an 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. In addition, length of stay has been the focus of both policy papers and a cottage industry of ED management consulting as an measure of ED throughput and a determinant of waiting times (McHugh et al., 2011; Thompson et al., 1996). In this study, I am able to measure length of stay from the arrival at the pod to entry of the discharge order, which is unaffected by inpatient bed availability, patient home transportation, or clinical care or patient adherence after ED visits. In addition to length of stay, I also measure formal inputs to production, recorded in detail as orders with precise timestamps for diagnostic tests and treatment.

As 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. Hospital admissions from the ED have received increasing attention as a source of rising system costs (Schuur and Venkatesh, 2012; Forster et al., 2003), and appropriate discharge decisions are clearly crucial for the efficiency of a delivery system. I accordingly measure total direct costs, including costs incurred both by formal utilization in the ED and during a subsequent admission.\(^{10}\) 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, let alone within the control of the ED physician, which

\(^{10}\)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).
reduces the precision of their estimated effects.

2.4 Patient Observable Characteristics

I observe a rich set of patient observable characteristics that likely provide more information than the physician can observe at the time she accepts each patient. When patients arrive to the ED, they are evaluated by a triage nurse, and assigned an Emergency Severity Index (ESI), which is an index ranging 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,” which is a phrase that describes why the patient arrived at the ED. Physicians do not interact with patients prior to accepting them, and because patient beds are behind curtains, physicians also generally do not have visual information about a patient’s appearance.

I observe all the information communicated to physicians via the computer interface prior to patient acceptance. In addition, I observe patient insurance status, language, race, and zip code of residence. I also observe rich diagnostic information, including 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). These patient characteristics are generally unobserved prior to patient acceptance, although race might be guessed by a patient’s last name, and similarly comorbidities might be guessed by a patient’s chief complaint. Diagnostic codes of course are also partly determined by patient care.

2.5 Descriptive Evidence

Figure 3 shows nonparametric 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 15-minute bin of arrival time interacted with each 15-minute bin of length of stay. Some findings are apparent from this plot of visits. First, very few patients are seen in the last hour to hour and a half prior to EOS. 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. Second, lengths of stay are shorter for patients who arrive and are accepted by a physician closer to EOS than for
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 3 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

ED physicians make three types of decisions. 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, including not only formal diagnostic tests and treatment but also clinical observation over time. They then produce a discharge decision, a key product of their work, based on their best knowledge of the patient’s health, either for admission or home discharge.

As an ED physician nears EOS, time becomes more privately costly, because the physician may go home after EOS only if she has completed patient care. This has implications for all three decisions above: For patients under their care, they will change the inputs of patient care, and may admit patients they normally would have discharged home. They will also be less likely to assume the care of new patients near EOS. The work environment, in which physicians may overlap shifts with a peer, modifies these effects via the distribution of workload, but the EOS (or the time from peer arrival to EOS) does not have any first-best implications for efficiency.

3.1 Model Setup

I develop a stylized model to formalize these predictions. Consider a physician at the arrival of a patient at time $t$, in the work environment $\mathcal{E}_t \equiv (W_t, W'_t)$. $W_t \equiv (t, w_t)$ captures the start time of the physician’s shift, $t$, and her current workload, $w_t$. $W'_t \equiv (t', w_t')$ describes similar
information for a potential peer who may start his shift in the same managerial location as the index physician’s. 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 \( z \) 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 \( q(z) \), 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:

\[
    u(t, \mathcal{E}_t; \theta; a, z, d) = \begin{cases} 
        O(\mathcal{E}_t; \theta), & a = 0 \\
        V(\theta, d) - c_r(t, \tau, \tilde{t}) - c_z z, & a = 1 
    \end{cases} \tag{1}
\]

\( O(\mathcal{E}_t; \theta) \) is the value of the “outside option” if \( a = 0 \), which captures concerns about the patient and the peer (e.g., social incentives). Importantly, \( O(\mathcal{E}_t; \theta) \) may depend on whether a peer is present or will arrive soon to accept the patient. Otherwise, if \( a = 1 \),

\[
    u(\cdot) = V(\theta, d) - c_r(t, \tau, \tilde{t}) - c_z z. 
\]

\( V(\theta, d) \) is the utility due to discharge decision \( d \) for patient with health state \( \theta \); \( c_r(t, \tau, \tilde{t}) \) is time cost of care, which privately depends on the EOS, \( \tilde{t} \); and \( c_z z \) is the test-treatment cost of care. Note that \( \tilde{t} \) only appears in \( c_r(t, \tau, \tilde{t}) \) and not in \( \mathcal{E}_t \), which captures the work environment in first-best terms. I discuss \( O(\mathcal{E}_t; \theta) \), \( V(\theta, d) \), and \( c_r(t, \tau, \tilde{t}) \) 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 implications for resource utilization and patient safety. Admitting a patient who could have been safely discharged home incurs unnecessary costs. On the other hand, discharging a patient home carries the risk that the patient may have a serious illness that could progress when unmonitored as an outpatient and perhaps lead to death.

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 \(p > p^* < \frac{1}{2}\):

\[
E[V|d = 0, p = p^*] = E[V|d = 1, p = p^*]
\]

\[
p^*V(1, 0) + (1 - p^*)V(0, 0) = p^*V(1, 1) + (1 - p^*)V(0, 1)
\]

\[
\frac{1 - p^*}{p^*} = \frac{V(1, 1) - V(1, 0)}{V(0, 0) - V(0, 1)} > 1.
\]

The primary objective of patient care is to observe \(\theta\) for the discharge decision, since physicians prefer to discharge patients appropriately.\(^{11}\) Care leading to a discharge decision is produced not only by formal diagnostic tests and treatment but also by clinical observation and reasoning over time. Testing and treatment are often (and ideally) sequentially ordered and require time (i.e., formal resource utilization is often a complement to treatment time in production), yet a barrage of testing and treatment, especially if ordered earlier, can substitute for time spent on clinical observation and reasoning. Testing, treatment, and time can all entail costs to the physician, but time could privately become more costly as the physician nears EOS, because time spent on patient care past EOS prolongs the time the physician must stay at work.

In the model above, I consider inputs to patient care as means to increase \(q\), the probability that the physician observes \(\theta\) before deciding \(d\). I impose the following production relationships: \(q\) is increasing and concave with respect to inputs to patient care \(z (\tau_1, \tau_2, \text{and } z)\). \(\tau_1\) and \(z\) are substitutes in production: \(\partial^2 q / (\partial \tau_1 \partial z) < 0\). However, \(\tau_2\) and \(z\) are complements: \(\partial^2 q / (\partial \tau_2 \partial z) > 0\).

\(^{11}\)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.
Because effective time per patient is reduced with higher workload \( w_t \), \( \partial^2 q / (\partial \tau_1 \partial w_t) < 0 \) and \( \partial^2 q / (\partial \tau_2 \partial w_t) < 0 \), in contrast to formal inputs, for which I make the normalizing assumption \( \partial^2 q / (\partial z \partial w_t) = 0 \). Finally, I assume that the cost of time \( \tau = \tau_1 + \tau_2 \) depends on whether it extends beyond the EOS: \( c_\tau (t, \tau, \bar{t}) = c^0_\tau (\tau) + \tilde{c}_\tau (\tau + t - \bar{t}) \), where \( c^0_\tau (\cdot) \) and \( \tilde{c}_\tau (\cdot) \) are both increasing functions, possibly with \( \tilde{c}_\tau (\cdot) = 0 \) for \( \tau + t - \bar{t} < 0 \), and \( c_\tau (\cdot) \) is weakly convex.\(^{12}\)

Given these assumptions, I predict that 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 \( \tau_1 \) or \( \tau_2 \) is decreased: Intensity decreases when \( \tau_2 \) is decreased, but it increases when \( \tau_1 \) is decreased. 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 this 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 decision to assume the care of a new patient arriving near EOS. This decision has obvious implications for the distribution of work across time and physicians.

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 \( a = 1 \),

\[
\max_z E \left[ \max_d V (\theta, d) \right] - c_\tau (t, \tau, \bar{t}) - c_z z,
\]

where

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

\(^{12}\)Time may be more costly as EOS approaches but prior to EOS if the physician needs to reduce her workload or attend to other duties (e.g., dictating charts) before going home.
and expected utility under $a = 0$, $E [O (\mathcal{E}_t; \theta)]$. Note that as time becomes more costly near EOS, reoptimization implies both higher input costs and lower $E [\max_d V (\theta, d)]$ (less appropriate discharges) by lower $q$. If the peer is present in the location ($t > t'$), then I assume that the patient is assigned to the physician with the greater net utility for accepting the patient: \footnote{The microfoundation for this relies on subgame-perfect reasoning similar to Rubinstein (1982) in a war of attrition (Bliss and Nalebuff, 1984), in which both physicians when present will know which physician has the lower net cost.}

\[
\max_{\theta} E \left[ \max_d u (t, \mathcal{E}_t; \theta, 1, z, d) \right] - E [O (\mathcal{E}_t; \theta)] .
\] (2)

Several countervailing forces influence the net expected utility between $a = 1$ and $a = 0$ in Equation (2) as the physician nears her EOS. As EOS approaches, expected utility decreases for $a = 1$ as her time costs of caring for the patient increase. If the peer is present at $t$, he can likely provide better care because he incurs smaller time costs, since $t' > t$, and because $w'_t > w_t$ when he begins his shift. However, if the peer has not yet arrived ($t < t'$), the patient could also worse off by waiting for care, an effect greater for some (e.g., sicker) patients than others. By $O (\mathcal{E}_t; \theta)$, I allow the physician both to consider patient care by the peer and a host of other considerations, such as financial incentives and what the peer experiences or thinks (e.g., social incentives).

The overlapping transition duration $o_b = \bar{t} - t'$ has implications for the distribution of work and for patient care. In longer transitions, the physician has less time pressure caring for patients she accepted prior to her peer’s arrival, because her peer arrives earlier relative to her EOS. Further, once her peer has arrived, the physician is more comfortable with declining patients, because these patients do not have to wait. This allows her to decompress her workload $w_t$ prior to EOS at $\bar{t}$. Recall that lower $w_t$ mitigates the time cost distortion because it increases the effective time that the physician can devote per patient. So with longer transitions, as $t$ approaches $\bar{t}$, the distortion in patient care applies to fewer patients, since fewer patients are accepted, and is less severe for patients who are accepted, since $w_t$ is lower.

### 3.4 Efficiency Considerations

To be more precise about the implications of a private wedge in the cost of time due to EOS,
I simply consider a welfare function $W$ identical to physician utility when $a = 1$, except that time costs do not depend on EOS:

$$W(t, E_t; \theta; a, \hat{z}, \hat{d}) = V(\theta, \hat{d}) - c^0_r(\hat{\tau}) - c_z \hat{z} - c_w(\theta; t' - t) \mathbb{1}(a = 0),$$

where $\hat{z}$ and $\hat{d}$ denote that input and discharge decisions may be undertaken by either the physician or a peer, depending on whether the physician accepted the patient or not. Welfare does not depend on who makes input and discharge decisions, but only on the decisions themselves. If $a = 0$ and $t < t'$, I also consider costs to waiting $c_w(\cdot)$. This represents the fact that some patients in the ED by definition should be seen emergently, i.e., that the quality of care depends on timeliness.

The discussion above suggests that efficiency costs from the EOS distortion increase as the physician nears her EOS. The distortion arises by changing the mix of patient-care inputs and reducing the quality of the discharge decision. The design of transitions with greater overlap between shifts can reduce this distortion by decreasing the number of patients seen by the physician near EOS, thus reducing the number of patients subject to the distortion, and by decreasing the workload of the physician near EOS, thus reducing the time pressure and the size of the distortion. Another way to reduce the distortion is for the triage-nurse manager to externally reduce the number of patients assigned to a location as the physician nears EOS. Of course, this is balanced against the fact that available physician labor is not being used.

Note that if $c_r(t, \tau, \overline{t}) = c^0_r(\tau)$ and if $O(E_t; \theta) = W(t, E_t; \theta; a = 0, \hat{z}, \hat{d})$, then physicians will implement the first best. But if physicians do incur a private cost for time spent beyond EOS, then they will see fewer patients near EOS than implied in the first best. However, if the ED were to impose that physicians see patients they should under $c_r(t, \tau, \overline{t}) = c^0_r(\tau)$, there is a general cost incurred from the distortion in patient care as long as the ED cannot impose the optimal $\hat{z}$ and $\hat{d}$ (e.g., because $p$ is privately observed by physicians and not ED administration). Thus, it follows that the ED should not assign some patients who would have been seen in the first best, particularly if the potential distortion in care is large.\textsuperscript{14}

\textsuperscript{14}In this paper, I restrict attention to these mechanisms of defining schedules and assigning patients within schedules, focusing on quantity of work. Another policy might be to fix the price of work by overtime, although
4 Effect on Patient Acceptance

In this section, I first describe the effect of approaching EOS on whether a physician accepts a new patient. As explained above, 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 structures. Because I observe physician behavior in shifts with different durations of overlap near EOS \( o_b \equiv \bar{t} - t' \), I can also verify the prediction that the decision to accept a patient also depends on what the patient entails. Specifically, it depends on whether there is or will soon be another physician present, which is a direct mechanism through which shift structure influences the distribution of work.

Figure 4 presents the hourly average rates of new patient visits, with each panel representing shifts with a different \( o_b \), for the index physician, for the pod inclusive of the index physician and any other physician who may be transitioning on or off, and for the entire 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, defined as shifts with \( o_b > 0 \), the average rates of patient flow show two consistent relationships with time. First, patient flow declines precipitously in the hour prior to the transitioning physician’s arrival to the location. Second, patient flow declines close to zero in the two to three hours prior to EOS. If there is sufficient time between EOS and the beginning of the transitioning physician’s shift, patient flow is relatively constant but diminished during that length of time. In terminal shifts, where \( o_b = 0 \), the decline in patient flow begins earlier, at least four hours prior to EOS.

Also shown in Figure 4, the flow of patients to transitioning peers generally at least makes up for the decline in flow for the index physician. That is, patients continue to arrive at the pod at similar or greater rates prior to the peer’s transitioning shift, and patients who are not accepted by the index physician may wait up to an hour to be seen by another physician (the this generally will not result in \( c_r (t, \tau, \bar{t}) = c_r^0 (\tau) \), 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.
peer who has not yet arrived). Finally, Figure 4 plots the flow of patients to the entire ED, showing that there is generally always a background flow of patients to other pods, even when flow decreases for the index physician. The overall ED flow appears more stable in shifts with greater observations and variation in EOS across times of the day (e.g., $o_b = 1$ and $o_b = 6$).

These relationships in transitioned shifts are remarkably consistent, over different $o_b$, 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. Furthermore, the reduction in the acceptance of patients prior to the arrival of a transitioning peer suggests anticipatory behavior, in which the index physician transfers work across time to a peer who has not yet arrived. The anticipatory decline is more pronounced in shifts with transitions of three hours or less, compared to shifts with transitions of four or six hours, suggesting that at least part of the decline is motivated by the EOS effect.

For terminal shifts, it bears mentioning that the long decline in patient flow rates is by construction due to the triage nurse assigning fewer patients to the terminal location. Patients who are not seen by a physician four hours prior to her EOS in a terminal shift would essentially have to wait at least four hours to be seen by any physician, which could be unacceptable, either in terms of patient health or social norms. This reduction in patient assignment to a terminal-shift location suggests that the triage nurse spares physicians from work near EOS, perhaps in part because the location is also scheduled to close at the physicians’ EOS.

5 Effect on Patient Care

In this section, I next evaluate the effect of approaching EOS for the care of patients who are accepted. If time becomes more costly near or past EOS, then physicians will spend less time on patients they accept. Patient care involves a mix of inputs, including both time and formal tests and treatment, and physicians may distort their orders for formal utilization either upward or downward, depending on whether time and formal inputs are net substitutes or complements,
respectively. Finally, as physicians produce less information in patient care for the discharge decision, they may resort to admitting more patients as EOS approaches.

Because patients are both assigned to locations by the triage nurse and accepted by physicians once assigned, there is scope for patient selection near EOS, which is in itself of interest. I address this issue in two ways. First, I consider selection on observables, including some patient characteristics which are unobserved by the physician at the time of acceptance. Second, for the effect on length of stay, I use a hazard model to estimate the increased hazard of discharge around EOS, regardless of when the patient arrived and was accepted. In this section, I also use variation in shift lengths to separate EOS effects from “fatigue,” which I consider as an effect relative to beginning a shift.

5.1 Main EOS Effects

As my main analysis, I address the following question: What is the effect of a patient’s arrival near a physician’s EOS on that patient’s care by that physician? I primarily use variation within the same health care providers working in different settings – in different locations and on different days of the week and month-year combinations – to control for provider skill and quality of interactions, time-invariant location unobservables, unobservables that vary depending on time. These could include unobserved characteristics for patients arriving at different times of the day, days of the week, months of the year, or different years, as well as for patients who are usually assigned to one location over another. Unobservable differences in the operation of the ED that vary over time or across location, such as the availability of hospital resources, can also be controlled for. Although I address patient selection more directly later, I also control for a rich set of patient characteristics. I finally use variation in shift lengths to control for time relative to the beginning of shifts.

In the full specification, I estimate the following equation:

\[
Y_{ijkpt} = \sum_{m=-6}^{-1} \alpha_m 1(|t - t_j| = m) + \sum_{n\in N} \gamma_n 1(|t - t_j| = n) + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \varepsilon_{ijkpt},
\]
where outcome $Y_{ijkpt}$ is indexed for patient $i$, physician $j$ (in shift from $t_j$ to $t_j$), assisting team $k$ (including the resident or physician assistant, and the nurse), physical pod $p$, and arrival time $t$. The coefficients of interest in Equation (3) are $\{\alpha_m\}$, which indicate 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 $(t - t_j)$, patient characteristics $(X_{it})$, pod identities $(\zeta_p)$, a sum of fixed effects across time categories $(\eta_t)$ (for month-year, day of the week, and hour of the day), and physician-team identities $(\nu_{jk})$.

Table 3 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 3 and plotted in Panel A of Figure 5, 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 3 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. This effect is separately identified from EOS effects because I observe shifts of different lengths. This difference, about 0.13 in the last hour prior to EOS, also appears to account for only a minor portion of the overall effect.

Table 4 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.\textsuperscript{15} These patients are also

\textsuperscript{15}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.
5.7% 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 5.

5.2 A Closer Look at Patient Care

Together, the effects above are consistent with an increasing cost of time near EOS, a resulting distortion in care, and less information produced so that risk-averse physicians are more likely to admit patients. These effects are most pronounced in the last hour prior to EOS. Conceptually, almost by construction, an increase in the number of orders as length of stay suggests that formal utilization is on net a substitute for time in patient care.

The conceptual framework in Section 3 divides length of stay into two components, $\tau_1$ and $\tau_2$, that purely substitute for and complement with formal utilization, respectively. Roughly speaking, $\tau_1$ can be thought of time spent on clinical monitoring and reasoning that is a substitute for brute-force utilization (e.g., serial abdominal examination as opposed to an abdominal CT scan). On the other hand, $\tau_2$ is time needed to follow up on utilization (e.g., time spent waiting for a radiologist to read the abdominal CT scan). In practice, $\tau_1$ and $\tau_2$ are not neatly divided, but some intuitive distinctions can be made: I consider time spent prior to the first formal order as part of $\tau_1$ and time spent after the last formal order as part of $\tau_2$. Time in between the first and last orders could belong to either $\tau_1$ or $\tau_2$, but the spacing of orders over time often reflects a sequencing that is closely related to clinical monitoring and reasoning, $\tau_1$.

I therefore measure length of stay in three time components: the time between pod arrival and the first order, the time between the first order and the last (non-discharge) order, and the time between the last order and the discharge order. Specifying these time components as relative shares of overall length of stay, I estimate a fractional logit model (Papke and Wooldridge, 1996) using similar regressors as in Equation (3).

I present marginal effects from this model for each hour prior to EOS in Figure 6. In Panel A of the figure, I scale time shares by the median predicted length of stay in each hour prior to
EOS according (3). In Panel B, I simply plot the shares proportional to length of stay. Although the duration of each of these components decreases as EOS approaches, their proportions remain relatively constant until the second to third hour prior to EOS, when the proportions for the time prior to the first order and the inter-order time both decrease. These results suggest that a distortion that reduces the relative proportion of \( \tau_1 \), particularly in the last hour prior to EOS, and is consistent with the increase in formal utilization (net substitution) in the last hour shown in Table 4 and Figure 5.

5.3 Patient Selection

I next address the question of patient selection relative to EOS. The results from Equation (3) control for observable patient characteristics and for unobservable characteristics of patients who may arrive to the ED across different absolute time categories (e.g., hour of the day or day of the week) and of patients who may arrive to different locations in the ED invariant with time. Here I evaluate variation in observable patient characteristics correlated with time relative to a physician’s EOS.

As described in Section 2.4, there is a formal system of communicating patient severity in the ED that is publicly and fully observed, and ED physicians otherwise accept patients with relatively little additional information (e.g., they generally do not see patients until after accepting them). Nevertheless I also exploit the full rich set of observed patient characteristics, including \textit{ex post} diagnoses and other characteristics such as insurance status that cannot be observed by the physician at the time she accepts the patient. This would capture information that does not appear in the formal system but could be communicated informally by nursing (e.g., “This patient has [formally undocumented] medical issues that need attention right now”), because patient medical issues are later documented \textit{ex post}.

In this analysis, I evaluate patient selection based on two sets of characteristics: those that could possibly be observed by a physician or triage nurse prior to treatment, \( X_{it}^{\text{prior}} \), and others that include rich diagnosis codes, insurance status, race, and language that would at best be incompletely observed until after patient care, \( X_{it}^{\text{full}} \). Separately for each set, I first estimate predicted outcomes measures for each patient visit by \( Y_{ijkpt} = \beta_{set} X_{it}^{set} + \varepsilon_{ijkpt} \), where \( set \in \)
\{prior, full\}. Next, I estimate the following regression describing the relationship between the predicted outcomes for selected patients, \( \hat{Y}_{ijkpt}^{\text{set}} = \beta^{\text{set}} X_{it}^{\text{set}} \), and the time of selection relative to EOS:

\[
\hat{Y}_{ijkpt}^{\text{set}} = \sum_{m=-6}^{-1} \alpha_m^{\text{set}} \mathbb{1} \left( \lfloor t_i - t_j \rfloor = m \right) + \sum_{n \in N} \gamma_n \mathbb{1} \left( \lfloor t - t_j \rfloor = n \right) + \zeta_p + \eta_t + \nu_{jk} + \epsilon_{ijkpt}, \tag{4}
\]

which is similar to Equation (3), except that \( \beta X_{it} \) is left out as a regressor. I interpret each coefficient \( \alpha_m^{\text{set}} \) as the effect on patient selection as described by a related outcome measure, predicted by \( X_{it}^{\text{set}} \), in the \( m \)th hour prior to EOS. Using the two sets of patient characteristics allows me to assess the degree of selection on unobservables at the time of patient assignment. Because detailed diagnoses are in fact endogenous to the intensity of patient care, estimates of selection based on the full set of patient characteristics may even be biased upward.

Figure 7 presents results 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. Selection appears to be in the direction of healthier or less resource-intensive patients as physicians near EOS: Patients who are expected to have shorter lengths of stay, lower frequencies of admissions, and incur lower costs and fewer orders are increasingly selected for as EOS approaches. Predicted length of stay is 5.4% lower in the last hour prior to EOS compared to greater than six hours prior to EOS, which is close to an order of magnitude smaller than the effects seen for actual length of stay. All outcomes show a monotonically decreasing relationship between predicted resource utilization and proximity to EOS, which is in contrast to actual outcomes for admission, costs, and orders.

In the Appendix, 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 indeed 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.

These results suggest the presence of selection but to a much smaller degree than to explain actual length of stay results. Furthermore, estimates using the predicted outcomes on the full set of patient characteristics, including detailed ex post diagnoses, suggest that selection on patient characteristics formally unobservable at the time is negligible. Therefore, the magnitude discrepancy between selection and realized reductions in length of stay is consistent with large behavioral changes in how physicians provide patient care, particularly in terms of time devoted, conditional on patient type. Furthermore, since monotonic relationships in all measures show that healthier and less resource-intensive are being chosen near EOS, selection points in the opposite direction of increases in admissions, costs, and orders in the last hour prior to EOS.

5.4 Effects Relative to Shift Beginning

The literature on shift work has been 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, it is more difficult to imagine significant fatigue near the end of a shift of nine hours, the modal shift length in this setting. Nonetheless, I am able to specifically address this issue by exploiting variation in shift length, which allows me to control for effects, such as fatigue, that are correlated with time since the beginning of shift. Assuming that, conditional on time since beginning of shift, fatigue is independent of time to EOS, I can interpret EOS effects free from fatigue.

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.\(^{16}\) 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 among three categories of

\(^{16}\)For other outcomes (orders, admission, and costs), the size of the EOS effect actually increases when controlling for time since shift beginning.
shift lengths. Given that the modal shift length is nine hours, I study shift 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 8 plots coefficients $\alpha_m$ from Equation (3) estimated separately for each category of shift length. Panel A has the coefficients plotted according to time relative to EOS and shows that the coefficients are largely similar across shift lengths and within hour prior to EOS. Panel B arranges the coefficients according to time relative to the shift beginning and illustrates that the EOS effect is largely independent of the time since beginning the shift.

5.5 Discharge Hazard over Time

Thus far, I have considered evidence relying on average measures of length of stay, but I have not yet used the full distribution in discharge times for a patient who has arrived at a given time prior to EOS. To illustrate the value of considering this information, recall that Figure 3 showed 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 3, I estimate a discrete-time logit hazard model for whether a patient is discharged or not at a given point in his stay:

$$H(t, \tau, i, j, k, p) = \sum_{m=-9}^{3} \alpha_m 1\left(\left|t - t_j\right| = m\right) + \sum_{n \in N} \gamma_n 1\left(\left|t - t_j\right| = n\right) + \eta_t + \zeta_\tau + E\left[Y_{ijkp} | X_{i, j, k, p, \tilde{\eta}_t}\right] + \varepsilon_it,$$

where using similar notation as before, $t$ is the arrival time of the patient, $\tau$ is the hour of the stay during which the patient may be discharged (conditional on not being discharged at $\tau - 1$), $i$ indexes the patient, $j$ and $k$ index the physician and other providers, and $p$ indexes the pod. 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 \( \zeta_{it} \). The error term \( \varepsilon_{it} \) is distributed as Type I extreme value. I also control for time relative to shift beginning \( (t - t_j) \), and I include a linear prediction for log length of stay conditional on patient characteristics, provider identities, pod identity, day of the week, and month-year interaction.

Figure 9 shows marginal effects, estimated from Equation (5), 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 patients are selected as those with conditions requiring further evaluation and treatment in the ED. After the fourth hour past EOS, exceedingly few patients remain, and these patients are likely to be transferred to the care of another physician.

6 Shift Structure, Workload, and Distortion

In this section, I return to shift structure in order to connect the two categories of strategic behavior that arise near EOS. Because time is privately costly near EOS, physicians are less likely to accept new patients, and they distort the care of patients they have accepted. These two moral hazards are connected because accepting patients increases workload, which decreases the effective time a physician can spend per patient. Therefore, patient care will be more distorted in environments in which physicians cannot yet avoid new work and decompress workload until they are too close EOS. I evaluate this prediction in this section by measuring how workload and patient care distortions vary across shifts with transition durations with a peer at EOS.

The purpose of this analysis is twofold: First, in Sections 4 and 5, I present evidence for the causal effect of a patient’s arrival at a given time relative to EOS on acceptance and subsequent care, but interpreting this effect as moral hazard is non-trivial because there are first-best reasons
for a physician to stop accepting patients and change patient care before she ending work. In this section, I use this prediction to support the notion that I have identified strategic behavior, under the identifying assumption that the EOS by itself has no first-best implications for patient care.\textsuperscript{17} Second, this analysis directly speaks to the concept of using shift structure as a policy lever in the setting for moral hazard. While I exploit shift structure as an instrument that varies workload near EOS, I note that the concept of a tradeoff between the extensive and intensive margins of moral hazard, via workload, is a general concept that can be influenced by a variety of other levers, including financial and social incentives.

\section*{6.1 Patient Censuses over Time}

I measure workload as the number of patients under the care of a given physician ("censuses") at a given time. Censuses are implied by numbers and times of patients accepted by a physician as well as numbers and times of those being discharged by that physician. The change in censuses over time therefore reflects the net flow of work: Increasing censuses indicate that the rate of influx of new patients outstrips the rate of discharge of patients completing their care, while decreasing censuses indicate the opposite.

Figure 10 shows unadjusted averages for census counts at each 30-minute interval, for the nine hours prior to EOS, in different shift types categorized by $o_b$. The time course of censuses for all shift types is generally characterized by an increase followed by a decrease. In shifts that start as a transitioning shift, beginning in a location that is already opened and staffed by a previous physician, censuses begin at a non-zero number. On average, censuses start at around two patients at the beginning of all shift types, except for the two shift types with $o_b = 2$, both of which do not transition from another shift. They peak at around eight to ten patients, except for shifts with $o_b = 6$, which peak much earlier with censuses closer to seven patients.

The timing of the census peak depends on $o_b$ of the shift. With longer transitions in substantially transitioned shifts ($o_b \geq 4$), the peak coincides essentially exactly at the beginning of the transition time. As physicians begin these transitions, the physician who will end her shift

\textsuperscript{17}In particular, I assume that the EOS has no first-best implications on efficiency, conditional on volume of work, time since beginning work, and time since a peer's arrival. This is consistent the institutional fact that the EOS is simply the time after which a physician may go home if work is complete.
sooner (but still in the relatively distant future) begins to accept patients at a rate that is slower than her discharge of patients. However, for transitioned shifts with $o_b \leq 3$, the census peak occurs at times progressively earlier than the beginning of transition. For example, for minimally transitioned shifts ($o_b = 1$), the peak occurs three hours prior to the beginning of transition, or four hours prior to EOS. Terminal shifts also peak in the average census around four hours prior to EOS.

Finally, all shifts technically end with patients remaining on the census. The number of patients remaining on census in the last 30 minutes prior to EOS is consistently close to four, on average among shifts with each $o_b$, with the sole exception of minimally transitioned shifts with $o_b = 1$. Minimally transitioned shifts in contrast have censuses of about six, essentially a 50% greater workload immediately prior to EOS than in the other shifts. For all shifts, the census is greater at the end than in the beginning. This implies that physicians end their shifts with more work in terms of patient count, which they must complete prior to leaving, than unattended work they begin their shifts with.

These results add further detail to the dynamic nature of physician workload, which is affected by shift design. As shown above, the transition overlapping near EOS changes both the time to discharge old patients and the rate of accepting new patients. These two factors are captured in the summary statistic of patient censuses under the care of physicians over the time course of a shift and across shift types. In order to leave work, physicians must decompress their workload prior to EOS and eventually bring their workload to zero before they leave work.

In substantially transitioned shifts, they begin decompressing their workload at the beginning of transition, since they have sufficient time from that point until EOS. In terminal shifts, because the location closes at the same time as EOS, workload also manages to decompress to the same census by EOS as in substantially transitioned shifts. However, in minimally transitioned shifts, new patients arrive at an undiminished rate to the location, and physicians have less time and scope to decompress their workload because the peer arrives late.

6.2 Heterogeneous Effects across Shift Types

I finally explore the interaction between workload and the distortion in patient care near EOS.
A higher workload, as measured by a higher number of patients to care for, further decreases the effective time that a physician can devote to each patient. Therefore, if physicians are unable to decompress their workload near EOS, then EOS distortions in patient care due to the increasing cost of time should be greater.

I have shown above that shift structure, particularly the transitions between shifts, directly influences the distribution of work across workers and time. I therefore examine the effect of workload on patient care near EOS by estimating heterogeneity in EOS effects across shift types. I consider three categories of shift types—terminal shifts, minimally transitioned shifts, and substantially transitioned shifts—that differ in the overlap near EOS, $o_b$. I categorize any shift after which there is no subsequent shift in the same location as a terminal shift. Shfits with a subsequent shift are transitioned shifts. A shift with at least four hours of overlapping transition ($o_b \geq 4$) is substantially transitioned, while a shift with only one hour of transition ($o_b = 1$) is minimally transitioned.\(^{18}\)

The analysis proceeds similarly as in Section 5, which estimates average effects across all shift types, except that I exploit additional identifying variation across different shift-type categories. I observe this variation within the same physicians and other providers, within the same patient characteristics, and within the same times of the day, days of the week, and month-year interactions. I estimate

\[
Y_{ijkpt} = -1 \sum_{r=-6}^{-1} \sum_{s \in S} \alpha_{rs} \mathbf{1}(\left\lfloor t_i - t_j \right\rfloor = r) \mathbf{1}(\text{Shift}_{j,t} \in s) + \beta X_{it} + \zeta_p + \eta_t + \nu_{jk} + \epsilon_{ijkpt},
\]

which is very similar to Equation (3) but additionally interacts the hourly EOS effects by categories of shift types corresponding to terminal shifts, minimally transitioned shifts, and substantially transitioned shifts. I normalize coefficients so that, in each of the shift categories, the reference category remains times greater than six hours prior to EOS.

Figure 11 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, shown in Panel A, is largely

\(^{18}\)While I observe shifts with $o_b \in \{2, 3\}$, they entail very few observations, as listed in Table 2. Results are unchanged whether I include these observations and, if so, whether I consider them as belonging to the minimally or substantially transitioned shift category.
similar among shift categories. All three shift categories show a substantial decline in length of stay for patient arrival times that approach EOS. However, EOS effects are notably absent in substantially transitioned shifts for orders, admission probability, and total costs, shown in Panels B to D. In contrast, minimally transitioned shifts have the largest increase in admissions and total costs at EOS, while a slightly greater number of orders are written at EOS in terminal shifts.

These results show that, although the EOS effect on length of stay is similar across shift categories, other effects on patient care appear qualitatively different between shifts with no or small transitions versus those with substantial transitions. When there is a substantial transition period, physicians do not distort care with more orders, admit patients more frequently, or incur more total costs for patients near EOS. The main structural difference between shift categories is in the accrual of workload over several hours prior to EOS, while differences in EOS effects become apparent in the last hour prior to EOS. This is consistent with an accumulating effect of the distribution of patients on the care of future patients, through the mechanism of workload and the scarcity of time per patient.

7 Discussion on Efficiency

In this paper, I find results consistent with physicians responding to increasing time costs near EOS: They are less likely to accept new patients; they discharge the patients they do accept near EOS more quickly but utilize more resources and are more likely to admit these patients. Under the assumption that the EOS by itself (aside from patient volume, time since beginning work, and time since peer arrival) does not imply any first-best reasons to change care, these responses appear to be distortions away from first-best efficiency.

This conclusion, however, assumes that physicians otherwise internalize the appropriate social costs in their delivery of patient care and discharge decision-making. In practice, physicians may have other private incentives that distort care. If there are countervailing distortions that induce physicians to increase patient lengths of stay (e.g., Chan, 2014), then the EOS effect may improve efficiency by reducing lengths of stay, consistent with the theory of the second best (Lipsey and
Lancaster, 1956). Nonetheless, if these countervailing distortions are constant with approaching EOS, then the EOS effect may still reduce efficiency within a certain proximity to EOS. More broadly, physicians likely do not internalize the appropriate social costs for any input in patient care. For example, physicians incur large material and personnel costs whenever they order MRI scans, but the private cost is relatively cheap with a few keystrokes in the computer order entry system. Thus, formal resource utilization is probably greater than socially optimal when compared to clinical activities performed by only physician, such as interviewing, examining, and discussing decisions with the patient. These clinical activities tend to be time-intensive for the physician relative to activities requiring formal orders. Under this underlying pattern of inefficiency, the EOS effect would exacerbate the overutilization of formal resources.

The general point is that schedules may create a private wedge in the value of time near EOS, with two consequences: First, physicians will want to avoid new work near EOS. Second, they will distort the care of patients they do accept, and this distortion is likely to be more costly. There are a wide variety of levers that act to reduce the first moral hazard. For example, the ED may incentivize physicians more strongly by paying them a bonus for each patient seen. More subtly, social incentives within the ED also encourage physicians to see more patients, because they would otherwise be leaving work for their peers. Both of these examples, as levers that reduce the first moral hazard of avoiding new work, could potentially worsen overall efficiency through the second moral hazard of distortion in care. If distortion in care is more difficult to observe, it would be second-best optimal to allow physicians near EOS to avoid seeing some new patients. Of note, the policy of tapering work before EOS (even in terminal shifts, where the tapering is done entirely by the triage nurse) is largely consistent with this second-best policy being implemented.

8 Conclusion

In this paper, I examine the behavior of ED physicians working in shifts. I describe evidence consistent with increasing costs of time as a consequence of scheduling work. Increased time costs near EOS result in two strategic responses by physicians, given that they are expected
to complete work prior to leaving. 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 the end of shift (EOS) approaches. Physicians reduce the time that it takes for them to discharge patients by 59% in the last hour prior to EOS. As the input of time becomes more costly, physicians also modify the mix of other inputs in patient care, by substituting to more formal diagnostic tests and treatment in the last hour prior to EOS. Finally, as time becomes too scarce for precise diagnosis and sufficient management, patients are more likely to be admitted. Thus, physicians incur 21% more costs for patients arriving in the last hour prior to EOS.

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 and must be distributed *ex post*. Unlike machines, workers who are nominally on duty are subject to behavioral incentives that may distort their choice and performance of work. Although shift work is pervasive in many settings, including health care, discussion on the organization of firms with shift work has been surprisingly silent on the behavioral effect of ending shifts. I show that this effect can be quite large not only in terms of time, but also in terms of distortions in other inputs and in the main product of discharge decisions. Using variation in the time between a peer’s arrival and the index physician’s EOS, I show a tradeoff between these two margins of moral hazard. This suggests a second-best optimality to (potentially large) “presenteeism” as a policy near EOS, in which workers are required to be present but are not given work.
References


Table 1: Sample Definition

<table>
<thead>
<tr>
<th>Sample Description or Step</th>
<th>Variables Added</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Raw visit data</td>
<td>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)</td>
<td>442,244</td>
</tr>
<tr>
<td>2. Drop visits with patients leaving before being assigned by physician or discharged</td>
<td>426,899</td>
<td></td>
</tr>
<tr>
<td>3. Merge with physician order data and bed audit data</td>
<td>Detailed physician orders with timestamps for medication, intravenous fluids, laboratory tests, radiology tests, and nursing orders; timestamps for bed movements</td>
<td>411,198</td>
</tr>
<tr>
<td>4. Merge with pod schedules</td>
<td>Shift types, start times, end times, and managerial locations</td>
<td>398,563</td>
</tr>
<tr>
<td>5. Identify visits with physician of record in visit data matching with schedules</td>
<td>372,224</td>
<td></td>
</tr>
</tbody>
</table>

**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 2: Shift Type Observation Numbers

<table>
<thead>
<tr>
<th>Shift type</th>
<th>Shifts</th>
<th>Hours</th>
<th>Potential Patients</th>
<th>Actual Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>(7, 0, 1)</td>
<td>95</td>
<td>665</td>
<td>1,645</td>
<td>1,160</td>
</tr>
<tr>
<td>(7, 1, 0)</td>
<td>237</td>
<td>1,659</td>
<td>6,674</td>
<td>2,597</td>
</tr>
<tr>
<td>(7, 1, 1)</td>
<td>101</td>
<td>707</td>
<td>4,281</td>
<td>1,783</td>
</tr>
<tr>
<td>(8, 0, 1)</td>
<td>319</td>
<td>2,552</td>
<td>8,453</td>
<td>4,952</td>
</tr>
<tr>
<td>(8, 1, 0)</td>
<td>174</td>
<td>1,392</td>
<td>7,440</td>
<td>1,981</td>
</tr>
<tr>
<td>(9, 0, 1)</td>
<td>3,453</td>
<td>30,879</td>
<td>84,292</td>
<td>58,589</td>
</tr>
<tr>
<td>(9, 0, 2)</td>
<td>325</td>
<td>2,349</td>
<td>6,411</td>
<td>4,541</td>
</tr>
<tr>
<td>(9, 0, 4)</td>
<td>408</td>
<td>2,898</td>
<td>9,326</td>
<td>4,839</td>
</tr>
<tr>
<td>(9, 0, 6)</td>
<td>364</td>
<td>3,276</td>
<td>16,186</td>
<td>5,899</td>
</tr>
<tr>
<td>(9, 1, 0)</td>
<td>3,414</td>
<td>30,528</td>
<td>118,030</td>
<td>59,897</td>
</tr>
<tr>
<td>(9, 1, 1)</td>
<td>2,909</td>
<td>26,181</td>
<td>116,108</td>
<td>54,221</td>
</tr>
<tr>
<td>(9, 1, 4)</td>
<td>2,249</td>
<td>19,170</td>
<td>80,279</td>
<td>28,694</td>
</tr>
<tr>
<td>(9, 1, 5)</td>
<td>60</td>
<td>540</td>
<td>2,554</td>
<td>892</td>
</tr>
<tr>
<td>(9, 1, 6)</td>
<td>211</td>
<td>1,899</td>
<td>8,157</td>
<td>2,524</td>
</tr>
<tr>
<td>(9, 2, 0)</td>
<td>464</td>
<td>3,294</td>
<td>12,027</td>
<td>6,317</td>
</tr>
<tr>
<td>(9, 3, 1)</td>
<td>485</td>
<td>3,277</td>
<td>17,013</td>
<td>6,699</td>
</tr>
<tr>
<td>(9, 3, 3)</td>
<td>60</td>
<td>540</td>
<td>3,226</td>
<td>1,089</td>
</tr>
<tr>
<td>(9, 4, 0)</td>
<td>347</td>
<td>2,347</td>
<td>9,966</td>
<td>3,994</td>
</tr>
<tr>
<td>(9, 4, 1)</td>
<td>212</td>
<td>1,908</td>
<td>8,974</td>
<td>3,370</td>
</tr>
<tr>
<td>(9, 4, 3)</td>
<td>426</td>
<td>2,752</td>
<td>16,730</td>
<td>5,344</td>
</tr>
<tr>
<td>(9, 4, 4)</td>
<td>772</td>
<td>5,094</td>
<td>26,094</td>
<td>9,413</td>
</tr>
<tr>
<td>(9, 4, 6)</td>
<td>2,141</td>
<td>19,269</td>
<td>99,726</td>
<td>29,007</td>
</tr>
<tr>
<td>(9, 5, 3)</td>
<td>60</td>
<td>540</td>
<td>2,851</td>
<td>1,043</td>
</tr>
<tr>
<td>(9, 6, 0)</td>
<td>634</td>
<td>5,706</td>
<td>34,943</td>
<td>9,244</td>
</tr>
<tr>
<td>(9, 6, 1)</td>
<td>1,504</td>
<td>13,536</td>
<td>61,197</td>
<td>21,861</td>
</tr>
<tr>
<td>(9, 6, 4)</td>
<td>575</td>
<td>5,175</td>
<td>31,088</td>
<td>9,597</td>
</tr>
<tr>
<td>(9, 9, 1)</td>
<td>353</td>
<td>3,177</td>
<td>15,965</td>
<td>4,598</td>
</tr>
<tr>
<td>(10, 0, 0)</td>
<td>176</td>
<td>1,760</td>
<td>4,812</td>
<td>2,578</td>
</tr>
<tr>
<td>(10, 0, 1)</td>
<td>243</td>
<td>2,430</td>
<td>5,783</td>
<td>4,615</td>
</tr>
<tr>
<td>(10, 0, 2)</td>
<td>137</td>
<td>1,040</td>
<td>2,631</td>
<td>1,901</td>
</tr>
<tr>
<td>(10, 0, 4)</td>
<td>139</td>
<td>1,050</td>
<td>3,616</td>
<td>2,378</td>
</tr>
<tr>
<td>(10, 1, 0)</td>
<td>277</td>
<td>2,770</td>
<td>9,092</td>
<td>4,401</td>
</tr>
<tr>
<td>(10, 4, 0)</td>
<td>139</td>
<td>1,050</td>
<td>4,335</td>
<td>1,834</td>
</tr>
<tr>
<td>(12, 0, 0)</td>
<td>142</td>
<td>1,704</td>
<td>4,119</td>
<td>2,423</td>
</tr>
<tr>
<td>(12, 4, 9)</td>
<td>319</td>
<td>3,828</td>
<td>16,490</td>
<td>5,566</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>23,924</td>
<td>206,942</td>
<td>860,544</td>
<td>369,841</td>
</tr>
</tbody>
</table>

**Note:** This table lists the number of observations for each shift type, each defined as <l, o_a, o_b>, where l is the shift length in hours, o_a is the overlap in hours with a previous shift, and o_b 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 <l, o_a, o_b> in progress), and actual patients (patients who are treated by a physician on a shift of type <l, o_a, o_b>).
**Table 3: End of Shift Effect on Log Length of Stay**

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

**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 4: End of Shift Effect on Other Outcomes

<table>
<thead>
<tr>
<th>Hour prior to EOS</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Order count</td>
<td>Inpatient admission</td>
<td>Log total cost</td>
<td>30-day mortality</td>
<td>14-day bounce-back</td>
</tr>
<tr>
<td>Last hour</td>
<td>1.411**</td>
<td>0.057**</td>
<td>0.208**</td>
<td>-0.003</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td>(0.562)</td>
<td>(0.024)</td>
<td>(0.080)</td>
<td>(0.008)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Second hour</td>
<td>-0.093</td>
<td>0.000</td>
<td>0.027</td>
<td>-0.001</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>(0.302)</td>
<td>(0.013)</td>
<td>(0.043)</td>
<td>(0.004)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Third hour</td>
<td>-0.003</td>
<td>0.002</td>
<td>0.009</td>
<td>-0.005</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(0.249)</td>
<td>(0.011)</td>
<td>(0.036)</td>
<td>(0.004)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Fourth hour</td>
<td>0.167</td>
<td>0.004</td>
<td>0.029</td>
<td>-0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.207)</td>
<td>(0.009)</td>
<td>(0.030)</td>
<td>(0.003)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Fifth hour</td>
<td>0.239</td>
<td>-0.004</td>
<td>0.034</td>
<td>-0.002</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.171)</td>
<td>(0.007)</td>
<td>(0.024)</td>
<td>(0.003)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Sixth hour</td>
<td>0.192</td>
<td>-0.007</td>
<td>-0.006</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.006)</td>
<td>(0.019)</td>
<td>(0.002)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>371,421</td>
<td>371,421</td>
<td>366,219</td>
<td>371,421</td>
<td>371,421</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.531</td>
<td>0.459</td>
<td>0.472</td>
<td>0.295</td>
<td>-0.044</td>
</tr>
<tr>
<td>Sample mean log length of stay (log hours)</td>
<td>13.518</td>
<td>0.269</td>
<td>6.750</td>
<td>0.018</td>
<td>0.060</td>
</tr>
</tbody>
</table>

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. Controls are as described for Table 3. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level.
Figure 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 2: Shift Variation

Note: This figure illustrates the variation in observations across shift types. Panel A plots shifts by shift ending time and shift length. Panel B plots shifts by shift ending time and the length of overlapping transition at the end of shift.
Figure 3: Density of Visits on Arrival Time and Length of Stay

**Panel A: Visit Count**

**Panel B: Density within Arrival Time**

**Note:** This figure plots the distribution of visits over arrival times relative to EOS and length of stay. Panel A plots visit counts within 15-minute intervals of arrival time and length of stay. Panel B plots the density of visits, conditional on arrival time.
Figure 4: 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 blue circles; patient visits for the location are shown in open red 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.
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 6: 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 7: Patient Selection Relative to End of Shift

A: Length of Stay

B: Orders

C: Inpatient Admission

D: Costs

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 (closed red circles) and on the full set of characteristics, including endogenous diagnoses (open red circles). Coefficients are estimated for predicted outcome using Equation (4). For reference, adjusted effects on actual outcomes from Figure 5 are shown in closed blue circles. The reference category is any time greater than six hours prior to EOS.
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 9: 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 (5). Marginal effects are shown in closed circles; 95% confidence intervals are shown in dashed lines.
Figure 10: 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 11: 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 (6). Estimates for terminal shifts (no overlap) are shown in open triangles; estimates for minimally transitioned shifts (one-hour overlap) are shown in open circles; and estimates for substantially transitioned shifts (four or more hours of overlap) are shown in closed circles.