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

David C. Chan†

October 26, 2017

Abstract

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

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

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

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

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

If workers overvalue their leisure time relative to other consequences of their workplace actions, schedules induce two distortions near end of shift (EOS): First, on an extensive margin, workers “slack off” by accepting fewer tasks than socially optimal. Second, on an intensive margin, workers may rush to complete their work, spending less time than socially optimal on assigned tasks near EOS. These effects are particularly important in work that is uncertain, time-sensitive, and information-rich:\(^2\) For example, if tasks are uncertain and time-sensitive, as they are in many environments with scheduling, then they cannot be pre-assigned. Performance on information-based tasks is often difficult to evaluate and therefore non-contractible. Finally, once a worker is assigned a task, worker-task specificity, for example due to tacit knowledge or personal relationships intrinsic to the task, makes passing the task to another worker costly.

In this paper, I study the effects of schedules on the behavior of emergency department (ED) physicians working in shifts. Shifts ending at different times, particularly due to changes in the shift schedule, allow me to separate effects related to shift work from differences due to the time of day. Shifts of different lengths allow separating these effects from “fatigue,” which I consider to

---

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

\(^2\) Examples of such workplaces could include large-scale construction, management consulting, and software engineering. A number of online worker scheduling services have emerged, and case studies of client firms across industries can be found, for example, at [https://www.shiftplanning.com/casestudies](https://www.shiftplanning.com/casestudies). The industry need not be 24-7, although the size of the economy involved in 24-7 activities has grown (Presser, 2003), and the portion of the economy with irregular or non-standard work times has grown significantly (Beers, 2000; Katz and Krueger, 2016). Key questions in these settings is how to schedule worker availability and how to assign tasks to workers.
depend on the time since the beginning of shift. Physicians work in virtually all types of shifts. I show that physicians accept fewer patients near EOS. For patients they are assigned, I also show that physicians shorten the duration of care (“length of stay”) in the ED and increase formal utilization, inpatient admissions, and hospital costs as the time of patient arrival approaches EOS. To support the claim that I have identified causal effects of time to EOS on patient care, I utilize patient characteristics plausibly unobserved at the time of assignment and quasi-random variation in the propensity of assignment to a shift nearing EOS as a function of patient arrival times at the ED.

To interpret changes in patient care as distortions, I use another source of variation from shift structure: the overlapping time between when a peer arrives on a new shift and when the index physician reaches EOS. My assumption is that, conditional on the volume of work, the time from the beginning of the shift, and the time from the peer’s arrival, the EOS should have no bearing on efficient patient-care decisions. In other words, conditional on these other characteristics of the work environment, the EOS is merely an arbitrarily varying rule stipulating when physicians may go home if work is complete. I show that distortions on the intensive margin of patient care are greatest when physicians have the least time to offload work onto a peer before EOS. In fact, there is no increased utilization or admissions when overlap is four or more hours. Further, distortions appear concentrated in shifts ending during daylight hours, when non-work plans may be costlier to defer. When EOS is during daylight, physicians are less likely to write orders past EOS, and they shorten patient stays to a greater extent when approaching EOS, consistent with physicians seeking leisure.

This evidence suggests a policy tradeoff between the extensive and intensive margins of distortion. On the extensive margin, workers “slack off” by accepting fewer patients. Although assigning more patients to physicians near EOS would reduce slacking off, it would increase the workload near EOS and the pressure for physicians to rush, inefficiently substituting other inputs for time. I analyze this tradeoff in a stylized theoretical model featuring schedules, worker-task specificity, and non-contractable performance of tasks. This model show that workers prefer fewer

---

3The idea of time per effective work is related to work by Coviello et al. (2014), who discuss the effect of dividing time among tasks, although with a single worker who works indefinitely. The time for completing a project mechanically is lower when fewer projects are active because time is divided among fewer projects.
tasks (i.e., “slack off”) relative to the first-best assignment near EOS, but when performance on
the tasks cannot be contracted upon, the (second-best) optimal assignment policy still allows
some slacking off.

To assess the observed patient-assignment policy relative to counterfactual policies along
the dimension of slacking off, I specify and estimate a discrete choice dynamic programming
model. In this model, physicians care about how they discharge individual patients and the
number of patients they are left with at EOS. As physicians proceed through their shifts, their
choices to discharge patients are influenced by expectations of future patient assignments. The
model allows me to simulate costs of physician time, patient time, and hospital resources under
counterfactual assignment policies. Assigning more patients near EOS such that physicians stay
an additional hour induces an additional $3,300 in hospital spending per shift; physicians also
reveal that they are willing to spend more than $790 in hospital dollars per each hour of leisure
saved, which is more than six times greater than the market wage. Consistent with the high
marginal rate of technical substitution between physician time and hospital resources, overall
costs rise steeply under counterfactual policies that assign more patients near EOS but change
relatively little under policies that assign fewer patients. Interestingly, I find that the observed
assignment policy approximately minimizes overall costs.

This paper contributes to two strands of literature. First, a central economic question is how
to induce workers to work efficiently, analyzed through the lens of incomplete contracts and the
principal-agent problem (Simon, 1947; Hart and Holmstrom, 1987). Following seminal papers
that evaluate a manager’s second-best optimal policy under hidden action or information, (e.g.,
Shapiro and Stiglitz, 1984; Aghion and Tirole, 1997; Milgrom and Roberts, 1988), I apply this
framework to the design of scheduling and assignment, and I find that work assignment should
be lower than first-best near EOS. This paper also contributes to an empirical literature on
the relationship between workplace design and productivity. As work environments featuring
flexible or irregular hours grow in prominence, recent empirical work has examined the effects
of such workplaces on productivity (Bloom et al., 2015) and workers’ underlying preferences for
such arrangements (Moen et al., 2016; Mas and Pallais, 2016).

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

4
Second, this paper sheds empirical light on the balance between extrinsic and intrinsic motivation (e.g., Benabou and Tirole, 2003). While workers no doubt care about their income and leisure, a now-substantial literature in economics recognizes that workers care about the “mission” of their job. In medical care, where information is continuous, multidimensional, and difficult to communicate, it would be extremely difficult to design incentives to provide the right care for patients if physicians only cared about income and leisure. By construction, salaries and schedules provide an environment in which extrinsic motives are muted relative to intrinsic ones, but the boundaries of schedules present a unique opportunity to study the tradeoff between private and intrinsic mission-oriented goals.

The issues I study in this paper are particularly relevant to health care delivery, which has experienced broad changes in the use of labor over the last few decades. Technological advances have dramatically increased the number of diagnostic and therapeutic decisions that should be made in rapid order from a patient’s presentation. Further, changes in work and society, including the emergence of dual-earner families, have driven worker preferences for more predictable yet flexible hours (e.g., Goldin, 2014; Presser, 2003). Thus, increasingly, health care is delivered by organizations, and schedules play an important role in assigning uncertain work (e.g., Briscoe, 2006; Casalino et al., 2003). These changes of course have parallels in other industries, which also feature increasingly interrelated and complex production.

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

---

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

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

2.1 Shift Work

I study a large, academic, tertiary-care ED in the US with a high frequency of patient visits. Like in virtually all other EDs around the country, work is organized by shifts. In the study sample from June 2005 to December 2012, shifts range from seven to twelve hours in length (ℓ). Shifts also differ in overlap with a previous shift (φ) or with a subsequent shift (π) in the same location. I observe 23,990 shifts in 35 different shift types summarized by (ℓ, φ, π) (Table A-7.1).

For physicians working in these shifts, the end of shift (EOS) is simply the time after which they are allowed to go home if they have completed their work. Because I focus on behavior at EOS, I pay special attention to π. This overlap is the time prior to EOS during which a physician shares new work with another physician who has begun work in the same location.7 “Location” refers to a set of beds in the ED in which a physician may treat patients. This managerial definition may differ from broader physical areas, or “pods,” where physicians may see each other but may not share the same beds. That is, a pod may contain more than one managerial location. During my sample period, I observe two to three pods, with a new pod opening in May 2011, that at various times were divided into two to five managerial locations.

In the study period, the ED underwent 15 different shift schedule changes at the location-week level. Within each regime, the pattern of shifts could differ across day of the week (see Figure 1). As is common in scheduled work, shift times were designed around estimated workload needs, and schedule changes reflected changes in the flow of patients to ED. Some shift regime changes were merely minor tweaks in the times of specific shifts, while others involved larger changes. Shifts are scheduled many months in advance, and physicians are expected to work in all types of shifts at all times and locations. Physicians may only request rare specific shifts off, such as holidays and vacation days, and shift trades are rare. During a shift, physicians cannot control the volume of patients arriving to the ED or the patient types that the triage nurse assigns to beds. Throughout the entire study period, physicians were exposed to the same financial incentives: They were paid a clinical salary based on the number of shifts they work.

---

7I distinguish between shifts that end with the closure of a patient location, or “terminal shifts” with π = 0, and those continuing patient care with another shift in the same location, or “transitioned shifts” with π > 0.
with a 10% bonus based on clinical revenue (measured by Relative Value Units, or RVUs, per hour) and modified by research, teaching, and administrative metrics.\textsuperscript{8} Although their salaries are based on numbers of shifts worked, physicians are not compensated for time worked past EOS.\textsuperscript{9}

### 2.2 Patient Care

After arrival at the ED, patients are assigned to a bed by a triage nurse. This assignment determines the managerial location for the patient and therefore the one or more physicians who may assume care for the patient. Once the patient arrives in a bed, a physician may sign up for that patient, if the patient is in her managerial location. Physicians are expected to complete work on any patient for whom they have assumed care, in order to reduce information loss with hand-offs (e.g., Apker et al., 2007), except in uncommon cases where the patient is expected to stay much longer in the ED. Because of this, physicians report often staying two to three hours past EOS.\textsuperscript{10} For patients arriving near EOS, physicians may opt not to start work and leave the patient for another physician. This option is more acceptable if this physician peer will arrive soon or has already arrived in the same location.

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

\textsuperscript{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.

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

\textsuperscript{10}In shifts with greater overlap, which have become more common, physicians report staying shorter amounts of time, but still up to one hour past EOS. Quantitative evidence using attending physician orders is presented in Figure A-7.1.
For physicians in the ED, the concept of patient discharge is a matter of discretion. Patient care is usually expected to continue after discharge, in either outpatient or inpatient settings. The key criterion for completion of work – or discharge – is whether the physician believes that sufficient information has been gathered for a discharge decision out of the ED. This decision is often made with incomplete diagnosis and treatment. Rather, the physician may decide to discharge a patient home with outpatient follow-up after “ruling out” serious medical conditions, or the physician may admit the patient for inpatient care if the patient could still possibly have a serious condition that would make discharge home unsafe.\(^\text{11}\)

Physicians may gather the information they need to make the discharge decision in several ways. Formal diagnostic tests are an obvious way to gain more information on a patient’s clinical condition. Treatment can also inform possible diagnoses by patient response, such as response to bronchodilators for suspected asthma. But time – for a careful history-taking, physical examination, serial monitoring, or a well-planned sequence of formal tests and treatment – remains an important input in the production of information. Diagnostic tests and treatments can be complements or substitutes for time: Formal tests (e.g., CT and MRI scans) take time to complete and can thus prolong the length of stay, but testing can also substitute for a careful questioning or serial monitoring to gather information more rapidly.

### 2.3 Observations and Outcomes

From June 2005 to December 2012, I observe 442,244 raw patient visits to the ED. I combine visit data with detailed timestamped data on physician orders, patient bed locations, and physician schedules to yield a working sample of 372,224 observations. Details of the sample definition process are described in Table A-7.2. In the sample, I observe the identities of 102 physicians, 1,146 residents and physician assistants, and 393 nurses.

Because I focus on behavior near EOS, I present in Figure 2 the key variation across the 23,990 shifts in the time of day for EOS, shift length, and the overlap with another shift at EOS. Table A-7.1 lists the underlying number of observations for each shift type, in terms of hours.

---

\(^{11}\)In this ED, there is yet a third discharge destination to “ED observation,” if the patient meets certain criteria that make discharge either home or to inpatient unclear and justify watching the patient in the ED for a substantial period of time (usually overnight) to watch clinical progress.
potential patients who arrive during a time when a shift of that type is in progress, and actual patients who are seen by a physician working in a shift of that type.

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

Since the primary product of ED care is the physician’s discharge decision, I focus on the decision to admit a patient as a key outcome measure, which has has also received attention as a source of rising system costs (Schuur and Venkatesh, 2012; Forster et al., 2003). I accordingly measure total direct costs, which are the hospital’s internal measures of costs incurred both in the ED and possibly during a subsequent admission. Finally, I measure thirty-day mortality, occurring in 2% of the sample visits, and return visits to the ED within 14 days (“bounce-backs”), occurring in 7% of the sample (Lerman and Kobernick, 1987). However, these latter outcomes are less strongly influenced by the ED physician and depend on a host of factors outside the ED and hospital system, reducing the precision of their estimated effects.

2.4 Patient Observable Characteristics

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

In addition to elements displayed via the ED computer interface, I observe patient language, race, zip code of residence, and diagnostic information. The last characteristic of diagnoses is only 12

---

12 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). These costs are internal valuations of actual resources. They are not charges and are unrelated to billing or revenue source.
incompletely known by physicians (or anyone) prior to assignment (via the “chief complaint”),
especially since physicians do not interact with patients or examine their medical records prior
to accepting them. I codify the diagnostic information into 30 Elixhauser indicators based on
diagnostic ICD-9 codes for comorbidities (e.g., renal disease, cardiac arrhythmias) that have
been validated for predicting clinical outcomes using administrative data (Elixhauser et al.,
1998). Diagnostic codes are recorded after patients are seen by the physician and discharged and
therefore are also partly determined by patient care.

3 Patient Assignment

In this section, I describe patient assignment near EOS, a function of both triage nurse choices to
assign patients to locations and physician choices to “accept” patients in their location. Figure
3 presents the hourly average rates of new patient assignments in 30-minute bins relative to
EOS. Each panel represents shifts with a different EOS overlap, $\sigma$, and shows assignments for
the index physician (patients accepted) and for the location inclusive of the index physician
(patients assigned by the triage nurse). Physicians are generally assigned two or three new
patients in an hour, and assignment rates are highest near the beginning of shift. For $\sigma > 0$,
assignment rates show two relationships with time. First, patient flow declines in the hour prior
to the transitioning peer’s arrival at the location. Second, patient flow declines close to zero in
the two to three hours prior to EOS. If there is sufficient $\sigma$, patient flow is relatively constant
but diminished in that duration. For $\sigma = 0$, the decline in patient flow begins earlier, at least
four hours prior to EOS. Also in Figure 3, patients who are not accepted by the index physician
may wait up to an hour to be seen by a peer yet to arrive, but patient flow to transitioning
peers generally at least makes up for the decline in flow for the index physician. That is, despite
decreases in patient assignment to the leaving physician, patients continue to arrive at the pod at
similar or greater rates prior to the peer’s transitioning shift.

The earlier arrival of peers allows for earlier reductions in patient assignment relative to EOS.
The reductions are in fact prior to peer arrival, especially in shifts with shorter transitions,
suggesting forward-looking behavior. For terminal shifts with no other physician working near
EOS in the same location, the long decline in patient flow rates results from the triage nurse assigning fewer patients to the location. Thus, assignment to physicians in general and “slacking off” in particular is achieved both by coworkers sharing a location and by triage nurse assignment to locations.

4 Effect on Patient Care

4.1 Identification and Balance

The panel nature of the data allows for me to control for time categories (e.g., time of the day or day of the week) because shifts start at different times. Furthermore, variation in shift lengths identifies the EOS effect separately from fatigue or other effects that depend on time relative to the beginning of shifts.\(^\text{13}\) Finally, I observe the same physicians and support staff (i.e., physician assistants, nurses, and residents) and can therefore control for their identities as a given physician approaches EOS. However, given that the number of assigned patients clearly declines as physicians near EOS, an important question is how the EOS effect on patient care is identified separately from patient selection correlated with time to EOS.

Figure 4 shows statistics of characteristics for patients assigned to physicians in 30-minute bins relative to EOS. Compared to numbers of patients assigned (Figure 3), and compared to variation within bins, patient characteristics are relatively constant across bins. There is a slight trend toward younger patients being assigned to physicians nearing EOS. Figure 5 similarly shows that the distribution of predicted length of stay, using \textit{ex ante} characteristics of age, sex, ESI, race, and language that are presumably observable at the time of assignment, is quite stable across time relative to EOS.\(^\text{14}\) This descriptive evidence suggests that assignment policies relative to EOS are primarily based on patient numbers rather than patient characteristics.

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

\textbf{Assumption 1 (Excludable Characteristics).} \textit{Conditional on ex ante patient characteristics, time of arrival, pod, and providers, patient potential outcomes are mean independent of assigned

\(^{13}\text{In alternative models, I also control for cubic splines of total number of patients seen prior to the index patient’s arrival. Results (not shown) are essentially identical with these additional controls.}\)

\(^{14}\text{Appendix A-1.1 quantifies this selection in terms of predicted outcomes.}\)
time relative to EOS.

**Assumption 2 (Random Arrival Times).** Conditional on time categories of arrival (e.g., day of week, time of day), pod, and providers, patient potential outcomes are mean independent of ED arrival times with different propensities for assignment to times relative to EOS.

The intuition behind Assumption 1 is that assignment within the ED operates through patient numbers and (to a much lesser extent) on *ex ante* patient characteristics. Other patient characteristics that are correlated with potential outcomes are mostly unknown to physicians before assignment and, under this assumption, excluded from assignment policies. This assumption is testable to the extent that I observe *ex post* clinical characteristics and can show that, conditional on *ex ante* patient characteristics, time categories, and staff identities, there exists no correlation between clinical characteristics and assignment to times relative to EOS. The intuition behind Assumption 2 is that, although ED staff may influence patient assignment within the ED, patients arrive at the ED without any systematic selection toward times when there may or may not be a physician near EOS. Specifically, variation in shift schedules within a time category of ED arrival drives the propensity of being assigned to a physician near EOS but is mean independent of potential outcomes of the arriving patients.

I assess the plausibility of each assumption in a regression framework. To assess Assumption 1, I regress presumably excludable patient characteristics based on Elixhauser diagnoses – specifically (i) predicted length of stay by diagnoses and (ii) count of diagnoses – on hour relative to EOS, controlling for *ex ante* patient characteristics, time categories (hour relative to beginning of shift, hour of the day, day of the week, and month-year interactions), pod, and provider identities. To evaluate Assumption 2, I regress both *ex ante* and *ex post* patient characteristics on hourly propensities for assignment to a physician at a time relative to her EOS, controlling for

---

15. This assumption is strengthened by the institutional fact that triage in the ED is supposed to be sufficiently summarized by the *ex ante* characteristic of ESI (Tanabe et al., 2004). Physicians are discouraged from further assessing patients prior to accepting them.

16. I also condition on pod and providers because time of arrival at triage is not always observed in the data, and I use to time at *ED floor* is my preferred measure of arrival time. The time between patient arrival at triage and arrival at ED floor may differ across patient types. For example, for patients sent to a 24-hour pod vs. a partial-day pod may be sicker; patients sent to a skilled physician may also be sicker. By conditioning on physician and pod identities, I assert that differences in the endogenous time between triage and ED floor that are predictive of patient outcomes are only correlated with physician and pod identities. In practice, however, controlling for physician and pod identities does not matter much for balance.
time categories, pod, and provider identities. I consider (i) predicted length of stay by all patient characteristics, (ii) age, (iii) male sex, (iv) ESI, (v) white race, (vi) black race, (vii) English language, (viii) Spanish language, and (ix) count of Elixhauser indices. Appendix A-1.2 provides further details.

Table 1 reports results consistent with conditional random assignment under either Assumption 1 (the first two columns) or Assumption 1 (the remaining columns). None of the regressions yield jointly significant coefficients on hour relative to EOS. Furthermore, Table 1 shows that variation in any of the patient characteristics, residualized by covariates depending on the assumption, remains quite large. In fact, if Assumption 1 were violated, and if patients with the lowest predicted length of stay based on ex post clinical characteristics were perfectly sorted to hours closest to EOS when more than one physician is available, the bias in the last hour of shift would be $-0.177$, more than 17 standard errors apart from the corresponding coefficient in the corresponding balance test (Column 1 of Table 1). Table A-1.1 presents additional balance results for other predicted outcomes, assessing both Assumptions 1 and 2.

4.2 Main EOS Effects

In the full specification, I estimate the following equation:

$$Y_{it} = \alpha_{m(i,t)} + \gamma_{m(i,t)} + \beta X_i + \eta T_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \epsilon_{it}. \tag{1}$$

Outcome $Y_{it}$ is indexed for patient visit $i$ at time $t$, and the object of interest is arrival hour $m(i,t) = \lceil t(i) - t \rceil$ prior to EOS, where seven or greater hours prior to EOS is the omitted category. I control for time relative to the shift beginning $m(i,t) = \lfloor t - t(i) \rfloor$, patient characteristics $X_i$, time categories $T_t$ (for month-year, day of the week, and hour of the day), pod $p(i)$, and physician $j(i)$ and team $k(i)$ (i.e., resident or physician assistant, and nurse) identities.

Table 2 shows results for log length of stay from versions of Equation (1) with varying sets of controls. All models estimate highly significant and increasingly negative coefficients for approaching time to EOS, with visits seven or more hours prior to EOS being the reference category. By the last hour prior to EOS, versions of Equation (1) estimate effects on log length
of stay ranging from $-0.53$ to $-0.72$. The full model, shown in Column 5 of Table 2 and plotted in Panel A of Figure 6, estimates an effect on log length of stay of $-0.59$ in the last hour and serves as the baseline model for this paper.

Results are essentially indistinguishable whether or not all patient characteristics or only *ex ante* characteristics are included (Columns 5 and 6 in Table 2), as suggested by Assumption 1. More generally, results are qualitatively unchanged regardless of whether I control for *any* patient characteristics, time categories, pod dummies, provider identities, or time relative to shift beginning. For example, the difference between Columns 4 and 5 columns represents the effect of time relative to shift beginning, which can include fatigue and is separately identified from EOS effects due to variation in shift lengths. This difference, about 0.13 in the last hour prior to EOS, also accounts for only a minor portion of the overall effect.\(^\text{17}\)

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

\[
Y^*_{it} = \frac{1}{N_t} \sum_{i \in I(t)} Y^*_{it} \alpha_{\bar{m}} P_{\bar{m}} (t) + \varepsilon_t, \tag{2}
\]

where $Y^*_t \equiv \frac{1}{N_t} \sum_{i \in I(t)} Y^*_{it}$ is the average of residualized length of stay $Y^*_{it}$ for the set $I(t)$ of $N_t$ patients arriving at $t$, and $P_{\bar{m}} (t)$ is the fraction of these patients being assigned to a shift with $\bar{m}$ hours prior to its end.\(^\text{18}\) In this specification each observation is an arrival hour. Results in Column 7 and 8, which respectively either include or omit *all* patient characteristics, closely match each other and thus support Assumption 2.

In Appendix A-1.4, I adopt an approach by Altonji et al. (2005) to quantify selection on unobservables necessary to explain the observed EOS effects. Using the intuition that estimates change little regardless of controls, I find that normalized selection on unobservables would need

\(^{17}\)See Appendix A-2 for more direct results on effects relative to shift beginning.\(^{18}\)As argued by Chetty et al. (2014), $Y^*_{it}$ is calculated using *within* $\bar{m}$ variation. Section A-1.3 presents an approach and corresponding results that closely follow Chetty et al. (2014) in a way that estimates a single measure of “forecast bias” and additionally accounts for within-shift endogeneity by estimating “jack-knifed” predicted $\hat{\varepsilon}_{\bar{m}}$ for shift using data that does not include that shift. As suggested elsewhere, results in that section fail to reject the null hypothesis of no forecast bias.
to be 475 times greater than normalized selection on observables in order to explain the effect of the last hour before EOS on length of stay.

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

5 Shift Overlap, Workload, and Distortion

I evaluate how workload and patient-care effects vary across shifts with varying overlap near EOS, for two purposes: First, this supports the interpretation that EOS effects reflect inefficiency, under the assumption that the EOS by itself has no first-best implications for patient care, conditional on volume of work, time after beginning work, and time after a peer’s arrival. Second, this analysis uses shift structure as a concrete example of patient assignment as a policy lever with efficiency tradeoffs: Assigning fewer patients near EOS leaves physicians idle, but assigning more patients worsens the EOS distortion in patient care.

5.1 Patient Censuses over Time

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

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

19 This suggests that formal orders are a net substitute for time. See Appendix A-3 for more direct results supporting this hypothesis.
the count of visits arriving prior to $t$ at $t'(i)$ and staying past $t$ until $t'(i) + \tau(i)$, where $\tau(i)$ is length of stay.

Figure A-7.3 shows unadjusted census averages in 30-minute intervals in different shift types by EOS overlap, $\overline{\sigma}$. Average censuses start at around two patients, representing unstaffed patients from the previous shift, except for shift types with $\sigma = 2$, which happen not to transition from another shift (i.e., $\sigma = 0$). Patients remain on the census at EOS: Approximately four patients remain on the average census in the last 30 minutes prior to EOS, with the exception of shifts with $\sigma = 1$, which have censuses of about six.

5.2 EOS Effects by Shift Overlap

I next consider how patient-care EOS effects may differ by shift overlap. With smaller $\overline{\sigma}$, EOS defines earlier times relative to peer arrival after which physicians are allowed to go home. Larger patient-care effects with small $\overline{\sigma}$, conditional on time from beginning of shift, are consistent with distortionary care. Further, the interaction reflects an intuitive tradeoff between extensive and intensive margins of distortion: Patient care will be less distorted with larger $\overline{\sigma}$, but this increases slacking off.

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

$$Y_{it} = \left( \alpha_{m(i,t)} + \kappa_{\overline{\sigma}} \right) \mathbf{1}(\overline{\sigma}(i) \in \overline{\sigma}) + \gamma_{m(i,t)} + \beta X_i + \eta T_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_{it}, \quad (4)$$

which is similar to Equation (1) but interacts the hourly EOS effects with the categories $\overline{\sigma}$, where $\overline{\sigma}(i)$ is a function assigning visit $i$ to overlap $\overline{\sigma}$ of the shift to which the visit is assigned. In each of the overlap categories, the reference category of $m(i,t)$ includes times that are seven hours or more before EOS.

Figure 7 shows interacted EOS effects on length of stay, orders, admission, and total costs. The EOS effect on length of stay is largely similar across shift categories (Panel A). All three

---

20 While I observe shifts with $\overline{\sigma} \in \{2, 3\}$, they entail very few observations, as listed in Table A-7.1. Results are essentially unchanged whether I omit these observations or consider them as belonging to the minimally transitioned shift category.
shift categories show a substantial decline in length of stay as EOS approaches. However, EOS effects are absent in shifts with $\sigma \geq 2$ for orders, admission probability, and total costs (Panels B to D). In contrast, shifts with $\sigma \leq 1$ show large increases in orders, admissions, and total costs at EOS.

### 5.3 Effective Time per Patient

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

$$\frac{\tau(i)}{\bar{w}(i)} = \tau(i) \left[ \frac{1}{\tau(i)} \int_{t}^{\tau(i)+\tau(i)} w(j(i), \tilde{I}) \, d\tilde{I} \right]^{-1}, \quad (5)$$

where census $w(j, t)$ is defined by Equation (3).

I regress the log of workload-adjusted length of stay using Equation (1).\(^{21}\) The last column of Table 3 shows that workload-adjusted length of stay decreases significantly only in the last hour prior to EOS. Thus, adjusting length of stay for workload reconciles previous results in which length of stay progressively decreases as EOS approaches, but orders, admissions, and costs increase only in the last hour. In Table 4, I also show that workload-adjusted length of stay is only decreased in the last hour of shift when $\sigma \leq 1$. When $\sigma \geq 2$, workload-adjusted length of stay does not decrease near EOS and, if anything, slightly increases prior to the last hour of shift.\(^{22}\) These relationships suggests workload-adjusted length of stay as a relevant measure of

---

\(^{21}\) This is different than controlling for current census; results in Table 2 are unchanged when flexible splines of current census are included in Equation (1). Instead, workload-adjusted length of stay solely captures future actions by the physician, including future censuses. Otherwise including future censuses as covariates in a regression framework would be problematic.

\(^{22}\) If anything, workload-adjusted length of stay slightly increases prior to EOS when $\sigma \geq 2$. Such increases do not appear to be associated with changes in other outcomes of orders, admissions, or costs, which could be consistent with increases in length of stay for strategic purposes, or “foot-dragging,” as discussed in Chan (2016).
time that enters into the physician production function, particularly when considering distortions in orders, admissions, and costs relative to EOS.

5.4 Time of Shift End

Finally, I investigate whether EOS effects differ according to the shift end time of day. Specifically, physician leisure is likely to be more valuable at the margin during daytime, when non-work activity may be planned and costlier to defer. I specify two categories $T$ of shift end times: shifts with EOS during daytime (from 6:00 a.m. to 8:00 p.m.) and the remainder of shifts. I then estimate the following regression:

$$Y_{it} = \left( \alpha \overline{m}_{i,t} + \kappa_T \right) \mathbf{1} \left( \overline{t}(i) \in T \right) + \gamma_{m(i,t)} + \beta T_t + \eta X_i + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_{it},$$

where $Y_{it}$ is log workload-adjusted length of stay, and the reference category of $\overline{m}(i,t)$ includes times that are seven hours or more before EOS.

As shown in Figure 8, the EOS effect on workload-adjusted length of stay is concentrated in shifts ending during daytime but not in other shifts. This is consistent with the benchmark distortion mechanism of leisure-seeking when it is costlier to defer non-work activity. Further, in Figure A-7.2, I show that physicians are less likely to continue to write orders past EOS in shifts ending during daytime, despite higher remaining patient loads at EOS in these shifts.

6 Stylized Model of Optimal Assignment

I introduce a simple model to consider how physician decisions may be distorted under work schedules. The key distortionary elements of the model are the following: (1) Workers have private information about their tasks; (2) workers care more about their own income and leisure relative to the social consequences of their actions; and (3) ex post worker-task specificity prevents workers from simply passing off tasks at EOS (Briscoe, 2007; Goldin, 2014). While the assignment

23Mas and Pallais (2016) find that workers strongly dislike employers setting schedules on short notice and have a preference for “regular hours,” ending at 5:00 p.m. A natural reason for this is that workers value non-work activities that require coordinating with others. Although activities done by oneself, such as sleeping, is valuable, these activities are less costly to defer by one or two hours, if unforeseen circumstances at work occur.
of tasks (patients) is observable, subsequent performance on the tasks is not contractible. This implies a second-best assignment policy that takes this into account.

6.1 Model Setup

Consider a physician in a shiftwork arrangement: She has a contract to work on a shift until EOS \( t \) or whenever she discharges her last patient, whichever is later, and she will receive a lump-sum payment \( y \) for this. Now consider a patient arriving at time \( t < t \). The relevant welfare parameters of her work environment is captured by \( E_t \), which includes, among other things, the start time of the physician’s shift, her workload, \( w_t \), and the start time and workload of a potential peer who may take the patient instead. The patient’s underlying health state, \( \theta \in \{0, 1\} \) for whether the patient is healthy (\( \theta = 0 \)) or sick (\( \theta = 1 \)), is unknown at this point, but \( \Pr (\theta = 1) = p \) is publicly known. The timing is as follows:

1. The physician may be assigned the patient (\( a = 1 \)) or not (\( a = 0 \)).

2. If she is assigned the patient, she observes private information \( I \) so that \( \Pr (\theta = 1|I) = p' \), and \( |p' - \theta| < |p - \theta| \).\(^{24}\) She decides on patient care inputs: time \( \tau \) and formal tests and treatments \( z \).

3. The physician observes \( \hat{\theta} = \theta \) with probability \( q (\tau, z) \in (0, 1) \) and no information (\( \hat{\theta} = \emptyset \)) with probability \( 1 - q (\tau, z) \). She decides \( d \in \{0, 1\} \), 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:

\[
U = \begin{cases} 
    y + \lambda O (\theta; E_t), & a = 0 \\
    y - \hat{c}_x (\tau) + \lambda (v (\theta, d) - c (\tau, z)), & a = 1 
\end{cases}
\]  

\[(7)\]

Utility is stated in dollar terms, where physician income \( y \) does not depend on her actions.\(^{25}\)

\(^{24}\)I rule out private information before patient acceptance in this model. This is generally consistent with the institutional setting, and I examine this empirically as Assumption 1 below.

\(^{25}\)In scheduled work \( y \) mostly depends \textit{ex ante} availability, not \textit{ex post} time past EOS. This model can accommodate some rewards correlated with staying past EOS (e.g., financial incentives for seeing more patients, social recognition); all that it requires is that physicians are relatively uncompensated for leisure.
\( O(\theta; \mathcal{E}_t) \) is the value of the “outside option” if \( a = 0 \), which depends on \( \theta \) and the work environment \( \mathcal{E}_t \). \( v(\theta, d) \) is the value of making discharge decision \( d \) for patient with health \( \theta \). \( c(\tau, z) \) is the cost of patient care inputs, from which I separate \( \tilde{c}_\tau(\tau) \), the cost of foregone leisure if the physician stays past EOS. \( \lambda \in (0, 1) \), and \( 1 - \lambda \) is the wedge by which the physician undervalues the mission of patient care.

To be clear about the wedge, first consider the social welfare function as equivalent to Equation (7), except without \( \lambda \) (i.e., \( \lambda = 1 \)). As \( \lambda \to 1 \), physician utility approaches social welfare, and the agency problem disappears. As \( \lambda \to 0 \), utility approaches the standard labor supply model in which workers only care about consumption and leisure. If \( \lambda = 0 \) (which I rule out), the physician would have no incentive to make the right decisions (despite observing \( \mathcal{I} \) and sometimes \( \theta \)).

### 6.2 Patient Care

I first examine EOS effects on patient care and the discharge decision, conditional on assignment \( (a = 1) \). Discharge decisions have important efficiency implications for resource utilization and patient health. Formally, patients with \( \theta = 0 \) should be discharged home, while those with \( \theta = 1 \) should be admitted: \( v(0, 0) > v(0, 1) \) and \( v(1, 1) > v(1, 0) \). Discharging a sick patient home is particularly harmful: \( v(1, 1) - v(1, 0) > v(0, 0) - v(0, 1) \).\(^{26}\) Because of this last fact, if \( \theta \) remains unobserved, the physician will admit if and only if \( p' > p^* \), where \( p^* < \frac{1}{2} \).\(^{27}\) In other words, the realized discharge decision \( d \) is a function of \( \hat{\theta} \):

\[
d(\hat{\theta}) = \begin{cases} 
\theta, & \hat{\theta} = \theta \\
1(p' > p^*), & \hat{\theta} = \emptyset 
\end{cases}
\]

The probability \( q \) of observing \( \theta \) is in turn a function of patient care inputs \( (\tau, z) \).\(^{28}\) \( q(\tau, z; w_t) \) is increasing and concave with respect to both \( \tau \) and \( z \) and also depends on workload of

---

\(^{26}\)I assume that the physician values discharge actions relative to the patient’s health state the same as the social planner does. An additional wedge of defensive medicine would further increase the physician’s \( v(1, 1) - v(1, 0) \) relative to \( v(0, 0) - v(0, 1) \) but not the social planner’s valuations. An EOS distortion in patient care, in concert with this distortion, would likely be larger, given concavity in \( q(\tau, z) \) described formally below.

\(^{27}\)This can be straightforwardly shown by noting that \( E[V|d = 0, p' = p^*] = E[V|d = 1, p' = p^*] \).

\(^{28}\)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, except that if \( z^* \) is increasing in \( p \), then physicians will be less likely to accept ex ante sicker patients.
the physician $w_t$. $\tau$ and $z$ may be net substitutes ($\partial^2 q / (\partial \tau \partial z) < 0$) or net complements ($\partial^2 q / (\partial \tau \partial z) > 0$) in production. Effective time per patient is reduced with higher workload $w_t$: $\partial^2 q / (\partial \tau \partial w_t) < 0$. This contrasts with formal inputs, for which I make the normalizing assumption $\partial^2 q / (\partial z \partial w_t) = 0$. The intuition behind this is that with more patients, a physician has to divide her time and attention between them, but formal utilization can be ordered with the click of a mouse.\footnote{In its most straightforward form, effective time per patient can be viewed as the concept behind “workload-adjusted length of stay”, in Section 5.3, which is length of stay divided by the average number of patients on census during a given patient’s length of stay. This concept reflects that the cost of a patient’s length of stay as an input does not depend on how many other patients a physician concurrently cares for, but the effectiveness of this time as an input into $q$ is reduced by workload.} By the normalizing assumption, I focus attention on substitutability or complementarity between time and formal utilization.

**Proposition 1.** Denote inputs in Section 6.1 that maximize expected utility in Equation (7), conditional on patient assignment ($a = 1$), as $(\tau^* (t), z^* (t))$. Denote corresponding inputs that maximize welfare as $(\tau^{FB} (t), z^{FB} (t))$. Assume that $F'_p (p^*) < \frac{1}{2}$. 

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

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

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

*Proof.* See Appendix A-4.1.

As the physician nears EOS, she will shorten length of stay $\tau$. The intensity of diagnostic tests and treatments may increase or decrease, depending on whether $\tau$ and $z$ are net substitutes or complements, respectively. Finally, she observes $\theta$ with lower probability $q$. This increases admissions, as long as $F'_{p^*} (p^*) < \frac{1}{2}$, where $F'_{p^*} (\cdot)$ is the c.d.f. of $p'$ conditional on $p$ and $a^* = 1$. Note that distortions in discharges are due solely through distortions in $(\tau^* (t), z^* (t))$, which lower $q$: conditional on $(\tau, z)$, physicians make the right decision on $d$.\footnote{In its most straightforward form, effective time per patient can be viewed as the concept behind “workload-adjusted length of stay”, in Section 5.3, which is length of stay divided by the average number of patients on census during a given patient’s length of stay. This concept reflects that the cost of a patient’s length of stay as an input does not depend on how many other patients a physician concurrently cares for, but the effectiveness of this time as an input into $q$ is reduced by workload.}
6.3 Patient Assignment

I next consider the case in which the physician is allowed to maximize her utility by choosing whether to accept the new patient, \( a \in \{0, 1\} \). The physician will accept the patient if and only if the expected utility of the outside option, \( E[O(\theta; \xi_t)] \), is below some threshold \( Q^* \). We can define similar thresholds that maximize welfare in the first-best case in which the social planner can implement \( (\tau^{FB}, z^{FB}) \) and in second-best case in which the social planner has no knowledge of appropriate patient-care inputs, denoted as \( Q^{FB} \) and \( Q^{SB} \), respectively.

**Proposition 2.** Consider \( a^* \) as the patient assignment in Section 6.1 that maximizes expected utility in Equation (7), \( a^{FB} \) as the assignment that maximizes expected welfare when optimal \( (\tau^{FB}, z^{FB}) \) is publicly known and contractible, and \( a^{SB} \) as the assignment that maximizes expected welfare when \( (\tau^{FB}, z^{FB}) \) is either publicly unknown or non-contractible. Assignment will follow threshold rules in which assignment occurs if and only if \( E[O(\theta; \xi_t)] \) is greater than a threshold. The respective threshold rules are \( Q^* \), \( Q^{FB} \), and \( Q^{SB} \), where \( Q^* < Q^{SB} < Q^{FB} \). \( Q^{FB} - Q^{SB} \) and \( Q^{SB} - Q^* \) increase as \( t \to \bar{t} \) decreases or as \( \lambda \) decreases.

**Proof.** See Appendix A-4.2.

There are obvious reasons for even first-best assignment to decrease near EOS: As \( t \to \bar{t} \), the outside option \( O(\theta; \xi_t) \) increases because a peer is more likely to be arriving soon or already present, and welfare-relevant costs of producing \( q \) change as the physician approaches EOS. Beyond this, however, assignment decided by the physician, \( a^* \), will be lower than the first-best assignment because she overvalues her leisure relative to other welfare relevant concerns (i.e., \( Q^* < Q^{FB} \)). In the second-best policy, the social planner maximizes welfare with patient assignment, but under the constraint that the physician still has control of inputs and will continue to choose \( (\tau^*, z^*) \). The second-best threshold \( Q^{SB} \) will be higher than \( Q^* \), because the social planner does not overvalue the physician’s leisure. But \( Q^{SB} \) will be lower than \( Q^{FB} \), because overvalued leisure still distorts patient care, and the social planner cannot implement appropriate patient care. The greater the patient-care distortions, the closer \( Q^{SB} \) will be to \( Q^* \) than to \( Q^{FB} \).
7 Counterfactual Assignment Policies

Because the assignment of patients is observable, it is natural to conceive ex ante policies of patient assignment, while downstream patient care decisions are non-contractible because patient types are not in general publicly observed. In Section 6, I show in a highly stylized conceptual framework that the second best policy involves some slacking off. In this section, I take this concept to the data, simulating outcomes under counterfactual assignment policies along the margin of slacking off. I consider patient assignment as a sufficient statistic for a wider range of managerial policies (e.g., rules, shift overlap, financial incentives) that may influence how patients are assigned but do not separately specify or incentivize how patients are to be treated.

7.1 Simulation Approach

At a high level, there are three components of simulating outcomes under counterfactual assignment policies. First, I specify and estimate a model physician discharge choices over multiple patients as a dynamic discrete choice problem. Second, I consider counterfactual assignment policies, and using the dynamic programming framework, determine physician discharge probabilities under these counterfactuals. Third, I simulate assignments and discharges under each counterfactual, and use these simulated observations to impute welfare-relevant costs of physician time, patient time, and hospital resources. I provide further details in Appendix A-5.

Consider an ED physician \( j \) at time \( t \) in state \( S \), which includes shift characteristics that the physician is working in, time categories (e.g., time of day, day of week), time relative to EOS, and the set of patients (and their characteristics) already under her care or the care of another physician present. I model the doctor’s decision to discharge patients, conditional on \( S \), as a dynamic discrete choice problem. In short time intervals, I assume that doctor may discharge at most one patient under her care, \( i \in I \) (including \( i = \emptyset \), or no discharge of any patient), and this decision may be represented as a Bellman equation:

\[
V (S) = E \left[ \max_{i \in I} \left\{ u (i, S) + \delta \int_{S'} V (S') \, dF (S'|i, S) \right\} \right],
\]

where \( V (S) \) is the value function of the dynamic programming problem, \( u (i, S) \) is the stochastic
utility flow of choosing \(i\) in state \(S\), \(\delta\) is the discount factor, and \(F(S'|i, S)\) is the Markov transition distribution function. The physician chooses \(i\) that maximizes the stochastic utility flow and the expected value function in the next period. I specify the flow utility as

\[
u (i, S) = \bar{u} (i, S) + \varepsilon_i ,
\]

where \(\bar{u} (i, S) = b (i, S) \theta_u\) is a linear combination of basis splines \(b (i, S)\), and \(\varepsilon_i\) is an i.i.d. Type I extreme value shock. \(b (i, S)\) are splines of patient \(i\)'s predicted length of stay and of the deviation in current length of stay from this prediction. Importantly, I exclude time to EOS from entering into \(\bar{u} (i, S)\).

I first estimate the transition function, \(F (S'|i, S)\), focusing on uncertainty in the assignment of new patients. I estimate the number of patients assigned to a physician in state \(S\) using an ordered logit model and each patient’s predicted length of stay using OLS. To estimate \(\theta_u\) in the utility flow function, I use tools from the dynamic discrete choice literature. Specifically, the log likelihood of observing actual patient discharges (i.e., indicators \(d (i, t)\) for whether patient \(i\) was discharged by physician \(j\) at time \(t\)) is

\[
\log L = \sum_{j,t} \sum_{i \in \mathcal{I}(j,t)} d (i, t) \log \tilde{Pr} (d (i, t) | i, S_{j,t}) ,
\]

where \(\tilde{Pr} (d (i, t) | i, S_{j,t})\) is the implied conditional choice probability under parameters \(\theta_u\). Dropping subscripts for simplicity, the probability of discharge choice \(i\) conditional on \(S\) is

\[
Pr (i | S) = \frac{\exp (\bar{u} (i, S))}{\sum_{i' \in I} \exp (\bar{u} (i', S))} , \tag{9}
\]

where \(\bar{u} (i, S) = \bar{u} (i, S) + \delta \int_S V (S') dF (S'|i, S)\). If \(S\) were in a low-dimensional space, I could solve for \(\theta_u\) using a nested fixed-point algorithm. However, because \(S\) is high-dimensional, I use a constrained optimization approach proposed by Barwick and Pathak (2015), building on the Mathematical Program with Equilibrium Constraints (MPEC) concept proposed by Su and Judd (2012). Rather than solving for \(V (S)\) using backward induction, this approach approximates \(V (S)\) using splines and incorporates a Bellman equation identity as constraints for the maximum
likelihood problem.

Estimates of the underlying utility flow function, \( \pi(i, S) \), are robust across functional form specifications of both \( \pi(i, S) \) and \( V(S) \); parameter results are presented in Table A-5.1. Discharge probabilities implied by the model fit the underlying data well, as shown in Figure 9. Appendix A-5.1 provides further details on identification, estimation, robustness, and goodness of fit of the dynamic programming problem.

With estimates of \( \theta_u \) in hand, the exercise is to consider counterfactual assignment policies that change \( F(S'|i, S) \). I consider counterfactual assignment policies along the margin of slacking off, adjusting the way time to EOS is considered, as parameterized by a scalar \( \Delta \) measured in hours, further described in Appendix A-5.2. If \( \Delta < 0 \), then fewer patients will be assigned; if \( \Delta > 0 \), then more patients will be assigned. Under a counterfactual \( F(S'|i, S) \), I recompute the Bellman equation, which gives conditional choice probabilities by Equation (9). Physicians may dynamically respond to counterfactual assignment policies, for example by further delaying the discharge of patients to forestall new patient assignments if new assignments are more likely near EOS (Chan, 2016).

Given counterfactual assignment policies and conditional choice probabilities for each \( \Delta \), I simulate patient assignments and discharges. I then use these counterfactual sequences of assignments and discharges to impute welfare-relevant costs of physician time, patient time, and hospital resources, for each counterfactual \( \Delta \) and simulation \( r = 1, \ldots, 20 \).\(^{30}\)

\[
\text{Costs}_{\Delta,r} = \text{PhysicianTime}_{\Delta,r} + \text{PatientTime}_{\Delta,r} + \text{HospitalResources}_{\Delta,r}.
\]

PhysicianTime\(_{\Delta,r} \) captures additional wages that the ED must pay in order to meet patient flow and is calculated as the number of physician hours needed to overall patient flow. Two considerations may increase PhysicianTime\(_{\Delta,r} \). First, if fewer patients are assigned before EOS, a peer must arrive earlier because backlog occurs earlier. On the other hand, if more patients are

---

\(^{30}\) This makes the conservative assumption that patient health is unaffected despite EOS distortions in time, formal utilization, and admissions. In sample, recall that I find no effect on mortality or bounce-backs (Table 3). Out of sample, with greater assignments near EOS, the EOS distortion may also worsen health outcomes, which implies even greater importance to reduce assignments near EOS. However, this should not matter for \( \Delta \) close to 0 and therefore for statements about optimal assignment if the optimal assignment policy occurs close to the observed assignment regime.
assigned prior to EOS, the index physician must stay later past EOS, and this foregone leisure is also valuable. I value physician time with a base-case wage of $120/hour, which is close to actual wages in this ED and national averages of hourly pay. To reflect that, all else equal, patients prefer shorter stays, I calculate PatientTimeCosts_\Delta as the total time spent by patients in the ED, valued at $20/hour. I calculate HospitalResources_\Delta by using previously estimated causal effects on length of stay and hospital resources, in Sections 4 and 5. Under an additional assumption that hospital resources are affected via the EOS mechanism of reducing lengths of stay, these two effects imply a marginal rate of technical substitution between time and hospital resources. This is nonparametrically identified, although based on results in Section 5.3, I choose to use workload-adjusted length of stay as the relevant measure impacting hospital costs. Further details are given in Appendix A-5.3.

7.2 Results

Increasing assignment near EOS results in large increases in hospital-resource costs, which dominate any physician-time savings from having the next physician arrive later. For example, an assignment policy that results in physicians staying an extra hour past EOS by assigning them about four additional patients also results in an increase of $3,700 in hospital spending per shift, for a net increase of $3,300 in total costs, compared to a policy that assigns those four patients to a physician who is not ending shift. Figure 10 shows average changes in total costs per shift, defined by Equation (10), under counterfactual policies, where policies are shown in terms of changes in the number of patients assigned. The observed pattern of assignment — implemented by shift overlap, triage nurse assignment, and physician responses — approximately minimizes total costs relative to counterfactual policies, implying that it is close to second-best optimal. Because hospital resources are potentially costlier, increasing assignments increases

---

31 The elasticity estimate is motivated by the fact that both observed total costs (Figure 6) and observed workload-adjusted length of stay (Figure A-7.4) increase only in the last hour prior to EOS, in Section 5.2. In simulated data, I calculate workload-adjusted length of stay decreases by 18.1% in the last hour of shift when \( \Delta = 0 \), an estimate very close to but more conservative than based on actual data in Table A-7.3. Since total costs increase by 20.8% in the last hour prior to EOS, I calculate the elasticity as \( 20.8\% / -18.1\% = -1.15 \). More detail is given in Appendix A-5.4.

32 Changes in costs with respect to changes in patients assigned is easier to understand than the policy index \( \Delta \), since \( \Delta \) is only maximum amount of change in time in the counterfactual assignment policy (i.e., \( |\overline{t}(\Delta) - \overline{t}| \leq \Delta \)). This can be appreciated in plots of curtailed assignment policies in Figure A-5.1, in which \( \Delta = -2 \) is still very similar to \( \Delta = 0 \).
costs more dramatically than does reducing assignments.

The intuition behind this robust finding is that, in the actual assignment regime $A_0$, very few patients are assigned in the last hour of shift, and this is precisely when costs start increasing. Reducing assignment near EOS would not substantially change in absolute terms the number of patients with EOS-induced distortions in hospital resources, but of course would increase physician time to cover shift changes. On the other hand, increasing assignment near EOS is much costlier for two reasons. First, patient care costs are much higher than physician wages. Second, increasing assignment not only exposes more patients to EOS cost distortions but also worsens the distortion per patient if the relevant time input measure is time per patient (e.g., workload-adjusted length of stay). Thus, if a planner were uncertain about when the optimal assignment policy should begin reducing assignments, it would be cheaper and less risky to pay for idle physician time as opposed to risking patient care distortions.

Another way to use the simulation is to assess the implicit tradeoff physicians make between foregoing leisure and increasing resource-utilization costs. At each point in time relative to normal completion, I compute the dollar value of extra hospital-resource costs incurred per leisure hour gained, shown in Figure 11 (details in Appendix A-5.4.1). At times prior to normal completion (but after EOS), the value of incremental leisure appears low, below the market wage of $120 per hour. However, the value of leisure quickly rises above market wage as work completion time increases. At about one hour past normal completion, physicians are willing to expend $790 in hospital resources, six times above the market wage, in order to avoid an additional hour at work.33

8 Discussion

The main focus of this paper is to assess a simple but, to my knowledge, unexplored consequence of work schedules: When work past scheduled availability is undercompensated, workers will

33Under a strict interpretation of Equation (7) and no (assumed) worsening of patient health, this implies that $\lambda$ is decreasing after EOS and that $\lambda = $120/$790 = 0.15$ at this point. In another dimension of heterogeneity, in Section 5.4 and Figure 8, I show reduced-form evidence that the EOS effect on workload-adjusted length of stay is entirely in shifts ending during daytime hours between 6:00 a.m. and 8:00 p.m., which is consistent with leisure being is more highly valued during these times, when others are awake and non-work activities are costlier to defer.
avoid new work, and the use of time for work will be distorted, possibly in costly ways. While these effects are illustrated concretely in the setting of health care, there are several general points of interpretation. I discuss some of these briefly here.

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

**Social and Behavioral Mechanisms of Distortion.** It is standard to assume that workers care about their own income and leisure more than the productive consequences of their workplace actions. Therefore, a natural interpretation of distortions near EOS is that they arise from strategic behavior, or moral hazard. However, other mechanisms could lead to the same welfare-reducing distortions. For example, social norms may pressure workers not to stay too long after EOS (e.g., doing so would signal incompetence), so that moving EOS too early without tapering work generates the same inefficient time pressure. Workers may take schedules as a contractual “reference point” (Hart and Moore, 2008), and there may even be accepted routines (e.g., sign-out rounds) that reinforce this sense. All of these mechanisms would have equivalent implications on how availability should be scheduled and worked assigned, but some mechanisms (e.g., social

---

34 Other definitions have described presenteeism as showing up to work while ill.
36 A related distortion caused by social incentives is that ED physicians care more about their peers than inpatient doctors they would be admitting their patients to. This is one (distortionary) source of client-worker specificity. In Appendix A-6 (Figures A-6.2 and A-6.1), I show that EOS distortions are more likely under certain physician types and physician-peer relationships, e.g., slower physicians and physicians handing off to a senior peer, a faster peer, or a peer with whom they are unfamiliar.
Workers in practice may underestimate the time it takes to complete work, paying too much attention to when EOS is officially set when planning after-work activities. Even rational workers would view uncertainty with respect to the time they leave work as costly, if they have plans that cannot be easily changed in the short term, such as picking up children from school or having a dinner reservation. This implies that the canonical model of a constant value of leisure is misspecified, and that short-term reductions in “leisure” are much costlier than prearranged work, as has been noted in the press with practices such as “flexible” or “just-in-time” scheduling that benefit employers but could be costly for employees (e.g., Greenhouse, 2014). Accounting for this would imply even greater welfare benefits from slacking off to ensure that workers leave close to a prespecified time.

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

---

37 In Figure A-7.1, I show evidence that physicians continue to write orders past EOS, consistent with the qualitative report from this ED that most physicians stay two to three hours past EOS (including activities such as dictating notes, conferring with other staff, and notifying patients, in addition to order-writing). In addition, Section 5.4 discusses evidence that EOS effects and reductions in attending-physician order-writing are both concentrated in shifts ending during daytime, when non-work activity is costlier to defer. This is consistent with the benchmark distortion mechanism of leisure-seeking.

38 Transfer costs have been discussed in the economics literature as “worker-task specificity” (e.g., Goldin, 2014) and in the medical literature manifesting as difficulties in passing information between providers transitioning care (e.g., Apker et al., 2007). As an example of information-driven transfer costs, a physician may find it difficult to instruct another physician to “discharge the patient home if his signs and symptoms are x% better,” if this information are multidimensional and understood only by personal examination and questioning. Communicating everything the physician knows about a patient is generally difficult. As an example of behavioral-driven transfer costs, physicians may not want to burden a colleague even if it is efficient to do so.
much. Further, there are likely general behavioral foundations for transfer costs that could prove difficult to eliminate.\textsuperscript{39}

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

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

\section{9 Conclusion}

I examine ED physicians working in shifts and find evidence consistent with behavioral distortions due to scheduled work: On an extensive margin, physicians are less likely to accept new patients

\textsuperscript{39}Polanyi (1966) has described tacit knowledge that is difficult to communicate, which could apply to complex patient care, at least in the sense that successful communication is costly both for the sender and the receiver (Dewatripont and Tirole, 2005). Although patient care shared with nurses and residents who overlap with attending physicians changing shifts could reduce transfer costs, the extent of this is also limited by knowledge and communication issues with attending physicians responsible for the discharge decision.
near EOS. On an intensive margin, physicians complete their work earlier as end of shift (EOS) approaches. As the input of time becomes costlier, physicians modify the mix of inputs in patient care, and as they produce less information for discharge decisions, they are more likely to admit patients.

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

References


Table 1: Balance Tests

<table>
<thead>
<tr>
<th>Hour prior to EOS</th>
<th>Assumption 1</th>
<th>Assumption 2</th>
<th>Assumption 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted log LOS</td>
<td>Diagnostic count</td>
<td>Predicted log LOS</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Last hour</td>
<td>0.003</td>
<td>0.008</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.070)</td>
<td>(0.066)</td>
</tr>
<tr>
<td>Second hour</td>
<td>-0.002</td>
<td>-0.028</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.037)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Third hour</td>
<td>0.002</td>
<td>0.001</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.031)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Fourth hour</td>
<td>0.004</td>
<td>0.014</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.026)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Fifth hour</td>
<td>0.004</td>
<td>0.011</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.021)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Sixth hour</td>
<td>0.000</td>
<td>-0.010</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.017)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>F-test p-value</td>
<td>0.598</td>
<td>0.699</td>
<td>0.852</td>
</tr>
</tbody>
</table>

Resid. char. distrib.

| Mean | 1.078 | 1.012 | 1.079 | 47.65 | 2.826 | 0.495 | 0.101 | 1.013 |
| 10th percentile | 0.985 | 0.381 | 0.969 | 41.69 | 2.596 | 0.329 | 0.017 | 0.681 |
| 90th percentile | 1.167 | 1.623 | 1.188 | 53.72 | 3.058 | 0.664 | 0.190 | 1.356 |

**Note:** This table assesses balance with respect to two identifying assumptions. Columns 1 and 2 assess Assumption 1, based on Equation (A-1.2), controlling for ex ante patient characteristics time categories, pod, and providers. The remaining columns assess Assumption 2, based on Equation (A-1.3), controlling for time categories, pod, and providers. Further details are given in Appendix A-1.2. Coefficient estimates and standard errors in parentheses are given for each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. The p-value of the F-test that all coefficients are jointly 0 is also given for each model. The last three rows report summary statistics of the variation in residualized patient characteristics, either across all patients (Columns 1 and 2) or across patient averages at the hourly level (remaining columns). * denotes significance at 10% level. Black race and English-speaking were also assessed for Assumption 2 and similarly yielded null results but are omitted from the table for brevity. Table A-1.1 reports results for other predicted outcomes.
Table 2: End of Shift Effect on Log Length of Stay

<table>
<thead>
<tr>
<th>Hour prior to EOS</th>
<th>(1) OLS</th>
<th>(2) OLS</th>
<th>(3) OLS</th>
<th>(4) OLS</th>
<th>(5) OLS</th>
<th>(6) OLS</th>
<th>(7) Between</th>
<th>(8) Between</th>
</tr>
</thead>
<tbody>
<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>
<td>-0.584***</td>
<td>-0.612***</td>
<td>-0.626***</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>
<td>(0.050)</td>
<td>(0.078)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>Second hour</td>
<td>-0.316***</td>
<td>-0.282***</td>
<td>-0.330***</td>
<td>-0.461***</td>
<td>-0.287***</td>
<td>-0.287***</td>
<td>-0.325***</td>
<td>-0.344***</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>
<td>(0.026)</td>
<td>(0.033)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>Third hour</td>
<td>-0.139***</td>
<td>-0.129***</td>
<td>-0.161***</td>
<td>-0.260***</td>
<td>-0.123***</td>
<td>-0.122***</td>
<td>-0.139***</td>
<td>-0.157***</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>
<td>(0.022)</td>
<td>(0.025)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Fourth hour</td>
<td>-0.112***</td>
<td>-0.092***</td>
<td>-0.111***</td>
<td>-0.173***</td>
<td>-0.091***</td>
<td>-0.089***</td>
<td>-0.098***</td>
<td>-0.107***</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>
<td>(0.018)</td>
<td>(0.021)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Fifth hour</td>
<td>-0.070***</td>
<td>-0.055***</td>
<td>-0.078***</td>
<td>-0.120***</td>
<td>-0.023</td>
<td>-0.022</td>
<td>-0.037**</td>
<td>-0.047**</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>
<td>(0.015)</td>
<td>(0.017)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Sixth hour</td>
<td>-0.065***</td>
<td>-0.048***</td>
<td>-0.057***</td>
<td>-0.090***</td>
<td>-0.010</td>
<td>-0.010</td>
<td>-0.015</td>
<td>-0.025***</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>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.015)</td>
</tr>
</tbody>
</table>

Patient characteristics
None, All, All, All, All, Ex ante, All, None

Time and pod
N, N, Y, Y, Y, Y, Y, Y

Physician-resident-nurse identities
N, N, N, Y, Y, Y, Y, Y

Time relative to shift beginning
N, N, N, N, Y, Y, Y, Y

Observations
371,107, 371,107, 371,107, 371,107, 371,107, 371,107, 68,044, 68,044

Adjusted R-squared
0.008, 0.189, 0.211, 0.400, 0.410, 0.404, 0.061, 0.064

Sample outcome
1.050, 1.050, 1.050, 1.050, 1.050, 1.050, 1.074, 1.074

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

<table>
<thead>
<tr>
<th>Hour prior to EOS</th>
<th>(1) Order count</th>
<th>(2) Inpatient admission</th>
<th>(3) Log total cost</th>
<th>(4) 30-day mortality</th>
<th>(5) 14-day bounce-back</th>
<th>(6) Workload-adjusted log LOS</th>
</tr>
</thead>
<tbody>
<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>
<td>-0.144***</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>
<td>(0.051)</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>
<td>0.015</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>
<td>(0.027)</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>
<td>0.090***</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>
<td>(0.022)</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>
<td>0.036*</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>
<td>(0.018)</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>
<td>0.037**</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>
<td>(0.015)</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>
<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>
<td>(0.012)</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>
<td>371,148</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>
<td>0.476</td>
</tr>
<tr>
<td>Sample mean outcome</td>
<td>13.518</td>
<td>0.269</td>
<td>6.750</td>
<td>0.018</td>
<td>0.060</td>
<td>-0.904</td>
</tr>
</tbody>
</table>

Note: This table reports coefficient estimates of Equation (1) regressing other outcome variables, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Workload-adjusted length of stay (LOS) is calculated by Equation (5), and workload-adjusted log LOS is the log of this value. Controls are the same as in Column 5 in Table 2. Standard errors are in parentheses. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level.
Table 4: Effect on Workload-adjusted Log Length of Stay by Shift Overlap

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

**Note:** This table reports coefficient estimates and standard errors in parentheses for EOS effects on workload-adjusted log length of stay, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Model 1 is estimated by Equation (4), while models 2 and 3 are estimated separately by Equation (1) on subsamples of the data according to $\bar{\sigma}$. All three models are estimated with a full set of controls, as in Column 5 in Table 2. Workload-adjusted length of stay (LOS) is calculated by Equation (5), and workload-adjusted log LOS is the log of this value. * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. This table is continued by Table A-7.3.
Note: Filled areas in vertical lines represent hours scheduled for a shift for a single physician. Hours when there is more than one physician present are represented by horizontally adjacent filled areas.
Figure 2: Shift Variation

Note: This figure illustrates the variation in observations across shift types. Panel A plots shifts by shift ending time and shift length. Panel B plots shifts by shift ending time and the length of overlapping transition at the end of shift. The sizes of circles in both panels represent the number of shifts out of 23,990 total in each two-dimensional category.
Figure 3: Flow of Patient Visits over Time

Note: This figure shows unadjusted average hourly rates of patient visits for each 30-minute interval relative to end of shift (EOS). Each panel shows results for shifts with a given EOS overlap time. Patient visits for the index physician are shown in closed circles; patient visits for the location are shown in open circles. Subsequent shift starting times are marked with a vertical line.
Figure 4: Patient Characteristics over Time

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

Note: This figure shows quantiles of predicted log length of stay for patients assigned to physicians in each 30-minute bin relative to EOS. Log length of stay is predicted by cubic splines of age, sex, indicators for ESI, indicators for language, and indicators for race. The solid line shows medians; dashed lines show 25th and 75th percentiles; and short-dashed lines show 5th and 95th percentiles.
Figure 6: End of Shift Effects

A: Length of Stay

B: Orders

C: Inpatient Admission

D: Costs

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 (4), and results are the same as in Column 5 in Table 2 for length of stay and Table 3 for the other outcomes. The reference category is any time greater than six hours prior to EOS. Bracketed dashed lines represent 95% confidence intervals for each estimate.
Figure 7: End of Shift Effects by Shift Overlap

Note: This figure shows end of shift (EOS) effects by EOS overlap times on length of stay (Panel A), orders (Panel B), inpatient admissions (Panel C), and costs (Panel D). Each outcome is estimated separately using Equation (4). Estimates for terminal shifts ($\bar{\sigma} = 0$) are shown in open triangles; estimates for minimally transitioned shifts ($\bar{\sigma} = 1$) are shown in open circles; and estimates for substantially transitioned shifts ($\bar{\sigma} \geq 2$) are shown in closed circles. The noted $p$-values in each panel are for a test that the coefficient of last hour of shift interacted with $\bar{\sigma} \leq 1$ is different than the coefficient of last hour of shift interacted with $\bar{\sigma} \geq 2$ (these coefficients are slightly different than those shown in the figure in that there are only two categories of shift overlap instead of three).
Figure 8: End of Shift Effect by Whether Last Hour is During Daytime

Note: This figure shows estimates of effects of hour relative to end of shift (EOS), using Equation (6), conditioning on whether the shift ends during daytime (6 a.m. to 8 p.m.) (solid dots) or not (hollow dots). Roughly half of the shifts during during daytime, with 11,183 shifts ending during daytime and the remainder, 11,344 shifts, ending during any other hour. The p-values of F-tests on the difference between (a) all hourly coefficients corresponding to daytime vs. not and between (b) the last-hour of shift coefficients corresponding to daytime vs. not are both less than 0.01.
Note: This figure shows the fit of the discharge policies with actual discharge decisions in a sample of 792,687 patient-time observations. Each panel evaluates the fit between the actual data (solid dots), a flexible multinomial logit “static” model of discharges (hollow dots), and a dynamic model of discharges with restrictions on what can enter into the utility flow (triangles). The dynamic model corresponds to the simplest model in Table A-5.1, Model 1, although the fit and the estimated parameters do not qualitatively differ across specifications. The y-axis in all panels is the probability or average of whether a patient \( i \) still undischarged as of time \( t - 1 \) could be discharged at \( t \) by doctor \( j \) on shift with certain state characteristics. The \( x \)-axis is some characteristic of \( (i, j, t) \): hours that \( (j, t) \) is away from EOS in Panel A; the number of patients on census, or \( \|I(j, t)\| \), in Panel B; predicted length of stay according to \( i \)'s characteristics and time categories of \( t \) in Panel C; the deviation in current log length of stay from the predicted log length of stay in Panel D. Each marker potentially represents 1/40th of the data if the \( x \)-axis is sufficiently continuous. The dynamic programming approach is described in greater detail in Appendix A-5.1.
Figure 10: Simulated Change in Total Cost per Shift under Counterfactual Assignment

Note: This figure plots simulated changes in total cost for each counterfactual assignment regime, as described in Section 7 and Appendix A-5. Assignment policies may either increase or decrease assignment, as illustrated in Figure A-5.1. Counterfactual costs depend on a dynamic programming model of physician discharge decisions. These results are generated from parameters of the dynamic programming model corresponding to Model 1 in Table A-5.1, although results do not qualitatively differ across model specifications. The x-axis is the change in total patients assigned per shift as a result of an assignment regime ("0" corresponds to the actual assignment regime). Daily costs include physician-time, patient-time, and hospital-resource costs. Each marker is an average over 20 simulations in an assignment policy $\mathcal{A}_\Delta$, and each simulation contains 280,000 to 460,000 observations, depending on whether $\mathcal{A}_\Delta$ decreases or increases the number of patients assigned relative to the actual assignments.
Figure 11: Value of Leisure in Dollars of Patient Care

Note: This figure plots the imputed value of leisure, revealed by increases in resource-utilization costs that shorten the time for completion of work under simulated counterfactual assignment policies. For each counterfactual assignment policy, data are simulated using the baseline counterfactual discharge policy $D_\Delta$ and then again for a counterfactual discharge policy $D_{\Delta'}$ that is insensitive to time relative to EOS. The difference between these two discharge policies yields a tradeoff of resource-utilization costs for shortened work completion time. The ratio between these two represents the value of leisure, denominated in hospital-resource dollars, and is plotted on the $y$-axis. These results are generated from parameters of the dynamic programming model corresponding to Model 1 in Table A-5.1, although results do not qualitatively differ across model specifications. The $x$-axis is the relative time of work completion, compared to assumed work completion under the observed assignment and discharge policies, $A_0$ and $D_0$, respectively. The horizontal line drawn at $120$ is the effective hourly wage for an hour of a physician’s (scheduled) time. Details are described in Appendix A-5.4.1.
Appendix

A-1 Identification under Patient Selection

A-1.1 Quantifying Selection in Terms of Predicted Outcomes

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

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

$$\hat{Y}_{it}^{set} = \alpha_{m(i,t)} + \gamma_{m(i,t)} + \eta T + \zeta p(i) + \nu j(i,k(i)) + \epsilon_{it},$$  \hspace{1cm} (A-1.1)

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

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

A-1.2 Assessing Identifying Assumptions

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

I assess Assumption 1 by considering an *ex post* characteristic $X_{\text{post}}$ and estimating the following regression:

$$X_{it}^{\text{post}} = \alpha_{m(i,t)} + \gamma_{m(i,t)} + \beta X_i^{\text{ante}} + \eta T_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_t. \quad (A-1.2)$$

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

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

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

$$X_{it} = \sum_{m=6}^{1} \alpha_m P_{m} (t) + \eta T_t + \zeta_{p(i)} + \nu_{j(i),k(i)} + \varepsilon_t. \quad (A-1.3)$$

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

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

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

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

$$Q_t = \sum_i 1 (t_i^a = t) \sum_{m=1}^{6} \hat{\alpha}_{m}^{s} 1 (\bar{t} (J (i, t), t) - t) = m) 1 (S (J (i, t), t) = s)$$

(A-1.4)

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

A-1.3.1 Forecast Bias

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

$$Y_t = \frac{\sum_i 1 (t_i^a = t) \tilde{Y}_{it}}{\sum_i 1 (t_i^a = t)}$$

(A-1.5)

where

$$\tilde{Y}_{it} = Y_{it} - \left[ \hat{\gamma}_{m(i,t)} + \hat{\beta} X_i + \hat{\eta} T_t + \hat{\zeta}_{p(i)} + \hat{\nu}_j (i) \right].$$

(A-1.6)

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

$$Y_{it} = \alpha_{m(i,t)} + \gamma_{m(i,t)} + \beta X_i + \eta T_t + \zeta_{p(i)} + \nu_{k(i)} + \varepsilon_{it},$$

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

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

$$Y_t = a + b Q_t + \chi_t.$$

(A-1.7)

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

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

under the assumption that

$$
\text{Cov} (Q_t, \chi_t) = 0.
$$

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

Column 1 of Table A-1.2 reports estimates of $b$ from Equation (A-1.7). The point estimate of $b$ is 1.029 with a robust standard error of 0.060 (clustered at each hour of $t$), which reflects tight estimation indistinguishable from 1 (i.e., I cannot reject the hypothesis of $\hat{\alpha}_{m}^{-s} - \alpha_{m} = 0$).

That is, under the assumption in (A-1.8), I cannot reject the null of no unobservable selection across physicians within time (i.e., Assumption 1). Panel A of Figure A-1.2 plots the relationship between $Y_t$ and $Q_t$ nonparametrically, dividing the data into 20 equal-sized groups (“vigintiles”) according to $Q_t$. This plot nonparametrically represents the expectation function of $Y_t$ conditional on $Q_t$. The relationship is highly linear, with slope close to 1.

### A-1.3.2 Observable Selection between Hours

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

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

$$
\hat{Y}^{set}_t = \frac{\sum_i 1 \left( t_i^a = t \right) \hat{Y}^{ante}_i}{\sum_i 1 \left( t_i^a = t \right)}, \quad (A-1.9)
$$

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

A-4
The regression
\[ \hat{Y}_t^{set} = a + b^{set}Q_t + \chi_t \] (A-1.10)
quantifies the degree of selection across hours, as predicted by characteristics \( X^{set} \): Under (A-1.8), \( b^{set} = \text{Cov} \left( E \left[ \hat{\alpha}_{m|t} | t \right], \hat{Y}_t^{set} \right) / \text{Var} \left( E \left[ \hat{\alpha}_{ms|t} | t \right] \right) \) for \( m = \lceil \bar{t} (j, t) - t \rceil \) and set \( s = S (j, t) \). Although the assumption in Equation (A-1.8) is not directly testable, a lack of observable selection (\( b^{set} \) is indistinguishable from 0) supports this assumption.

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

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

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

**A-1.4.1 Conceptual Framework**

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

\[ Y = \sum_m \alpha_m A_m + \Omega \Gamma + \xi, \] (A-1.11)

where I omit subscripts for simplicity. I define \( A_m \equiv 1 \left( \lceil \bar{t} (j, t) - t \rceil = m \right) \) for whether the time \( t \) that patient \( i \) was assigned to physician \( j \) was in the \( m \)th hour from \( j \)'s EOS. \( \alpha_m \) is the causal effect of a patient being assigned in the \( m \)th hour prior to EOS. \( \Omega \) is the full set of other variables, both observed and unobserved, that determine outcome \( Y \), while \( W \) includes only observed patient, time, and provider characteristics (to be distinguished from \( X_{it} \) in Equation (1), which only includes patient characteristics). \( \Gamma \) is the causal effect of \( \Omega \) on \( Y \). \( \Gamma_W \) is the subvector of \( \Gamma \) that corresponds to \( W \) within \( \Omega \), and \( \xi \) is an index of the unobserved variables.
Since variables in $\mathbf{W}$ are likely correlated with $\xi$, rewrite Equation (A-1.11) as

$$Y = \sum_m \alpha_m A_m + \mathbf{W}' \gamma_W + \varepsilon,$$  \hspace{1cm} (A-1.12)

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

**A-1.4.2 Measure of Selection on Unobservables**

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

$$\frac{E[\varepsilon|A_m = 1] - E[\varepsilon|A_m = 0]}{\text{Var}(\varepsilon)} = \frac{E[\mathbf{W}' \gamma_W|A_m = 1] - E[\mathbf{W}' \gamma_W|A_m = 0]}{\text{Var}(\mathbf{W}' \gamma_W)}.$$ \hspace{1cm} (A-1.13)

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

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

**A-1.4.3 Estimation of Potential Bias**

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

$$A_m = \mathbf{W}' \beta_W^m + A_m^*,$$ \hspace{1cm} (A-1.14)

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

$$Y = \sum_m \alpha_m A_m^* + \mathbf{W}' \left( \gamma_W + \sum_m \alpha_m \beta_W^m \right) + \varepsilon.$$
This leads to a statement of the potential bias due to selection on unobservables:

\[
\lim\alpha_m \approx \alpha_m + \frac{\text{Cov}(A^*_m, \varepsilon)}{\text{Var}(A^*_m)} = \alpha_m + \frac{\text{Var}(A_m)}{\text{Var}(A^*_m)} (E[\varepsilon | A_m = 1] - E[\varepsilon | A_m = 0]),
\]

From Equation (A-1.13), the bias can be stated in terms of \(E[W'\gamma_W | A_m = 1] - E[W'\gamma_W | A_m = 0] \):

\[
\text{Bias} = \frac{\text{Var}(A_m) \text{Var}(\varepsilon)}{\text{Var}(A_m) \text{Var}(W'\gamma_W)} \left( E[W'\gamma_W | A_m = 1] - E[W'\gamma_W | A_m = 0] \right)
\]

Under the null hypothesis that \(\alpha_m = 0\), \(\gamma_W\) can be consistently estimated by Equation (A-1.11). I can then arrive at a consistent estimate of bias in Equation (A-1.15) with the following procedure, with results shown in Table A-1.3: For each \(m \in \{-6, \ldots, -1\}\), I define \(A_m\) over all observations and empirically calculate \(\text{Var}(A_m)\). I also calculate \(\text{Var}(A^*_m)\) after estimating Equation (A-1.14) for each \(m\). Similarly, I estimate \(\text{Var}(\varepsilon) = 0.160\) and \(\text{Var}(W'\gamma_W) = 0.580\) from Equation (A-1.12). Equation (A-1.12) also allows me to form an estimate of selection on observables, \(\hat{E}[W'\gamma_W | A_m = 1] - \hat{E}[W'\gamma_W | A_m = 0]\), for each \(m\). Using the condition in Equation (A-1.13) that normalized selection on unobservables is bounded by normalized selection on observables, I then calculate an upper bound of the bias due to selection on unobservables with Equation (A-1.15). As shown in Table A-1.3, the upper bound of the bias in \(\hat{\alpha}_1\), the effect of arriving in the last hour of shift on the length of stay, estimated by Equation (1), is \(-0.00124\). Given that \(\hat{\alpha}_1 = -0.5873\), this implies that normalized selection on unobservables would have to be 475 times greater than normalized selection on observables. As a comparison, in their example of the impact of Catholic school on educational attainment, Altonji et al. (2005) argue that selection on unobservables is highly unlikely with a ratio 3.55.

A-2 Effects Relative to Shift Beginning

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

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

A-3 Time Components of Length of Stay

In Section 4, length of stay decreases while formal utilization increases near EOS. This suggests
that formal utilization is a net substitute for time in patient care. In this appendix, I further
examine this hypothesis by a closer look at the time components of length of stay. In practice,
time is not neatly divided into pure substitute or complement components with formal utilization
(call these components $\tau_1$ and $\tau_2$, respectively), but some intuitive distinctions can be made:
Time before the first formal order likely belongs to $\tau_1$ (e.g., time spent interviewing the patient
or performing serial abdominal examination as opposed to CT scan). Time after the last formal
order likely belongs to $\tau_2$, reflecting time needed to follow up on utilization (e.g., waiting for CT
scan report). Although time in between the first and last orders could belong to either $\tau_1$ or $\tau_2$,
the spacing of these orders often reflects clinical monitoring and reasoning more closely related
to $\tau_1$.

Measuring length of stay in three component shares – time between pod arrival and first
order, time between first and last (non-discharge) orders, and time between last and discharge
orders – I estimate a fractional logit model (Papke and Wooldridge, 1996) using similar regressors
as in Equation (1). Figure A-3.1 presents results of marginal effects relative to EOS. Panel A
scales time shares by the median predicted length of stay in each hour prior to EOS according
(1); Panel B simply plots the unscaled proportional shares. These proportions remain relatively
unchanged except for the last hour prior to EOS, when the proportions for time prior to first order
and inter-order time both decrease. These results suggest relative reductions in $\tau_1$, particularly
in the last hour prior to EOS, and are consistent with the increase in formal utilization (net
substitution) in the last hour shown in Table 3 and Figure 6.

A-4 Stylized Model Proposition Proofs

A-4.1 Proof of Proposition 1

Denote inputs in Section 6.1 that maximize expected utility in Equation (7), conditional on patient
assignment ($a = 1$), as $(\tau^* (t), z^* (t))$. Denote corresponding inputs that maximize welfare as
($\tau_{FB}(t), z_{FB}(t)$). Assume that $F_{p'}(p^*) < \frac{1}{2}$.

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

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

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

Define $\Delta^\theta_\theta = v(1, 1) - v(1, 0)$ and $\Delta^\theta_\theta = v(0, 0) - v(0, 1)$. Using this notation, note that

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

which implies that

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

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

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

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

The corresponding first-order condition with respect to $z$ is

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

where

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

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

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

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

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

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

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

A-4.2 Proof of Proposition 2

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

Proof. The expected utility under \( a = 0 \) is \( E [O (\theta; \varepsilon_t)] \), and the expected utility under \( a = 1 \) is

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

where

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

A-10
Note that at the time of patient assignment, the physician only knows the (publicly known) probability that \( \theta = 1, p \), instead of the probability \( p' \) privately observed by the physician after assignment.

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

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

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

\[ \qed \]

### A-5 Counterfactual Simulations

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

#### A-5.1 Dynamic Programming Model

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

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

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

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

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

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

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

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

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

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

$$V(S) = \gamma + \log \left[ \sum_{i \in I}\exp(\overline{v}(i, S)) \right]$$

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

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

A-5.1.1 Estimating the Transition Probability Function

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

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

\[
 w (j, t) = \sum_{i} 1 \left( t \geq t^a (i) \right) 1 \left( t \leq t^d (i) \right) 1 (j = J (i)), \tag{A-5.5}
\]

where \( t^a (i) \) is the arrival (assignment) time of \( i \), \( t^d (i) \) is the corresponding discharge order time, and \( J (i) \) is the physician corresponding to \( i \). This model represents the assignment policy function \( A_0 (S) \).

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

### A-5.1.2 Estimating Utility Flow Parameters

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

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

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

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

\[
 \hat{\theta}_V = \arg \max_{\theta_V} \left\| h (S) \theta_V - \log \left[ \sum_{i \in I} \exp \left( b (i, S) \hat{\theta}_u + \delta E \left[ h \left( S' | i, S \right) \right] \theta_V \right) \right] \right\| \tag{A-5.6}
\]
by nonlinear least squares across observations \( \{(j,t)\} \). In practice, I implement Equation (A-5.4) by setting the constraint

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

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

4. Perform constrained maximum likelihood estimation, where the relevant data include the indicator variables \( \{d(i,t)\} \), such that \( d(i,t) = 1 \left( i_{j(i),t} = i^*_{j(i),t} \right) \), and the set of basis functions \( \{b(i,S_{j,t})\} \) and \( \{E[h(S_{j,t}) | i,S_{j,t}]\} \) for each physician-time \((j,t)\) and choice \(i\). The log likelihood of the data is

\[
\log L = \sum_{j,t} \sum_{i \in I(j,t)} d(i,t) \log \hat{Pr}(d(i,t) | i,S_{j,t}), \tag{A-5.8}
\]

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

A-5.1.3 Identification

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

In the conceptual framework (Section 6), I also consider foregone leisure as relevant for physician utility. However, because I do not observe the actual time that a physician goes home, and because patients are never assigned past EOS, I conveniently ignore modeling this utility explicitly.\(^{40}\) Instead, I include the number of patients remaining on the physician’s census at

\(^{40}\)In an alternative specification of the dynamic programming problem, I model future utility flows from simulated outcomes, as in Hotz et al. (1994), including the times that physicians are likely to go home, based on a simple rule that they are likely to go home between when all but one or two patients have been discharged (as assumed in Section A-5.4). This results in utility estimates that suggest that leisure time is quite important: staying an extra hour is two to five times as important as deviating from ideal (predicted) length of stay by 50%
EOS as a state variable in the value function. As argued above, because patients are never assigned past EOS, counterfactual assignment policies and therefore conditional discharge choice probabilities only differ prior to EOS.

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

### A-5.1.4 Specification and Model Fit

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

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

In Figure A-5.2, I further evaluate model fit by average shift outcomes with respect to time to EOS. In particular, I evaluate fit by the following averages for patients arriving in each 30-minute period with respect to EOS: the number of patients arriving in the period; the length of stay (and its log); and the average workload during a patient’s length of stay, or \( \bar{w}(i) \) defined in for a single patient. However, this specification does not allow me to calculate counterfactual discharge policies.
Equation (5). The panels of this figure evaluating length of stay and average workload are based on the same simulation algorithm, described in Appendix A-5.3, used to evaluate counterfactual assignment policies. This figure shows that the fit is reasonably good using the conditional choice probabilities implied by Equation (A-5.3) in the dynamic model, although length of stay (particularly, its log) does not seem to decrease by as much in the last two hours prior to EOS.

### A-5.2 Counterfactual Assignment Policies

The counterfactual assignment policies \( A_\Delta (S) \equiv A_0 (\overline{m} (\Delta), S^-) \) are constructed by modifying time to EOS. That is, for a true time to EOS \( m \), the assignment function considers a modified time to EOS \( m (\Delta) \) that is a function of the scalar parameter \( \Delta \):

\[
\overline{m} (\Delta) = \begin{cases} 
\max (\Delta, \overline{m}) & \Delta > 0 \\
\overline{m}, & \Delta = 0 \\
\overline{m} (1 - \min (\max (|\Delta| - \overline{m}, 0), 1)) & \Delta < 0
\end{cases}
\] (A-5.9)

In this convenient parameterization, \( \overline{m} (\Delta) \) is at most \( \Delta \) greater than \( \overline{m} \) if \( \Delta > 0 \) and at most \( \Delta \) smaller than \( \overline{m} \) if \( \Delta < 0 \). In the case where \( \Delta < 0 \), the function is slightly more complicated so that it is continuous; starting at \( \overline{m} = |\Delta| \), \( \overline{m} (\Delta) = \overline{m} \) and decreases to \( \overline{m} (\Delta) = 0 \) by \( \overline{m} = |\Delta| - 1 \). \( A_\Delta (S) \) correspondingly increases assignments relative to \( A_0 (S) \) for \( \Delta > 0 \) and decreases assignments for \( \Delta < 0 \). Figure A-5.1 shows both the time modification function \( \overline{m} (\Delta) \) as well as the corresponding counterfactual assignment policies \( A_\Delta (S) \) for \( \Delta \in \{-4, -2, 2, 4\} \).

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

### A-5.3 Simulation

For a given assignment policy \( A_\Delta (S) \) and corresponding discharge policy \( D_\Delta (S) \), I simulate patient arrivals and discharges to create a set of patient arrival and discharge observation. Specifically, I follow this procedure for each simulation \( r \):

1. Start \( t \) at three hours before the beginning of each shift \( s \). Set \( w_{\Delta, r} (j, t - 1) = 0 \).
2. Determine new assignments at \( t \) for each \( s \).
   
   (a) Simulate \( a_{\Delta, r} (s, t) \) new assignments for \( s \) at \( t \), using \( A_\Delta \). Denote each of these new assignments with an unused \( i \notin (\text{outside}) \), note that \( t_{\Delta, r} (i) = t \), and also simulate predicted log length of stay for each \( i \).
(b) Assign patients $i \in I_{outside}$ where $t_{outside}^a (i) = t$ to the relevant shifts $s$.

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

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

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

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

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

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

### A-5.4 Imputing Cost Outcomes

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

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

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

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

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

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

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

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

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

\[
\log \left( \frac{\tau(i) / w(i)}{\tau_{\Delta,r}(i)} \right) = \alpha_{m(i,t)}^{\Delta,r} + g(m(i,t))' \gamma_{g}^{\Delta,r} + \varepsilon_{i}^{\Delta,r},
\]

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

In simulated data with \(\Delta = 0\), I estimate \(\hat{\alpha}_1 = 0.150, \hat{\alpha}_2 = 0.590\), which implies that workload-adjusted length of stay decreases by 18.1% in the last hour of shift under the observed assignment policy. Note that this difference is slightly smaller (more conservative) than that implied by coefficients \(\hat{\alpha}_1 = -0.232\) and \(\hat{\alpha}_2 = -0.069\) estimated without simulation using actual data (Table A-7.3 and Figure A-7.4). Given that total costs increase by 20.8% in the last hour prior to EOS, I estimate the elasticity of hospital-resource costs to workload-adjusted length of stay, for decreases in workload-adjusted length of stay that are 5.9% below baseline, as 20.8%/−18.1% = −1.15. I thus calculate hospital-resource costs as

\[
\text{HospitalResources}_{\Delta,r} = \sum_s \sum_m \sum_{t=\xi(s)} \sum \left[ \left[ t(s) - t \right] = m \right] a_{\Delta,r}(s, t) \times \exp \left( \text{BaseLogCosts} - 1.15 \cdot \min \left( 0, \hat{\alpha}_m^{\Delta,r} - \hat{\alpha}_2 \right) \right),
\]

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

### A-5.4.1 Imputing the Value of Leisure

I can also impute the revealed value of leisure in terms of hospital-resource costs by calculating the ratio between extra hospital-resource costs incurred and leisure time gained as a result of the physician discharge behavior near EOS. The discharge function \( D_0(S) \) increases the discharge hazard as \( t \) approaches \( T(s) \), shortening workload-adjusted length of stay and increasing hospital-resource costs. I examine what discharges would look like if not influenced by EOS behavior by modifying \( t \) in the discharge function. That is, I consider a modified discharge function \( D(m, S^-) \equiv D_0(\max(4, m), S^-) \) that does change as a function of time relative to EOS, at least in the four hours prior to EOS.

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

\[
\text{LeisureValue}_{\Delta,r} = -\frac{\text{HospitalResources}_{\Delta,r} | A_\Delta, D_\Delta - \text{HospitalResources}_{\Delta,r} | A_\Delta, D_\Delta}{\text{WorkCompletionTime}_{\Delta,r} | A_\Delta, D_\Delta - \text{WorkCompletionTime}_{\Delta,r} | A_\Delta, D_\Delta}.
\]

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

### A-6 Heterogeneous Effects by Physician and Peer Types

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

\[
Y_{it} = \alpha_{m(i,t)}^{Type(i)} + \gamma_{m(i,t)} + \beta X_i + \eta T_t + \kappa_{Type(i)} + \zeta_{p(i)} + \nu_{j(i),k(i)} + \epsilon_{it}, \tag{A-6.1}
\]

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

The first set of categories refer to the physician \( j(i) \) at \( t \): whether \( j(i) \) is male, whether the \( j(i) \) is older than average at \( t \), whether \( j(i) \) has greater tenure than average at \( t \), and whether \( j(i) \) is faster than average. The last characteristic is estimated from a regression of log length of stay on patient characteristics, time categories, pod identities, and (the object of interest) physician dummies. Results in Figure A-6.1 show roughly similar results, regardless of physician type, except that faster physicians are much less likely to have EOS distortions.
The second set of categories refer to the relationship between physician \( j(i) \) and the peer \( j^- (i) \) of the subsequent shift in the same managerial location, if there is one. This peer will potentially assume the care of remaining patients who are not seen or who need to be transferred, and he or she also observes the index physician ending shift to a greater degree than any other physician. I consider whether \( (j(i), j^- (i)) \) are of the same sex, whether \( j(i) \) has less tenure than \( j^- (i) \), whether \( (j(i), j^- (i)) \) are “familiar” (i.e., they have previously worked more than 60 hours in the same location), and whether \( j(i) \) is slower than \( j^- (i) \). Results in Figure A-6.2 show that physicians are more likely to have EOS distortions when working with a senior peer, when they are not familiar with the peer, and when the peer is faster.

A-7 Additional Results

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

- Table A-7.2 describes the process of constructing the sample, including the number of observations in each step.
- Table A-7.1 lists the number of observations for each shift type. Observations are counted in terms of unique shifts, hours, potential patients (who could be assigned to a shift of that shift type at time of arrival), and actual patients (who are assigned to a shift of that shift type).
- Table A-7.3 reports coefficients for EOS effects on workload-adjusted length of stay, as a continuation of Table 4. Results in this table only control for time relative to shift beginning. I use these more parsimonious regressions to operationalize workload-adjusted length of stay as the key substitute for hospital-resource costs in the structural model in Section 7, in which simulating the full set of covariates would be impractical. Results are estimated on both actual and simulated data.
- Figure A-7.1 shows evidence on how long physicians stay past EOS in terms of the share of shifts in which an order written by the attending physician of record (AOR) is yet to be written, out of shifts in which an order written by any attending physician is yet to be written.
- Figure A-7.2, Panel A, shows the AOR order share from Figure A-7.1 at one hour past EOS for shifts ending at each hour of the day. Panel B of the figure shows the corresponding census of patients remaining at EOS.
- Figure A-7.3 shows average patient counts (“censuses”) for physicians in shifts with different overlap \( \delta \).
- Figure A-7.4 shows coefficients for EOS effects on workload-adjusted length of stay, reported in Table A-7.3, estimated on both actual and simulated data.
Table A-1.1: Balance Tests, Other Predicted Outcomes

<table>
<thead>
<tr>
<th>Predicted outcome</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>A. Assumption 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours prior to EOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test p-value</td>
<td>0.748</td>
<td>0.867</td>
<td>0.855</td>
<td>0.579</td>
<td>0.208</td>
</tr>
<tr>
<td>Resid. char. distrib.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>13.88</td>
<td>0.274</td>
<td>6.777</td>
<td>0.019</td>
<td>0.060</td>
</tr>
<tr>
<td>10th percentile</td>
<td>12.16</td>
<td>0.186</td>
<td>6.487</td>
<td>0.008</td>
<td>0.053</td>
</tr>
<tr>
<td>90th percentile</td>
<td>15.46</td>
<td>0.352</td>
<td>7.035</td>
<td>0.027</td>
<td>0.067</td>
</tr>
<tr>
<td>B. Assumption 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours prior to EOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test p-value</td>
<td>0.740</td>
<td>0.869</td>
<td>0.833</td>
<td>0.793</td>
<td>0.962</td>
</tr>
<tr>
<td>Resid. char. distrib.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>13.88</td>
<td>0.274</td>
<td>6.777</td>
<td>0.019</td>
<td>0.060</td>
</tr>
<tr>
<td>10th percentile</td>
<td>10.29</td>
<td>0.132</td>
<td>6.256</td>
<td>0.003</td>
<td>0.043</td>
</tr>
<tr>
<td>90th percentile</td>
<td>17.43</td>
<td>0.413</td>
<td>7.286</td>
<td>0.031</td>
<td>0.075</td>
</tr>
</tbody>
</table>

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

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

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

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

Note: This table reports estimates in a procedure based on Altonji et al. (2005) to calculate potential bias from selection on unobservables, as described in Appendix A-1.4. Selection is modeled for whether a patient is assigned in the $m$th hour prior to EOS ($A_m$) by Equation (A-1.14), the residual of which is $A_m^\gamma$. Selection on observables is defined as $\hat{\alpha}_m = \left(\frac{\text{Var}(W') \gamma W}{\text{Var}(A_m)}\right)$, which states that normalized selection on unobservables is at most equal in magnitude to normalized selection on observables, and an upper bound of bias from selection on unobservables is calculated from Equation (A-1.15). I use $\text{Var}(\varepsilon) = 0.160$ and $\text{Var}(W' \gamma W) = 0.580$ in this calculation. $\hat{\alpha}_m$ is estimated by Equation (1); for convenience, results are repeated from Column 5 of Table 2. Finally $\hat{\alpha}_m$ is stated as a multiple of the bias upper bound in the last column of this table.
Table A-5.1: Dynamic Model Utility Parameter Estimates

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

**Note:** This table shows estimated utility flow parameters, with standard errors in parentheses. For incurred by discharging a patient in discrete dynamic programming models. The utility of not discharging any patient is normalized to 0. All models include utility flows for splines of deviation in current log length of stay from predicted log length of stay (or “deviation LOS”) and linear predicted log length of stay. To interpret the spline coefficients on deviation LOS, the first five rows report estimates for the utility of discharging a patient at deviation LOS values of (−1.5, −1, −0.5, 0, 0.5), representing approximately equally spaced quantiles in the data. These estimates are linear combinations of the spline coefficient estimates, and standard errors for these linear combinations are calculated using the delta method. Log workload-adjusted length of stay of a discharged patient is included in the utility flow of models 2, 4, and 5; hours from last discharge is included in the utility flow of models 3 to 5. Model 5 is similar to Model 4 but includes cubic splines for predicted log length of stay and hours from last discharge. In all models, the value function at EOS is only a function of the number of patients remaining on census (or “census”) and the average patient arrival time for these patients relative to EOS. Value functions at times prior to EOS are otherwise constrained by Equation (A-5.7) from the dynamic programming Bellman equation; in all models, I fit value functions with the census, the sum of deviation LOS (across patients on census), and the census interacted with time to EOS and hours from last discharge. Details of the estimation procedure are given in Appendix A-5.1. All models are fit on a sample of 792,687 patient-interval observations, corresponding to 115,674 physician-interval choice sets (groups). Log likelihoods are given in the last row. For comparison, a fully flexible static logit model (i.e., with no restrictions on what enters the utility flow) had a log likelihood of −77,071.90. Graphical fit of discharge probabilities is given in Figure 9.
Table A-7.1: Shift Type Observation Numbers

<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>6,996</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>
</tbody>
</table>

**Total** | 23,924 | 206,942| 860,544            | 369,841          |

**Note:** This table lists the number of observations for each shift type, each defined as ⟨ℓ, o, o⟩, where ℓ is the shift length in hours, o is the overlap in hours with a previous shift, and o is the overlap in hours with a subsequent shift in the same location. Observations are counted in terms of unique shifts, hours, potential patients (patients who arrive at the ED during a time when there is a shift of type ⟨ℓ, o, o⟩ in progress), and actual patients (patients who are treated by a physician on a shift of type ⟨ℓ, o, o⟩).
### Table A-7.2: 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></td>
<td>426,899</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></td>
<td>372,224</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 A-7.3: Effect on Workload-adjusted Length of Stay by Shift Overlap

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All $\bar{\sigma}$</td>
<td>$\bar{\sigma} \leq 1$</td>
<td>$\bar{\sigma} &gt; 1$</td>
<td>All $\bar{\sigma}$</td>
<td>$\bar{\sigma} \leq 1$</td>
<td>$\bar{\sigma} &gt; 1$</td>
</tr>
<tr>
<td>Hour prior to EOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last hour</td>
<td>-0.232***</td>
<td>-0.339***</td>
<td>0.031</td>
<td>-0.200***</td>
<td>-0.240***</td>
<td>-0.053</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.05)</td>
<td>(0.061)</td>
<td>(0.037)</td>
<td>(0.049)</td>
<td>(0.063)</td>
</tr>
<tr>
<td></td>
<td>-0.069***</td>
<td>-0.089***</td>
<td>0.168***</td>
<td>-0.067***</td>
<td>-0.064**</td>
<td>0.101***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.025)</td>
<td>(0.035)</td>
<td>(0.019)</td>
<td>(0.025)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>Second hour</td>
<td>-0.016</td>
<td>-0.027</td>
<td>0.176***</td>
<td>-0.013</td>
<td>-0.007</td>
<td>0.122***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.020)</td>
<td>(0.028)</td>
<td>(0.015)</td>
<td>(0.020)</td>
<td>(0.029)</td>
</tr>
<tr>
<td></td>
<td>-0.077***</td>
<td>-0.076***</td>
<td>0.087***</td>
<td>-0.052***</td>
<td>-0.044**</td>
<td>0.075***</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.026)</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Third hour</td>
<td>-0.052***</td>
<td>-0.048***</td>
<td>0.046***</td>
<td>-0.037***</td>
<td>-0.029**</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.022)</td>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.022)</td>
</tr>
<tr>
<td></td>
<td>-0.032***</td>
<td>-0.028***</td>
<td>0.013</td>
<td>-0.028***</td>
<td>-0.021**</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.014)</td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Fourth hour</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control for time relative to shift beginning</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Patient, provider, and other time controls</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Sample</td>
<td>Full, $\bar{\sigma}$</td>
<td>$\bar{\sigma} \leq 1$, $\bar{\sigma} &gt; 1$</td>
<td>Full, $\bar{\sigma}$</td>
<td>$\bar{\sigma} \leq 1$, $\bar{\sigma} &gt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>334,955</td>
<td>231,576</td>
<td>101,657</td>
<td>334,783</td>
<td>231,710</td>
<td>101,692</td>
</tr>
<tr>
<td>Adjusted $R$-squared</td>
<td>0.010</td>
<td>0.011</td>
<td>0.001</td>
<td>0.009</td>
<td>0.011</td>
<td>0.001</td>
</tr>
<tr>
<td>Sample mean outcome</td>
<td>-0.920</td>
<td>-0.987</td>
<td>-0.789</td>
<td>-0.927</td>
<td>-0.973</td>
<td>-0.798</td>
</tr>
</tbody>
</table>

*Note:* This table is a continuation of Table 4, reporting coefficient estimates and standard errors in parentheses for EOS effects on workload-adjusted length of stay, for arrival at each hour prior to end of shift (EOS), where arrival greater than six hours is the reference period. Models (1) to (3) are estimated on actual data, while Models (4) to (6) are estimated on simulated data. Models also differ by which shifts, based on overlap $\bar{\sigma}$, are included. All models are estimated with Equation (A-5.11), which controls for time relative to shift beginning but not for other variables, in order to facilitate comparison between actual and simulated data. Workload-adjusted length of stay is calculated by Equation (5). * denotes significance at 10% level, ** denotes significance at 5% level, and *** denotes significance at 1% level. Results are graphically shown in Figure A-7.4.
Figure A-1.1: Patient Selection on Observables Relative to End of Shift

Note: This figure shows selection on observables for each hour prior to end of shift (EOS) on length of stay (Panel A), orders (Panel B), inpatient admissions (Panel C), and costs (Panel D). Each outcome is predicted based on patient characteristics observable prior to acceptance (age, sex, ESI) (closed circles) and on the full set of characteristics usually unobservable until after patient acceptance (e.g., 29 Elixhauser indices, race, language) (short-dashed line, open circles). Coefficients are estimated for predicted outcome using Equation (A-1.1). For reference, adjusted effects on actual outcomes from Figure 6 are shown with the dashed line. The reference category is any time greater than six hours prior to EOS.
Figure A-1.2: Assessing Selection Bias Using Arrival Times

Note: This figure is the graphical representation of Table A-1.2, using only variation in arrival times to assess selection bias. It shows binned scatterplots of actual (residualized) log length of stay (Panel A), log length of stay predicted on “ex ante” characteristics possibly observable to physicians prior to assignment (Panel B), and log length of stay predicted on all characteristics including clinical diagnoses usually observable only after assignment (Panel C). Predicted and actual log lengths of stay are all averaged within hour cell and weighted by visit. The core data for the x-axis on all three panels is the log length of stay predicted by the times to EOS, defined by Equation (A-1.4) as $Q_t$. $Q_t$ is calculated as follows: First, coefficients on time relative to EOS are calculated from (1) using leave-shift-out sampling. Next, these coefficients are averaged across shifts in process at hour $t$, weighted by visits. To calculate residualized actual log length of stay (Panel A), I subtract expected log length of stay based on all covariates listed in the note for Table 2, except for time to EOS, using only variation within time to EOS. To calculate predicted log length of stay by patient characteristics (Panels B and C), I residualize the characteristics by time categories and use within-EOS-time variation to predict log length of stay. Patient characteristics and time categories are described in the note for Table 2, respectively. To construct each of the binned scatterplots, I demean values on the x- and y-axis, separate the data into 20 equal-sized groups (by patient visits) ordered by $A_t$, then plot the mean value within each bin. Solid lines show the best linear fit by OLS on the underlying microdata, clustered by hour (coefficients and standard errors are given as notes in each panel, also given in Table A-1.2). Details are given in Appendix A-1.3.
Figure A-2.1: Effects on Length of Stay by Shift Length

A: EOS

B: Shift Beginning

Note: This figure shows coefficients from Equation (1) estimated separately for shifts of seven or eight hours in length (open circles), nine hours in length (closed circles), and ten hours in length (open triangles). Panel A arranges estimates by hours relative to end of shift (EOS). Panel B arranges estimates by hours relative to shift beginning.
Figure A-3.1: Time Components

Note: This figure plots time components of length of stay as a function of hours relative to end of shift (EOS): time from pod arrival to first order (open circles), time from first to last (non-discharge) order (open triangles), and time from last order to discharge order (closed circles). Panel B shows marginal effects from a fractional logit model on these shares. Panel A represents these results as time in hours, incorporating results on the EOS effect on length of stay.
Figure A-5.1: Example Counterfactual Assignment Policies

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

Note: This figure shows the fit of simulated outcomes with actual outcomes observed on 22,434 shifts with non-missing shift types out of the universe of 24,499 shifts. Outcomes are plotted against the x-axis of hours relative to EOS, in 30-minute intervals. Panel A evaluates the fit of actual and simulated patient assignments at each point in time, given the actual census of patients (i.e., discharges are not modeled). Panels B to D evaluate the fit between actual outcomes (solid dots) and those simulated by two discharge models, for any patient assigned in the relevant 30-minute interval: a fully flexible “static” model of conditional discharge probabilities (hollow dots), and a dynamic model of the discharge probabilities with restrictions on what can enter into the utility flow (triangles). The dynamic model, described in greater detail in Appendix A-5.1, corresponds to the simplest model in Table A-5.1, Model 1, although the fit and the estimated parameters do not qualitatively differ across specifications. The simulation algorithm is given in Section A-5.3. Panel D (“Census”) reflects actual and simulated average census during the length of stay of a patient \(i\), \(w(i)\) in Equation (5).
Figure A-6.1: End of Shift Effect by Physician Type

Note: This figure shows estimates of effects of hour relative to end of shift (EOS) for physicians of different types, using Equation (A-6.1). In each panel, estimates for physicians of the stated type are shown in solid dots, while estimates for physicians not of the stated type are shown in hollow dots. Panel A shows male vs. female; Panel B shows physicians who are older than average (about 39 years old) vs. not; Panel C shows physicians who have higher tenure at the ED (about 5.5 years) vs. not; Panel C shows physicians who have are faster than average (i.e., have a lower fixed effect in a regression of length of stay) vs. not. p-values for the significance of the difference between the last hour effect for physicians with and without the characteristic are given as notes in the lower right.
Figure A-6.2: End of Shift Effect by Physician-Peer Relationship

A: Same Sex

B: Peer Higher Tenure

C: Familiar

D: Peer Faster

Note: This figure shows estimates of effects of hour relative to end of shift (EOS) for different types of relationships between physicians and the peer in the subsequent shift, if there is one. Estimates are from Equation (A-6.1). In each panel, estimates for physician-peer relationships of the stated type are shown in solid dots, while relationships not of the stated type are shown in hollow dots. Panel A shows physician-peer pairs of the same vs. different sex; Panel B shows physicians who have higher vs. lower tenure than their peer; Panel C shows physicians who have worked more vs. less than 60 hours together; Panel D shows physicians who are faster vs. slower than their peer. p-values for the significance of the difference between the last hour effect for pairs with and without the characteristic are given as notes in the lower right.
Figure A-7.1: Attending Physician Order-writing over Time

Note: This figure shows the activity of order-writing by the attending physician of record (AOR), conditional on an order written by any attending physician, at various points in time relative to end of shift (EOS). The AOR is the physician on the bill for patient care, corresponding to the physician whose shift is matched to a patient visit. The AOR order share on the y-axis is the number of shifts in which there exists an AOR order at a later time divided by the number of shifts in which there exists any attending physician order at a later time, as a function of time relative to EOS.
Figure A-7.2: Order-writing and Censuses by Shift Ending Time

Note: This figure shows the attending of record (AOR) order share, defined in Figure A-7.1, at one hour past end of shift (EOS) (Panel A) and the average number of patients remaining on census at EOS (Panel B), for shifts ending at various times of the day. The shaded gray area indicates “daytime” hours between 6:00 a.m. and 8:00 p.m.
Figure A-7.3: Censuses over Time

Note: This figure plots average censuses over time relative to the end of shift (EOS). Each panel shows results for physicians in shifts with a given EOS overlap time. Subsequent shift starting times are marked with a vertical line.
Figure A-7.4: Workload-adjusted Length of Stay

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