To Batch or Not to Batch? Impact of Admission Batching on Emergency Department Boarding Time and Physician Productivity

Arshya Feizi
Boston University, afeizi@bu.edu

Anita Tucker
Boston University, altucker@bu.edu

Jillian Berry Jaeker
jillian.berry29@gmail.com

William Evan Baker
Boston Medical Center, wbaker@bu.edu

We study the behavior of batching by discretionary workers in the first stage of a two-tier queuing system, and explore the trade-off it causes between their productivity and second stage wait times. Specifically, we focus on the behavior of batching admissions by emergency department (ED) physicians. Using data from a large hospital, we show that the probability of batching admissions is increasing in the hour of an ED physician's shift, and that batched patients experience a 4.7% longer delay from hospital admission to receiving an inpatient bed. Using a mediation analysis we show that this effect is partially due to the increase in the coefficient of variation of inpatient bed-requests caused by batching. However, we also find that batching admissions is associated with an average of 10.0% more patients seen in a shift, and a 2.6 minute reduction a physician's average throughput time. An important implication of our work is that workers may induce delays in downstream stages, caused by practices that increase their productivity.

Key words: Emergency Department, Batching, Boarding, Productivity, Healthcare operations, Behavior, Empirical, Queueing

---

1. Introduction

Operations management literature has employed queuing models extensively in manufacturing and service settings to optimize capacity and wait times. Nonetheless, understanding the behavior of workers is a crucial aspect of designing and operating successful queuing systems. In this work, we focus on the batching behavior of servers in a two-tier queuing system. Specifically, we focus on the servers in the first stage, who batch referrals to the second stage; we explore the trade-off it causes between worker productivity in the first stage and wait times due to batched arrivals in the second stage. To our knowledge, this paper is among the first to study this trade-off.

We investigate this question empirically in an emergency department (ED) setting. We choose this setting for two reasons. First, an ED is essentially a two-tier queuing system; arriving patients are queued to be roomed in the ED, and those who require further hospital care are admitted, and are queued to receive an inpatient bed. Second, ED overcrowding is a worldwide problem, and part of the solution is improving patient flow in the ED (Morley et al. 2018). The main reason for ED overcrowding is ED boarding, which is defined as the delay from when an ED patient is admitted to the hospital for inpatient care until she physically departs the ED (ACEP 2018, Morley et al. 2018). Our paper focuses on the behavior of batching admissions by ED physicians, a practice that generates batched arrivals to the queue for inpatient beds. We explore when physicians are more likely to batch admissions, and its impact on boarding times and their shift-level productivity. Specifically, we ask the following research questions:

1. Do physicians batch their admissions toward the end of their shifts?
2. Does batching admissions negatively affect boarding times?
3. Does batching admissions improve physician productivity?

We collect 25 months of detailed, retrospective patient and ED physician shift data from a single urban academic hospital with one of the busiest EDs in the US (Mackenzie Bean 2018). Arriving patients are first triaged, and roomed primarily on a first-come-first-serve (FCFS) basis according to their severity level. Physicians sign up to treat patients once they are roomed. Once a patient undergoes the required evaluation and treatment, their attending physician determines whether the patient needs to be admitted to the hospital for inpatient care or discharged. If she admits the patient, a “bed request” is submitted electronically, along with some details about the patient’s condition. Physicians may decide to postpone this task to attend to other patients, especially when the ED is busy. Once a bed is requested, the responsibility of managing patient flow is transferred to departments outside of the ED. First the bed control department must assign an inpatient bed; this process primarily depends on the inpatient unit capacity and the number of admitted patients awaiting transfer. Next, the cleaning staff must ensure that the inpatient bed and room is ready, and patient transport must be scheduled to transfer the patient from the ED to the inpatient bed. Finally, the ED physician signs the patient out and then the patient leaves the ED.

Occasionally, a physician cannot make a disposition decision (i.e. discharge or admit) by the end of their shift. In these cases, the patient is handed off to the next physician. However, there is a disincentive to hand-off patients because the second physician in our study ED is not credited for the patient's care. We hypothesize that physicians start their shifts with attending to many patients, but toward the end of their shifts, they wrap up the tasks at hand to avoid handing their patients off to the next physician. This process may result in having to admit multiple patients in a relatively short period of time or in other words, *back-to-back*. We refer to this behavior as admission batching.

We define two or more admissions to be back-to-back (i.e batched) if they occur within a threshold of 9.1 minutes of one another. The threshold is derived empirically, and is in line with our observations of the ED, and the experience of our ED physician partner. We find that the probability of batching admissions is increasing in the hour of shift, and in fact, physicians are $4.8\times$ more likely to batch their admissions in the final two hours of their shift compared to the beginning of their shift. Next, we find the impact of batching admissions on the boarding times of batched patients, and on physicians' shift-level productivity.

We follow Shi et al. (2015), and divide boarding time into two portions: the delay from when a bed is requested until one is assigned (pre-allocation delay), and the delay from when the bed is assigned until when the patient physically departs the ED (post-allocation delay). We focus on pre-allocation delay to the hospital's medical units as our key measure of boarding time. Drawing from queuing and healthcare operations literature, we hypothesize, and show, that batching admissions increases pre-allocation delay. Specifically, batched patients experience a pre-allocation delay increase of ~ 7 minutes, or 4.7%, on average. We further show that this increase is, in part, due to the increase in coefficient of variation (CV) of bed requests. In a counterfactual analysis we estimate that average system-level pre-allocation delay may be reduced by a theoretical maximum of 15% if batching is eliminated. Furthermore, we find that the impact of batching on pre-allocation delay in our empirical results are in the same order of magnitude as the predictions of queueing models with batched arrivals.

To study the impact of admission batching on physician productivity, we consider two productivity measures used in the healthcare operations literature, and relevant in practice: time to making a disposition decision (akin to discharge rate used in Song et al. (2015) and Song et al. (2018)), and number of patients served (KC et al. 2020). We find that in shifts with at least one batch of admissions, the physician attended to an average of 2.1 additional patients in their shift, equivalent to a 10.0% increase in number of patients served. Also, their shift-level average time to disposition decision decreases by $\sim$2.6 minutes.

This work makes several contributions to theory and practice. First, much of the analytical work in healthcare operations management literature that studies ED boarding has primarily employed queuing models with time-homogeneous or -non-homogeneous Poisson arrivals (or admissions), with a focus on either capacity-related or patient-flow-related issues, with the goal of better matching inpatient bed supply with demand over the course of a day (Green 2002, Best et al. 2015, Véricourt and Jennings 2011, Yankovic and Green 2011, Dai and Shi 2019, Shi et al. 2015, Whitt and Zhang 2017). Our findings complement this literature. We focus on a behavior of ED physicians, batching admissions, and empirically show that it exacerbates boarding times. This is important because it demonstrates that in a multistage queue, such as a hospital's ED, server behavior may distort an assumed exogenous interarrival distribution to the system when passing units to downstream stages, and add to its inherent variance. Moreover, to our knowledge, our paper is the first to empirically quantify the impact of batching on wait times, and tie these results with analytical work on queueing systems with batch arrivals. Second, this paper contributes to the behavioral operations literature on the impact of batching tasks on worker productivity. Whereas Ibanez et al. (2017) shows a decrease in productivity when workers try to batch similar tasks, we demonstrate that batching may be associated with an increase in productivity. Third, from a practical lens, we find that boarding is partly caused by ED physicians' work practices, something that is under the ED's control. That said, we also show that batching admissions is associated with an increase in shift-level productivity. This implies that the impact of batching admissions on ED overcrowding is a trade-off between longer boarding times vs. shorter patient throughput times. Ultimately however, we find that overall ED length of stay (LOS) is longer in physician shifts with batched admissions, and that patients admitted in a batch also experience a longer LOS.

2. Literature Review

The medical literature has investigated and offered solutions to ED overcrowding by either reducing ED demand (e.g., by improved access to primary-care physicians), expediting diagnosis and treatment process (e.g., the provider-in-triage model), or shortening the discharge/boarding process (e.g., setting ED LOS targets) (Moineddin et al. 2011, Napoli et al. 2020, Ngo et al. 2018, Jones et al. 2017, Forero et al. 2010). See Morley et al. (2018) for a thorough review.

Our paper focuses on the behavior of batching admissions by ED physicians, the gatekeepers to inpatient care. We study the trade-off between higher physician productivity and increased boarding times caused by this behavior. As such, our research is most relevant to three streams of operations management literature: 1) healthcare operations, with emphasis on queuing models that investigate ED patient flow and boarding, 2) behavioral queuing, with a focus on the effect of server behavior on system performance, and 3) queues with batched arrivals, and batching behavior by servers.

Many papers in the operations management literature have developed models for capacity planning on a strategic level, with the goal of timely service to patients (Green 2002, Yankovic and Green 2011, Yom-Tov 2010). For instance, Best et al. (2015) finds the optimal number of hospital wings considering the trade-off between the boarding time advantage offered by capacity pooling vs. the shorter length of stay when patients receive focused care in specialized wings. In a similar context, Véricourt and Jennings (2011) finds the optimal number of nurses considering how frequently patients require nursing services.

Other scholars have studied bed allocation and discharge policies, considering that system dynamics at the level of hour-to-hour resolution is crucial for operational planning in practice (Armony et al. 2015, Powell et al. 2012, Shi et al. 2015, Chan et al. 2017, Saghafian et al. 2012). For example, Chan et al. (2017) finds the optimal number and timing of physician rounds in an inpatient unit, noting that patients need a physician to evaluate them before discharge. Most relevant to our paper is Shi et al. (2015), which uses a stochastic network model to capture the complex dynamics of hospital operations on an hourly level, and uses it to study discharge policies that smooth ED patient flow, and predict boarding times. Most modeling papers assume that admissions (or arrivals) follow a time-homogeneous or -non-homogeneous process, where $CV = 1$. While this assumption is tractable for modeling purposes, it does not capture the system dynamics due to server behavior.

The behavior of servers in queuing systems has been shown to vary with queue structure (Song et al. 2015, Wang and Zhou 2018) and system utilization (or workload) (KC and Terwiesch 2012, Berry Jaeker and Tucker 2016). For instance, counter to the prediction of theoretical models, Song et al. (2015) shows that wait times are lower when the ED assigns patients to physicians upon arrival to the waiting room (dedicated queue) instead of the time when they are roomed (pooled queue). This is because physicians in a dedicated queue were able to manage their workload more efficiently, and became more productive. Similar results were obtained in Wang and Zhou (2018) in a supermarket setting. A few studies have also investigated server behavior in two-tier queuing systems, similar to our ED setting. For instance, Batt and Terwiesch (2016) shows that to alleviate downstream congestion under high demand, triage nurses in the first stage of ED care initiate diagnostic tests which are normally handled by the ED physician, in the second stage of ED care. Freeman et al. (2017) finds that midwives in a maternity department, who act as gatekeepers to specialists, increase their referral rate in response to increased workload. Our study adds to this literature by highlighting the trade-off between increased productivity and longer wait times, when gatekeepers batch referrals to the second stage of a two-tier queuing system.

The characteristics of queues with batched arrivals (e.g., $M^X/M/c$ queue) have been investigated extensively, analytically (Burke 1975, Sakasegawa 1977, Eikeboom and Tijms 1987, Van Ommeren 1990, Lee and Srinivasan 1989, Pang and Whitt 2012, Yao 1985). A recent application of such model is demonstrated in the context of optimizing blood testing procedures, which is performed in batches (Bar-Lev et al. 2017). However, studies of batching *behavior* by workers, and whether it is operationally beneficial, are relatively new. Closely related to our work is Dobson et al. (2012) that studies batching by ED residents. Inspired by the observation that ED residents visit multiple patients (batch visits) before discussing each individual case with their attending physician, Dobson et al. (2012) uses a queuing model to study its impact on ED throughput time. It proposes a three-stage tandem queueing system with two distinct servers: a resident and an attending physician. Each patient (pronoun *he*) receives service in three stages. In the first stage, the resident examines the patient. In the second stage, he enters a concurrent queue where he waits for a consultation between the resident and the attending physician. In the third stage, the patient enters the attending physician's queue and waits to receive the final examination after which he leaves the system. When the attending physician is busy, the resident has the option to wait for the attending physician to become free to perform a consultation, or alternatively, may continue to take on new patients from the first queue and present them to the attending physician in a batch. Dobson et al. (2012) shows that a throughput-maximizing policy involves batching, albeit the batch size for number of patients presented to the attending physician for a single consultation must be below a threshold. Our empirical results are in line with the findings of Dobson et al. (2012) in that we show that batching may be beneficial for physician productivity (i.e., throughput time). However, we also find that batching admissions may negatively affect the wait times of downstream queues. An important gap we address in this body of work is to *empirically* study a queueing system with batched arrivals, as suggested by Pang and Whitt (2012).

Ibanez et al. (2017) shows that when faced with a queue of images to read, radiologists are likely to modify the work sequence from a first-come-first-serve policy (FIFO), to spending time to find similar tasks in the queue, and performing those tasks in a batch. This behavior is shown to be detrimental to worker productivity since the deviation cost outweighs the benefit of batching, resulting in a higher average throughput time. In another empirical study, Meng et al. (2018) finds that nurses batch patient visits to distant rooms in the ED to minimize walking, which results in longer wait times between nurse visits and lower satisfaction for these patients. We contribute to this literature by demonstrating that batching admissions by ED physicians is more likely to occur towards the end of their shifts; and unlike (Ibanez et al. 2017), we find that this behavior increases physician productivity, but at the cost of longer boarding times for the batched patients.

3. Clinical Setting

Our data comes from a large 567-bed academic institution with a level 1 trauma center, which consists of two distinct campuses. The hospital's ED served over 135,000 patients in 2018, making it one of the top 10 busiest EDs in the United States (Mackenzie Bean 2018).

Similar to most EDs, patients arriving at our study ED are first registered, triaged, and assigned an Emergency Severity Index (ESI) score from 1 (most severe) to 5 (least severe) based on their medical condition (Gilboy et al. 2011). After triage, walk-in patients – which constitute about 70% of ED arrivals – wait in the waiting room to be placed in an ED bed. We refer to being placed in a room as "being roomed". Patients arriving via ambulance (about 30% of patients) are roomed within minutes of arriving to the hospital. Waiting room patients with lower ESI scores have higher priority, and within a given ESI score, patients are mostly served on a first-come-first-serve basis[1].

Patients are roomed in a particular "pod". The ED consists of 4 pods, where each pod is an area in the ED staffed by a team of caregivers. Although each pod has a limited number of beds, it is common for triage nurses to place patients in a gurney in the hallway spaces within the pods when the ED is busy. Hallway placement is a common surge capacity policy across many EDs in the US (Stiffler and Wilber 2015). This implies that in practice, the ED capacity is not constrained by its number of beds, as most ED arrivals are accommodated in a timely manner.

Staff scheduling is such that each pod is managed by *at least* one physician in the busier hours of the day, in addition to 3–6 nurses. In the quieter hours such as after midnight and weekends, one physician may be responsible for managing patient flow in more than one pod. However, staffing levels are relatively constant for a given shift. In other words, for a given day of week and time of day, staffing levels do not vary. Noteworthy is that physicians are rotated, so they are placed relatively uniformly across the different sections of the ED.

Physicians sign up for patients after they are roomed. There are no direct financial incentives for physicians to take on new patients. In addition, although their productivity does contribute to their relative value units (RVUs) which are tracked, it is not heavily emphasized on an individual basis; but is instead a group performance metric. That said, the institutional culture is such that attending physicians feel ownership of their designated pods, and responsibility for managing patient flow. Also worth noting is that there are no policies that limit the number of patients that a physician can take.

Once a patient is picked up by a physician, they are diagnosed and treated. Most diagnosis tests are ordered shortly after the initial evaluation, and in some cases, additional tests are ordered during their ED course. Finally, the physician must make a disposition decision; she must either discharge the patient to go home (or to another facility), or admit the patient to the hospital for further care by electronically submitting an inpatient "bed request", along with details about the patient's condition. A point to note is that often residents assist physicians. It is possible that the resident submits the patient's bed request, and the attending physician "cosigns" the admission. This policy facilitates admissions and prevents delays, especially when ED workload is high. Cosigns need not be instantaneous. For example, a physician may cosign an admission even 24 hours after it occurred. The bed request will be under the attending physician's name, and not the resident. This detail is important because it justifies admissions with timestamps that are after the physician's assigned shift has ended, as we later discuss in section 5.2.

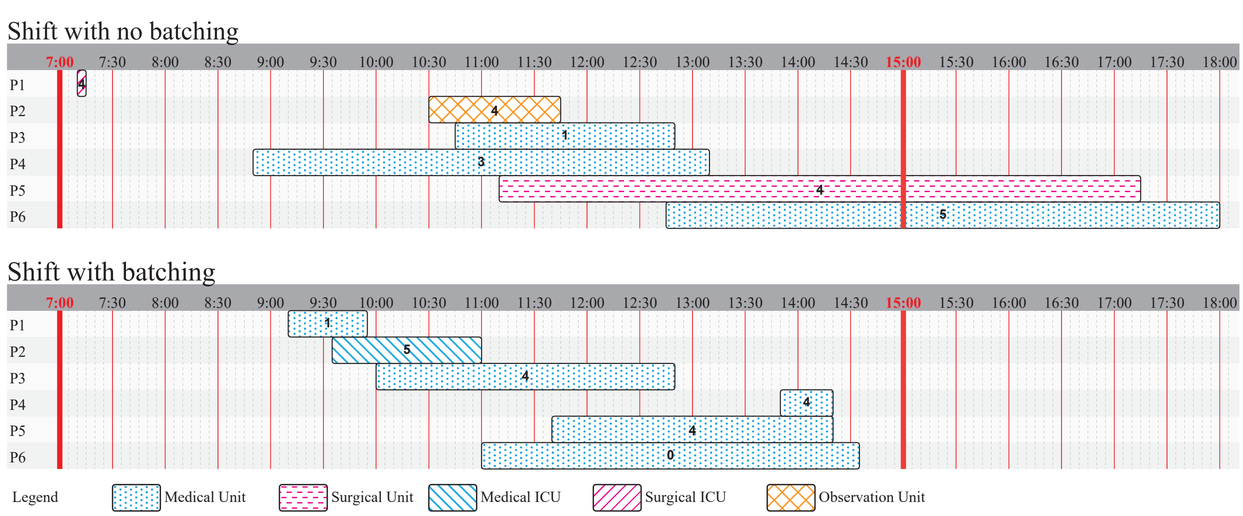
A patient's admission process involves reviewing charts and/or pulling test results, and the median time-to-admission, defined as the time interval from when they are roomed until they are admitted, is ~ 170 minutes. Nevertheless, physicians typically have prior beliefs on which patient(s) will end up being admitted. As such, reviewing test results is more of a *confirmation* on a prior belief than providing *all information* about the patient (Li et al. 2020)[2]. That said, some cases will require some critical data. For instance, a patient presenting to the ED with abdominal pain requires a computerized tomography (CT) scan test to be examined to ascertain if they have appendicitis and need to be admitted, or can be discharged with minimal risk of readmission. Although we do not have detailed patient-level data for the turnaround times (TATs) of all tests[3], from reviewing limited data we know that the median TATs for imaging exams (e.g. radiology or CT), electrocardiograms (EKGs) and ultrasounds, and laboratory tests (e.g. blood or urine) are ~ 90 minutes, < 15 minutes, and ~ 60 minutes, respectively. Also, there is little variation in test TAT. Furthermore, a patient's test results may be available, but the physician is busy with other patients and therefore cannot read the results immediately.

If a patient's disposition cannot be determined by the time their physician's shift ends, the patient is handed off to the next physician coming in. Such instances may occur either because the patient's condition requires additional observation, or sometimes, because they are still awaiting important test results. Interestingly, although the new physician may even end up performing much of the work, she does not "take credit" for the patient, and therefore, in most cases, will not even add herself to the patient's caregiver team. This is because the patient is billed under the first attending who signed-up for the patient, who, by convention, completes a provider note. Therefore, there is a disincentive among physicians to sign up for patients toward the end of their shifts because they know the incoming physician will be available soon, and can "start fresh" with a patient and be the primary provider on the case. Rather, toward the end of shifts physicians aim to wrap up the cases they already have, which may result in batching of admissions.

To further illustrate how batching admissions may occur, Fig. 1 shows the timeline of patient admissions in two 7am – 3pm physician shifts from our data which are similar in terms of number of patients admitted, discharged, and the average ESI scores of the patients. Each patient's bar represents the time from when they were roomed until when they were admitted, rounded up to the nearest five minutes. Also, the number of diagnosis tests for each patient is noted on each bar. The physician in the top panel spreads admissions more evenly, and conversely, the physician in the bottom panel admits three of her patients (patients #4, #5, #6) toward the end of her shift. (Note that patient #6 in the bottom panel had no diagnosis tests.) This anecdote further suggests that the number of tests does not necessarily determine whether or not a patient is admitted in a batch.

Once a bed is requested, the bed control unit, staffed by 1–3 people depending on the time of day, must assign an inpatient bed to the patient. This takes an average of 1.2 hours for an ICU bed and 2.1 hours for a general ward bed. The bed control unit also notifies an inpatient provider that she is to receive a patient. Noteworthy is that medical ward patient bed placement is performed by clerical staff from the bed control admitting office, whereas ICU patient bed placement is done by a senior level bed control nurse and uses a different bed allocation process. Once notified, the inpatient provider will then call

Figure 1 (Color online) Example of two shifts, with and without admissions batching.



the ED physician and signal her readiness to receive the new patient. Occasionally, the inpatient caregiver is a physician-in-training or resident who is ineligible to admit patients if they have reached their daily quota (also known as *admission cap*); in this case, the unit is said to be "capped". These capping rules are enforced at the national-level based on the trainee's specialty (ACGME 2019), and are designed to manage resident workload and improve patient safety (Thanarajasingam et al. 2012). Patients admitted to a unit which is capped may experience a longer boarding time since they will have to wait for a physician, or even a shift change, to be transferred. Once the inpatient bed and the providers are ready, finally, the ED physician performs patient sign-out, and the patient is ready to leave the ED.

4. Hypothesis Development

4.1. Likelihood of Batching Admissions Toward End of Physician Shifts

The literature on worker productivity in healthcare settings provides compelling evidence that workers increase their service rate toward the end of their shifts (Deo and Jain 2018, Batt et al. 2019, Chan 2018). For example, Chan (2018) finds that ED physicians accept fewer patients and rush to complete their work toward the end of their shifts. Batt et al. (2019) finds similar results, and that physicians attempt to either discharge or admit the patients in their care to avoid handing them off. Given the patient flow process and institutional culture of our study ED explained in section 3, we posit that at the beginning of the shift, a physician is fresh, and likely to accept as many patients as they can in an effort to facilitate patient flow in the ED. In addition, after visiting and treating a patient, she may decide to examine others before returning to her desk to fill out orders for tests and consults for the set of patients. Conversely, toward the end of the shift, physicians slow down on taking new patients, and speed up to complete the pending tasks for the remaining patients in their care. This behavior is analogous to reordering tasks in a queue, and executing similar tasks in a batch (Ibanez et al. 2017). As a result, we theorize that physicians are more likely to admit multiple patients back-to-back toward the end of their shift. Therefore:

Hypothesis 1. *The likelihood of batching admissions is higher toward the end of a physician's shift.*

4.2. Impact of Batching Admissions on Pre-allocation Delay

We hypothesize that batches of admissions are more likely to form toward the end of physician shifts. Considering that most of the shifts in our ED setting are non-overlapping (most physicians rotate at 7am, 3pm, and 11pm), any form of batching behavior on an individual-level due to the hour in shift will be amplified to a system-level because other physicians are most likely batching around the same time.

When the bed control unit is faced with a sharp increase in admission rate, it may become overloaded and operate past the point of speeding up (Berry Jaeker and Tucker 2016). Specifically, to assign a bed, the bed manager must first communicate with the ward manager that they would be receiving a new patient. This is a serial process if the patients in the batch are to be placed in different wards. In other words, multiple beds cannot be assigned at once, and there is little a bed manager can do to expedite the process. Furthermore, there may not be sufficient staff in a ward to accommodate a batch of patients if the unit is capped. Therefore, we theorize that a spike in admissions caused by physician-level batching will result in an increase in pre-allocation delay. Stated formally:

HYPOTHESIS 2. *Patients who are admitted in a batch experience a longer pre-allocation delay.*

4.3. Impact of Batching Admissions on Physician Productivity

Similar to the idea of reducing setup times per unit of inventory in batch manufacturing (Drexl and Kimms 1997), we posit that physicians batch their admission decisions to minimize the setup costs associated with switching between tasks – whether the cost is physical (e.g. Meng et al. 2018) or cognitive work (e.g. Staats and Gino 2012). For instance, the task of evaluating patients involves walking to patient beds, interacting with patients, and ordering diagnostic tests. In contrast, the task of completing a case comprises discharging or admitting the patient and also, discussing the assessment and plan with the care team, and executing other communications as appropriate (with the patient, her/his outpatient providers, family, etc). Signing up for multiple patients to evaluate, and subsequently, batching admissions when possible, may facilitate productivity through serving more patients at once (KC 2014), while avoiding interruptions (Froehle and White 2014). Therefore, we theorize that physicians who engage in batching admissions serve more patients in their shift.

Hypothesis 3A. *More patients are served in shifts with batched admissions.*

Arguably, not all patient admissions can be postponed to the end of a shift. For instance, patients who need immediate care cannot be delayed. As such, we hypothesize that physicians are selective as to *which* admissions are postponed to the end of a shift. KC et al. (2020) shows that under increased workload, physicians are more likely to select easy tasks – defined by less severe patients in that study – to complete, which increases their shift productivity. Building on this, we posit that physicians are more likely to complete non-complex cases (i.e. those with an easy decision to discharge or admit) on an individual basis, but delay the admission decision of complex cases (i.e. those with a more difficult decision) to complete in a batch. Taken collectively, we theorize that:

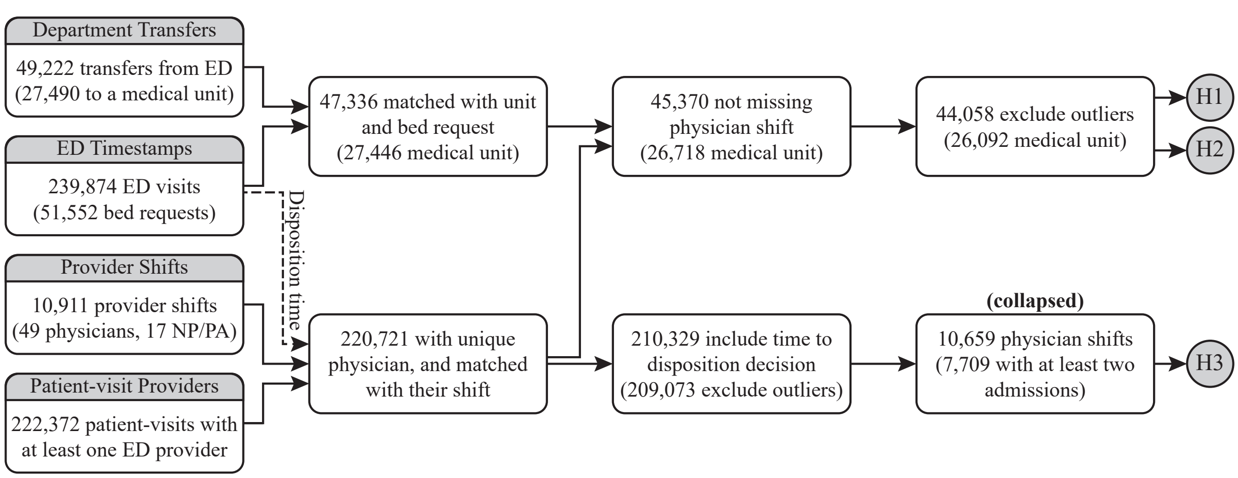
Hypothesis 3B. *Average disposition time is lower in shifts with batched admissions.*

5. Data and Variable Description

5.1. Data Sources and Cleaning Procedure

Our data consists of patient-level data and physician shift data from 1 June 2016 to 31 July 2018. It consists of four separate data sets. We summarize each data set below, and describe the cleaning process. A summary illustration of the descriptions and data cleaning procedure is given in Fig. 2.

Figure 2 Data sources and data cleaning process.



1) ED timestamps: The first data set consists of patient-level data for all ED patients. This data includes the date and time of day for: when they arrived, when they were roomed, the tests performed in the ED, when they were admitted (i.e. a bed request was issued), when an inpatient bed was assigned, and when they physically departed the ED. It also includes their ESI score, the hospital service they required, and their visit class (inpatient or observation). In addition, our ED patient data includes demographic information: age (in years), race (9 categories, coded as Black vs. non-Black), gender (male/female/unspecified), insurance type (4 main categories). After excluding patients who left the ED prior to being roomed, the study period comprises 239,874 patient visits in the ED (108,931 unique patients), where an inpatient bed was requested on 51,552 of these visits (29,574 unique patients). Additional details are provided in table EC.24 and Fig. EC.2 in the electronic companion.

2) **Provider shifts:** The second data set consists of the dates and times of ED caregiver shifts. It includes 10,911 physician shifts during the study period. There are a total of 66 caregivers in this data set: 49 physicians and 17 nurse practitioners (NPs) or physician assistants (PAs). Over 90% of the ED physician work shifts in our study period are 8 hours long, and are mostly divided into the following: 7am – 3pm (20.4%), 3pm – 11pm (20.6%), 8am – 4pm (18.2%), 4pm – 12am (11.9%), and 11pm – 7am (14.9%). Each physician in the data set has an median of 250 shifts.

3) **Patient providers:** The third data set is the patient provider data, which consists of all the caregivers (including physicians, NPs, PAs, nurses, residents, surgeons, consultants, therapists and etc.) who served a patient during her visit, and timestamps for when their service to the patient began and ended. We match 222,372 patient visits with at least one ED provider.

4) **Department transfers:** The final data set includes timestamps for arrivals and departures to and from all hospital departments for all patient visits. This data allows us to physically track the location of patients from the time they entered the hospital to the time they were discharged. Out of the patients who started their clinical pathway from the ED, 49,222 were transferred to inpatient care units, 27,490 of whom where admitted to a medical unit.

Boarding time consists of two segments: the delay between bed request and bed assignment, known as "pre-allocation delay", and the delay between bed assignment and physical departure from the ED, known as "post-allocation delay" (Shi et al. 2015). We use pre-allocation delay as the dependent variable in our analysis of boarding times in hypothesis

2 because it is closely related to the operations of the ED and the state of the hospital. For example, hospital occupancy levels or weekends are important predictors of pre-allocation delay. Also, we are interested in how the batching behavior of ED physicians impacts boarding times; and pre-allocation delay – which starts with the physician requesting a bed – is more proximal to physicians' behaviors than is post-allocation delay. In other words, pre-allocation delay is determined by the bed management unit's capacity to assign beds, which is partly influenced by the timing of bed requests by ED physicians. In contrast, despite the operational significance of post-allocation delay, its underlying causes are primarily due to inpatient operations not present in our data (Shi et al. 2014).

When estimating pre-allocation delay, we focus on patients who were admitted to a general medical ward, rather than those admitted to an ICU, surgical ward, or observation ward since this approach offers the cleanest estimation of pre-allocation delay. The reasons are as follows. First, medical admissions comprise the largest group of patients. $\sim 56\%$ of the patients admitted to the hospital from the ED fall into this category. (See table EC.26 of electronic companion for full breakdown of admissions and average ED boarding time for admissions to each unit.) Those patients also have long boarding times in general. Thus, they are the subset of patients for whom ED boarding poses the biggest problem. Second, there are inherent differences in the types of units; (i) ICU bed allocation requires senior level bed control nurse, (ii) observation units operate differently than medical units, and (iii) most admissions to surgical units are not directly from the ED ($\sim 70\%$), and the admission-related operations of the surgical units are heavily dependent on scheduling of surgeries, which is not the focus of this paper. Furthermore, off-service placement can affect boarding time and is primarily a problem of medical patients being placed on surgical wards (Dai and Shi 2019, Song et al. 2020). Thus, off-service placement, which is not the focus of our study, could inflate pre-allocation delay if we include surgical ward placements.

As depicted in Fig. 2, first we merge the ED timestamp data with the department transfers data; this results in a data set of 47,336 patient visits, where the patient was transferred to an inpatient care unit in response to an ED bed request, and their destination unit and pre-allocation delay is known. A sub-sample of 1,886 patients (44 medical) in our sample were dropped because they were missing either a "bed request" or "bed-assigned" timestamp, so we could not calculate pre-allocation delay for these observations[4]. Complete summary statistics for admitted vs. non-admitted patients are given in table EC.27 in the electronic companion.

Next, we match each patient visit's provider(s) with their shifts. For patient visits with more than one ED provider ($\sim 10\%$ of patient visits), we further use the exact disposition time from the ED time stamp data to identify the unique person who discharged or admitted the patient by matching the time frame the physician served the patient with the disposition time. This results in a data set of 220,721 patient visits for which the provider at the time of disposition, and her shift, is known. This data set is used to determine whether an admission was done in a batch with other admissions, and find at what point in the shift it took place. Note that all admitted patients – and not only medical – are included when determining whether the admission was done in a batch. Out of the observations in this data set, 210,329 patients were not missing their time to disposition decision, defined as the time interval from when a patient is roomed to when their physician makes their disposition decision (discharge or admit). To test whether batching admissions impacts physicians' throughput time, we perform our analysis on a physician-shift level, where the key dependent variable of interest is the average time to disposition decision in a shift. We trim time to disposition decision on a patient-level by removing the observations in the top and bottom 1% of its distribution to eliminate outliers, then we collapse the data set at the physician-shift level. This results in a sample of 10,659 physician shifts, out of which 7,709 shifts had at least two admitted patients. We use this sample to test hypothesis 3. Note that we exclude shifts with under two admissions because if a shift has fewer than two admissions it is not possible for physicians to batch admissions.

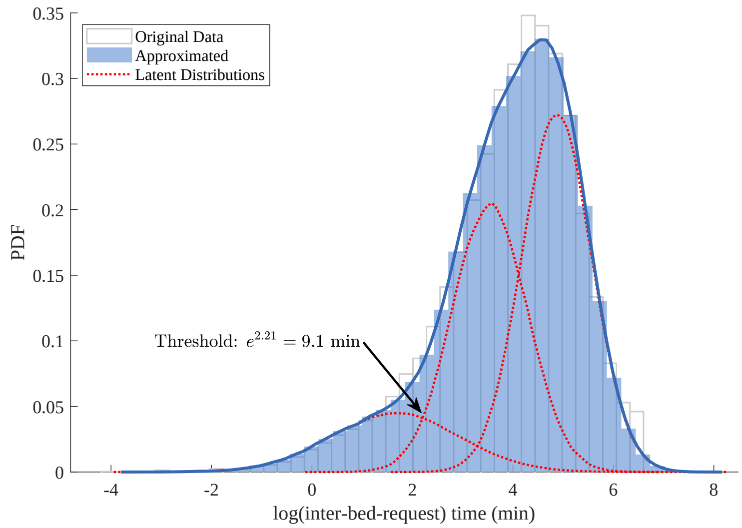
Finally, we match the "department transfers with timestamps" data set, with "patient provider with shifts" data set, to obtain a sample of 45,370 admitted patient visits, where the physician who admitted them and her shift, and whether or not they were admitted in a batch, is known. After trimming the top and bottom 1% of pre-allocation delay, and removing observations where the hospital occupancy was under 20% at the time of admission ($\sim 1\%$ of data), we arrive at a sample of 44,058 admitted patient observations, out of which 26,092 are admitted to a medical unit. We use this sample to test hypotheses 1 and 2.

5.2. Independent Variables

There are three key independent variables used in our study: a binary variable to determine whether an admission was done in a batch, the time left in a shift when an admission occurred, and a binary variable on a shift-level that determines if batching occurred in that shift.

5.2.1. Batched Admissions: Batching admissions is admitting patients "back-to-back". Since there is no universal definition for how high a physician's admission rate must be for a batch to be formed, or even any objective way of knowing from our data if a physician batched her admissions, we rely on an empirical estimation. We use a Gaussian mixture model (GMM) for this purpose. Mixture models are probabilistic tools to identify latent (i.e. unobserved) sub-samples within a larger sample; for example, to identify unobserved customer segments within a large group of shoppers (Fiebig et al. 2010). In our paper, we assume that the distribution of time between bed requests for physicians is a combination of multiple latent distributions, one of which belongs to batched admissions. Then, our goal is to find the distribution of inter-bed-request times for batched admissions, and determine the empirical inter-bed-request time threshold, below which two or more patients are assumed to be admitted in a batch. To execute this idea, we first find the empirical distribution of intervals between successive admissions (i.e. inter-bed-request time) during a physician's shift, and log-transform it to reduce its skew, and allow for negative values in the distribution. Next, we find the optimal number and parameters for Gaussian distributions that estimate the log-transformed distribution using the expectation maximization (EM) algorithm (Bishop 2006, ch. 9). As depicted in Fig. 3, we find that three distributions provide the best approximation. We assume that the distribution with the lowest mean is the inter-bed-request time distribution of batched admissions. We set the batching threshold to 9.1 minutes, equivalent to the time interval below which the probability of an inter-bed-request belonging to the batched distribution exceeds 50%. This threshold is practically in-line with our observations at the ED, and the experience of our ED physician partner. $\sim 18\%$ of the patients in our data are admitted in a batch. In $\sim 80\%$ of the cases the batch size is two, and in $\sim 97\%$ of the cases the batch size is less than four. We show the robustness of our results to defining a batch based on a minimum batch size of three in section EC.1.5, and defining a batch from *all* disposition times (instead of only admissions) for hypotheses 1 and 3 in section EC.1.4 of the electronic companion.

Figure 3 (Color online) Distribution of inter-bed-request time and approximated distribution using GMM



Note. Goodness-of-fit plots are provided in Fig. EC.3 of the electronic companion.

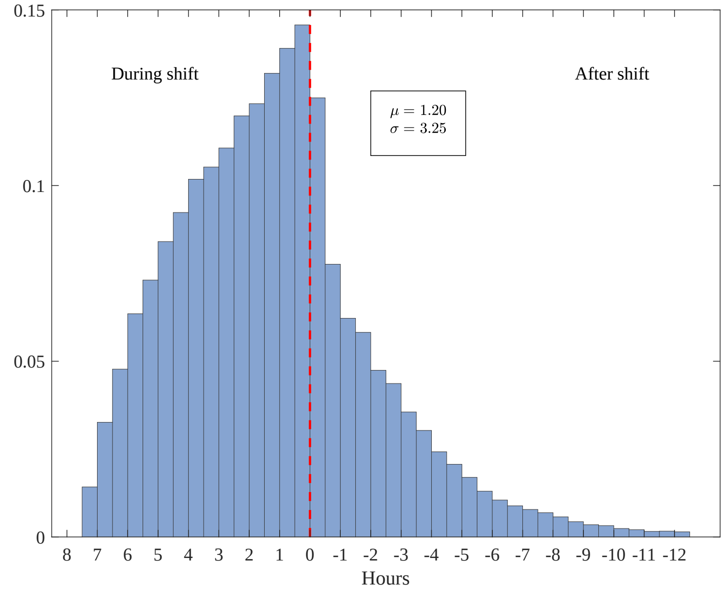
5.2.2. Time Left in Shift: To test hypothesis 1, we must show the probability of a physician batching her admissions over the course of her shift. We define "time left in shift (TLS)" for each admitted patient as:

$$TLS_i = ShiftEnd_{js} - BedRequest_{ij}$$

where $BedRequest_{ij}$ is the date/time server j admitted patient i, and $ShiftEnd_{js}$ is the date/time server j's shift s ended. This variable captures how much time was left in a server's shift when she admitted a patient.

The histogram for TLS is shown in Fig. 4, where the red dashed line at zero indicates the end of a shift. Note that the number of bed requests increases towards the end of the shift ($TLS = 0$), but decreases once the shift ends. The negative values of TLS suggest that the bed request occurred *after* the server's shift ended. While this may seem counterintuitive at first, it is actually quite plausible for two reasons. First, many times a resident submits the bed request, and the attending physician cosigns it, even after their assigned shift has ended (Arndt et al. 2017). Second, most often physicians do not add themselves to the caregiver team of patients handed off to them since they are not credited for the patient. Therefore, the patient is not admitted under the second physician's name, and in the data, it appears as though the first physician admitted the patient after their assigned shift had ended.

Figure 4 (Color online) Histogram of time left in shift when requesting inpatient bed



Note. The x-axis is reversed to aid readability.

5.2.3. Shift with Batching: To find support for hypothesis 3, we must show that when physicians batch their admissions their productivity increases. We define the variable, “shift with batching (SB)”, as:

$$SB_{js} = \mathbb{1}[N_{Batch,js} \geq 2]$$

where $N_{Batch,js}$ is the number of patients admitted in a batch for physician j at shift s and $\mathbb{1}[.]$ is an indicator function. Hence, SB_{js} is a binary variable equal to one if there are at least two patients admitted in a batch. Note that by definition, two is the minimum number of patients required for a single batch to be formed. Out of the 7,709 physician shifts with at least two admissions (see Fig. 2), $\sim 43\%$ are a batched shift. We identify the impact of SB on the average number of patients served in the shift (hypothesis 3A) and average throughput time (hypothesis 3B).

5.3. Dependent Variables

5.3.1. Pre-allocation Delay: Hypothesis 2 studies ED boarding, which is the delay between being admitted to the hospital and leaving the ED. For the reasons explained in section 5.1, we use pre-allocation delay as our key dependent variable. We log transform pre-allocation delay (hereafter denoted by $LPAD$) in our regression analysis because the distribution is otherwise right-skewed (Song et al. 2015). The empirical PDFs of pre-allocation delay, post-allocation delay, and boarding time are shown in Fig. EC.4 for admissions to the medical units.

5.3.2. Shift Productivity: We use two measures of shift productivity: 1) the total number of patients seen in shift, denoted by N_{pts}, similar to KC et al. (2020); and 2) patient throughput time, similar to Song et al. (2015), but averaged on a physician-shift level.

Average shift throughput time, denoted by $\overline{STT}$, is calculated as follows:

$$\overline{STT}_{js} = \frac{1}{N^D_{pts,js}} \sum_{i=1}^{N^D_{pts,js}} (Disposition_{ijs} - Roomed_{ijs})$$

where $Disposition_{ijs}$ is the disposition date/time, and $Roomed_{ijs}$ is rooming date/time of patient i, served by physician j during shift s. Disposition time is equivalent to the time of discharge decision for patients who are discharged, and time of bed request for patients who are admitted. $N^D_{pts,js}$ is the number of patients discharged or admitted *during* the shift. We exclude patients admitted or discharged after the shift when calculating average shift throughput time because after-shift patients are not completed by the original physician, and the behavior of another physician on another shift may impact their disposition time.

We use both measures of productivity since they complement each other. Considering only the number of patients served could be misleading because it does not capture *how long* it took to serve the patients. In other words, serving more patients does not automatically translate to higher patient flow if they are not discharged or admitted in a timely manner. Likewise, considering average shift throughput time, alone, may not necessarily indicate higher productivity since it does not capture the number of patients on whom the average is generated.

5.4. Control Variables

We employ five sets of controls in our analysis:

Date/time controls: We use date/time variables, including the hour of day, year, week of year, and day of week of admission, to control for trends and seasonal changes in demand. In addition, these controls absorb the variation in staffing levels. Specifically, although the number of workers varies from shift to shift, for any given shift (e.g. morning shift on a weekday), it does not vary. Hence, the date/time variables also serve as a proxy to control for staffing levels.

Ward Controls: First, we control for the actual ward to which the patient was admitted. This also accounts for any unobserved variables related to the site of some departments located in a second, discontinuous campus.

We also control for the possibility that the unit was capped at the time of admission. To do this we use the proxy of the ward's number of new patients (N_{6h}) in the 6 hours prior to the patient's bed request. Since our data does not include the residents' shifts, we are unable to precisely determine whether a team or unit was capped at the time of bed requests. The idea behind our proxy variable is that if a unit was capped, the average number of new patients to that unit should be lower. We include this variable to control for the possibility that knowledge of the unit being capped may influence a physician's timing of a bed request.

Finally, we control for the occupancy level of the destination unit at the time of a patient's bed request, denoted by Occ, which directly impacts boarding times. To find occupancy levels, similar to Berry Jaeker and Tucker (2016), we pool the capacity of each unit (i.e. medical, surgical, surgical ICU, medical ICU, and observation), and calculate its occupancy level by dividing the number of occupied beds at the time of bed request by the maximum number of occupied beds in that month.

Medical Controls: Certain medical conditions could increase the likelihood of batching and also be associated with longer waits. For example, physicians could prioritize the admission of urgent patients and delay admission of less severe patients, which could result in multiple low-severity patients being admitted at once. To address this, we include several control variables related to patients' medical conditions: ESI score, age, number of tests performed in the ED (N_{tst}) as a measure of diagnosis complexity, their Charlson comorbidity index (CI) (Charlson et al. 1994), the inpatient service they required, and whether they were admitted as an inpatient or observation patient. We also include the number of ICU visits (N_{ICU}) for the patient visit in the duration of their hospital visit as another proxy for how sick they were when arriving to the ED (Song et al. 2020).

Shift Controls: We use shift controls when evaluating hypotheses 1 and 3. The idea behind these variables is to control for idiosyncratic features of shift structures, physicians, and workload within shifts that may increase the probability of batching admissions, or impact physician productivity. For instance, batching is arguably more probable during the 7am – 3pm shift because of higher demand, or the number of patients served is higher when there are more patients roomed. For each physician-shift the controls include: length of shift in hours ($Hour$), hours since last shift (H_{last}), hours to next shift (H_{next}), the shift time (morning, afternoon or overnight), number of patients roomed in shift (N_{roomed}), waiting room census (WRC), and the admitting physician. In testing hypothesis 1, we also include the order of the patient within the patients seen on that shift. In testing hypothesis 3, we also include the numerical medical controls averaged at a shift-level, as an indicator of patient sickness for the shift; these include $\overline{N_{tst}}$, $\overline{ESI}$, $\overline{N_{ICU}}$, $\overline{CI}$, $\overline{AGE}$ and $\overline{WRC}$.

Patient Controls: These comprise of the patient's race (coded as Black or non-Black), gender and insurance information. Table EC.28 includes the summary statistics of patient controls, specifically for those who were admitted to medical units.

Table 1 provides summary statistics of the continuous patient-level variables and their correlations, for the sample of patients who were admitted to medical units. Similarly, table 2 provides the same information for shift-level variables, for the sample of physician-shifts which have at least two admissions.

Table 1 Summary statistics and correlation table for ED patients admitted to medical units

		μ	σ	1	2	3	4	5	6	7	8	9	10	11	12
1	$LPAD$	0.34	1.18	1.00											
2	BCH	0.18	0.39	-0.00	1.00										
3	TLS	1.15	3.70	0.07*	0.05*	1.00									
4	Occ	85.6	6.69	0.24*	-0.04*	-0.05*	1.00								
5	N_{6h}	1.25	1.42	-0.08*	0.01*	-0.03*	-0.02*	1.00							
6	Age	58.6	17.6	0.02*	0.02*	0.03*	-0.03*	-0.02*	1.00						
7	ESI	2.51	0.56	0.01*	-0.01*	-0.02*	-0.03*	0.04*	-0.07*	1.00					
8	N_{tst}	3.61	1.27	-0.00	0.01	-0.13*	0.03*	-0.03*	0.19*	-0.19*	1.00				
9	CI	3.05	2.95	0.03*	0.00	0.06*	-0.02*	-0.06*	0.34*	-0.04*	0.06*	1.00			
10	N_{ICU}	0.05	0.29	-0.02*	-0.00	-0.01	-0.02*	-0.02*	0.02*	-0.03*	0.05*	0.02*	1.00		
11	Black	0.68	0.47	-0.01	-0.00	-0.00	0.00	-0.01	0.01	-0.00	0.01*	0.09*	-0.02*	1.00	
12	Male	0.53	0.50	-0.00	0.00	0.03*	-0.00	-0.01*	-0.09*	-0.03*	-0.04*	-0.00	0.02*	-0.04*	1.00
13	WRC	104.8	75.5	-0.01*	0.02*	0.07*	-0.13*	0.12*	0.01	0.01*	-0.01	-0.00	0.00	-0.00	-0.01

Note: $^*p < 0.05$, $N = 25,565$

6. Econometric Models and Results

6.1. Batching Admissions Toward End of Shifts

6.1.1. Model We run the logistic regression in equation 1 on a patient-level to find the likelihood of a patient being admitted in a batch as a function of a physician's time left in her hourly shift. We only include admissions that occurred during the shift (observations

Table 2 Summary statistics and correlation table for shift-level variables

		μ	σ	1	2	3	4	5	6	7	8	9	10	11	12
1	N_{pts}	20.9	5.84	1.00											
2	STT^{a}	2.67	0.57	-0.06*	1.00										
3	SB	0.37	0.48	0.17*	0.12*	1.00									
4	$Hour$	7.91	0.53	0.20*	-0.07*	-0.04*	1.00								
5	N_{roomed}	112.7	38.5	-0.09*	-0.13*	-0.01	0.44*	1.00							
6	$\overline{AGE}$	49.3	5.74	-0.14*	0.20*	0.18*	-0.01	0.27*	1.00						
7	$\overline{N_{tst}}$	2.38	0.55	-0.23*	0.42*	0.27*	-0.20*	0.05*	0.45*	1.00					
8	$\overline{ESI}$	2.81	0.26	0.14*	-0.40*	-0.24*	0.24*	-0.00	-0.44*	-0.73*	1.00				
9	$\overline{CI}$	1.69	0.64	-0.10*	0.20*	0.16*	-0.07*	0.13*	0.53*	0.41*	-0.41*	1.00			
10	$\overline{N_{ICU}}$	0.07	0.10	-0.16*	-0.03*	0.12*	-0.08*	-0.04*	0.19*	0.39*	-0.37*	0.15*	1.00		
11	H_{last}	71.1	115.0	-0.00	0.00	0.00	0.00	0.08*	0.027*	0.02*	-0.01	0.02*	-0.02	1.00	
12	H_{next}	71.8	108.2	-0.02	-0.01	-0.00	-0.01	0.05*	0.01	0.02	-0.00	0.01	-0.01	0.17*	1.00
13	$\overline{WRC}$	99.27	72.24	-0.02*	0.01	0.05*	0.15*	0.35*	0.11*	0.07*	-0.04*	0.03*	0.01	0.01	-0.02

Note: $^*p < 0.05$, $N = 7,709$. [a]unit is hours

with $TLS > 0$) because as described in section 3, for admissions that occurred after the shift (i.e. $TLS < 0$) the physician either cosigned the order after their shift ended, or the patient was handed off and admitted by a second physician. However, for the sake of comparison, we run a separate model in which we include all observations. Hypothesis 1 holds if $\beta_1 < 0$ because it indicates that as we approach the end of a shift, the probability of batching increases.

$$ln(\frac{Prob(BCH)_i}{1 - Prob(BCH)_i}) = \beta_0 + \beta_1 TLS + \boldsymbol{\gamma X_i} \quad (1)$$

The control vector $\boldsymbol{X}$ includes shift, date/time, ward, medical and patient controls. Importantly, recall that the shift controls include the admitting physician fixed effects to adjust the estimates for unobserved person-specific practice features of the physicians, and number of patients roomed in shift as a proxy for demand because the likelihood of batching admissions could be higher in the busier hours of the ED. We also cluster the standard errors by physician to account for the possibility that error terms could be correlated within each physician.

6.1.2. Results Results for the logit model are presented in Table 3. (The complete results of the models are provided in table EC.30 of the electronic companion.)

The coefficient of TLS in model (a) is $\beta_1 = -0.1543$ ($p < 0.01$), and increases in magnitude to $\beta_1 = -0.2060$ ($p < 0.01$) when we add all controls. The negative coefficient provides support for hypothesis 1: the likelihood of batching admissions increases by an average of $e^{0.2060} - 1 = 22.9\%$ per hour closer to the end of a shift. Note that individual batching tendencies are accounted for in this analysis since we control for the admitting physician.

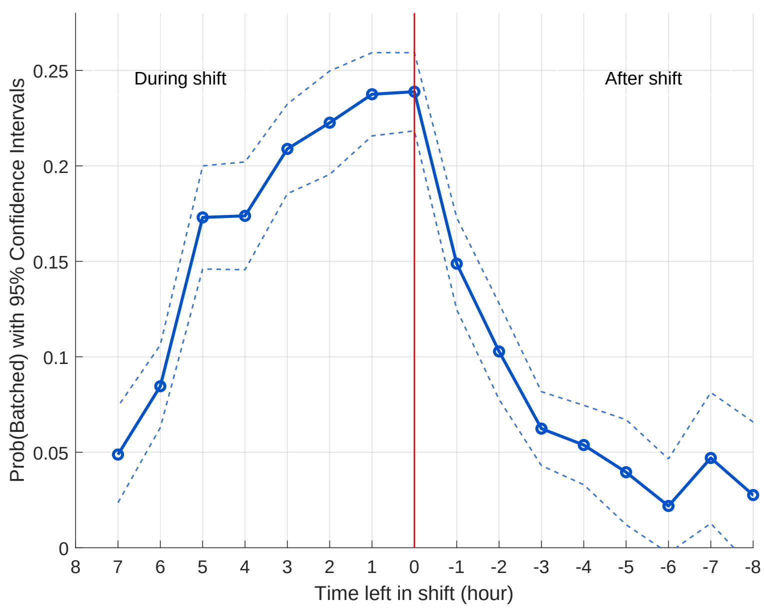
Table 3 Hypothesis 1: Likelihood of batching admissions towards end of shifts

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(a)	(b)	(c)
TLS	-0.1543**	-0.2060**	0.0889**
	(0.0093)	(0.0183)	(0.0104)
Controls	NO	All controls	All controls
N	31,563	30,466	42,034
R^2	0.0146	0.0423	0.0374
LL	-15750.0	-14888.5	-19272.6

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Standard errors clustered by physician. Drop in N in model (b) is due to missing values in controls. Model (c) includes all observations (both $TLS < 0$ and $TLS \geq 0$).

When we include during- and after-shift admissions (all observations) in model (c), the coefficient on TLS becomes positive $\beta_1 = 0.0889$ ($p < 0.01$), which *disproves* the hypothesis. To explore the large difference between the results of models (b) and (c), we examine the hour-by-hour behavior of ED physicians. This also relaxes the assumed linear relationship between the likelihood of batching and time left in shift in the earlier OLS model. We replace the continuous TLS with the discrete $HoursLeft = \lfloor TLS \rfloor$, and rerun model (c) to find the marginal effect of each hour left in a shift on the probability of batching, for up to 8 hours after the shift has ended. The result is plotted in Fig. 5. It is clear that

Figure 5 Hourly probability of batching admissions



Note. The x-axis is reversed to aid readability.

the probability of batching in the earliest hour of the shift ($TLS = 7$) is nearly 5%, and increases to about 24% in the final two hours of the shift. (Recall from table 1 that the base probability of a patient being admitted in a batch is 18%.) This shows that toward the end of their shifts, physicians are about 4.8× more likely to batch admissions compared to the start, and $\sim 1.3\times$ more likely compared to the baseline probability.

One explanation for this result could be the possibility of unequal shift times. For example, in a hypothetical situation where physicians had either 4-hour or 8-hour shifts, even if patients were uniformly admitted throughout shifts, there would still be more admissions in the $0-4$ hour interval, simply because all shifts contain this interval, but not a $4-8$ hour interval. However, in our data this is rarely the case; 96.0% of physician work shifts are 7 or 8 hours long, so unequal shift lengths do not explain our finding.

In summary, regardless of whether TLS is defined as a continuous or discrete hourly variable, we find that the probability of batching is increasing toward end of shifts, which supports hypothesis 1.

6.2. Impact of Batching Admissions on Pre-allocation Delay

6.2.1. Model To test the impact of batching admissions on pre-allocation delay, we run the ordinary least squares (OLS) model in equation 2 on a patient-level:

$$LPAD_i = \beta_0 + \beta_1 BCH_i + \boldsymbol{\gamma} \boldsymbol{X_i} + \epsilon_i \tag{2}$$

where, $\boldsymbol{X}$ is a control vector that includes date/time, ward, medical, and patient controls. We also control for the admitting physician. Hypothesis 2 is supported if $\beta_1 > 0$.

Arguably, there exists an endogeneity concern due to selection into treatment. It is possible that less sick patients are selected to be batched, and due to their less severe condition, their pre-allocation delay is longer. Therefore, to ensure that β_1 has a causal interpretation, we first check the covariate balance between the batched and unbatched patients. We find that the normative difference of all covariates between the treated and untreated samples are under 0.1; typically, a difference below the threshold of 0.1 or 0.2 indicates that the two samples are statistically identical (Batt et al. 2019). Hence, there is no substantial evidence that suggests that patients who were batched were not chosen at random. Table EC.29 in the electronic companion shows the complete covariate balance.

6.2.2. Results Table 4 shows the results for the model presented above, before and after adding the controls. In model (b) which includes controls, $\beta_1 = 0.0455$, $(p < 0.01)$ which means that batched patients experience an average of 4.7% increase in pre-allocation delay, equivalent to 7 minutes. In terms of magnitude, this is equivalent to the pre-allocation delay increase caused by a 1.2% increase in hospital occupancy $(0.0455/0.0376 = 1.2)$. Complete regression results are presented in table EC.31 of the electronic companion.

Table 4 Hypothesis 2: Impact of batching admissions on pre-allocation delay

Dep. Variable:	ln(Pre-allocation Delay)	
	(a)	(b)
BCH	-0.008	0.0455^{**}
	(0.0188)	(0.0168)
Occ		0.0376^{**}
		(0.0016)
Controls	NO	YES
N	25,565	25,565
R^2	0.0000	0.2172
F	0.18	45.40

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

We show the robustness of our results to (i) including admissions to all units in our sample (and not only medical units), and (ii) using boarding time, and post-allocation delay as dependent variables in section EC.1.1.

6.2.3. Mechanism As one possible mechanism, we test whether the coefficient of variation (CV) of bed requests, defined as standard deviation/mean of inter-bed-request times, mediates the relationship between batching admissions and pre-allocation delay. We posit that a spike in inpatient bed demand caused by batching admissions on a physician-level distorts the bed request distribution on a system-level, and injects variance in the overall inter-bed-request time distribution observed by the bed management unit. Consequently, according to queuing theory, an increase in interarrival CV increases the average wait times (Whitt 1993).

An empirical challenge is approximating the inter-bed-request CV at a patient-level. First, we sort the data by time of bed request, and calculate all inter-bed-request times.

Next, we Winsorize the top 1% of inter-bed-request times by replacing the inter-bed-request time of those observations with the average value. This prevents long time gaps in admissions from heavily impacting our CV calculation, and is shown to produce a more robust estimate of sample variance (Wilcox 2011)[5]. Then for each focal patient i who is assumed to be admitted at time t_0, we generate the set of inter-bed-request times, defined by $T_i = \{\Delta t_n = t_n - t_{n-1} : -4 \leq n \leq 0\}$, on which we calculate CV. In summary, we calculate each patient's bed request CV on the set of 5 most recent admissions, which includes the focal patient. We select a group of 5 observations because over 95% of the batches in our data have a size less than 5, and to capture the effect of batching on CV, we must include at least one non-batched observation in the group. Also, the disadvantage of *increasing* the group size is that the time span over which CV is calculated is increasing. When 5 observations are used, the median time span is 1.2 hours. Also, about 95% of the time spans fall below 3 hours, which is reasonable for estimation since the longest time spans occur in early morning hour when the hospital is not busy. The advantage of this method in calculating CV is twofold. First, this method ensures that the focal patient is always the last person in their respective set in the time domain. For instance, if observations are aggregated by hour (or any other fixed unit of time as in Shi et al. (2015), Ahuja et al. (2021)), the CV experienced by each patient will most likely depend on *future* bed requests in the same hour. Second, each patient's CV is generated from an equal number of observations. Therefore, any numerical biases in CV calculation due to the number of observations in the sample is consistent across all observations, which is essential for identification. Specifically, consistent with observations in Shi et al. (2015), the numerical calculation of CV is a function of the number of observations used in the sample, even if the underlying sample is independent and identically distributed (IID). Also, for a positively-skewed distribution – such as inter-bed-request times – the sample CV is negatively biased, and the magnitude of the bias is non-linearly decreasing in sample size (Breunig 2001). In our empirical analysis, if each patient's CV is *more biased* when the ED is not busy (i.e. fewer observations) and *less biased* when the ED is busy (i.e. more observations), a positive correlation between CV and batching may emerge due to the fact that both admission batching and the numerical error in calculating CV are correlated with how busy the ED is. A higher CV will also most likely be associated with longer pre-allocation delay (again, because the ED is busy). Hence, if numerical bias in CV calculation is inconsistent across the observations we may detect (non-existent) effects caused by the way CV is measured. Noteworthy is that it is difficult to mitigate this problem by simply controlling for the number of observations in our analysis due to the non-linear relationship between sample CV and sample size.

Summary statistics for CV are given in Table 5.

Table 5 Summary statistics of CV

	μ	σ	min	p10	p25	p50	p75	p90	max
CV	0.870	0.289	0.094	0.523	0.668	0.845	1.045	1.254	2.029

Note. $N = 25,565$. Slight drop in number of observations is because we remove patients with missing controls.

To test the mediation effect of CV, we follow the product-of-coefficients method proposed in Preacher and Hayes (2008) and used by others (Dietvorst et al. 2018, Boulding et al. 2017, Kim et al. 2019). Having established the total effect of batching on pre-allocation delay *without* the mediator (CV) in the OLS model (table 4), we now estimate the regressions in equations 3 and 4 on a patient level, and allow for the error terms to co-vary (seemingly unrelated regression). The control vector $\boldsymbol{X}$ is identical to that used in equation 2.

$$CV = \alpha_0 + aBCH \qquad\qquad + \alpha_2 OCC_p + \boldsymbol{\gamma X_i} + \mu_i \tag{3}$$

$$LPAD = \beta_0 + c'BCH + bCV + \beta_2 OCC_p + \boldsymbol{\theta X_i} + \nu_i \tag{4}$$

The indirect effect of BCH on $LPAD$ (mediation) is captured by $a \times b$: the impact of batching on CV in equation 3 times the impact of CV on $LPAD$ from equation 4. The direct effect is c', and the total effect is $c = c' + a \times b$. We bootstrap this procedure 5000 times, as recommended (Preacher and Hayes 2008). We find that $a = 0.0713$ ($p < 0.01$), $b = 0.0555$ ($p < 0.05$), $a \times b = 0.0040$, ($p < 0.05$), and $c' = 0.0416$, ($p < 0.05$). The mediation level of CV is therefore derived from $\frac{a \times b}{a \times b + c'} = 8.8\%$. Altogether, this analysis shows that patients admitted in a batch suffered an average of 4.7% longer pre-allocation delay (7 minutes), where 8.8% is because of the increase in CV. Complete regression results are presented in table EC.32. Recall that an admission batch is determined at a physician-level, whereas CV is determined at a system-level. Hence, the significant impact of batching on CV, given by a, is an indication of individual behaviors impacting system-level performance in aggregate.

Noteworthy is that measurement errors in the mediator variable, caused by estimating CV using a small sample size will only attenuate the estimated indirect effect and inflate its standard errors (Fritz et al. 2016, Wooldridge 2015). Hence, the above analysis serves as a conservative estimate of the mediating role of CV in the impact of batching on pre-allocation delay. That said, our results remain fairly robust when we estimate patient-level CV using a range of sample sizes ($2 \leq n \leq 11$) as we show in section EC.1.3 of the electronic companion.

6.3. Impact of Batching Admissions on Physician Productivity

6.3.1. Model We are interested in testing the impact of batching admissions on physician productivity (hypothesis 3). Recall from section 5.1 that for this purpose, the treatment in which we are interested is a binary variable, "shift with batching" (SB), which is defined on a physician-shift level, and is equal to one if at least one batched admission occurs in the shift. We run the model below, where the unit of analysis is a physician-shift.

$$Y_{js} = \beta_0 + \beta_1 SB_{js} + \boldsymbol{\theta X_j} + \epsilon_{js} \tag{5}$$

In this model, Y is either average shift throughput time (i.e. average time to making a disposition decision) or number of patients seen in shift. The control vector $\boldsymbol{X}$ represents the shift controls as described in section 5.1, and ϵ is the error term, clustered by physician. Hypotheses 3A and 3B are supported if $\beta_1 > 0$ and $\beta_1 < 0$, respectively.

An endogeneity concern due to selection into treatment exists the model described in equation 5. Specifically, shifts with batching admissions may be busier shifts, where admissions naturally occur at a higher rate; also, shifts which have less sick patients are less likely to have admissions, and therefore batching; but at the same time, throughput rates may be higher since the decision to discharge is easier and faster. Therefore, we must first match shifts with and without batching with respect to the shift workload. We do so using coarsened element matching (CEM).

Unlike propensity score matching, CEM does not require a selection model to determine, in our case, the probability that a physician will batch admissions in their shift, which may be theoretically difficult to set up and specify correctly (Blackwell et al. 2009). The CEM method stratifies the set of desired covariates into bins, similar to a histogram. Next, it drops the strata that do not contain at least one observation with one treated and one untreated observation. Finally, the observations are weighted based on the proportion of treated observations in their stratum (Imbens and Wooldridge 2009). We run the specification in equation 5 using a weighted OLS on CEM-matched observations (i.e. the doubly robust estimator Robins and Ritov 1997).

6.3.2. Results Table 6 provides a summary of the covariate balance before and after CEM matching. Prior to matching, note that the standardized difference in a few of the covariates is greater than 0.2, indicating imbalance in the original data, which may bias our estimates. However, the matched sample is well-balanced in all covariates, suggesting that we have successfully generated a sample in which the probability that a shift is batched can be considered random. Estimating our model on this sample reduces the possibility that our results are driven by selection into treatment rather than differences between shifts with and without batching.

Table 6 Covariate balance in full and matched samples, used to test hypothesis 3

Sample	Full sample No weights			CEM-matched sample CEM-weighted		
Variable	SB=0	SB=1	Norm. Diff.	SB=0	SB=1	Norm. Diff.
Mean age	48.65	50.38	0.22	49.57	50.36	0.02
Shift length	7.91	7.94	0.04	7.95	7.95	0.00
Mean number of ED tests	2.29	2.53	0.31	2.51	2.51	0.00
Number of patients roomed	111.7	114.5	0.05	114.44	114.62	0.00
Mean ESI	2.85	2.75	0.26	2.77	2.75	0.01
Mean Charlson Score	1.63	1.79	0.17	1.71	1.79	0.05
Mean number of ICU visits	0.06	0.08	0.13	0.07	0.08	0.03
hours since last shift	70.75	71.54	0.00	63.59	65.74	0.02
hours to next shift	71.59	71.81	0.00	65.74	65.51	0.00
Morning Shift	0.43	0.36	0.10	0.35	0.35	0.00
Afternoon Shift	0.35	0.48	0.18	0.48	0.48	0.00
Overnight Shift	0.22	0.17	0.10	0.17	0.17	0.00
Weekend shift	0.22	0.25	0.04	0.21	0.25	0.07
Waiting room census	95.43	106.05	0.10	106.32	106.34	0.00
N	4723	2844		4331	2729	

Results for the specification given in equation 5 are provided in table 7. (See table EC.33 of electronic companion for complete results.) Results that assess the impact of shift with batching on the number of patients served in shift and throughput time are provided in models (a)-(b), and models (c)-(d), respectively. Models (a) and (c) show baseline OLS results with no matching for comparison, and models (b) and (d) show the results after CEM-matching. According to model (b), an average of $\beta_1 = 2.1$ ($p < 0.01$) more patients were served in shifts where the physician batched their admissions; this is $2.1/20.9 = 10.0\%$

higher than shifts with no batching. This provides support for hypothesis 3A. Likewise, according to model (d), the coefficient $\beta_1 = -0.0431$ ($p < 0.01$) indicates that in shifts with admissions batching the average throughput time of the physician is lower by 2.6 minutes[6]. This lends support for hypothesis 3B. Note that the coefficients in the OLS models are not very different than their CEM-matched counterparts. That said, the coefficient decrease in model (b), and magnitude increase in model (d) demonstrates that the decision to batch admissions could have been, in part, driven by patient demand or the sickness of patients served.

Table 7 Hypothesis 3: Impact of batching admissions on physician productivity

Dep. Variable:	Number of Patients Seen		Average Throughput Time	
	(a) unweighted OLS	(b) weighted CEM-matched	(c) unweighted OLS	(d) weighted CEM-matched
SB	2.2957** (0.1092)	2.1340** (0.1138)	-0.0264^ (0.0138)	-0.0431** (0.0143)
Shift Length	1.8243** (0.1824)	2.0065** (0.2059)	0.2119** (0.0149)	0.2229** (0.0195)
Mean Num. of ED tests	-1.8324** (0.1472)	-1.9750** (0.1730)	0.3029** (0.0164)	0.3030** (0.0202)
N	7567	7060	7567	7060
R^2	0.3369	0.3422	0.2328	0.2005

Note. $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. 90 out of 7,709 physician shifts had none of the admissions occur during the shift, and were therefore removed from the sample. An additional 52 physician shifts were missing either hours to next shift or hours since last shift, and hence were eliminated from the sample. Controls included in all models.

In section EC.1.2, we show the robustness of our results to using the actual number of batched admissions during the shift instead of a binary measure (SB) as the independent variable, and using shift-level time to discharge and time to admit as dependent variables.

7. Post-hoc and Counterfactual Analyses

7.1. Connection to Theory of Queues with Batched Arrivals

In this section, we relate our empirical findings on the impact of batching admissions on pre-allocation delay with the theoretical literature on queues with batched arrivals (Hanschke 2006, Yao 1985, Pang and Whitt 2012). Most relevant to our work is Yao (1985) that provides a closed-form approximation for the expected wait time in a queue with batched arrivals. We use this approximation as a baseline to estimate: 1) what would happen in our system if batching did not occur; 2) the extent to which our empirical results align with the predictions of queueing models.

In a queue with batched arrivals, the average wait time (W_q) in steady state is calculated from Little's Law, as follows:

$$W_q = \frac{L_q}{\lambda_B m_g} \tag{6}$$

where λ_B is the arrival rate of batches, m_g is the average batch size, and L_q is the queue length. Yao (1985, Expression 13) provides a two-moment approximation for L_q in a $GI^X/G/c$ queue in equilibrium. This expression can be simplified for a $M^X/M/c$ case where $c > 1$ as:

$$L_q \approx \left[\frac{\rho^{\sqrt{2c+1}}}{1-\rho}\right]\left[\frac{m_g(1+c_g^2)+1}{2}\right] + \left[\frac{m_g(1+c_g^2)-1}{2}\right]\rho^{\sqrt{0.5(c+1)}} \tag{7}$$

where c_g^2 is the squared CV (SCV) of the batch size distribution, and ρ is the queue traffic intensity, defined by

$$\rho = \frac{\lambda_B m_g}{c\mu}$$

where μ is the mean service rate.

We are able to estimate all parameters in equation 7 with the exception of number of servers, c, as shown in table 8. The challenge with estimating c is that it is not clear how many inpatient beds are assigned to ED patients at any given time. Namely, the medical units see patients coming in from both the ED and transfers from other units. Moreover, recall that a unit may be capped and the inpatient team cannot accept any new incoming patients (see section 3). Due to these challenges, we conduct our analysis for a *range* of c.

Table 8 Empirical estimate of parameters in equation 7

m_g	c_g^2	$1/\lambda_B$ (min)	$1/\mu$ (min)	W_q (min)
1.11	0.40^2	47.5	3594.0	152.1

To estimate pre-allocation delay in a system *without* batching, we set the arrival rate to $\lambda = m_g\lambda_B$ for a fair comparison. This implies that the total traffic for systems with and without batched arrivals are similar. Fig. 6a illustrates the results for this comparison as a function of number of beds, c. The horizontal line represents the baseline average pre-allocation delay in our setting ($W_q = 152.1$ minutes). We observe that our system's pre-allocation delay corresponds to a bed capacity of ≈ 92, which is about 50% of the average monthly bed occupancy we see in our data. At $c = 92$, the average wait time in a system without batched arrivals is $\sim 15\%$ lower than that of a system with batched arrivals, such as the ED in our setting. This implies that if batching admissions were eliminated, it would have a significant impact on the overall system-level boarding times.

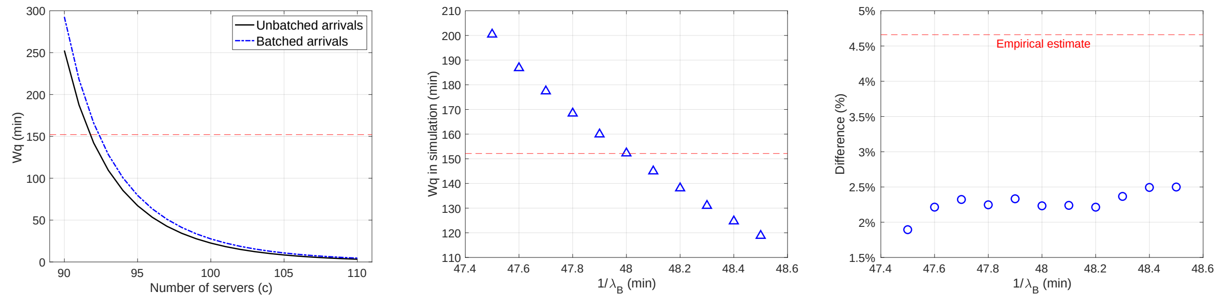
Next, we compare our empirical estimates to predictions from queuing models with batched arrivals. For this, we run a discrete-event simulation of a $M^x/M/c$ queue to compare the wait times of arrivals in a batch with those who are not batched. We derive the probability mass function of batch size distribution from our data, and generate batches of size g which arrive with rate λ_B as inputs to the simulation. Also, to mimic the ED, we set $c = 92$, and $1/\mu = 3594.0$. The simulation output is the mean wait time difference between inputs of size $g > 1$ (i.e. batched arrivals) and those of size $g = 1$ (i.e. unbatched arrivals).

We run the simulation for a time equivalent of > 20 years (1.2×10^7 minutes), and discard the first ~ 5 years (i.e. 2.6×10^6 minutes) of data to ensure that our results are taken from when the queue is operating in a steady state. We run the simulation for a range of $1/\lambda_B$ (i.e. $1/\lambda_B$) to account for the fact that our estimate of c may not be entirely accurate, and we may further tune the simulation by slightly adjusting the batch arrival rate. Fig. 6b shows that at $1/\lambda_B = 48.0$ the simulated W_q is very close to our data.

The simulated wait time difference between batched and unbatched arrivals are depicted in Fig. 6c. The horizontal line at $\delta = 4.6\%$ shows the baseline empirical estimate from section 6.2. We observe that our empirical estimate of the impact of batching on pre-allocation delay is slightly greater than the prediction of a queueing model. This is most likely due to the many factors involved in the data generation process in the ED which are not captured in a time-invariant $M^x/M/c$ model. First, note that we simulate the medical units, but there are other units (e.g. ICUs) in the hospital, and bed requests to those units may be prioritized over medical units. Second, our simulation does not consider the important day-of-week and hourly effects, which are correlated with admission rates, staffing levels, and hospital occupancy, all of which impact pre-allocation delay.

Overall, our counterfactual simulation and comparison with models of queues with batched arrivals highlight two points. First, there is significant potential for reducing pre-allocation delay if batching on an individual-level is eliminated. Second, our empirical estimate of the impact of batching on pre-allocation delay is greater than the prediction of a queueing model; albeit, they are in the same order. An important advantage of our

Figure 6 Counterfactual analysis results



(a) ***Yao (1985)*** W_q ***estimate*** **(b)** ***Simulated*** W_q ***as function of*** $1/\lambda_B$ **(c)** ***Simulated wait time difference between batched and unbatched***

discrete-event simulation is that we are also able to model small time gaps between the arrival times of the patients who make up a batch (Pang and Whitt 2012) to better mimic the ED admission process; that said, we do not find any significant differences in the results.

7.2. Post-hoc Analysis

Below we provide an overall summary of our two post-hoc analyses, but refer readers to section EC.2 of the electronic companion for additional details.

7.2.1. Net effect of batching on ED length of stay: First we conduct a patient-level and shift-level analysis to assess the overall impact of batching on ED LOS, defined as the time from when a patient is roomed, until the time they depart the ED (which includes boarding time for admitted patients).

First we explore the net effect of batching on ED LOS on a patient level, which includes only a sample of admitted patients. We find that the net effect of batching admissions on ED LOS is positive and negative for patients admitted during and after the shift, respectively. Specifically, patients admitted in a batch during a shift suffer both a longer time to admission, and a longer boarding time, which adds to total ED LOS. However, although patients admitted after shift by a second physician (or through a cosigning procedure) experienced an overall longer ED length of stay, those who appeared in a batch were less affected due to an expedited admission.

To complement the patient-level analysis, we also investigate the impact of batching on the average LOS of all patients treated in a shift. In other words, given that batching admissions in a shift boosts productivity at a cost of longer boarding, we study which effect is dominating. We find that in shifts with batching the overall LOS of patients is higher than shifts without batching.

In summary, our analysis shows that the system-level cost of batching admissions is greater than the individual-level productivity gain. This suggests that batching admissions is detrimental to ED overcrowding, since it increases the average ED LOS.

7.2.2. Batching mechanism: Or second post-hoc analysis explores how batches form by focusing on the time it takes for the physician to admit a patient. We find that batched patients experienced a longer time to admission compared to unbatched patients. Our physician partner suggests that the delay may be due to postponing tasks at various points in the shift. Specifically, many times physicians order tests, but postpone the task of reviewing them in order to attend to other patients who have not yet been diagnosed. This behavior may result in delaying the admission of some patients to attend to a higher number of patients.

8. Discussion and Conclusion

We study the behavior of batching admissions by ED physicians, and the trade-off it causes between productivity and boarding times. We focus on this behavior because not only is it a known - yet understudied - phenomenon (Vose et al. 2014, Etzler 2019), but also, our data allows us to measure its impact through the time patients spend in a queue to receive an inpatient bed (i.e. pre-allocation delay) and the shift-level performance of the physicians. We define two or more patients to be admitted in a batch if they are admitted within 9.1 minutes of each other, a threshold derived empirically and in-line with practice. First, we show that the probability of batching admissions increases toward the end of physician shifts. We further show that physicians who batch their admissions may benefit from higher productivity, measured by the number of patients seen in a shift, and their average time to disposition decision. On the other hand, we also find that patients who are admitted in a batch experience a longer pre-allocation delay, in part because of the increase in bed request CV. Specifically, we show that when physicians batch their admissions at an individual level, it increases the bed request CV observed at a system level, especially since most shifts are overlapping. Finally, our counterfactual analysis suggests that by eliminating batching, pre-allocation delay may be reduced by a theoretical maximum of 15%. In summary, our work shows that ED physicians may batch their admissions in order to be more productive; however, this comes at a cost of longer boarding times for the batched patients.

8.1. Implications for Theory

The causes and consequences of batching behavior by workers is a relatively understudied area. By examining when batching occurs, and the trade-off it causes between wait time and worker productivity, our paper makes several contributions to the literature.

Worker behavior and productivity: First, we add to the literature on worker behavior by showing that physicians are likely to batch their admissions toward the end of their shifts. Although we do not rigorously study its mechanism(s), our post-hoc analysis suggests that batching may be caused in part by postponing admissions to attend to other patients, or as suggested in Chan (2018), a byproduct of speeding up toward end of shifts. However, we do not find a change in the actual probability of a patient *being* admitted as a function of hour in shift (results not shown for brevity), which is aligned with the findings of Batt et al. (2019). In addition, Ibanez et al. (2017) shows the detrimental effect of server batching on average task completion time when workers actively search to find similar tasks to complete together. In contrast, we find that batching admissions may actually be beneficial for physicians' shift-level productivity, albeit at the cost of longer boarding times for the batched patients. That said, as Dobson et al. (2012) analytically shows, there may be an upper bound to how much batching is beneficial for improving throughput times. For example, if physicians in our study ED batch too much by postponing too many admissions, then we could observe a drop in productivity as a result of ED overcrowding and too much multi-tasking (KC 2014).

Behavioral queueing: Second, we contribute to the literature on the behavior of servers in a queuing system (Delasay et al. 2019, 2016, Schultz et al. 1998, 2003). Interestingly, Pang and Whitt (2012) calls for an empirical study of queueing systems with batched arrivals. Our work empirically demonstrates that servers in the first stage of a two-tier queuing system may batch their referrals to the second stage, which, in turn, systematically increases wait times. Relatedly, prior studies of the causes of hospital boarding focus on capacity-related causes, which may be costly to increase; or process-driven causes, which may be difficult to change. In contrast, we illustrate that the behavioral modification of reducing admission batching is an effective lever for reducing boarding times. We highlight that batching increases pre-allocation delay, partly because it increases arrival CV of the second stage. This is important since most models of hospital operations assume time-homogeneous or -non-homogeneous Poisson arrivals for tractability, and do not capture the impact of CV variations due to server behavior, or the effect(s) of batching.

Queuing models with batched arrivals: Finally, the batching behavior of physicians allows us to study queues with batched arrivals, empirically. To our knowledge, most papers in this area are analytical, and their results have not been tested in an empirical setting. In our counterfactual analysis, we compare our empirical results to the approximation presented in Yao (1985) and show that (i) our findings are in the same order as the prediction of queueing models; and (ii) by eliminating batching, pre-allocation delay may be reduced significantly.

8.2. Implications for Practice

Emergency department (ED) overcrowding is a worldwide problem, and its solution boils down to improving patient flow in the hospital. (See Morley et al. (2018) for a thorough review.) As such, our study offers important implications for hospital managers and ED practitioners.

First, we show that common behaviors such as batching admissions in the ED may cause unintended inefficiencies such as longer boarding times, the key cause of ED overcrowding (ACEP 2018). To our knowledge, prior studies have only focused on the capacity-related causes of ED boarding, and have not investigated the behaviors that may exacerbate boarding times. While adding inpatient bed capacity is costly, managers may improve the ED flow by employing levers to encourage staff to change their behavior. That said, we also show that batching admissions is associated with a decrease in average throughput time, and an increase in the number of patients served, at a shift-level. This suggests that the impact of batching admissions on ED overcrowding is ultimately a trade-off between increased boarding times for batched patients and improved physician productivity. Interestingly, our post-hoc analysis shows that the net effect of batching on ED LOS is positive (i.e. longer LOS), which implies that overall, batching is detrimental to ED overcrowding in our setting.

Second, we quantify the impact of batching admissions on pre-allocation delay. The impact of batching admissions on pre-allocation delay is equivalent to the increase caused by an occupancy increase of 1.2% (see Table 4). Considering that our study setting has an average monthly maximum occupancy of 188.2 medical beds, the waiting cost of batching admissions is equivalent to having an average of 2.3 fewer beds. Noteworthy is that redoing our analysis with boarding time – the addition of pre- and post-allocation delay – as the dependent variable in our analysis yields similar results. (This is shown in the robustness section of the electronic companion.) Interestingly, Bair et al. (2010) shows that adding two additional inpatient beds decreases the proportion of overcrowded days by 6.6%, and retrieves one additional left-without-being-seen (LWOBS) patient per day using a discrete-event simulation model on a hospital similar in size to our study setting. Hence, reducing batching may therefore help our study hospital achieve its goal of reducing its score on the national ED crowding scale (NEDOCS) and the rate of LWOBS patients.

Another implication of our work is that workers may inadvertently cause delays in downstream stages due to behaviors that seemingly increase their productivity. This is important from a performance assessment perspective. Many organizations use a "relative performance feedback" with the goal of improving individual worker productivity by means of generating competition and sharing best practices among workers (Song et al. 2018). Incorporating metrics such as boarding time into this feedback could create incentives for workers to become more consciously aware of the downstream consequences of their practice styles.

8.3. Limitations and Future Work

This paper has several limitations that may be remedied by future work. First, like many empirical studies, our data is limited to one setting. Other papers may extend or replicate our findings in other industries or organizations. Second, our data does not allow us to concretely tease out the mechanisms of batching admissions. Admissions batching could be in part the result of other upstream care processes such as lab, radiology, consulting, or even nurse batching. For example, Shi et al. (2014) observes that physicians make their rounds in the ICU every morning, after which a batch of bed requests are submitted to transfer some patients to the general ward units. Although our post-hoc analysis hints that postponing certain tasks at various points in the shift may be a reason batches form, future research could study batch formation in greater detail, which would lead to a better understanding of the productivity/wait time trade-off associated with batching. Third, other studies may extend this paper by exploring other behaviors or biases that systematically add variance to either processing times or interarrival times in a multi-stage queue. Fourth, our counterfactual analysis does not consider time-varying arrival rates. Specifically, in many aspects our ED setting is similar in context to the models presented in Daw and Pender (2019) and Pang and Whitt (2012). Both of these papers model a time-varying infinite-server queue with batched arrivals. Although infinite-server queuing models are helpful in healthcare settings to provide insights on capacity planning and understanding sources of variations (Worthington et al. 2020), but we cannot infer waiting times from such models as by design, queue length is zero with infinite servers. Future queueing models may provide approximations for the dynamics of time-varying queues with batched arrivals, and finite servers. Finally, our work empirically demonstrates that batching increases CV on a system-level. Future modelling papers may incorporate changes in CV as a result of worker behavior in stochastic models of inpatient operations, capacity-planning, or even, in determining optimal shift schedules (e.g. Bhulai et al. 2008).

8.4. Conclusion

Batching behavior by discretionary workers is an understudied area. Using data from an ED, a two-tier queuing system, we show that batching by workers in the first stage might increase wait times for the second stage. On the other hand, we find that batching may be beneficial for workers' shift-level productivity. Therefore, the wait-time/productivity trade-off in batching exists, and must be considered. In our setting, ultimately, we find that the net effect of batching on overall shift-level ED LOS is positive (i.e. longer LOS), and also, batched patients suffer from an increased boarding time, which, in turn, suggests a lower quality of care.

Endnotes

1. Our data shows that this rule is violated $\sim 18\%$ of the time. One possible reason could be related to mental heuristics of decision makers when routing patients in the ED (e.g. Ding et al. 2019). Since physicians sign up for patients only after they are roomed, and do not interact with

patients prior to this, we do not believe that these violations drive any part of our results. We show the robustness of our results to removing these patients in section EC.1.6

2. In our data, a basic decision tree model shows that just by looking at age, comorbidities, acuity level and the specific tests performed (e.g. ultra-sound, x-ray, etc.) one is able to predict which patient will be admitted with $\sim 85\%$ and $\sim 80\%$ accuracy in a balanced training and test data set, respectively.

3. Either the "test order time" or "test result received" timestamps are missing from most of the data.

4. We find no evidence that points to any selection bias in the observations that are missing timestamps throughout the data cleaning process.

5. Our main result holds even we do not Winsorize inter-bed-request time.

6. Note that we did not log-transform average throughput time because its distribution was not heavily skewed due to the central limit theorem

References

ACEP (2018) Definition of Boarded Patient. `https://www.acep.org/patient-care/policy-statements/definition-of-boarded-patient/`, online; accessed 29 January 2014.

ACGME (2019) ACGME Program Requirements for Graduate Medical Education in Internal Medicine. `https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/140_InternalMedicine_2019.pdf`.

Ahuja V, Alvarez C, Staats BR (2021) An operations approach for reducing glycemic variability: Evidence from a large primary care setting. *SMU Cox School of Business Research Paper* (19-13).

Armony M, Israelit S, Mandelbaum A, Marmor YN, Tseytlin Y, Yom-Tov GB (2015) On patient flow in hospitals: A data-based queueing-science perspective. *Stochastic Systems* 5(1):146–194.

Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan WJ, Sinsky CA, Gilchrist VJ (2017) Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *The Annals of Family Medicine* 15(5):419–426.

Bair AE, Song WT, Chen YC, Morris BA (2010) The impact of inpatient boarding on ed efficiency: a discrete-event simulation study. *Journal of Medical Systems* 34(5):919–929.

Bar-Lev SK, Boxma O, Perry D, Vastazos LP (2017) Analysis and optimization of blood-testing procedures. *Probability in the Engineering and Informational Sciences* 31(3):330.

Batt RJ, KC DS, Staats BR, Patterson BW (2019) The effects of discrete work shifts on a nonterminating service system. *Production and Operations Management* 28(6):1528–1544.

Batt RJ, Terwiesch C (2016) Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Science* 63(11):3531–3551.

Berry Jaeker JA, Tucker AL (2016) Past the point of speeding up: The negative effects of workload saturation on efficiency and patient severity. *Management Science* 63(4):1042–1062.

Best TJ, Sandıkçı B, Eisenstein DD, Meltzer DO (2015) Managing hospital inpatient bed capacity through partitioning care into focused wings. *Manufacturing & Service Operations Management* 17(2):157–176.

Bhulai S, Koole G, Pot A (2008) Simple methods for shift scheduling in multiskill call centers. *Manufacturing & Service Operations Management* 10(3):411–420.

Bishop CM (2006) *Pattern recognition and machine learning* (springer).

Blackwell M, Iacus S, King G, Porro G (2009) cem: Coarsened exact matching in stata. *The Stata Journal* 9(4):524–546.

Boulding W, Guha A, Staelin R (2017) Do we really need to change the decision maker? counterintuitive escalation of commitment results in real options contexts. *Management Science* 63(10):3459–3472.

Breunig R (2001) An almost unbiased estimator of the coefficient of variation. *Economics Letters* 70(1):15–19.

Burke P (1975) Delays in single-server queues with batch input. *Operations Research* 23(4):830–833.

Chan CW, Dong J, Green LV (2017) Queues with time-varying arrivals and inspections with applications to hospital discharge policies. *Operations Research* 65(2):469–495.

Chan DC (2018) The efficiency of slacking off: Evidence from the emergency department. *Econometrica* 86(3):997–1030.

Charlson M, Szatrowski TP, Peterson J, Gold J (1994) Validation of a combined comorbidity index. *Journal of Clinical Epidemiology* 47(11):1245–1251.

Dai J, Shi P (2019) Inpatient overflow: An approximate dynamic programming approach. *Manufacturing & Service Operations Management* 21(4):894–911.

Daw A, Pender J (2019) On the distributions of infinite server queues with batch arrivals. *Queueing Systems* 91(3):367–401.

Delasay M, Ingolfsson A, Kolfal B (2016) Modeling load and overwork effects in queueing systems with adaptive service rates. *Operations Research* 64(4):867–885.

Delasay M, Ingolfsson A, Kolfal B, Schultz K (2019) Load effect on service times. *European Journal of Operational Research* 279(3):673–686.

Deo S, Jain A (2018) Slow first, fast later: Temporal speed-up in service episodes of finite duration. *Production and Operations Management* .

Dietvorst BJ, Simmons JP, Massey C (2018) Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science* 64(3):1155–1170.

Ding Y, Park E, Nagarajan M, Grafstein E (2019) Patient prioritization in emergency department triage systems: An empirical study of the Canadian triage and acuity scale (CTAS). *Manufacturing & Service Operations Management* 21(4):723–741.

Dobson G, Lee HH, Sainathan A, Tilson V (2012) A queueing model to evaluate the impact of patient "batching" on throughput and flow time in a medical teaching facility. *Manufacturing & Service Operations Management* 14(4):584–599.

Drexl A, Kimms A (1997) Lot sizing and scheduling—survey and extensions. *European Journal of Operational Research* 99(2):221–235.

Eikeboom A, Tijms H (1987) Waiting-time percentiles in the multi-server Mx/G/c queue with batch arrivals. *Probability in the Engineering and Informational Sciences* 1(1):75–96.

Etzler P (2019) Ready, set, stall? three strategies to reduce ed boarding and get patients to the most appropriate care setting faster. URL `https://www.carelogistics.com/blog/ready-set-stall-three-strategies-to-reduce-ed-boarding-and-get-patients-faster-to-the-most-appropriate-care-setting`.

Fiebig DG, Keane MP, Louviere J, Wasi N (2010) The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science* 29(3):393–421.

Forero R, Hillman KM, McCarthy S, Fatovich DM, Joseph AP, Richardson DB (2010) Access block and ED overcrowding. *Emergency Medicine Australasia* 22(2):119–135.

Freeman M, Savva N, Scholtes S (2017) Gatekeepers at work: An empirical analysis of a maternity unit. *Management Science* 63(10):3147–3167.

Fritz MS, Kenny DA, MacKinnon DP (2016) The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivariate Behavioral Research* 51(5):681–697.

Froehle CM, White DL (2014) Interruption and forgetting in knowledge-intensive service environments. *Production and Operations Management* 23(4):704–722.

Gilboy N, Tanabe T, Travers D, Rosenau AM (2011) Emergency severity index (ESI): A triage tool for emergency department. *Rockville, MD: Agency for Healthcare Research and Quality* .

Green LV (2002) How many hospital beds? *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* 39(4):400–412.

Hanschke T (2006) Approximations for the mean queue length of the GIX/G (b, b)/c queue. *Operations Research Letters* 34(2):205–213.

Ibanez MR, Clark JR, Huckman RS, Staats BR (2017) Discretionary task ordering: Queue management in radiological services. *Management Science* 64(9):4389–4407.

Imbens GW, Wooldridge JM (2009) Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47(1):5–86.

Jones P, Wells S, Harper A, Le Fevre J, Stewart J, Curtis E, Reid P, Ameratunga S (2017) Impact of a national time target for ED length of stay on patient outcomes. *The New Zealand Medical* .

KC DS (2014) Does multitasking improve performance? evidence from the emergency department. *Manufacturing & Service Operations Management* 16(2):168–183.

KC DS, Staats BR, Kouchaki M, Gino F (2020) Task selection and workload: A focus on completing easy tasks hurts performance. *Management Science* .

KC DS, Terwiesch C (2012) An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* 14(1):50–65.

Kim T, John LK, Rogers T, Norton MI (2019) Procedural justice and the risks of consumer voting. *Management Science* 65(11):5234–5251.

Lee HS, Srinivasan MM (1989) Control policies for the MX/G/1 queueing system. *Management Science* 35(6):708–721.

Li W, Sun Z, Hong LJ (2020) Who is next: Patient prioritization under emergency department blocking. *Available at SSRN 3590979* .

Mackenzie Bean (2018) 13 hospitals with the most er visits — 2018. `https://www.beckershospitalreview.com/rankings-and-ratings/hospitals-with-the-most-er-visits-2018.html`.

Meng L, Batt R, Terwiesch C (2018) The impact of facility layout on service worker behavior: An empirical study of nurses in the emergency department. *Available at SSRN 3143883* .

Moineddin R, Meaney C, Agha M, Zagorski B, Glazier RH (2011) Modeling factors influencing the demand for emergency department services in Ontario: a comparison of methods. *BMC Emergency Medicine* 11(1):13.

Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L (2018) Emergency department crowding: A systematic review of causes, consequences and solutions. *PloS one* 13(8):e0203316.

Napoli AM, Shihab Ali AL, Baird J (2020) Boarding is associated with reduced emergency department efficiency that is not mitigated by a provider in triage. *Western Journal of Emergency Medicine* 21(3):647.

Ngo H, Forero R, Mountain D, Fatovich D, Man WN, Sprivulis P, Mohsin M, Toloo S, Celenza A, Fitzgerald G, et al. (2018) Impact of the four-hour rule in Western Australian hospitals: Trend analysis of a large record linkage study 2002-2013. *PLOS One* 13(3):e0193902.

Pang G, Whitt W (2012) Infinite-server queues with batch arrivals and dependent service times. *Probability in the Engineering and Informational Sciences* 26(2):197–220.

Powell ES, Khare RK, Venkatesh AK, Van Roo BD, Adams JG, Reinhardt G (2012) The relationship between inpatient discharge timing and emergency department boarding. *The Journal of Emergency Medicine* 42(2):186–196.

Preacher KJ, Hayes AF (2008) Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* 40(3):879–891.

Robins JM, Ritov Y (1997) Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* 16(3):285–319.

Saghafian S, Hopp WJ, Van Oyen MP, Desmond JS, Kronick SL (2012) Patient streaming as a mechanism for improving responsiveness in emergency departments. *Operations Research* 60(5):1080–1097.

Sakasegawa H (1977) An approximation formula $L_q \simeq \alpha\rho^\beta/(1-\rho)$. *Annals of the Institute of Statistical Mathematics* 29(1):67–75.

Schultz KL, Juran DC, Boudreau JW, McClain JO, Thomas LJ (1998) Modeling and worker motivation in JIT production systems. *Management Science* 44(12-part-1):1595–1607.

Schultz KL, McClain JO, Thomas LJ (2003) Overcoming the dark side of worker flexibility. *Journal of Operations Management* 21(1):81–92.

Shi P, Chou MC, Dai J, Ding D, Sim J (2015) Models and insights for hospital inpatient operations: Time-dependent ED boarding time. *Management Science* 62(1):1–28.

Shi P, Dai J, Ding D, Ang SK, Chou M, Jin X, Sim J (2014) Patient flow from emergency department to inpatient wards: Empirical observations from a singaporean hospital. *Available at SSRN 2517050* .

Song H, Tucker AL, Graue R, Moravick S, Yang JJ (2020) Capacity pooling in hospitals: The hidden consequences of off-service placement. *Management Science* 66(9):3825–3842.

Song H, Tucker AL, Murrell KL (2015) The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science* 61(12):3032–3053.

Song H, Tucker AL, Murrell KL, Vinson DR (2018) Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices. *Management Science* 64(6):2628–2649.

Staats BR, Gino F (2012) Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Science* 58(6):1141–1159.

Stiffler KA, Wilber ST (2015) Hallway patients reduce overall emergency department satisfaction. *The Journal of Emergency Medicine* 49(2):211–216.

Thanarajasingam U, McDonald FS, Halvorsen AJ, Naessens JM, Cabanela RL, Johnson MG, Daniels PR, Williams AW, Reed DA (2012) Service census caps and unit-based admissions: resident workload, conference attendance, duty hour compliance, and patient safety. *Mayo Clinic Proceedings* 87(4):320–327.

Van Ommeren JC (1990) Simple approximations for the batch-arrival MX/G/1 queue. *Operations research* 38(4):678–685.

Véricourt Fd, Jennings OB (2011) Nurse staffing in medical units: A queueing perspective. *Operations Research* 59(6):1320–1331.

Vose C, Reichard C, Pool S, Snyder M, Burmeister D (2014) Using lean to improve a segment of emergency department flow. *JONA: The Journal of Nursing Administration* 44(11):558–563.

Wang J, Zhou YP (2018) Impact of queue configuration on service time: Evidence from a supermarket. *Management Science* 64(7):3055–3075.

Whitt W (1993) Approximations for the GI/G/m queue. *Production and Operations Management* 2(2):114–161.

Whitt W, Zhang X (2017) A data-driven model of an emergency department. *Operations Research for Health Care* 12:1–15.

Wilcox RR (2011) *Introduction to robust estimation and hypothesis testing* (Academic Press).

Wooldridge JM (2015) *Introductory econometrics: A modern approach* (Cengage Learning).

Worthington D, Utley M, Suen D (2020) Infinite-server queueing models of demand in healthcare: A review of applications and ideas for further work. *Journal of the Operational Research Society* 71(8):1145–1160.

Yankovic N, Green LV (2011) Identifying good nursing levels: A queuing approach. *Operations Research* 59(4):942–955.

Yao DD (1985) Some results for the queues Mx/M/c and GIx/G/c. *Operations Research Letters* 4(2):79–83.

Yom-Tov GB (2010) Queues in hospitals: Queueing networks with reentering customers in the QED regime. *The Technion-Israel Institute of Technology Ph.D. Thesis* .

Electronic Companion

EC.1. Robustness Checks

In this section, we test the robustness of our results to some of the key assumptions.

EC.1.1. Robustness check for hypothesis 2

We rerun the model in equation 2, where the dependent variable is boarding time, and post-allocation delay, and compare with the main results. We also rerun the same model and include medical, surgical, medical ICU, surgical ICU, and observation units in our sample. Results are given in table EC.1.

Table EC.1 Impact of batching admissions on pre-allocation delay, post-allocation delay, and boarding

Dep. Variable:	(a) (main) Pre	(b) Post	(c) Boarding	(d) Pre	(e) Post	(f) Boarding
BCH	0.0455^{**}	0.0231^{*}	0.0324^{**}	0.0262	0.0273^{*}	0.0287^{**}
	(0.0168)	(0.0110)	(0.0087)	(0.0152)	(0.0059)	(0.0045)
Occ	0.0376^{**}	-0.0009	0.0182^{**}	0.0217^{*}	0.0025	0.0108^{*}
	(0.0016)	(0.0009)	(0.0008)	(0.0058)	(0.0012)	(0.0029)
Units	Medical	Medical	Medical	All units	All units	All units
N	25,565	25,115	25,549	42,560	41,690	42,538
R^2	0.2172	0.0697	0.2174	0.2473	0.0662	0.2614

$^{**}p < 0.01$, $^{*}p < 0.05$, $\hat{}p < 0.1$, robust standard errors shown in parentheses. All dependent variables are log-transformed. Error terms clustered by unit type in models (d)–(f).

The impact of batching on pre-allocation delay is not significant at a 5% or 10% level when we include all units in our sample. This is most likely due to the differences in the bed assignment procedure across units. Excluding the observation unit (Obs) results in a significant result at a 10% level, and further excluding the surgical unit (Sur) results in a significant result at a 5% level. Table EC.2 shows this.

EC.1.2. Robustness check for hypothesis 3

We conduct several robustness checks for hypothesis 3 and rerun the model shown in equation 5. First, we use the number of batched admissions during the shift (N_{BCH}) as the independent variable instead of a binary indicator, SB. The idea is to utilize a non-binary measure of batching, and evaluate the marginal effect of each batched admission. Second,

Table EC.2 Impact of batching admissions on pre-allocation delay, post-allocation delay, and boarding after excluding observation, and observation and surgery units, respectively

Dep. Variable:	(a) Pre	(b) Post	(c) Boarding	(d) Pre	(e) Post	(f) Boarding
BCH	$0.0343^{\wedge}$	0.0276^{*}	0.0319^{**}	0.0394^{*}	$0.0314^{\wedge}$	0.0335^{**}
	(0.0115)	(0.0073)	(0.0031)	(0.0081)	(0.0101)	(0.0014)
Occ	0.0308^{**}	0.0022	0.0153^{**}	0.0311^{*}	0.0032	0.0158^{*}
	(0.0040)	(0.0025)	(0.0018)	(0.0054)	(0.0036)	(0.0021)
Units	Ex. Obs	Ex. Obs	Ex. Obs	Ex. Obs & Sur	Ex. Obs & Sur	Ex. Obs & Sur
N	37,888	37,171	37,866	30,793	30,221	30,775
R^2	0.2185	0.0574	0.2143	0.2195	0.0584	0.2138

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. All dependent variables are log-transformed. Error terms clustered by unit type.

we use time to admission as the dependent variable of equation 5. Note that this also replicates the post-hoc result as described in section EC.2.3 on a shift-level. Specifically, in section EC.2.3 we find that batched patients suffer a longer time to admission. By extension, the average time to admission should be longer in shifts with batching in comparison with shifts with no batching. As a final robustness check, we rerun the model in equation 5 with time to discharge as the dependent variable. The idea is to test whether the decrease in shift throughput time is due to a decrease in discharge time. In other words, since throughput time is lower in shifts with batched admissions (hypothesis 3B) while batched admissions take longer, then the average time to discharge must be *lower* in shifts with batching. In short, faster discharges must be the underlying reason for higher shift productivity.

Summary statistics and correlations of shift with batching, number of batched admissions during shift, time to admit and time to discharge are provided in table EC.3.

Table EC.3 Summary statistics and correlation table for shift with batching, number of batched admissions during shift, time to admit and time to discharge

		mean	SD	min	max	1	2	3
1	Shift with batching	0.37	0.48	0	1	1		
2	Number of batched admissions	0.96	1.43	0	11	0.87^{*}	1	
3	Time to admit[a]	2.58	0.93	0.10	8.05	0.04^{*}	0.04^{*}	1
4	Time to discharge[a]	2.87	0.81	0.21	11.53	0.10^{*}	0.10^{*}	0.11^{*}

Note: $^{*}p < 0.05$, $N = 7,619$. [a]unit is hours

The results are provided in table EC.4. In model (a) the coefficients of number batches in shift is significant in the hypothesized direction. Namely, average throughput time is lower in shifts with batching, while the number of patients served is higher. Also, models (b) and (c) show that shift with batching is positively associated with time to admit ($\beta = 0.0742$, $p < 0.01$), and negatively associated with time to discharge ($\beta = -0.0424$, $p < 0.05$). This further lends support to hypothesis 3.

Table EC.4 Robustness of hypothesis 3

	Model (a)		Model (b)	Model (c)
Dep. Variable:	N_{pts}	STT	Time to admit	Time to discharge
Shift with Batching			0.0742**	-0.0424*
			(0.0202)	(0.0171)
Num. Batched	0.8469**	-0.0182**		
	(0.0407)	(0.0050)		
Shift Length	1.9977**	0.2231**	0.2025**	0.2000**
	(0.1963)	(0.0194)	(0.0458)	(0.0326)
Num. ED Tests	-2.0350**	0.3042**	0.2076**	0.4364**
	(0.1755)	(0.0204)	(0.0429)	(0.0253)
Num. Roomed in Shift	0.0863**	0.0008^	-0.0016	0.0015*
	(0.0066)	(0.0003)	(0.0010)	(0.0006)
ESI Score	1.0572*	0.0061	0.5142**	-0.3361**
	(0.5058)	(0.0495)	(0.0788)	(0.0630)
N	7060	7060	7060	6986
R^2	0.3541	0.2013	0.1047	0.2121

**$p < 0.01$, *$p < 0.05$, ^$p < 0.1$, robust standard errors shown in parentheses. Regressions done on a CEM-matched sample. In model (c), physician shifts with average time to discharge of over 5.5 hours were removed (top 1%). Covariate balance in all models is similar (or identical) to that shown in table 6

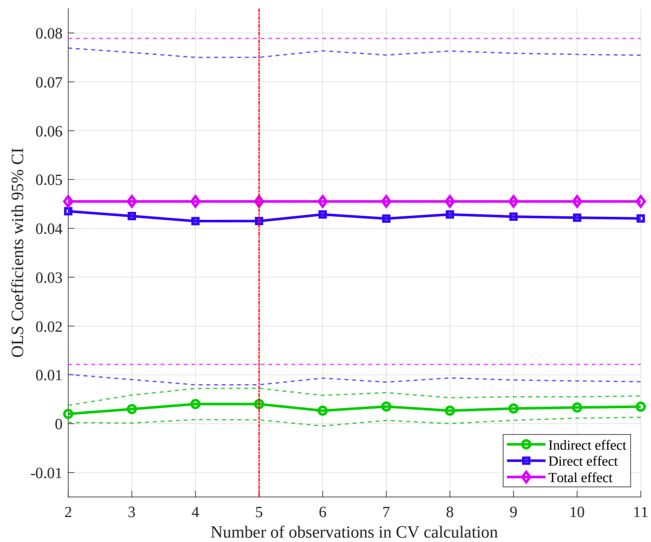
EC.1.3. Robustness check for the mediating effect of CV in hypothesis 2

As discussed in section 6.2.3, we calculated the individual CV experienced by each patient by grouping the latest 5 admissions, including the focal patient, and calculating the CV of inter-bed-request times on those observations. We test the robustness of our results to using a different number of observations. Fig. EC.1 shows the results of redoing the analysis with each CV definition. The x-axis is the number of observations used in calculating CV, and the y-axis shows the regression coefficients with their 95% confidence intervals for the indirect, direct, and total effect of batching on pre-allocation delay.

It is evident that the impact of batching on pre-allocation delay through the increase of CV (indirect effect) is positive and significant at the 5% level, except for when 6 observations are used to calculate CV. In that instance, the indirect effect is significant on a

10% level. Taken in total, we argue that the results are fairly robust to the number of observations chosen to calculate CV.

Figure EC.1 (Color online) Mediation coefficients with their 95% confidence intervals



As a second robustness check, we do not Winsorize the top 1% of inter-bed-request times, and redo our mediation analysis. We redo the mediation analysis as described in section 6.2.3. The result of our analysis, before and after Winsorizing the inter-bed-request times is shown in table EC.5. We observe that our results are robust to not Winsorising inter-bed-request times, as indicated by the positive and significant results in columns (d) and (e). The mediation level is $a \times b = 0.0715 \times 0.0461 = 0.0033$ ($p < 0.05$), and $c' = 0.0422$, ($p < 0.05$). The mediation level of CV is therefore derived from $\frac{a \times b}{a \times b + c'} = 7.3\%$. This result is close to the 8.8% mediation level calculated in the main body of the paper, when CV is calculated after Winsorizing the top 1% of inter-bed-request times.

While our main results are robust to using un-Winsorized inter-bed-request times, we believe that the better approach to calculating patient-level CV is in fact to Winsorize the top 1% of inter-bed-request times to eliminate the effect of very long intervals between bed requests (e.g. shift changes or less busy ED) from influencing the variation that ultimately influences pre-allocation delay (Wilcox 2011).

Table EC.5 Mediation results with and without Winsorizing inter-bed-request times when calculating CV

		With Winsorizing (main results)		No Winsorizing	
	(a)	(b)	(c)	(d)	(e)
Dep. Variable:	$LPAD$	CV	$LPAD$	CV	$LPAD$
Batch	0.0455^{**}	0.0713^{**}	0.0416^{*}	0.0715^{**}	0.0422^{*}
	(0.0168)	(0.0049)	(0.0169)	(0.005)	(0.0169)
CV			0.0555^{*}		0.0461^{*}
			(0.0228)		(0.0221)
N	25,565	25,565	25,565	25,565	25,565
R^2	0.2172	0.0173	0.2174	0.0157	0.2173

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

EC.1.4. Batching defined on disposition times (not admissions) in hypotheses 1 & 3

In this section we redo the analysis of hypotheses 1 and 3 with a new definition of batching which is based on *disposition times*, as opposed to *admission times*. In other words, we include discharged patients in the new definition of batching.

Following the same method described in section 5.2, we use a GMM to decipher the underlying latent distributions of the log inter-disposition-time distribution (i.e. time between successive disposition decisions in a physician-shift). We find that the threshold for batching is 6.8 minutes, and 31% of the dispositions are completed in a batch.

Hypothesis 1: We re-estimate equation 1, where BCH is replaced with the new definition. The results of our estimation are shown in table EC.6, similar in format to that of table 3.

Table EC.6 Robustness check for hypothesis 1: Batching dispositions towards end of shifts

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(a)	(b)	(c)
TLS	-0.1834^{**}	-0.1772^{**}	0.1128^{**}
	(0.0068)	(0.0182)	(0.0118)
Controls	NO	All controls	All controls
N	31,563	30,791	42,413
R^2	0.0230	0.0501	0.0458
LL	-20399.2	-19363.5	-25724.4

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Standard errors clustered by physician. Drop in N in model (b) is due to missing values in controls. Model (c) includes all observations (both $TLS < 0$ and $TLS \geq 0$).

The coefficient of TLS in model (b) (i.e. sample of patients admitted during the shift) is $\beta_1 = -0.1772$ ($p < 0.01$). The negative coefficient lends additional support for hypothesis 1.

Hypothesis 3: To check the robustness of our results to the new definition of batching (i.e. batching disposition decisions and not admissions), we follow the CEM-matching procedure as described in section 6.3, and rerun the analysis in a similar manner. However, there are three differences. First, since batching is defined on all patients – whether they are discharged or admitted – we should not expect the average disposition time to necessarily decrease. Specifically, since disposition decisions may be delayed, it is plausible that more patients are served, but the average disposition time is also longer. As a result, hypothesis 3B may not hold (which is not unexpected). Second, we match the shifts with and without batching on the number of "admitted" patients in the shift, and control for it in equation 5 because we want to make sure that our results are not *driven* by the number of admitted patients in a physician-shift. Third, our sample is defined based on the shifts with at least two *patients seen* in the shift instead of two *admissions* in the shift. This leaves us a sample size of 10,548 out of 10,659 physician shifts (see Fig. 2).

The CEM-matching results, and analysis outcomes are provided in tables EC.7 and EC.8, respectively. Table EC.7 demonstrates that prior to matching there were some imbalances in shifts with and without batching; however, after matching we have generated a balanced sample since all normative differences between covariates in shifts with and without batching are below a threshold of 0.2 (Batt et al. 2019).

The coefficient on SB is positive and significant in column (b) of table EC.8, which provides support for hypothesis 3A. However, hypothesis 3B is not supported in the new definition of shift with batching which is based on batching all dispositions, since the coefficient on SB in both columns (c) and (d) are positive. However, as discussed above, this only shows that batched dispositions are delayed, and does not imply lower productivity or contradictory results.

EC.1.5. Robustness to using at least 3 admissions to define an admission batch

We originally used at least two back-to-back admissions to define a batch. We check the robustness of our results to increasing the batch size to at least three. In other words, we define a batch of admissions when at least three patients are admitted within a short

Table EC.7 Covariate balance in full and matched samples, when batching is defined on all disposition decisions

Sample	Full sample No weights			CEM-matched sample CEM-weighted		
Variable	SB=0	SB=1	Norm. Diff.	SB=0	SB=1	Norm. Diff.
Mean age	47.53	47.46	0.01	47.95	47.50	0.01
Shift length	7.92	8.22	0.17	8.33	8.31	0.00
Mean number of ED tests	1.96	1.95	0.01	1.91	1.90	0.01
Number of patients roomed	128.3	123.8	0.08	133.7	133.6	0.00
Mean ESI	3.19	3.10	0.10	3.15	3.16	0.00
Mean Charlson Score	1.46	1.45	0.01	1.54	1.43	0.05
Mean number of ICU visits	0.06	0.05	0.05	0.05	0.04	0.06
hours since last shift	86.50	71.04	0.08	61.79	60.43	0.01
hours to next shift	86.29	71.70	0.09	64.65	59.44	0.04
Morning shift	0.61	0.47	0.21	0.55	0.54	0.00
Afternoon shift	0.33	0.37	0.06	0.38	0.38	0.00
Overnight shift	0.06	0.16	0.24	0.08	0.08	0.00
Weekend shift	0.12	0.22	0.19	0.10	0.16	0.09
Waiting room census	102.4	107.2	0.05	111.28	114.73	0.02
N	1162	9386		951	5164	

Table EC.8 Hypothesis 3: Impact of batching disposition decisions on physician productivity

Dep. Variable:	Number of Patients Seen		Average Throughput Time	
	(a) unweighted OLS	(b) weighted CEM-matched	(c) unweighted OLS	(d) weighted CEM-matched
SB	2.5517**	2.7215**	0.0886**	0.0419$^{\wedge}$
	(0.1543)	(0.2072)	(0.0174)	(0.0247)
Shift Length	1.1120**	1.0563**	0.0948**	0.0570**
	(0.0906)	(0.1137)	(0.0257)	(0.0206)
Mean Num. of ED tests	-3.1307**	-2.8130**	0.3889**	0.4454**
	(0.1270)	(0.1742)	(0.0157)	(0.0230)
Num. Patients admitted	1.2067**	1.2356**	-0.0177**	-0.0146*
	(0.0255)	(0.0404)	(0.0034)	(0.0058)
N	10,548	6,115	10,545	6,115
R^2	0.6101	0.5816	0.6350	0.6898

Note. $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Controls included in all models.

interval (9.1 minutes) of each other. By this definition, 5.8% of the patient admissions are done in a batch. In the remainder of the section, we show the robustness of our main results to the new definition.

Hypothesis 1 Table EC.9 shows the main results for hypothesis 1.

The coefficient of TLS in column (b) (i.e. sample of patients admitted during the shift) is $\beta_1 = -0.1700$ ($p < 0.01$), which provides support for hypothesis 1.

Hypothesis 2 Table EC.10 shows the main results for hypothesis 2. We observe that the coefficient of interest is $\beta_1 = 0.0929$ ($p < 0.01$), which provides support for our second hypothesis under the new definition of batching.

Table EC.9 Robustness check for hypothesis 1: Batch defined based on at least three patients

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(a)	(b)	(c)
TLS	-0.1798** (0.0165)	-0.1700** (0.0435)	0.1080** (0.0121)
Controls	NO	All controls	All controls
N	31,563	30,298	41,828
R^2	0.0150	0.0748	0.0645
LL	-7557.1	-6919.9	-8765.8

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Standard errors clustered by physician. Drop in N in model (b) is due to missing values in controls. Model (c) includes all observations (both $TLS < 0$ and $TLS \geq 0$).

Table EC.10 Hypothesis 2: Batch defined based on at least three patients

Dep. Variable:	ln(Pre-allocation Delay)	
	(a)	(b)
BCH	0.0361 (0.0302)	0.0929** (0.0273)
Occ		0.0376** (0.0016)
Controls	NO	YES
N	25,565	25,565
R^2	0.0000	0.2173
F	1.43	45.46

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Hypothesis 3 Following the method described in section 6.3, we first use CEM-matching to generate a balanced sample of shifts with and without batching, then run our analysis on the matched sample to test hypotheses 3A and 3B. The results for covariate balance before and after CEM-matching are shown in table EC.11, and the results for the impact of batching on physician productivity are given in table EC.12.

We observe that the sample is balanced as expected after CEM-matching, and the results in table EC.12 are consistent with the main results of the paper. In short, both hypotheses 3A and 3B are robust defining a batch based on at least three admissions.

EC.1.6. Robustness to removing patients who were not roomed on a first-come-first-serve basis

We check the robustness of our results to removing the subset of patients who were not roomed on a first-come-first-serve (FCFS) basis from our sample.

Table EC.11 Covariate balance in full and matched samples when batch size is at least 3

Sample	Full sample No weights			CEM-matched sample CEM-weighted		
Variable	SB=0	SB=1	Norm. Diff.	SB=0	SB=1	Norm. Diff.
Mean age	49.36	51.14	0.23	50.34	51.19	0.02
Shift length	7.92	7.93	0.02	7.94	7.94	0.00
Mean number of ED tests	2.40	2.59	0.27	2.58	2.58	0.00
Number of patients roomed	112.1	116.6	0.09	116.8	116.7	0.00
Mean ESI	2.80	2.72	0.24	2.74	2.72	0.01
Mean Charlson Score	1.71	1.83	0.14	1.78	1.84	0.03
Mean number of ICU visits	0.07	0.08	0.09	0.08	0.08	0.00
hours since last shift	71.66	68.47	0.02	67.60	67.05	0.00
hours to next shift	71.45	74.12	0.02	66.21	69.31	0.02
Morning shift	0.41	0.36	0.07	0.36	0.36	0.00
Afternoon shift	0.38	0.50	0.17	0.51	0.51	0.00
Overnight shift	0.21	0.14	0.14	0.14	0.14	0.00
Weekend shift	0.23	0.26	0.04	0.21	0.26	0.08
Waiting room census	99.26	106.8	0.07	111.4	106.6	0.03
N	6428	675		5257	659	

Table EC.12 Hypothesis 3: Impact of batching on physician productivity when batch size is at least 3

Dep. Variable:	Number of Patients Seen		Average Throughput Time	
	(a) unweighted OLS	(b) weighted CEM-matched	(c) unweighted OLS	(d) weighted CEM-matched
SB	2.4621**	2.2363**	-0.0361*	-0.0494**
	(0.1965)	(0.1977)	(0.0177)	(0.0173)
Shift Length	1.7761**	1.9884**	0.2063**	0.1926**
	(0.2072)	(0.3874)	(0.0150)	(0.0277)
Mean Num. of ED tests	-1.8985**	-2.1652**	0.3034**	0.3038**
	(0.1643)	(0.1922)	(0.0155)	(0.0262)
Num. Roomed in Shift	0.0794**	0.0893**	0.0004	0.0017**
	(0.0062)	(0.0076)	(0.0004)	(0.0006)
N	7,103	5,916	7,103	5,916
R^2	0.3269	0.3506	0.2100	0.1877

Note. $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Controls included in all models.

As described in section 3 patients arriving by ambulance are typically roomed right away, and self-arrivals are roomed on FCFS basis by their ESI score. However, this general rule is violated about 18% of the time, as shown in table EC.13. Moreover, table EC.14 provides a summary of what percentage of FCFS (and not FCFS patients) are admitted. By redoing our analysis without these patients, we build confidence that our results are not driven by a set of events or decisions prior to a patient being roomed.

Hypothesis 1 We re-estimate equation 1 without after removing the non-FCFS patients. The results of our estimation are shown in table EC.15, similar in format to that of table 3.

Table EC.13 Breakdown of patients roomed on a FCFS basis by mode of ED transportation

Transport	**FCFS** (%)	**not FCFS** (%)	**Total**
Ambulance	93.18	6.82	69,986
Self arrival	77.19	22.81	173,437
Other	89.33	10.67	9,913
Total	82.07	17.93	253,336

Table EC.14 Breakdown of admissions by whether patient was roomed on FCFS basis

	Discharged (%)	**Admitted** (%)	**Total**
FCFS	76.78	23.22	203,555
not FCFS	88.24	11.76	36,317
Total	78.51	21.49	239,872

Table EC.15 Robustness check for hypothesis 1: Removing non-FCFS patients

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(a)	(b)	(c)
TLS	-0.1526**	-0.2079**	0.0885**
	(0.0096)	(0.0195)	(0.0106)
Controls	NO	All controls	All controls
N	29,082	28,119	38,680
R^2	0.0143	0.0424	0.0370
LL	-14515.8	-13732.2	-17745.9

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Standard errors clustered by physician. Drop in N in model (b) is due to missing values in controls. Model (c) includes all observations (both $TLS < 0$ and $TLS \geq 0$).

The coefficient of TLS in model (b) (i.e. sample of patients admitted during the shift) is $\beta_1 = -0.2079$ ($p < 0.01$). The negative coefficient lends additional support for hypothesis 1.

Hypothesis 2 Table EC.16 shows the main results for hypothesis 2. We observe that the coefficient of interest is $\beta_1 = 0.0494$ ($p < 0.01$), which provides support for our second hypothesis after removing the non-FCFS patients.

Hypothesis 3 We first remove the non-FCFS patients from our original patient-level sample, then collapse our data set on a physician-shift level as described in section 5.1. Next, following the method described in section 6.3, we first use CEM-matching to generate a balanced sample of shifts with and without batching, then run our analysis on the matched sample to test hypotheses 3A and 3B. The results for covariate balance before and after CEM-matching are shown in table EC.17, and the results for the impact of batching on physician productivity are given in table EC.18.

Table EC.16 Hypothesis 2: Removing non-FCFS patients

Dep. Variable:	ln(Pre-allocation Delay)	
	(a)	(b)
BCH	-0.0033	0.0494^{**}
	(0.0195)	(0.0176)
Occ		0.0374^{**}
		(0.0016)
Controls	NO	YES
N	23,417	23,417
R^2	0.0000	0.2198
F	0.02	42.29

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

We observe that the sample is balanced as expected after CEM-matching, and the results in table EC.18 are consistent with the main results of the paper. In short, both hypotheses 3A and 3B are robust defining a batch based on at least three admissions.

Table EC.17 Covariate balance in full and matched samples; non-FCFS patients excluded from sample from which physician-shift sample is derived

Sample	Full sample No weights			CEM-matched sample CEM-weighted		
Variable	SB=0	SB=1	Norm. Diff.	SB=0	SB=1	Norm. Diff.
Mean age	49.29	51.09	0.21	50.30	51.06	0.02
Shift length	7.91	7.94	0.04	7.95	7.94	0.00
Mean number of ED tests	2.36	2.60	0.31	2.58	2.59	0.00
Number of patients roomed	111.6	114.4	0.05	114.3	114.4	0.00
Mean ESI	2.81	2.72	0.26	2.73	2.72	0.01
Mean Charlson Score	1.70	1.86	0.16	1.80	1.86	0.03
Mean number of ICU visits	0.07	0.09	0.14	0.08	0.09	0.02
hours since last shift	70.73	71.61	0.01	63.47	65.23	0.02
hours to next shift	71.60	71.88	0.00	66.62	66.35	0.00
Morning shift	0.43	0.36	0.11	0.35	0.35	0.00
Afternoon shift	0.35	0.48	0.19	0.48	0.48	0.00
Overnight shift	0.22	0.17	0.10	0.17	0.17	0.00
Weekend shift	0.22	0.25	0.04	0.21	0.25	0.07
Waiting room census	95.28	105.97	0.10	106.5	. 106.1	0.00
N	4684	2817		4267	2703	

Table EC.18 Hypothesis 3: Impact of batching on physician productivity, excluding non-FCFS patients

Dep. Variable:	Number of Patients Seen		Average Throughput Time	
	(a) unweighted OLS	(b) weighted CEM-matched	(c) unweighted OLS	(d) weighted CEM-matched
SB	2.2707**	2.1137**	-0.0238$^{\wedge}$	-0.0296*
	(0.1131)	(0.1149)	(0.0139)	(0.0138)
Shift Length	1.8324**	2.0092**	0.2301**	0.2177**
	(0.1830)	(0.2139)	(0.0172)	(0.0169)
Mean Num. of ED tests	-1.6998**	-1.8426**	0.2777**	0.2645**
	(0.1282)	(0.1435)	(0.0160)	(0.0211)
Num. Roomed in Shift	0.0777**	0.0853**	0.0002	0.0014**
	(0.0059)	(0.0067)	(0.0004)	(0.0005)
N	7,501	6,970	7,501	6,970
R^2	0.3345	0.3420	. 0.2093	0.1731

Note. $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Controls included in all models.

EC.2. Post-hoc: Impact of Batching Admissions on ED Length of Stay

As mentioned in section 7.2.1, we explore the overall effect of batching admissions on ED length of stay (LOS), defined as the time from when the patient is roomed until they depart the ED. We study this effect on a patient-level and shift-level separately, as discussed below.

EC.2.1. Patient-level: Impact of batching on LOS (sample of admitted patients)

We first study whether batching increases ED LOS on a patient-level. The question is whether patients admitted in a batch experience a longer ED LOS. To do so, we must also consider the heterogeneous effect of batching on ED LOS, depending on whether the patient was batched before or after the primary physician's shift had ended because arguably, if batching occurs after the shift then part of its impact on LOS could be because the patient was cosigned or handed-off. Table EC.19 provides summary statistics for ED LOS, calculated for patients admitted during and after the shift, respectively.

Table EC.19 Summary statistics for ED length of stay for admitted patients

	mean	sd	min	p50	p99	max	Count
During Shift	6.31	2.96	0.42	5.74	16.46	22.82	31,474
After Shift	9.41	4.03	0.67	8.66	21.33	23.26	11,645
Total	7.15	3.56	0.42	6.40	19.18	23.26	43,119

Note: Unit is hours. top 1% of ED LOS excluded to eliminate outliers.

We rerun equation 2, with log-transformed ED LOS as the dependent variable. To further explore the heterogeneous effect of batching on patients admitted before and after the shift, we add a binary control if the patient was admitted after the primary physician's shift ended ($AFTER$), and also include an interaction term between batching and $AFTER$. The results are given in table EC.20.

Prior to adding the after-shift and interaction controls (models 1 and 3), batching appears to lower ED LOS since the coefficient of batching BCH is negative. However, after adding the interaction terms (models 2 and 4), it is clear that patients admitted in a batch during the shift experienced a longer LOS ($\beta = 0.0197$ in model 2, $\beta = 0.0276$ in model 4), but experienced a shorter LOS when batched after the shift since the interaction coefficients of $BCH \times AFTER$ are greater than BCH.

Table EC.20 Post-hoc results: Patient-level analysis

Dep. Variable:	log(ED LOS)			
Sample	Admitted to medical units		Admitted to All units	
	(1)	(2)	(3)	(4)
BCH	-0.0322**	0.0197**	-0.0315**	0.0276*
	(0.0063)	(0.0069)	(0.0060)	(0.0080)
$AFTER$		0.3943**		0.4375**
		(0.0063)		(0.0363)
$BCH \times AFTER$		-0.1186**		-0.1382**
		(0.0150)		(0.0150)
N	25,513	25,513	42,388	42,388
R^2	0.1723	0.2826	0.2381	0.3538

$^{**}p < 0.01$, $^{*}p < 0.05$, $\hat{}p < 0.1$, robust standard errors shown in parentheses. Controls included in all models. Errors clustered by unit type in models 3 and 4

EC.2.2. Shift-level: Impact of batching on shift-level average LOS

We study the overall impact of batching admissions on shift-level performance. For this, we find the average ED LOS of patients treated in each physician shift, similar to the definition of shift throughput time, mentioned in section 5.3. However, the difference between the calculation of average shift throughput time and ED LOS is that when calculating average shift throughput time, we excluded after-shift patients to focus on the behavior of an individual physician. In contrast, in calculating average LOS, we include all patients, regardless of when they left the ED. This is an important distinction for two reasons: first, unlike a disposition decision, the actual LOS of patients may go well beyond an individual shift due to boarding times or other reasons related to the discharge process of the patients (which is outside of the scope of this paper); second, the LOS of patients is ultimately what matters when assessing ED overcrowding because regardless of where in their treatment process they are, patients present in the ED occupy resources, and contribute to crowding.

We conduct a similar analysis to section 6.3. Namely, the key independent variable in our analysis is a binary variable, shift with batching (SB), but the dependent variable in this case is average ED LOS ($\overline{LOS}$). Also, to account for the possibility that average ED LOS in a shift may be impacted by the patients who were either admitted or discharged after the shift had ended, we further control for the fraction of patients in the shift whose disposition occurred after the shift, $FRAC_{after}$, in the model shown in equation 5. Table EC.21 shows the summary statistics and correlations between $\overline{LOS}$, $FRAC_{after}$, and SB.

Table EC.21 Summary statistics and correlations between $\overline{LOS}$, $FRAC_{after}$, and SB

		mean	sd	min	max	(1)	(2)
1	$\overline{LOS}^{a}$	4.13	1.68	0.02	10.53	1	
2	$FRAC_{after}$	0.21	0.16	0	1	0.72*	1
3	SB	0.35	0.48	0	1	0.22*	0.01

Note: [a]Unit is hours.

The results of the model as described above are provided in table EC.22. Recall that we use CEM-matching to match shifts with and without batching on patient and shift characteristics. The results for the regression run on a weighted CEM-matched sample in table EC.22. We see that the coefficient on SB is positive and significant ($\beta = 0.0878$, $p < 0.01$). The coefficient suggests that the average LOS on shifts with batching is ~ 0.09 hours (~ 5.4 minutes) longer than shifts without batching. We conclude that batching increases the average LOS on a shift-level.

Table EC.22 Post-hoc results: Shift-level analysis

Dep. Variable:	$\overline{LOS}$
SB	0.0878**
	(0.0193)
$FRAC_{after}$	2.8342**
	(0.1053)
N	6,968
R^2	0.4634

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. All controls from section 6.3 included. Errors clustered by physician

EC.2.3. Potential Batch Formation Mechanism

We explore a potential batch-forming mechanism during the shift, and also provide a possible rationale for why batching may occur after the assigned shift has ended. For this, we focus on the time it takes to admit patients. If batching admissions is a byproduct of physicians speeding up towards end of shifts, then patients who are batched during a shift should experience a shorter time to admission. At the end of the shift, a few remaining patients might have lingering issues, such as pending lab reports or consults, who are later admitted by a resident and cosigned by the primary physician, or handed off to a second physician. In the latter case, most likely the primary physician would have had completed the majority of the diagnosis tasks, and considering the cultural disincentive to hand off patients, had provided detailed notes regarding the patient's conditions to the second physician (who is not credited with the patient's care) in an effort to facilitate their work as much as possible. Hence, the admission decision process for such patients is easier, and the second physician is able to admit multiple patients in a short period of time. Consequently, we expect that patients who are batched after the shift experience a shorter time to admission compared to unbatched after-shift admissions due to the expedited decision process.

To test this theory, we run the regression in equation EC.1 on a patient-level, with robust standard errors:

$$ln(AdmitTime) = \beta_0 + \beta_1 BCH + \beta_2 AFTER + \beta_3 BCH \times AFTER + \boldsymbol{\gamma X_i} + \epsilon_i \qquad \text{(EC.1)}$$

where $ln(AdmitTime)$ is the log-transformed time it takes to to admit a patient, and $AFTER$ is a binary variable equal to 1 if a patient is admitted after the shift, and 0 otherwise. $\boldsymbol{\gamma X}$ includes date/time, ward, medical and patient controls, and the admitting physician. Recall from section 6.2 that assignment of patients to the batched or unbatched groups is as good as random. Hence, coefficient β_1 captures the impact of batching during the shift on time to admission, and $\beta_1 + \beta_3$ captures the impact of batching after shift. The results are given in table EC.23. We find that $\beta_1 = 0.0578$ ($p < 0.05$), equivalent to an average of 9.1 minutes increase, and $\beta_1 + \beta_3 = -0.1618$ ($p < 0.01$), equivalent to an average of 52.2 minutes decrease, in time to admission for patients admitted during and after the physician shift, respectively.

This evidence suggests that patients who were batched after the shift were expedited, as expected, but those batched during the shift were *delayed*. Our physician partner suggests that the delay may be due to postponing tasks at various points in the shift. Specifically, many times physicians order tests, but postpone the task of reviewing them in order to attend to other patients who have not yet been diagnosed. This behavior may result in delaying the admission of some patients to attend to a higher number of patients.

We also find that $\beta_2 = 0.9090$ ($p < 0.01$), suggesting that the average time to admission is longer when patients are admitted after the shift has ended. This effect may most likely be attributed to the fact that patients with pending issues are either handed off or cosigned, and therefore, on average, they experience a longer time to admission.

Table EC.23 Results for equation EC.1

Dep. Variable:	ln(AdmitTime)
BCH	0.0578*
	(0.0150)
$AFTER$	0.9090**
	(0.0616)
$BCH \times AFTER$	-0.2196**
	(0.0105)
Controls	YES
N	42,256
R^2	0.4096

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses.

EC.3. Additional Summary Statistics, Tables and Figures

Table EC.24 Summary statistics of number of patients visits and bed requests per patient

Variable	mean	sd	min	p1	p50	p99	max
Number of visits	2.20	3.24	1	1	1	14	163
Bed requests	0.47	1.29	0	0	0	5	44
Bed requests if admitted at least once	1.74	1.98	1	1	1	10	44

Note. Statistics given from 108,931 unique patient visits, out of which 29,574 patients were admitted at least once

Table EC.25 Correlations between tests for admitted patients

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	CT	1.00						
(2)	EKG	0.08^{*}	1.00					
(3)	Imaging	0.34^{*}	0.22^{*}	1.00				
(4)	Other Imaging	0.05^{*}	0.03^{*}	0.06^{*}	1.00			
(5)	Ultrasound	-0.00	-0.10^{*}	0.16^{*}	0.03^{*}	1.00		
(6)	Lab	0.07^{*}	0.14^{*}	0.11^{*}	0.00	0.04^{*}	1.00	
(7)	X-ray	-0.03^{*}	0.33^{*}	0.62^{*}	-0.06	-0.12^{*}	0.08^{*}	1.00

$^{*}p < 0.05$

Figure EC.2 Number of bed requests per patient across all individual visits

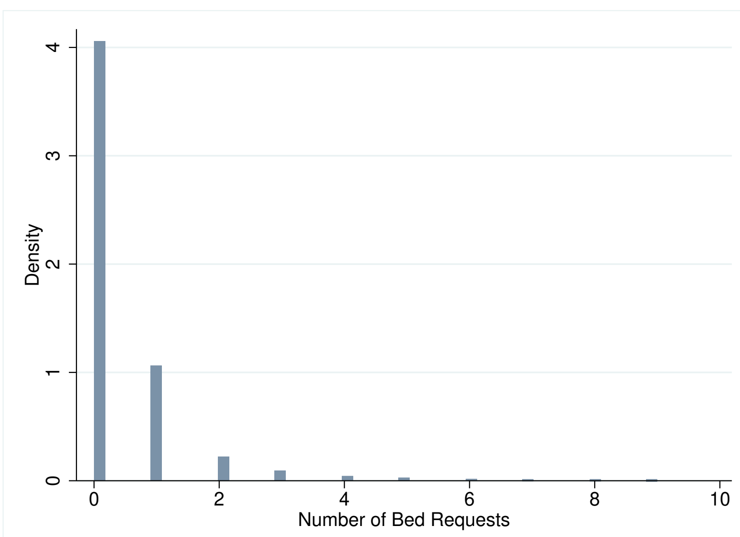
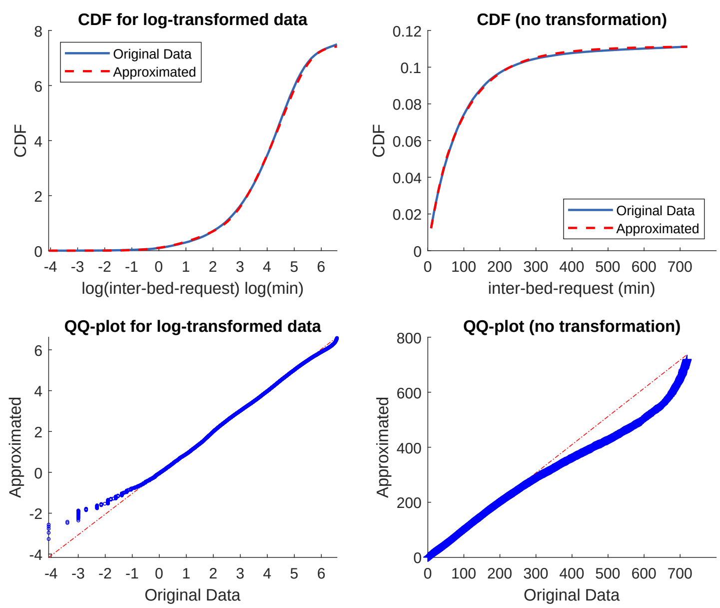


Table EC.26 Units to which Patients are Admitted from the ED

Unit	Count	Percentage	mean boarding time (min)
Medical Units	27,490	55.85	248.09
Surgical Units	7,669	15.58	227.09
Observation Units	6,269	12.74	132.03
Medical ICUs	4,126	8.38	180.39
Surgical ICUs	1,611	3.27	180.74
Operating Room	1,277	2.59	127.27
Pediatrics Unit	598	1.21	140.20
Labor and Delivery	119	0.24	122.86
Pediatric ICU	63	0.13	112.19
Total	49,222	100.00	

Our raw data does not include labor and delivery and pediatrics patients, and we exclude them from our analysis. This explains the very small number of admissions to these units.

Figure EC.3 Goodness of fit plots for GMM.



Note. GMM was applied to approximated the log-transformed inter-bed-request times. Also note that:

(i) the deviations from the 45°line in the Q-Q plots are not in the vicinity of the critical batching threshold;

(ii) the large deviation in Q-Q plot on the bottom right is because the original data is truncated at 720 minutes (∼6.5 log minutes) (see Fig. 3), and the model does not fit well at large log(inter-bed-request times); and since the approximated log-transformed data was transformed back using the exponent function to make this plot, the large values are grossly exaggerated. To illustrate this, note that the CDF plots fit well, and that the bump in the bottom left Q-Q plot between ∼5.5 – ∼6.5 on the x-axis, is the same deviation seen in the bottom right Q-Q plot.

Table EC.27 Summary statistics for admitted vs. not admitted ED patients

Variable	Not Admitted	Admitted	Total
Insurance class			**237,121**
Commercial	78.46%	21.54%	57,807
Medicaid	82.54%	17.46%	119,994
Medicare	63.07%	36.93%	36,106
Free Care & Other	91.29%	8.71%	23,214
Race			**239,874**
Black / African American	81.13%	18.87%	176,195
Non-Black	74.93%	25.07%	63,679
ESI Score			**237,959**
1	16.97%	83.03%	2,157
2	50.73%	49.27%	47,679
3	80.29%	19.71%	114,487
4	98.63%	1.37%	66,228
5	99.65%	0.35%	7,408
Gender			**239,873**
Male	78.73%	21.27%	120,231
Female	80.24%	19.76%	119,617
Unspecified	96.00	4.00	25
Year			**239,874**
2016	79.00%	21.00%	63,592
2017	79.29%	20.71%	110,428
2018	80.27%	19.73%	65,854
Day of Week			**239,874**
Sunday	79.00%	21.00%	29,258
Monday	79.19%	20.81%	37,797
Tuesday	79.57%	20.43%	36,600
Wednesday	79.91%	20.09%	35,832
Thursday	79.17%	20.83%	34,862
Friday	79.29%	20.71%	34,961
Saturday	80.27%	19.73%	30,564
Week			**239,874**
1	78.21%	21.79%	4,020
2	78.87%	21.13%	4,255
3	79.54%	20.46%	4,287
4	79.98%	20.02%	4,311
5	79.21%	20.79%	4,329
6	79.83%	20.17%	4,397
7	79.94%	20.06%	4,348
8	80.35%	19.65%	4,452
9	79.36%	20.64%	4,224
10	77.65%	22.35%	4,019

Continued on next page

a All percentages are based on a total of 239,874 ED patients.
b Total count less than 239,874 are due to missing values.

Table EC.27 Summary statistics for admitted vs. not admitted ED patients (Cont.)

Variable	Not Admitted	Admitted	Total
11	78.83%	21.17%	4,016
12	79.66%	20.34%	4,287
13	79.88%	20.12%	4,250
14	79.23%	20.77%	4,286
15	80.03%	19.97%	4,367
16	80.17%	19.83%	4,373
17	80.21%	19.79%	4,376
18	80.08%	19.92%	4,263
19	80.86%	19.14%	4,363
20	80.39%	19.61%	4,371
21	79.43%	20.57%	4,361
22	80.25%	19.75%	4,992
23	80.64%	19.36%	6,508
24	80.32%	19.68%	6,621
25	80.69%	19.31%	6,443
26	81.08%	18.92%	6,475
27	78.83%	21.17%	6,268
28	79.50%	20.50%	6,595
29	78.51%	21.49%	6,367
30	79.62%	20.38%	6,260
31	78.74%	21.26%	4,789
32	78.88%	21.12%	4,294
33	78.10%	21.90%	4,232
34	80.18%	19.82%	4,268
35	77.70%	22.30%	4,183
36	79.76%	20.24%	4,263
37	78.93%	21.07%	4,277
38	78.33%	21.67%	4,274
39	79.29%	20.71%	4,244
40	79.78%	20.22%	4,278
41	78.15%	21.85%	4,242
42	78.68%	21.32%	4,227
43	80.28%	19.72%	4,213
44	79.05%	20.95%	4,038
45	79.62%	20.38%	4,190
46	79.30%	20.70%	4,223
47	78.84%	21.16%	3,928
48	78.72%	21.28%	4,290
49	79.49%	20.51%	4,183
50	79.84%	20.16%	4,017
51	78.63%	21.37%	3,944
52	79.25%	20.75%	4,722

a All percentages are based on a total of 239,874 ED patients.
b Total count less than 239,874 are due to missing values.

Table EC.28 Summary statistics for patient controls

Variable	Percent	Total
Insurance class		**44,362**
Medicaid	42.43%	18,821
Medicare	28.38%	12,590
Commercial	25.17%	11,168
Free Care & Other	4.02%	1,783
Race		**44,716**
Black / African American	67.48%	30,173
Non-Black	32.52%	14,543
Gender		**44,716**
Male	52.41%	23,434
Female	47.59%	21,281
Unspecified	0.00%	1

Note. $N = 44,716$

Figure EC.4 Histogram of boarding time in minutes

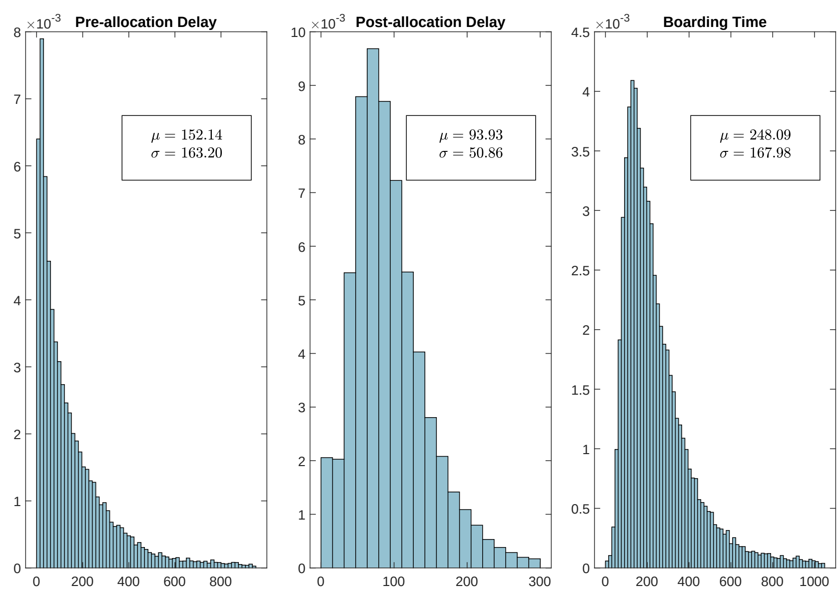


Table EC.29 Covariate balance between batched and unbatched admissions

Variable	Batched	Unbatched	Standardized Difference
Occupancy Level	85.71	85.09	0.07
Day of week			
Sunday	0.12	0.13	0.03
Monday	0.15	0.16	0.02
Tuesday	0.15	0.16	0.02
Wednesday	0.15	0.12	0.06
Thursday	0.15	0.15	0.00
Friday	0.16	0.15	0.01
Saturday	0.12	0.13	0.01
Year			
2016	0.29	0.28	0.01
2017	0.47	0.48	0.02
2018	0.25	0.23	0.02
Week			
1	0.02	0.02	0.01
2	0.02	0.02	0.01
3	0.02	0.02	0.00
4	0.02	0.02	0.01
5	0.02	0.02	0.01
6	0.02	0.02	0.00
7	0.02	0.02	0.01
8	0.02	0.02	0.01
9	0.02	0.02	0.03
10	0.02	0.02	0.01
11	0.02	0.01	0.02
12	0.02	0.02	0.00
13	0.02	0.02	0.01
14	0.02	0.02	0.01
15	0.01	0.01	0.01
16	0.01	0.01	0.01
17	0.01	0.02	0.01
18	0.02	0.02	0.00
19	0.02	0.01	0.03
20	0.02	0.02	0.02
21	0.02	0.02	0.01
22	0.02	0.02	0.01
23	0.03	0.03	0.01
24	0.03	0.03	0.01
25	0.03	0.03	0.01
26	0.03	0.03	0.01
27	0.02	0.03	0.02
28	0.03	0.02	0.02
29	0.03	0.03	0.00
30	0.03	0.02	0.01
31	0.02	0.02	0.01
32	0.02	0.02	0.01
33	0.02	0.01	0.04

Continued on next page

Note. Calculated for admissions to medical units

Table EC.29 Covariate balance between batched and unbatched admissions (Cont.)

Variable	Batched	Unbatched	Standardized Difference
34	0.02	0.02	0.01
35	0.02	0.02	0.00
36	0.02	0.02	0.02
37	0.02	0.02	0.03
38	0.02	0.02	0.00
39	0.02	0.02	0.01
40	0.02	0.02	0.01
41	0.02	0.02	0.03
42	0.02	0.02	0.00
43	0.02	0.02	0.01
44	0.02	0.01	0.02
45	0.02	0.02	0.02
46	0.02	0.02	0.01
47	0.02	0.02	0.01
48	0.02	0.02	0.04
49	0.02	0.02	0.01
50	0.02	0.02	0.01
51	0.02	0.02	0.01
52	0.02	0.02	0.00
Admitting Department ID			
Dept.1	0.15	0.16	0.02
Dept.2	0.07	0.07	0.00
Dept.3	0.09	0.09	0.00
Dept.4	0.10	0.09	0.01
Dept.5	0.04	0.04	0.00
Dept.6	0.20	0.19	0.02
Dept.7	0.21	0.21	0.01
Dept.8	0.11	0.11	0.00
Dept.9	0.03	0.03	0.01
Number of admissions in past 6 hours	1.23	1.28	0.02
AGE	58.46	59.16	0.03
ESI Score			
1	0.01	0.01	0.01
2	0.48	0.50	0.02
3	0.48	0.48	0.01
4	0.02	0.01	0.04
5	0.00	0.00	0.01
Number of ED tests	3.60	3.63	0.02
Hospital Service			
Cardiac ICU	0.00	0.00	0.02
Cardiology	0.06	0.06	0.01
Cardiology; General	0.00	0.00	0.01
Cardiothoracic Surgery	0.00	0.00	0.01
Dental	0.00	0.00	0.01
Emergency Medicine	0.00	0.00	0.00
Family Medicine	0.18	0.20	0.03
General Medicine	0.41	0.39	0.02

Continued on next page

Note. Calculated for admissions to medical units

Table EC.29 Covariate balance between batched and unbatched admissions (Cont.)

Variable	Batched	Unbatched	Standardized Difference
General Surgery	0.04	0.04	0.00
Geriatrics	0.07	0.07	0.01
Gynecology	0.00	0.00	0.00
Hematology/Oncology	0.04	0.04	0.00
Observation	0.01	0.01	0.02
Infectious Disease	0.05	0.04	0.02
Internal Medicine	0.02	0.02	0.00
Medical ICU	0.00	0.00	0.01
Neurology	0.04	0.04	0.01
Neurosurgery	0.00	0.00	0.01
Orthopedic Surgery	0.00	0.00	0.01
Orthopedics	0.01	0.00	0.01
Otolaryngology/ENT	0.00	0.00	0.01
Pulmonology	0.00	0.00	0.02
Renal	0.05	0.05	0.00
Stroke Neurology	0.00	0.00	0.00
Trauma	0.00	0.00	0.00
Urology	0.00	0.00	0.01
Vascular Surgery	0.00	0.00	0.02
Visit class			
Inpatient	0.65	0.63	0.03
Observation	0.35	0.37	0.03
Insurance class			
Commercial	0.24	0.25	0.01
Free Care & Other	0.04	0.03	0.01
Medicaid	0.41	0.42	0.01
Medicare	0.31	0.30	0.01
Race			
Black	0.32	0.32	0.00
Non-Black	0.68	0.68	0.00
Male	1.53	1.53	0.01
Charlson Comorbidity Index	3.05	3.07	0.00
Number of ICU Visits	0.05	0.04	0.01
Physician ID			
1	0.03	0.03	0.00
2	0.03	0.03	0.01
3	0.03	0.03	0.02
4	0.00	0.00	0.01
5	0.02	0.02	0.00
6	0.00	0.00	0.01
7	0.00	0.00	0.04
8	0.02	0.02	0.00
9	0.01	0.01	0.01
10	0.01	0.02	0.02
11	0.00	0.00	0.03
12	0.04	0.03	0.06
13	0.00	0.00	0.03

Continued on next page

Note. Calculated for admissions to medical units

Table EC.29 Covariate balance between batched and unbatched admissions (Cont.)

Variable	Batched	Unbatched	Standardized Difference
14	0.04	0.05	0.04
15	0.00	0.00	0.01
16	0.03	0.03	0.01
17	0.00	0.00	0.05
18	0.02	0.01	0.03
19	0.02	0.03	0.01
20	0.00	0.00	0.00
21	0.02	0.02	0.00
22	0.01	0.01	0.00
23	0.01	0.01	0.02
24	0.03	0.03	0.00
25	0.03	0.03	0.00
26	0.03	0.03	0.00
27	0.02	0.03	0.01
28	0.01	0.01	0.02
29	0.01	0.01	0.02
30	0.03	0.03	0.01
31	0.03	0.03	0.02
32	0.03	0.03	0.00
33	0.04	0.03	0.01
34	0.02	0.02	0.03
35	0.04	0.03	0.01
36	0.04	0.03	0.01
37	0.00	0.00	0.01
38	0.02	0.02	0.00
39	0.03	0.04	0.02
40	0.04	0.03	0.01
41	0.00	0.00	0.03
42	0.01	0.02	0.02
43	0.04	0.04	0.02
44	0.02	0.02	0.04
45	0.03	0.04	0.02
46	0.01	0.01	0.02
47	0.03	0.03	0.01
48	0.03	0.02	0.02
49	0.02	0.02	0.01
50	0.02	0.02	0.01
51	0.00	0.00	0.01
Number of Observations	20,864	4,701	

Note. Calculated for admissions to medical units

EC.4. Complete Regression Results

Table EC.30 Complete regression results for equation 1, used to test hypothesis 1

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(1)	(2)	(3)
TLS	-0.1543**	-0.2060**	0.0889**
	(0.0093)	(0.0183)	(0.0104)
Hour			
1		1.2463**	0.1985
		(0.2607)	(0.1784)
2		0.9672**	0.1291
		(0.2319)	(0.1856)
3		0.7746**	0.2018
		(0.2121)	(0.1713)
4		0.5516$^{\hat{}}$	0.2977
		(0.3084)	(0.2284)
5		0.3660	0.3253$^{\hat{}}$
		(0.2600)	(0.1874)
6		0.2466	0.4739*
		(0.2613)	(0.1857)
7		-0.0561	0.1377
		(0.3552)	(0.1867)
8		-0.2596	-0.4993
		(0.4567)	(0.3578)
9		0.2981	-0.2991
		(0.3123)	(0.2167)
10		0.4808	-0.2060
		(0.3008)	(0.2105)
11		0.6620*	0.2239
		(0.2794)	(0.1730)
12		0.7238**	0.5602**
		(0.2616)	(0.1623)
13		0.7050**	0.7972**
		(0.2567)	(0.1526)
14		0.5534*	0.9209**
		(0.2682)	(0.1548)
15		0.2022	0.6039**
		(0.2589)	(0.1533)
16		0.4187$^{\hat{}}$	0.4392**
		(0.2203)	(0.1372)
17		0.8551**	0.2385
		(0.2490)	(0.1597)
18		0.9774**	0.4176**
		(0.2087)	(0.1545)
19		0.8755**	0.5834**
		(0.2108)	(0.1516)
20		0.6663**	0.6093**
		(0.1947)	(0.1179)
21		0.5346**	0.7928**
		(0.1999)	(0.1443)
22		0.3945*	0.9477**
		(0.1950)	(0.1258)
23		0.2570	0.8532**
		(0.1960)	(0.1361)

Continued on next page

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\hat{}}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Table EC.30 Complete regression results for equation 1, used to test hypothesis 1 (Cont.)

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(1)	(2)	(3)
Day of Week			
Monday		-0.1864*	-0.0862
		(0.0808)	(0.0743)
Tuesday		-0.1848*	-0.0842
		(0.0743)	(0.0761)
Wednesday		-0.3602**	-0.2559**
		(0.0976)	(0.0759)
Thursday		-0.1446^	-0.0820
		(0.0759)	(0.0664)
Friday		-0.1195	-0.0571
		(0.0796)	(0.0746)
Saturday		-0.1226	-0.0178
		(0.0856)	(0.0780)
Year			
2017		-0.0392	-0.0127
		(0.0591)	(0.0624)
2018		-0.1726*	-0.1254
		(0.0781)	(0.0777)
Week			
2		-0.0592	-0.0542
		(0.1582)	(0.1430)
3		-0.0161	-0.0445
		(0.1751)	(0.1542)
4		-0.1866	-0.1748
		(0.1710)	(0.1621)
5		-0.2103	-0.1225
		(0.1988)	(0.1703)
6		-0.1420	-0.0684
		(0.1811)	(0.1400)
7		-0.1315	-0.1453
		(0.1930)	(0.1536)
8		-0.0792	-0.0899
		(0.1814)	(0.1586)
9		0.0485	0.0574
		(0.1615)	(0.1565)
10		0.0444	0.0600
		(0.1792)	(0.1597)
11		-0.3597^	-0.2793
		(0.1990)	(0.1883)
12		-0.2226	-0.1446
		(0.1715)	(0.1583)
13		-0.2016	-0.1591
		(0.1910)	(0.1826)
14		-0.3329^	-0.1982
		(0.1838)	(0.1620)
15		-0.1847	-0.1925
		(0.1766)	(0.1741)
16		-0.2179	-0.1916
		(0.2134)	(0.1669)
17		-0.1025	-0.1069
		(0.2015)	(0.1709)
18		-0.0980	-0.0965
		(0.1992)	(0.1684)
19		-0.4215*	-0.4831**
		(0.2065)	(0.1819)

Continued on next page

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Table EC.30 Complete regression results for equation 1, used to test hypothesis 1 (Cont.)

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(1)	(2)	(3)
20		0.1898	0.0533
		(0.1514)	(0.1642)
21		0.0688	0.0273
		(0.1829)	(0.1713)
22		-0.0632	-0.1018
		(0.1831)	(0.1474)
23		0.0136	-0.0699
		(0.1942)	(0.1597)
24		-0.2726	-0.3071$^{\wedge}$
		(0.1822)	(0.1673)
25		-0.1658	-0.2457
		(0.1826)	(0.1787)
26		-0.2703*	-0.1757
		(0.1362)	(0.1361)
27		0.0206	0.0265
		(0.1638)	(0.1445)
28		-0.2121	-0.2468$^{\wedge}$
		(0.1678)	(0.1449)
29		0.0158	-0.0191
		(0.1462)	(0.1539)
30		-0.3794**	-0.2821*
		(0.1400)	(0.1312)
31		-0.1452	-0.1638
		(0.1592)	(0.1276)
32		-0.4184*	-0.2904$^{\wedge}$
		(0.2105)	(0.1658)
33		-0.6168**	-0.5677**
		(0.2137)	(0.1802)
34		-0.1753	-0.2204
		(0.1937)	(0.1701)
35		0.0362	0.0240
		(0.1717)	(0.1521)
36		-0.0175	-0.0062
		(0.1538)	(0.1285)
37		0.1409	0.1716
		(0.1900)	(0.1549)
38		-0.1175	-0.1004
		(0.2215)	(0.2007)
39		0.0727	0.0818
		(0.1834)	(0.1766)
40		-0.2820	-0.2837$^{\wedge}$
		(0.1857)	(0.1705)
41		0.0048	0.1070
		(0.1418)	(0.1543)
42		-0.2864	-0.1973
		(0.2181)	(0.1964)
43		-0.2016	-0.1508
		(0.1998)	(0.1702)
44		-0.2283	-0.3200$^{\wedge}$
		(0.1927)	(0.1784)
45		-0.3163	-0.2675
		(0.2059)	(0.1782)
46		-0.1484	-0.2280
		(0.1825)	(0.1741)

Continued on next page

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Table EC.30 Complete regression results for equation 1, used to test hypothesis 1 (Cont.)

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(1)	(2)	(3)
47		-0.2838	-0.2337
		(0.2176)	(0.1728)
48		0.1455	0.1055
		(0.1729)	(0.1555)
49		-0.2319	-0.1422
		(0.2001)	(0.1665)
50		-0.3388	-0.1216
		(0.2254)	(0.1753)
51		-0.1604	-0.1002
		(0.1888)	(0.1843)
52		-0.0452	-0.1105
		(0.1867)	(0.1659)
Department ID			
Dept. 2		-0.0165	-0.0263
		(0.1104)	(0.0966)
Dept. 3		-0.0007	-0.0670
		(0.0849)	(0.0754)
Dept. 4		-0.0647	-0.0731
		(0.0888)	(0.0809)
Dept. 5		$0.1605^{\wedge}$	0.1276
		(0.0943)	(0.0859)
Dept. 6		-0.2067^{*}	-0.2177^{*}
		(0.0992)	(0.0880)
Dept. 7		0.0258	0.0074
		(0.1062)	(0.0927)
Dept. 12		0.0951	0.0905
		(0.1233)	(0.1152)
Dept. 13		-0.0136	0.0618
		(0.1313)	(0.1222)
Dept. 14		-0.1646	-0.1714
		(0.1548)	(0.1265)
Dept. 16		-0.0905	-0.0257
		(0.1858)	(0.1460)
Dept. 17		0.0694	$0.1927^{\wedge}$
		(0.1120)	(0.1026)
Dept. 18		-0.2859	-0.1055
		(0.2516)	(0.1909)
Dept. 19		0.0088	0.0466
		(0.1487)	(0.1253)
Dept. 23		0.0258	0.1597
		(0.2362)	(0.2199)
Dept. 24		-0.0033	-0.0697
		(0.1834)	(0.1625)
Dept. 25		0.0110	-0.0314
		(0.1472)	(0.1236)
Dept. 26		0.1287	0.0980
		(0.1879)	(0.1490)
Dept. 27		-0.1740	-0.1653
		(0.1696)	(0.1554)
Dept. 28		0.0911	0.0711
		(0.0766)	(0.0692)

Continued on next page

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Table EC.30 Complete regression results for equation 1, used to test hypothesis 1 (Cont.)

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(1)	(2)	(3)
Dept. 29		0.0695	0.1124
		(0.0802)	(0.0727)
Dept. 30		0.0573	0.1314
		(0.0876)	(0.0818)
Dept. 31		0.0118	0.0739
		(0.0934)	(0.0803)
Dept. 36		0.0279	0.0657
		(0.1095)	(0.0861)
Dept. 40		0.1007	0.1634
		(0.1425)	(0.1235)
N_{6h}		0.0079	0.0106
		(0.0126)	(0.0102)
AGE		0.0029**	0.0021*
		(0.0011)	(0.0009)
ESI Score			
2		0.0785	0.0586
		(0.0754)	(0.0624)
3		0.0443	0.0240
		(0.0824)	(0.0701)
4		-0.0721	-0.1585
		(0.1784)	(0.1396)
5		0.2335	-0.1127
		(0.6803)	(0.6367)
Number of ED tests		0.0370**	0.0316**
		(0.0141)	(0.0109)
Hospital Service			
Bariatric Surgery		-0.1852	0.2541
		(1.0045)	(0.7841)
CHF		0.8384	0.7922
		(0.9321)	(0.9483)
Cardiac ICU		0.4366	0.4171
		(0.3400)	(0.3091)
Cardiology		0.5841^	0.5885*
		(0.2994)	(0.2575)
Cardiology; General		0.8302*	0.7820*
		(0.3758)	(0.3181)
Cardiothoracic Surgery		0.0999	0.0969
		(0.3802)	(0.2984)
Critical Care Medicine		0.2162	-0.0323
		(0.7582)	(0.7206)
Critical Care Surgery		0.0000	-0.7656
		(.)	(1.2645)
Dental		0.8470**	0.6536*
		(0.3051)	(0.2767)
Emergency Medicine		0.6819	0.3665
		(0.4668)	(0.3947)
Family Medicine		0.6347*	0.6208**
		(0.2639)	(0.2410)
General Medicine		0.5133^	0.4469^
		(0.2682)	(0.2494)
General Surgery		0.6978**	0.5861*
		(0.2553)	(0.2424)

Continued on next page

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Table EC.30 Complete regression results for equation 1, used to test hypothesis 1 (Cont.)

Dep. Variable:		$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$	
	(1)	(2)	(3)
Geriatrics		0.5168^	0.4146
		(0.2870)	(0.2560)
Gynecology		0.8486**	0.6628*
		(0.2964)	(0.2772)
Hematology/Oncology		0.6180*	0.5341*
		(0.2941)	(0.2568)
Hospice		0.7933	0.5093
		(0.8419)	(0.9093)
Hospitalist		0.4965^	0.4149^
		(0.2774)	(0.2492)
Infectious Disease		0.3994	0.3189
		(0.2851)	(0.2609)
Internal Medicine		0.6431*	0.6352*
		(0.2869)	(0.2519)
Maternal-Fetal Medicine		1.7383**	1.6031**
		(0.5088)	(0.4237)
Medical ICU		0.5903^	0.4432
		(0.3078)	(0.2788)
Neuro ICU		0.8720	0.7034
		(0.8414)	(0.7600)
Neurology		0.4546^	0.4305^
		(0.2730)	(0.2428)
Neurosurgery		0.3100	0.4401^
		(0.3177)	(0.2383)
OMFS		1.5546*	1.2206*
		(0.7516)	(0.5400)
Obstetrics		0.0000	0.2244
		(.)	(1.1226)
Orthopedic Surgery		-1.0531	-0.3786
		(1.0763)	(0.7677)
Orthopedics		0.5385	0.4996
		(0.3423)	(0.3210)
Otolaryngology/ENT		0.4409	0.3502
		(0.4151)	(0.3772)
Plastic Surgery		0.7233	0.4249
		(0.8206)	(0.7747)
Pulmonology		0.8279**	0.7407*
		(0.3145)	(0.2926)
Renal		0.6058*	0.5359*
		(0.2841)	(0.2586)
Stroke Neurology		0.5772	0.4297
		(0.4970)	(0.4410)
Surgical Oncology		0.9189	0.3025
		(0.7896)	(0.8159)
Transplant		0.6150	0.8073^
		(0.5950)	(0.4550)
Trauma		0.5550	0.5715^
		(0.3801)	(0.3265)
Urology		0.3228	0.3613
		(0.3316)	(0.2400)
Visit class			
Emergency		0.9747	1.8124
		(2.0928)	(2.4969)
Inpatient		-0.3989	0.2030
		(0.5958)	(0.4856)

Continued on next page

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Table EC.30 Complete regression results for equation 1, used to test hypothesis 1 (Cont.)

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(1)	(2)	(3)
Observation		-0.3133	0.3073
		(0.5890)	(0.4806)
Charlson Comorbidity Index		-0.0040	-0.0002
		(0.0065)	(0.0055)
Num. ICU visits		0.0603	0.0385
		(0.0445)	(0.0426)
Insurance class			
Commercial		-0.0211	-0.0122
		(0.0396)	(0.0340)
Free Care & Other		-0.0789	-0.0778
		(0.0819)	(0.0724)
Medicaid		0.0501	$0.0587^{\wedge}$
		(0.0386)	(0.0350)
Race			
Black		0.0090	0.0115
		(0.0309)	(0.0265)
Male		0.0541	$0.0508^{\wedge}$
		(0.0330)	(0.0297)
Patient order		0.0123^{**}	0.0251^{**}
		(0.0027)	(0.0028)
Shift length		-0.2673^{**}	-0.3385^{**}
		(0.0973)	(0.0586)
Hours b/w shifts		-0.0002	-0.0003
		(0.0002)	(0.0002)
Hours to next shift		-0.0000	-0.0001
		(0.0002)	(0.0002)
Shift time			
7am – 3pm		0.5964^{**}	0.5023^{**}
		(0.1700)	(0.1432)
3pm – 11pm		0.9361^{**}	0.6041^{**}
		(0.1340)	(0.1077)
11pm – 7am		0.9413^{**}	0.7567^{**}
		(0.2588)	(0.1769)
8am – 4pm		0.4052^{*}	$0.2731^{\wedge}$
		(0.1625)	(0.1464)
4pm – 12am		0.8430^{**}	0.3596^{**}
		(0.1594)	(0.1247)
8am – 8pm		1.3682^{**}	1.7676^{**}
		(0.2849)	(0.5649)
N_{roomed}		0.0040^{*}	0.0021
		(0.0016)	(0.0015)
Physician ID			
2		0.0549^{**}	0.2001^{**}
		(0.0191)	(0.0167)
3		0.1861^{**}	0.1960^{**}
		(0.0196)	(0.0165)
4		1.7484^{**}	0.5578^{*}
		(0.3069)	(0.2689)
5		-0.0628^{*}	$0.0460^{\wedge}$
		(0.0285)	(0.0260)
6		2.0896^{**}	1.3054^{**}
		(0.1196)	(0.1004)
8		0.2589^{**}	0.2084^{**}
		(0.0295)	(0.0221)

Continued on next page

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Table EC.30 Complete regression results for equation 1, used to test hypothesis 1 (Cont.)

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(1)	(2)	(3)
10		0.2460^{**}	0.2759^{**}
		(0.0571)	(0.0510)
11		0.6118^{**}	0.5808^{**}
		(0.0505)	(0.0475)
14		-0.2585^{**}	-0.0620
		(0.0832)	(0.0727)
16		0.3313^{**}	0.3577^{**}
		(0.0226)	(0.0213)
17		$0.1817^{\hat{}}$	0.5334^{**}
		(0.0950)	(0.1007)
18		0.2128^{**}	0.1865^{**}
		(0.0216)	(0.0172)
20		-0.1606^{**}	-0.1164^{**}
		(0.0313)	(0.0285)
23		0.2319^{**}	0.2563^{**}
		(0.0187)	(0.0163)
24		0.5902^{**}	0.7754^{**}
		(0.2228)	(0.2013)
25		-0.0209	-0.0075
		(0.0305)	(0.0232)
26		0.1640^{*}	0.2050^{**}
		(0.0695)	(0.0653)
28		-0.2725^{**}	-0.1125^{**}
		(0.0360)	(0.0359)
30		0.0418^{*}	$0.0332^{\hat{}}$
		(0.0202)	(0.0177)
31		0.3308	0.2063
		(0.2154)	(0.1823)
33		-0.0846^{**}	0.0698^{**}
		(0.0205)	(0.0179)
34		0.0343	0.1127^{**}
		(0.0214)	(0.0185)
35		0.1587^{**}	0.3020^{**}
		(0.0317)	(0.0289)
37		-0.2690^{**}	0.0130
		(0.0638)	(0.0568)
38		0.0569	0.2610^{**}
		(0.0891)	(0.0767)
39		0.1241^{**}	0.1994^{**}
		(0.0257)	(0.0223)
41		0.1720^{**}	0.2474^{**}
		(0.0147)	(0.0142)
42		-0.0566^{**}	0.1048^{**}
		(0.0195)	(0.0140)
43		0.0088	0.2086^{**}
		(0.0756)	(0.0638)
44		0.1755	-0.2741
		(0.2282)	(0.2344)
45		-0.2202^{**}	-0.0959^{**}
		(0.0254)	(0.0212)
46		-0.0563^{**}	0.0666^{**}
		(0.0175)	(0.0141)
47		0.0031	0.0770^{**}
		(0.0161)	(0.0148)

Continued on next page

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\hat{}}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Table EC.30 Complete regression results for equation 1, used to test hypothesis 1 (Cont.)

Dep. Variable:	$ln(\frac{Prob(BCH)}{1-Prob(BCH)})$		
	(1)	(2)	(3)
48		-0.4449**	0.0197
		(0.0824)	(0.0743)
49		-0.1100**	0.1197**
		(0.0298)	(0.0284)
50		0.0854**	0.1607**
		(0.0206)	(0.0176)
51		-0.0159	-0.0099
		(0.0161)	(0.0135)
52		0.5347	0.2776
		(0.4157)	(0.3524)
53		0.4096**	0.4048**
		(0.0507)	(0.0428)
54		0.3406**	0.3740**
		(0.0227)	(0.0202)
55		-0.3417**	-0.1871**
		(0.0197)	(0.0160)
57		0.2629**	0.3044**
		(0.0180)	(0.0155)
58		0.0422	0.0988
		(0.0683)	(0.0634)
59		0.0869**	0.1364**
		(0.0196)	(0.0193)
60		-0.1241**	-0.1172**
		(0.0211)	(0.0156)
61		0.1031**	0.2697**
		(0.0266)	(0.0230)
63		0.1872**	0.1545**
		(0.0290)	(0.0280)
66		0.6413**	0.7386**
		(0.0593)	(0.0547)
Waiting Room Census		-0.0003	0.0001
		(0.0004)	(0.0004)
constant	-0.9545**	-0.5113	-0.7238
	(0.0396)	(1.6647)	(1.2311)
N	31,563	30,466	42,034
R^2	0.0146	0.0423	0.0374

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. Slight drop in N is because we removed observations with missing values in controls.

Table EC.31 Complete regression results for equation 2, used to test hypothesis 2

Dep. Var	ln(Pre-allocation Delay)		
	Coefficient	Robust Std. Err.	$P>\|t\|$
Batched	0.0455124	0.0168445	0.007
Occupancy Level	0.0375958	0.0015731	0.000
Day of Week			
Monday	0.4050379	0.0260696	0.000
Tuesday	0.5468499	0.0262246	0.000
Wednesday	0.5559771	0.0261306	0.000
Thursday	0.5225916	0.0264449	0.000
Friday	0.4261117	0.0260542	0.000
Saturday	0.1196312	0.0274652	0.000
Year			
2017	0.2337013	0.0180774	0.000
2018	0.2753042	0.0258205	0.000
Week			
2	0.5171903	0.0690266	0.000
3	0.2591884	0.0741493	0.000
4	0.4697640	0.0719630	0.000
5	0.3795603	0.0706094	0.000
6	0.3621223	0.0712532	0.000
7	0.4092073	0.0706700	0.000
8	0.4655966	0.0685935	0.000
9	0.3684871	0.0719108	0.000
10	0.3226328	0.0695479	0.000
11	0.1560167	0.0734296	0.034
12	0.4221708	0.0744392	0.000
13	0.2941115	0.0724340	0.000
14	0.3606874	0.0720997	0.000
15	0.4117799	0.0791758	0.000
16	0.4363444	0.0738206	0.000
17	0.3689570	0.0784630	0.000
18	0.8341249	0.0802166	0.000
19	0.5822762	0.0722076	0.000
20	0.2299459	0.0687625	0.001
21	0.1486223	0.0697309	0.033
22	0.1948489	0.0696075	0.005
23	0.2150848	0.0653020	0.001
24	0.1432892	0.0655662	0.029
25	0.1795884	0.0646749	0.005
26	0.1042187	0.0660294	0.114
27	0.0865263	0.0668576	0.196
28	0.3549980	0.0683001	0.000
29	0.3980474	0.0680497	0.000
30	0.2864453	0.0698006	0.000
31	0.0344705	0.0696865	0.621
32	0.1700022	0.0691368	0.014
33	0.4006781	0.0717460	0.000
34	0.2532105	0.0731904	0.001
35	0.2777397	0.0693387	0.000
36	0.1809084	0.0707244	0.011
37	0.3114307	0.0696271	0.000
38	0.5710913	0.0699150	0.000
39	0.3656532	0.0706912	0.000
40	0.2893295	0.0701674	0.000
41	0.3893246	0.0716134	0.000
42	0.3303540	0.0717387	0.000
43	0.5383705	0.0742510	0.000
44	0.2664092	0.0715232	0.000
45	0.0706203	0.0727157	0.331
46	0.1701430	0.0717117	0.018

Continued on next page

Table EC.31 Complete regression results for equation 2, used to test hypothesis 2 (Cont.)

Dep. Var	ln(Pre-allocation Delay)		
	Coefficient	Robust Std. Err.	$P > \|t\|$
47	0.0943117	0.0737511	0.201
48	0.1934601	0.0717965	0.007
49	0.3110722	0.0742228	0.000
50	0.3247829	0.0725798	0.000
51	0.5688549	0.0717654	0.000
52	0.2167284	0.0694747	0.002
Department ID			
Dept. 1	0.0359207	0.0540696	0.506
Dept. 2	0.1142687	0.0572238	0.046
Dept. 3	0.0848557	0.0572394	0.138
Dept. 4	0.1499802	0.0542510	0.006
Dept. 12	-0.2387147	0.0562848	0.000
Dept. 20	-0.0241518	0.0464875	0.603
Dept. 21	-0.0336879	0.0463571	0.467
Dept. 22	-0.1134339	0.0480944	0.018
Num. of Admissions in past 6 hours	-0.0201715	0.005957	0.001
Age	0.0002979	0.0004770	0.532
ESI Score			
2	-0.0349296	0.0735937	0.635
3	-0.0323837	0.0739038	0.661
4	-0.0069392	0.0855733	0.935
5	-0.5166008	0.2375281	0.030
Num. of ED tests	0.0148143	.0054292	0.006
Hospital Service			
Cardiac ICU	-0.1103303	0.1756113	0.530
Cardiology	-0.3914569	0.1312502	0.003
Cardiology; General	-0.3866919	0.1761999	0.028
Cardiothoracic Surgery	-0.0850190	0.1927435	0.659
Dental	0.0231749	0.2150668	0.914
Emergency Medicine	0.4838410	0.2025703	0.017
Family Medicine	0.0200381	0.1278584	0.875
General Medicine	0.1852439	0.1271949	0.145
General Surgery	-0.1686033	0.1329164	0.205
Geriatrics	0.0532234	0.1302948	0.683
Gynecology	-0.0514182	0.2070660	0.804
Hematology/Oncology	-0.4046328	0.1323596	0.002
Observation	0.0329188	0.1503661	0.827
Infectious Disease	0.1173184	0.1315358	0.372
Internal Medicine	0.0547566	0.1368525	0.689
Medical ICU	0.1031926	0.1732824	0.552
Neurology	-0.4298684	0.1337625	0.001
Neurosurgery	-0.1591677	0.2277216	0.485
Orthopedic Surgery	-0.5662670	0.2459593	0.021
Orthopedics	-0.5009076	0.1622067	0.002
Otolaryngology/ENT	0.4413316	0.2396370	0.066
Pulmonology	0.0197978	0.2076717	0.924
Renal	-0.2568958	0.1304578	0.049
Stroke Neurology	-0.1281449	0.2230440	0.566
Trauma	-0.1427615	0.2608363	0.584
Urology	0.0103598	0.2174762	0.962
Visit Class Name			
Inpatient	0.0510353	0.0141169	0.000
Charlson Comorbidity Index	0.0030246	0.0025537	0.236
Number of ICU Visits	-0.0395745	0.0290723	0.173
Insurance Class Name			
Commercial	-0.0114853	0.0180586	0.525

Continued on next page

Table EC.31 Complete regression results for equation 2, used to test hypothesis 2 (Cont.)

Dep. Var	ln(Pre-allocation Delay)		
	Coefficient	Robust Std. Err.	$P>\|t\|$
Free Care & Other	-0.0794205	0.0372689	0.033
Medicaid	0.0115673	0.0175949	0.511
Black	-0.0281309	0.0141837	0.047
Male	-0.0100987	0.0134276	0.452
Physician ID			
1	0.1395322	0.0524224	0.008
2	0.1317901	0.0545735	0.016
3	0.2618873	0.2732732	0.338
4	0.1676608	0.0581254	0.004
5	0.295919	0.1854214	0.111
6	0.3181683	0.1757103	0.070
7	0.1213606	0.0598760	0.043
8	0.0176113	0.0780007	0.821
9	0.1614573	0.0696570	0.020
10	0.5421596	0.2034469	0.008
11	0.0436492	0.0548474	0.426
12	0.3015158	0.1514488	0.047
13	0.1174845	0.0485626	0.016
14	-0.0455473	0.2132563	0.831
15	0.0369851	0.0543032	0.496
16	0.0964749	0.1509075	0.523
17	0.0934891	0.0605623	0.123
18	0.1054678	0.0549612	0.055
19	0.1795399	0.1353920	0.185
20	-0.0082009	0.0575204	0.887
21	0.1870359	0.0648595	0.004
22	0.0241390	0.0753402	0.749
23	0.0976460	0.0525518	0.063
24	0.0666713	0.0527779	0.207
25	0.1255183	0.0544510	0.021
26	0.1479217	0.0581137	0.011
27	0.1099399	0.0726607	0.130
28	0.054959	0.0914371	0.548
29	0.1137645	0.0539406	0.035
30	0.085421	0.0536121	0.111
31	0.0895213	0.0521010	0.086
32	0.1118129	0.0586950	0.057
33	0.1665473	0.0576445	0.004
34	0.1664226	0.0501332	0.001
35	0.0379938	0.0501529	0.449
36	0.3213601	0.1571181	0.041
37	0.1242991	0.0632185	0.049
38	0.1361543	0.0503702	0.007
39	0.138068	0.0500657	0.006
40	-0.3680421	0.1930742	0.057
41	0.0925708	0.0658396	0.160
42	0.1636991	0.0494803	0.001
43	0.0512053	0.0579851	0.377
44	0.0766899	0.0525695	0.145
45	0.1887383	0.0659201	0.004
46	0.1015419	0.0510074	0.047
47	0.2004422	0.0518214	0.000
48	0.1414625	0.0587760	0.016
49	0.0471828	0.0609879	0.439
50	0.3048952	0.2413666	0.207
Hour			
1	0.0071786	0.0644611	0.911
2	0.112472	0.0686990	0.102

Continued on next page

Table EC.31 Complete regression results for equation 2, used to test hypothesis 2 (Cont.)

Dep. Var	ln(Pre-allocation Delay)		
	Coefficient	Robust Std. Err.	$P > \|t\|$
3	0.1574426	0.0707528	0.026
4	0.1820054	0.0729789	0.013
5	0.5010637	0.0753983	0.000
6	0.7438274	0.0697244	0.000
7	0.8209719	0.0691315	0.000
8	0.7149861	0.0703279	0.000
9	0.7162758	0.0637654	0.000
10	0.6898545	0.0602885	0.000
11	0.683133	0.0566033	0.000
12	0.6911972	0.0553610	0.000
13	0.7058089	0.0543001	0.000
14	0.8291992	0.0526930	0.000
15	0.8687588	0.0520173	0.000
16	0.7330635	0.0541611	0.000
17	0.8185552	0.0521667	0.000
18	0.7479485	0.0515691	0.000
19	0.3485631	0.0529839	0.000
20	0.2008036	0.0538411	0.000
21	0.1297796	0.0542713	0.017
22	0.1168044	0.0539278	0.030
23	0.3387243	0.0544555	0.000
constant	-4.3985290	0.2131756	0.000

Table EC.32 Mediating role of CV in relationship between batching and pre-allocation delay

(a) Seemingly unrelated regression results

	Stage 1	Stage 2	Stage 3
Dep. Variable:	*LPAD*	*CV*	*LPAD*
Batch	0.0455^{**}	0.0713^{**}	0.0416^{*}
	(0.0168)	(0.0049)	(0.0169)
CV			0.0555^{*}
			(0.0228)
N	25,565	25,565	25,565
R^2	0.2172	0.0173	0.2174

$^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$
Standard errors shown in parentheses.

(b) 5000-Sample bootstrap results

Effect	Mean	Bias-corrected 95% CI
Indirect	0.003954	(0.00074, 0.00748)
Direct	0.041559	(0.00879, 0.07422)
Total	0.045512	(0.01299, 0.07812)

Note. $N = 25,565$

Table EC.33 Complete regression results for equation 5, used to test hypothesis 3

Dep. Variable:	Number of Patients Seen		Average Throughput Time	
	(1)	(2)	(3)	(4)
	unweighted OLS	weighted CEM-matched	unweighted OLS	weighted CEM-matched
Shift with batching	2.2957**	2.1340**	-0.0264^	-0.0431**
	(0.1092)	(0.1138)	(0.0138)	(0.0143)
Shift Length	1.8243**	2.0065**	0.2119**	0.2229**
	(0.1824)	(0.2059)	(0.0149)	(0.0195)
Physician ID				
2	-2.8986**	-2.8061**	0.3382**	0.2919**
	(0.0389)	(0.0444)	(0.0033)	(0.0039)
3	-2.3659**	-2.3656**	0.4678**	0.4362**
	(0.0567)	(0.0656)	(0.0058)	(0.0066)
4	1.2553*		0.4487**	
	(0.5225)		(0.0542)	
5	-2.7150**	-2.5731**	0.1824**	0.1825**
	(0.0717)	(0.0878)	(0.0066)	(0.0084)
6	-6.1945**	-5.9445**	1.1664**	1.1646**
	(0.1656)	(0.2523)	(0.0144)	(0.0259)
7	-13.3225**	-17.9131**	-0.7869**	-1.6447**
	(0.6942)	(0.9523)	(0.0682)	(0.0752)
8	-2.7643**	-2.2212**	0.3864**	0.3616**
	(0.0570)	(0.0729)	(0.0055)	(0.0078)
10	-0.7015*	-0.5153*	0.2546**	0.2284**
	(0.2816)	(0.2210)	(0.0164)	(0.0188)
11	0.7667**	0.9427**	0.1801**	0.1559**
	(0.1689)	(0.2567)	(0.0130)	(0.0274)
13	-9.7451**		-0.9830**	
	(0.8758)		(0.0807)	
14	-3.0498**	-3.2506**	0.3872**	0.3618**
	(0.2517)	(0.2712)	(0.0266)	(0.0294)
15	-16.8881**		-0.9075**	
	(0.6033)		(0.0574)	
16	-1.6849**	-1.6036**	0.4380**	0.3893**
	(0.0542)	(0.0533)	(0.0065)	(0.0064)
17	-1.9421**	-2.6279**	0.0925**	0.1464**
	(0.1447)	(0.1632)	(0.0130)	(0.0209)
18	0.4653**	0.5070**	0.2749**	0.2524**
	(0.0693)	(0.0631)	(0.0051)	(0.0053)
19	-7.3594**	-11.0391**	-0.6118**	-1.1223**
	(0.5697)	(1.0343)	(0.0578)	(0.0744)
20	-2.1656**	-2.3574**	0.1613**	0.0633**
	(0.0670)	(0.0911)	(0.0054)	(0.0107)
23	-1.0857**	-0.7093**	0.2281**	0.1908**
	(0.0435)	(0.0466)	(0.0026)	(0.0035)
24	-0.2019	9.3769**	0.1790**	0.6535**
	(0.5013)	(0.2569)	(0.0534)	(0.0340)
25	-1.9093**	-1.8572**	0.1944**	0.0947**
	(0.1170)	(0.1168)	(0.0076)	(0.0101)
26	-1.3881**	-0.8499**	0.3803**	0.2504**
	(0.1422)	(0.2353)	(0.0123)	(0.0247)
28	-2.1299**	-2.1725**	0.1938**	0.1395**
	(0.1016)	(0.1015)	(0.0070)	(0.0098)
29	-10.0433**		0.0829^	
	(0.5525)		(0.0447)	
30	-2.3239**	-2.6982**	0.2413**	0.2042**
	(0.0693)	(0.0706)	(0.0043)	(0.0059)

Continued on next page

Note. $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. 90 out of 7,709 physician shifts had none of the admissions occur during the shift, and were therefore removed from the sample. An additional 52 physician shifts were missing either hours to next shift or hours since last shift, and hence were eliminated from the sample. Controls included in all models.

Table EC.33 Complete regression results for equation 5, used to test hypothesis 3

Dep. Variable:	Number of Patients Seen		Average Throughput Time	
	(1)	(2)	(3)	(4)
	unweighted OLS	weighted CEM-matched	unweighted OLS	weighted CEM-matched
31	0.7798*	0.8424**	0.0799*	0.0088
	(0.2983)	(0.2995)	(0.0327)	(0.0374)
33	-1.1798**	-1.0589**	0.3761**	0.3426**
	(0.0484)	(0.0529)	(0.0039)	(0.0054)
34	-0.0731$^{\wedge}$	-0.2632**	0.0667**	0.0539**
	(0.0414)	(0.0376)	(0.0020)	(0.0034)
35	-3.3882**	-3.3752**	0.3936**	0.3406**
	(0.1152)	(0.1366)	(0.0099)	(0.0133)
37	-0.2778$^{\wedge}$	0.4599*	0.3615**	0.2743**
	(0.1408)	(0.2061)	(0.0122)	(0.0218)
38	-2.3009**	-2.5728**	0.2563**	0.2497**
	(0.2588)	(0.2896)	(0.0284)	(0.0349)
39	-1.1187**	-1.0972**	0.2191**	0.1962**
	(0.0779)	(0.0810)	(0.0041)	(0.0067)
41	-1.8247**	-1.6736**	0.1385**	0.1153**
	(0.0463)	(0.0460)	(0.0040)	(0.0047)
42	-1.6707**	-1.5411**	0.2471**	0.2159**
	(0.0515)	(0.0574)	(0.0037)	(0.0055)
43	-1.8840**	-1.9355**	0.1756**	0.1661**
	(0.2309)	(0.2439)	(0.0246)	(0.0274)
44	-12.1244**		-0.9476**	
	(0.6239)		(0.0621)	
45	-2.1663**	-2.2050**	0.1859**	0.1490**
	(0.0402)	(0.0476)	(0.0034)	(0.0051)
46	-1.7057**	-1.6322**	0.1887**	0.1546**
	(0.0329)	(0.0449)	(0.0035)	(0.0042)
47	-1.7819**	-1.6195**	0.1462**	0.1203**
	(0.0345)	(0.0464)	(0.0039)	(0.0054)
48	-0.5737**	0.3336	0.0145	-0.0495*
	(0.1213)	(0.2378)	(0.0098)	(0.0233)
49	-3.8134**	-3.9866**	0.5076**	0.4863**
	(0.0543)	(0.0544)	(0.0049)	(0.0053)
50	-0.3894**	-0.6413**	0.1988**	0.1666**
	(0.0578)	(0.0567)	(0.0057)	(0.0059)
51	-0.9483**	-1.0105**	0.1595**	0.1343**
	(0.0293)	(0.0380)	(0.0027)	(0.0039)
52	-9.7760**	-14.2864**	-0.9014**	-1.5843**
	(0.6540)	(0.9885)	(0.0596)	(0.0712)
53	0.8553**	0.9069**	0.3044**	0.2641**
	(0.1402)	(0.2071)	(0.0138)	(0.0199)
54	-3.3594**	-3.3010**	0.2033**	0.1823**
	(0.0974)	(0.1071)	(0.0088)	(0.0111)
55	-2.4659**	-2.2969**	0.2737**	0.2199**
	(0.0602)	(0.0604)	(0.0032)	(0.0057)
57	-2.1799**	-2.0768**	0.2362**	0.1969**
	(0.0758)	(0.0766)	(0.0061)	(0.0071)
58	-0.2918*	-0.5807*	0.1232**	0.0414
	(0.1397)	(0.2358)	(0.0124)	(0.0275)
59	0.3408**	0.1531**	0.0316**	0.0281**
	(0.0492)	(0.0555)	(0.0033)	(0.0047)
60	-0.6506**	-0.8393**	0.2050**	0.1939**
	(0.0320)	(0.0456)	(0.0024)	(0.0038)

Continued on next page

Note. $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. 90 out of 7,709 physician shifts had none of the admissions occur during the shift, and were therefore removed from the sample. An additional 52 physician shifts were missing either hours to next shift or hours since last shift, and hence were eliminated from the sample. Controls included in all models.

Table EC.33 Complete regression results for equation 5, used to test hypothesis 3

Dep. Variable:	Number of Patients Seen		Average Throughput Time	
	(1)	(2)	(3)	(4)
	unweighted OLS	weighted CEM-matched	unweighted OLS	weighted CEM-matched
61	-0.9954**	-1.0197**	0.4829**	0.4248**
	(0.0857)	(0.0864)	(0.0064)	(0.0082)
63	1.9006**	2.3209**	0.1096**	0.0908**
	(0.0561)	(0.0610)	(0.0042)	(0.0052)
66	-1.0041**	-1.5748**	0.0250*	-0.0125
	(0.1348)	(0.1483)	(0.0101)	(0.0121)
Number of ED tests	-1.8324**	-1.9750**	0.3029**	0.3030**
	(0.1472)	(0.1730)	(0.0164)	(0.0202)
Num. of patients roomed in shift	0.0763**	0.0868**	-0.0001	0.0008^
	(0.0058)	(0.0067)	(0.0004)	(0.0004)
Waiting Room Census	0.0028*	0.0023*	0.0000	-0.0001
	(0.0011)	(0.0011)	(0.0001)	(0.0001)
AM shift	-10.5251**	-11.4706**	-0.2781**	-0.3329**
	(0.5859)	(0.6549)	(0.0455)	(0.0508)
PM shift	-8.2957**	-8.6763**	-0.0591	-0.0697^
	(0.4831)	(0.5253)	(0.0357)	(0.0408)
Mean age	-0.0029	-0.0029	0.0036*	0.0034*
	(0.0124)	(0.0123)	(0.0014)	(0.0016)
Mean ESI Score	0.4809	0.9503^	-0.1433**	0.0086
	(0.4280)	(0.5093)	(0.0398)	(0.0498)
Mean Charlson Comorbidity Score	0.0592	0.0623	0.0160	0.0109
	(0.1144)	(0.1389)	(0.0136)	(0.0144)
Mean Num. of ICU visits	-6.0397**	-5.5455**	-1.3957**	-1.3012**
	(0.6159)	(0.6104)	(0.1069)	(0.1045)
Hours between shifts	0.0001	-0.0015	-0.0000	0.0001
	(0.0007)	(0.0011)	(0.0000)	(0.0001)
Hours to next shift	-0.0001	-0.0002	-0.0001	0.0001
	(0.0007)	(0.0013)	(0.0000)	(0.0001)
Weekend shift	5.1890**	5.4741**	0.0498*	0.0310
	(0.5638)	(0.5898)	(0.0194)	(0.0255)
constant	8.1722**	5.2937*	0.4727**	-0.0370
	(2.0768)	(2.1492)	(0.1739)	(0.1861)
N	7567.0000	7060.0000	7567.0000	7060.0000
R^2	0.3369	0.3422	0.2328	0.2005

Note. $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\wedge}p < 0.1$, robust standard errors shown in parentheses. 90 out of 7,709 physician shifts had none of the admissions occur during the shift, and were therefore removed from the sample. An additional 52 physician shifts were missing either hours to next shift or hours since last shift, and hence were eliminated from the sample. Controls included in all models.