# Gatekeeping, Fast and Slow: An Empirical Study of Referral Errors in the Emergency Department

Michael Freeman
INSEAD, 138676 Singapore, Republic of Singapore michael.freeman@insead.edu

Susan Robinson
Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, United Kingdom
susan.robinson@addenbrookes.nhs.uk

Stefan Scholtes
Judge Business School, University of Cambridge, Cambridge CB2 1AG, United Kingdom s.scholtes@jbs.cam.ac.uk

Using data from over 300,000 visits to an emergency department (ED), we study the accuracy of gatekeeping decisions – the choices that physicians make regarding patient discharge or admission to the hospital. In our study context, we focus specifically on the effectiveness of a second gatekeeping stage in the ED – a clinical decision unit (CDU). While only 9.9% of patients in our sample are routed through the CDU, we find that had the unit not been in place during the observation period, the rates of unnecessary hospitalization and wrongful patient discharge from the ED would have increased by 14.3% and 29.6%, respectively. We also find that the CDU is especially beneficial for patients with a high ex ante risk of experiencing unnecessary hospitalization, with the rate for the most high-risk patients reduced from 14.0% without the CDU to just 4.8% had all such patients been routed through the CDU. The appropriateness of referrals is therefore a key contributor to the CDU's effectiveness: We estimate that random allocation of patients in our study hospital to the CDU would have reduced the unit's effectiveness by more than half. Finally, we investigate a critical trade-off in designing a two-stage gatekeeping system: Resources must be split between the two stages, increasing congestion in the first stage when the second stage is enlarged. We demonstrate that in the study hospital, the combination of an ED and CDU performs better than a pooled system that combines the capacity of both stages to enlarge the ED but does not have a designated CDU. In fact, we estimate that in this specific case, reducing the size of the first-stage ED in order to expand CDU capacity from the current 9.9% of ED patients to 25% would further reduce unnecessary hospitalizations by up to 33%. We discuss the insights that these results provide as to the circumstances under which it may be advantageous to add a second stage to a gatekeeping system.

*Key words*: gatekeeping; congestion; referral error; healthcare: hospitals; emergency department; econometrics
*History*: March 30, 2020

## 1. Introduction

Physicians in hospital emergency departments (EDs) perform two complementary tasks. First, they provide direct patient care in order to stabilize acutely ill patients, relieve symptoms, and diagnose illnesses. Second, they act as gatekeepers of hospital beds, deciding whether a patient needs to be admitted to the hospital for further diagnosis and specialized treatment or can be safely discharged after receiving treatment in the ED.

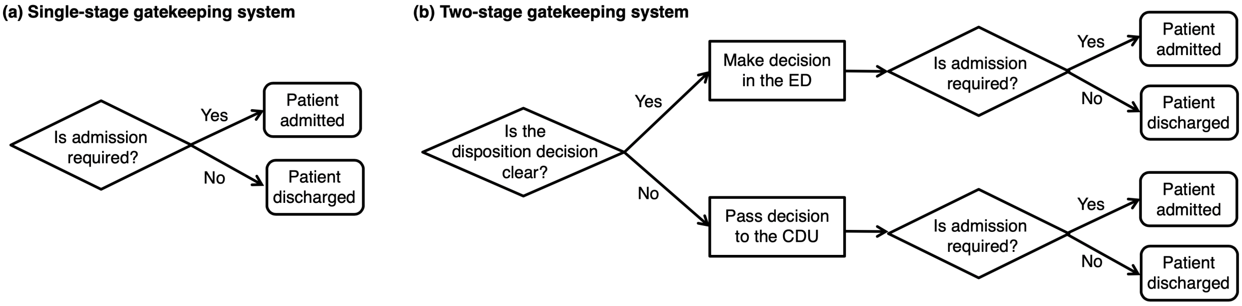
When these gatekeeping decisions are incorrect, then patients are put at risk and it is costly for the system. Wrongfully discharged patients will revisit the ED with the same complaint but in poorer health, and if they are then admitted to the hospital, they will often have a worse prognosis and need more resource-intensive interventions. On the other hand, patients who are unnecessarily admitted face the possibility of hospital-acquired infections, adverse events like falls or medication errors, and general physical and mental deterioration due to reduced mobility and an unfamiliar environment. As Inouye et al. (2008) point out, an unnecessary hospitalization *"can initiate the terminal downward spiral for an older person,"* resulting in *"delirium, falls, functional decline, institutionalization, and death."*

In fact, unnecessary admissions not only have consequences for the admitted patients, but they affect the safety and efficiency of the hospital as a whole in several ways. First, scarce resources are diverted from more vulnerable patients already in the hospital, which puts these patients at increased risk (Kuntz et al. 2015). Second, when beds are occupied unnecessarily, the hospital loses the flexibility that helps it respond to incoming patients appropriately (Song et al. 2019). Third, when a surge in emergency admissions pushes the hospital beyond its emergency bed capacity, pre-booked elective beds must be used as a buffer, leading to cancellations of elective patients and idling of expensive surgical resources (Freeman et al. 2019).

The decision-making process in the ED is complicated by the fact that patients arrive at random with a wide variety of symptom complexes, some that are life-threatening and require immediate attention and others that are non-urgent and for which treatment can be delayed without harm. Given that resources are often insufficient for all patients to be treated immediately, patients are triaged on arrival based on the severity of their pathologies, with the diagnosis and treatment of those most at risk prioritized. However, in order to ensure that lower-priority patients are still seen within a reasonable amount of time, regulators, payers, and hospital managers implement various schemes (e.g., incentives, time targets, publication of waiting time performance, online real-time waiting time displays) to prevent excessive wait times. While these schemes are designed to ensure that all patients are seen promptly, the time pressure this creates can force gatekeepers to make decisions with less information and hence make them more prone to error.

Some hospitals have addressed this problem by replacing the traditional single-stage gatekeeping process – in which all patients must be either admitted or discharged after assessment and treatment in the ED – with a two-stage approach, as illustrated in Figure 1. Under the two-stage approach, after stabilizing a patient and carrying out diagnostic tests, the ED physician makes a first-stage gatekeeping decision. If she is confident that she can make an accurate admission or discharge decision – and can do so within the normal time frame for an ED visit (up to four hours in our context) – then the normal ED process continues and the physician makes the final gatekeeping decision in the ED. If, on the other hand, the physician is not confident in making an accurate and

**Figure 1** Flow charts of the traditional single-stage gatekeeping process, i.e., a fast-only ED (left), and the proposed two-stage gatekeeping process, i.e., a fast-and-slow ED (right).



timely gatekeeping decision herself, then she can instead route the patient into a clinical decision unit (CDU). Once transferred into the CDU, the patient is cared for by an ED team dedicated to this unit and may stay for an extended period, typically up to 24 hours, during which they may receive further diagnostic tests and therapies before the final gatekeeping decision is made. Much like a hospital admission, transferring a patient to the CDU moves them "off the clock;" they are no longer subject to the same time pressures or targets as patients in the fast ED.

Unlike triage, which occurs upon the patient's arrival, this routing decision comes *after* the traditional ED assessment and treatment process. In effect, the two-stage gatekeeping system splits the ED into two different functions: a "fast ED," in which rapid decisions are made for those patients who clearly require admission to the hospital (e.g., cardiac arrest, stroke, hip fracture) or can be safely discharged (minor wounds, sprains, burns), and a "slow ED" – the CDU – where patients with symptom complexes and diagnoses that are more difficult to diagnose or resolve (e.g., undiagnosed chest pain, asymptomatic head trauma) can be transferred for further investigation and observation prior to the gatekeeping decision (see Table 1). Note that all patients pass through the fast ED, while some fraction will pass through both the fast and the slow EDs. In effect, this serves to separate the patients, ex post, into two categories: (i) "fast-only" patients, whose disposition decision (admission or discharge) is made in the fast ED, and (ii) "fast-and-slow" patients, who are initially treated in the fast ED but at some point are transferred into the slow ED where the disposition decision is made. In practice, this two-stage system should help ensure both the timely processing of patients in the fast ED and the reduction of gatekeeping errors for patients channeled through the slow ED.

The medical literature provides ample evidence of the medical benefits of routing patients with specific diseases, such as heart failure, through a CDU (see §2.2.1). However, little is known about how a CDU affects the core gatekeeping function of the ED or about the circumstances under which such a two-stage gatekeeping system is especially beneficial. Furthermore, the aggregate system-wide effects of introducing a CDU have not been properly explored. In particular, introducing a slow ED comes at a cost: Some of the staff who would normally work in the fast ED are now in the

**Table 1 Nomenclature.**

| Term | Meaning |
|---|---|
| Fast ED | The main ED, excluding the CDU. On arrival, all patients will enter the fast ED. |
| Slow ED | The CDU. Some fraction of patients will be transferred to the slow ED from the fast ED. |
| Fast-Only-ED | An ED that consists of a fast ED only (i.e., an ED without a CDU). |
| Fast-And-Slow-ED | An ED that consists of both a fast ED and a slow ED (i.e., an ED with a CDU). |

CDU, and this reduction in resources leads to more frequent congestion in the fast ED. Physicians are then forced to make a trade-off between the best treatment for the patient at hand and the need for fast gatekeeping decisions that increase throughput and reduce wait times for the patients still to be seen. Thus, while the two-stage system may reduce gatekeeping errors for those patients routed through the slow ED, it may also increase gatekeeping errors in the fast ED. It is therefore not clear that a combined fast-and-slow ED will outperform the traditional fast-only ED in terms of overall gatekeeping accuracy.

We address this system-level question empirically in this paper. Making use of a multi-year dataset of over 300,000 adult ED attendances in a large UK teaching hospital that maintained a CDU during the entire observation period, we present four findings. First, we confirm that the CDU reduces both types of gatekeeping error – unnecessary hospitalization and wrongful discharge. Using appropriate sample selection methods to account for non-random assignment of patients to the CDU, we estimate that if the subject ED had operated without a CDU during the observation period, the observed rates of unnecessary hospitalization and wrongful discharge would have increased by 14.3% (from 4.34% to 4.96%) and 29.6% (from 0.71% to 0.92%), respectively.

Second, we show that not all patients benefit from admission to the CDU. For lower-risk patients (i.e., those for whom the gatekeeping decision is more clear-cut and who are unlikely to be hospitalized unnecessarily), the accuracy of gatekeeping decisions in the CDU is no better than it is for those in the fast ED. In contrast, we find that higher-risk patients with a high ex ante chance of unnecessary hospitalization particularly benefit from the CDU. In fact, we estimate that without the CDU, 14.0% of the most high-risk patients would have been hospitalized unnecessarily, while this rate would drop to just 4.8% if all these high-risk patients had instead been routed through the CDU. This differential effect across patients suggests that in order for the CDU to function effectively, ED physicians must perform well in their role as first-stage gatekeepers, ensuring that it is those higher-risk patients who are referred to the CDU. We find evidence that this was the case in our study hospital. In particular, for those patients who were routed through the CDU, we estimate that the rate of unnecessary hospitalization would have increased by 118.6% (from 5.06% to 11.06%) and wrongful discharge by 195.2% (from 1.05% to 3.10%) if the study hospital had not operated a CDU.

Third, turning to the effect of congestion, we show that there is a trade-off between lower error rates for patients routed through the CDU and simultaneously higher error rates for those patients remaining in the smaller and therefore more frequently congested fast ED. Specifically, modeling the effect of congestion on gatekeeping errors for the ED without a CDU, we show that increasing ED congestion levels from low ($2\sigma$ below the mean) to high ($2\sigma$ above the mean) would increase the unnecessary hospitalization rate by 27.1% and decrease the wrongful discharge rate by 17.2%.

Fourth, we perform a counterfactual analysis to evaluate whether a single- or two-stage gatekeeping system would be optimal in our study ED. We show that in this hospital, the redeployment of resources from the CDU to the fast ED, which would reduce congestion in the fast ED and therefore reduce gatekeeping errors, would be significantly less effective than the CDU itself. In fact, our data suggest that if our study hospital shifted capacity from the fast ED in order to expand the slow ED to accommodate 25% of the most high-risk patients, it could reduce unnecessary hospital admissions by up to 33%.

Our data and analysis therefore provide evidence that implementing a two-stage gatekeeping system – a combination of a fast ED and a slow ED – can effectively reduce gatekeeping errors across the ED and thereby safeguard scarce hospital resources for the most vulnerable patients. This finding has immediate implications for the design and management of EDs, as it suggests that all of those with a sufficiently complex patient mix should be structured as fast-and-slow EDs.

## 2. Contribution to the Literature

The primary contribution of this paper is to the operations and healthcare management literature related to gatekeeping processes. In addition, some of our insights are relevant for several areas within the empirical healthcare OM literature and also to the medical literature that examines the role of short-stay observation units such as CDUs. In this section, we outline how the paper's contributions are positioned within these literature streams.

### 2.1. Operations literature

**2.1.1. Gatekeeping.** Gatekeeping systems are customer flow systems comprising multiple service levels, with customer progression from a lower to a higher level controlled by gatekeepers who have a dual role. These gatekeepers can (1) provide a range of services themselves and (2) may also refer a customer with more complex needs up to the next service level, which consists of more highly skilled and more costly providers (Shumsky and Pinker 2003). Early studies in the OM literature focused on systems with a single gatekeeping stage and examined economic models to understand how to incentivize a system-optimal referral rate from the gatekeeper to the specialist (Shumsky and Pinker 2003, Hasija et al. 2005). More recently, the framework has been extended and adapted to specific applications such as security-check queues (Zhang et al. 2011) and outsourcing decisions (Lee et al. 2012). This literature models gatekeepers as economic agents who maximize their time-averaged income from wages plus bonuses per-customer-diagnosed and per-customer-successfully-treated.

As robust as this economic modeling literature may be, however, its insights are not readily transferable to contexts where gatekeeping decisions are not economically motivated but may instead follow professional or social norms. This is likely the case in an ED with salaried physicians, and in such a context empirical or experimental studies may provide better insights into the behavior of gatekeeping systems. There have been few such studies to date (though exceptions exist, e.g., Freeman et al. 2017, Gorski et al. 2017), and the question of how the gatekeeping process can be modified to reduce referral errors (i.e., under- and over-referral to specialists) has not yet been addressed. As previously mentioned, however, it is important to understand referral errors because they are both costly and may worsen individual outcomes and system performance. Our paper is, to the best of our knowledge, the first in the OM literature to expand the notion of gatekeeping beyond the standard single-stage setup by empirically examining the benefits and trade-offs of a gatekeeping system comprising additional stages.

**2.1.2. Empirical healthcare operations.** This paper also contributes to a growing body of research within the OM literature that studies the impact of organizational factors on clinical, operational and financial outcomes in healthcare systems, such as mortality (e.g. KC and Terwiesch 2012, Kim et al. 2014, Kuntz et al. 2015), service times (e.g. KC and Terwiesch 2009, Berry Jaeker and Tucker 2017, Chan et al. 2017), and queue abandonment (Batt and Terwiesch 2015). Of particular relevance is the work on congestion in patient flow systems: In two studies of intensive care units (ICUs), KC and Terwiesch (2012) and Kim et al. (2014) show that ICU staff block admissions and discharge patients prematurely when their specialist unit becomes congested. While this behavior rations access to congested services so that the neediest patients can be treated, it also leads to deterioration in system performance, as evidenced by an increase in ICU readmission rates. In contrast to these studies, which focus on the specialist unit, we instead focus on how the upstream gatekeeping process can be *redesigned* in order to reduce referral errors and better ensure that congested downstream specialized services are reserved for those most in need.

Other empirical studies have also examined the effect of upstream congestion on gatekeeping decisions. For example, Freeman et al. (2017) show that midwives who act as gatekeepers to specialist obstetricians refer high-complexity patients to obstetricians at higher rates in the presence of congestion. Further, Gorski et al. (2017) show that hospital admission rates from the ED increase with congestion. Building on evidence from these studies, we highlight the trade-off that occurs as stages are added to the gatekeeping process: On one hand, the accuracy of referral decisions may improve for those patients who pass through the additional gatekeeping stages, but on the other hand, when resources are shifted in order to operate the new downstream stages, upstream congestion will increase. Since this upstream congestion may degrade the quality of gatekeepers' referral decisions, it is unclear whether adding additional stages is advantageous at all. To the best of our knowledge, our empirical study is the first to demonstrate that a multi-stage gatekeeping process can outperform the traditional single-stage process and to provide insights into the conditions under which the multi-stage structure is especially advantageous.

**2.1.3. Other operations literature.** This paper is also closely related to a series of recent analytical papers in the OM literature examining ED triage. While triaging has traditionally prioritized patients based on their level of urgency (FitzGerald et al. 2010), recent analytical studies have explored ways in which the basic triage process might be improved by segmenting patients along other dimensions. Chan et al. (2013), for example, develop a triage algorithm to allocate burn victims to beds based on their expected length of stay and comorbidity profile. Most relevant to our work are two modeling papers studying the ED triage process (Saghafian et al. 2012, 2014) that propose augmenting triage by segmenting ED patients based not only on severity but also on their (i) admission likelihood and (ii) clinical complexity. Saghafian et al. (2018) also use a modeling approach to identify the impact of allowing nurses to offload triage decisions to more experienced telemedical physicians, extending the standard single-stage triage process into a two-stage process. While our paper complements these studies with an empirical examination, our context differs in two important ways. First, a multi-stage gatekeeping process like the one we are studying channels patients into downstream gatekeeping stages during service itself, while triaging puts patients into a specific queue before providing services. Second, our outcomes of interest differ from the prevailing concerns (average cost and waiting time) and focus on gatekeeping referral errors.

## 2.2. Medical literature

**2.2.1. Healthcare literature on decisions units.** Finally, this paper contributes to the medical literature on CDUs by (i) investigating the gatekeeping role of these units and (ii) providing a system-level – rather than patient-level – study of their effect on the accuracy of gatekeeper referral decisions (i.e., admission and discharge) in the context of emergency medicine.

This literature has predominantly focused on patients with specific conditions (e.g., chest pain, asthma) and finds that when these patients are routed through observation units, their outcomes (e.g., mortality rates, return hospitalization rates, and other disease-specific outcomes) are equal to or better than those of patients admitted into inpatient units (Conley et al. 2016). Compared to conventional inpatient management, observation units are also associated with higher levels of patient satisfaction and significant cost savings (Cooke et al. 2003). However, there is only limited overall evidence demonstrating the effectiveness and safety of these units across all conditions (Galipeau et al. 2015). Indeed, one risk is that CDUs may become “dumping areas” for patients who ought to have been admitted or discharged (Brillman et al. 1995). These patients may then be worse off when admitted to the CDU, possibly making the CDU ineffective overall. Our study addresses this risk by assessing the impact of the CDU on the accuracy of disposition decisions for *all* patients routed through the unit.

A number of cross-sectional studies have also looked at the gatekeeping role of CDUs in reducing hospital admission rates, generally finding that the presence of an observation unit makes admission less likely (Roberts et al. 2010, Lo et al. 2014). Such studies, however, compare outcomes before and after the opening of a new CDU, which expands overall hospital capacity rather than repurposing capacity from the ED (e.g. Lo et al. 2014, Schull et al. 2012). It is not clear, then, whether it is the CDU or the increase in capacity driving the observed improvements in performance. Our analysis builds on these papers by comparing gatekeeping performance in the presence of a CDU with the performance of an expanded ED that draws resources away from the CDU.

Schull et al. (2012) also noted that admission rates alone are insufficient performance measures; rather, the appropriateness of the admission must also be taken into account, e.g., by considering reductions in short-stay (i.e., potentially avoidable) admissions. Our paper contributes to this stream of medical literature by showing that CDUs not only reduce admission rates but also reduce rates of *inappropriate* admissions and discharges. In other words, CDUs improve the gatekeeping accuracy of the ED.

**2.2.2. Contribution to practice.** A hospital's decision whether to invest in a CDU depends on the competing alternatives, which include expanded acute care ED capacity or additional inpatient capacity (Baugh et al. 2011). It has been argued that CDUs are not always clinically appropriate because they may absorb staff who could have remained in the ED (Cooke et al. 2003) and that CDUs therefore exacerbate ED overcrowding. Yet to our knowledge the inherent trade-off between an expanded ED (i.e., a fast-only ED) and a smaller ED with an integrated CDU (i.e., a fast-and-slow ED) has not been fully investigated. Taking a standard ED in the UK as the study site, our paper provides the first evidence that a combined fast-and-slow ED outperforms an expanded fast-only ED with respect to the ED's primary gatekeeping function. We also examine the circumstances under which this is likely to be the case.

## 3. Two-Stage Gatekeeping in Emergency Departments

In this section, we will describe the gatekeeping process in our study ED and explain the mechanisms by which the second gatekeeping stage – the slow ED, or CDU – helps ED physicians make better disposition decisions.

### 3.1. The emergency department

Our study ED is the hospital's single front door for all emergency patients, including those referred by primary care physicians. The observation period ran from 2006 through 2013, and at that time the ED was visited by 250 patients per day on average. On weekdays, patient care is overseen by five emergency medicine consultants (the most senior grade of emergency physician in the UK, the equivalent of an attending physician in the US) who work a staggered shift pattern. Consultants' responsibilities include treating patients and supervising a team of trainee doctors. On weekends, senior medical staffing is reduced to two consultants. In addition to their other duties, one consultant is nominated as the Emergency Physician in Charge (EPIC). This doctor focuses on operational issues, such as ensuring that the sickest patients are seen promptly, fast patient flow is maintained, and any problems (such as delayed tests or specialist reviews) are quickly identified, appropriately escalated, and resolved.

Upon arrival in the ED, patients meet a triage nurse who assesses their degree of urgency and infection risk. If the patient does not need immediate attention and is not infectious, she registers at reception and moves to the waiting area. Here, a nurse takes her vital signs and a brief history and may order preliminary diagnostic tests (e.g., blood, imaging). The patient then waits to be seen by a physician, typically a trainee doctor working under the supervision of an ED consultant. The physician examines the patient in a cubicle and may order additional diagnostic tests or consult a specialist in the hospital. If, after assessment, the physician determines that the patient requires a level of care beyond what the ED can provide, she can admit the patient to an acute bed in the hospital. Otherwise, the patient's symptoms are treated, then she is discharged and may be referred for an outpatient or primary care follow-up appointment. These disposition decisions made by trainee doctors are generally reviewed by a senior physician (a consultant or a specialist registrar with at least six years of experience), especially if the patient has a high-risk condition (e.g., atraumatic chest pain) or is in a specific demographic group (e.g., children under the age of one). Thus, the disposition decision rations access to scarce and expensive hospital inpatient beds and is the key gatekeeping activity in the ED (Blatchford and Capewell 1997).

### 3.2. Gatekeeping in the emergency department

Getting the disposition decision right can be challenging because patients present with a variety of complaints and symptoms. Some can be managed easily in the ED (e.g., wound suturing, casting, splinting) while others are complex and clearly require hospital admission for specialized, longer-term care (e.g., hip fracture, heart attack, stroke). Many patients, however, present with symptoms that could either be caused by a minor ailment or could be the sign of a more severe or even life-threatening condition (e.g., chest or abdominal pain). These patients require careful diagnosis prior to a disposition decision, yet medical diagnosis, particularly in the context of emergency medicine, is difficult and error-prone. In fact, Graber et al. (2005) estimate that one in ten medical diagnoses made in EDs are inaccurate, and these diagnostic errors are the leading cause of internal investigations and malpractice claims (Kachalia et al. 2007, Cosby et al. 2008).

The occurrence of gatekeeping errors in a medical context is especially salient due to the costs involved. Medical errors may also have a negative emotional impact on physicians (Christensen et al. 1992), can result in malpractice investigations and/or litigation (Studdert et al. 2006), and sometimes lead to reputation damage and peer disapproval (Leape 1994). The costs (financial or otherwise) that a physician associates with these concerns will affect how they trade off false positives (unnecessary hospitalizations) and false negatives (wrongful discharges) when making gatekeeping decisions.

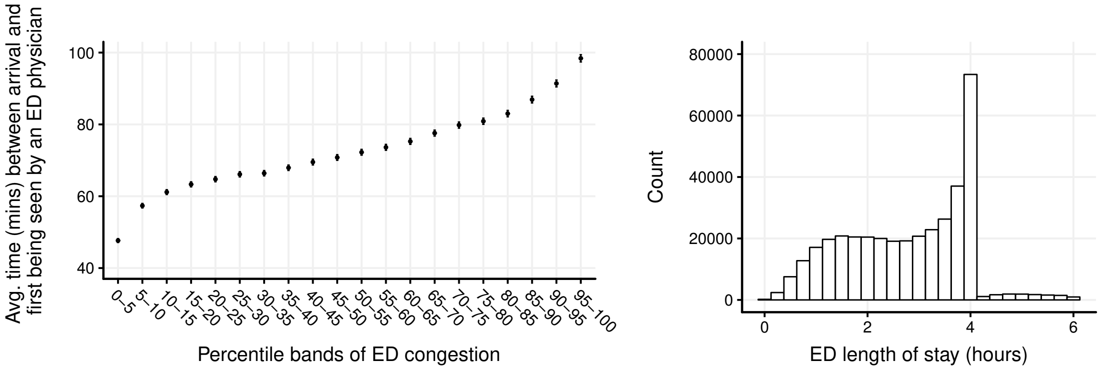
Compounding this cost issue, the current prevalence of overtreatment suggests that when faced with uncertainty, medical professionals often choose to do more rather than less (Gawande 2015). For example, unnecessary referrals to specialists are more common than missed referrals (Bunik et al. 2007). The threat of litigation is often cited as a cause of this phenomenon, and medical professionals have been shown to refer patients more frequently to higher intensity care when they perceive a risk of undertreatment (Shurtz 2013). In the stark words of a physician in our study hospital, *"No one has ever been sued for admitting a patient to the hospital."* ED physicians may therefore opt to admit patients (potentially unnecessarily) rather than discharge when in doubt.

This problem is exacerbated when physicians are exposed to increased congestion and must make decisions under time pressure and cognitive strain. In England, these pressures are intensified by the government's four-hour waiting time target, which requires 95% of patients to leave the ED within four hours of arrival. During our study period, failure to meet this target in any month attracted a fine of £200 per breach (NHS 2013), which could cost the study hospital between £75,000 (5% breaches) and £300,000 (20% breaches) per month. As a consequence, the four-hour target was taken seriously, with clinical staff making a special effort to discharge or transfer patients before their length of stay (LOS) in the ED passed the four-hour mark (see Figure 2 (right)). This policy, together with the fact that patients had to wait longer to see a physician during periods of increased congestion (see Figure 2 (left)), meant that when the ED was crowded, physicians had less time to spend with each patient.

### 3.3. The clinical decision unit

Our study hospital also operates a CDU. This is a dedicated bedded area that, while physically separate, is located next to, organizationally integrated with, and under the control of the main ED. It comprises two single-sex bays with four beds each and six patient seats in the center of the unit. The dedicated CDU staffing consists of a senior nurse and around-the-clock coverage provided by two registered nurses, a healthcare assistant, and a trainee doctor. Additionally, one of the emergency medicine consultants is nominated as the accountable CDU consultant. While working a regular shift in the ED, this consultant is also responsible for patients in the CDU, where she performs regular checks and is the first point of senior clinical contact for CDU staff. The CDU

**Figure 2** **(Left) Mean time between patient arrival in the ED and being seen by an ED physician as a function of ED congestion, with 95% confidence bands; (Right) Histogram of ED length of stay.**



consultant also makes disposition decisions, either discharging patients or admitting them to a specialist ward in the hospital.

According to our study hospital's operational policy (and consistent with medical guidelines and literature, e.g., Baugh et al. 2012): *"The CDU is a 24-hour facility that is used for patients who require a short period (<24 hours) of observation and/or treatment. It will also be used for patients who need a short admission for the diagnosis or exclusion of specific conditions, so enabling appropriate placement of patients. It is not a holding or overflow area."* The primary purpose of the CDU, then, is to bring together patients who are expected to be discharged within a reasonable time frame but who require additional diagnostics or therapies beyond their initial ED stay. These patients are sometimes referred to as "high-risk discharges" (Cooke et al. 2003).

To prevent the CDU from becoming backlogged with patients who would normally be admitted or discharged, admission to the CDU is tightly controlled and requires discussion with an ED consultant. This role is typically delegated to the EPIC who, as noted earlier, is responsible for overall ED patient flow. A key element of admission control is that non-ED physicians or hospital bed managers cannot use the CDU as overflow when hospital beds become scarce. All patients in the CDU must arrive there directly from the ED.

While the CDU is not meant to be used for admission avoidance and flow management, the EPIC does occasionally use it this way when the ED is crowded and bay capacity is needed for other patients. For example, patients waiting for test results might be moved to the CDU to create capacity in the ED. This can have the added benefit of maintaining compliance with the four-hour waiting time target because a move to the CDU takes the patient "off the clock."[1] However, the decision to admit a patient to the CDU is not made lightly, as it comes at a significant cost for the referring physician: Since a CDU admission is classed as a hospital admission for administrative

[1] While transfer to the CDU can in principle occur any time after a patient's arrival in the ED, in practice two-thirds of patient transfers from the ED into the CDU occur in the 30-minute period leading up to the four-hour-target, and 90% occur between two and four hours after arrival. Figure EC.4 of the e-companion provides a histogram of ED length of stay for patients not admitted versus those admitted to the CDU.

and reimbursement purposes, the referring ED physician must complete the hospital admission paperwork, which includes drug charts and nursing orders. A CDU nurse then completes the admission process, which typically takes around 45 minutes per patient.

## 4. Hypothesis Development

### 4.1. The effect of the CDU on gatekeeping accuracy

Figure 1 illustrates the key difference between an ED with (right) and without (left) a CDU, which is that in the presence of a CDU, ED physicians are not forced into making a dichotomous inpatient admission or discharge decision. Instead, an additional decision node is added to the traditional ED gatekeeping process: If the disposition decision is not clear, a patient can be classified as "requiring more work" and referred to the second stage (i.e., the CDU) of the gatekeeping system. Bearing this in mind, we now discuss a number of mechanisms through which we expect the CDU to affect the quality of disposition decisions made in the ED.

**Time.** As previously mentioned, a move to the CDU takes the patient "off the clock." This allows additional time for assessment and diagnosis and hence means that a better-informed referral decision can be made at a later stage. However, in regard to inpatient admission, the medical literature finds that CDUs have a variety of advantages that cannot be explained purely by the additional time for assessment, including lower costs, lower rehospitalization rates, and better patient satisfaction (Conley et al. 2016).[2] These unexpected advantages point to other potential benefits of a CDU referral, which we explore below.

**Culture of discharge.** The only patients that ED physicians should refer to the CDU are those who, if admitted to a hospital ward, would have a high probability of being discharged quickly. If a patient seems unlikely to be discharged quickly, admission to a hospital ward is more appropriate because it avoids subjecting the patient to an unnecessary additional transfer. For this reason, Cooke et al. (2003) attribute the benefits of a CDU to a "culture of rapid discharge." The idea is that admitting a patient to the CDU rather than an inpatient bed prevents them from mixing with other patients for whom a longer stay is appropriate, likely leading to increases in their own stay. This tendency towards discharge is confirmed in the literature, with Baugh et al. (2012) finding that *"approximately 80% of patients managed in the [CDU] are able to be safely discharged home."* The CDU's focus on rapid discharge belies a different mindset from that of the fast ED, where the emphasis is on quickly and effectively stabilizing and treating patients with acute needs, and from that on the hospital ward, where patients are expected to stay for several days.

[2] By controlling in our analysis for the time that a patient spends in the CDU, we also show that in our study ED, time is not the only mechanism through which the CDU affects the accuracy of disposition decisions – see §5.5 and §6.1 for more on the procedure that we use for this.

Note that this discharge-oriented mindset is not inconsistent with the earlier observation (see §3.3) that those ED patients referred to the CDU are also more likely to be high-risk discharges. In other words, it would be risky to discharge these patients from the ED precisely because they require further diagnostics or treatments to rule out or manage particular complications and conditions. While these complications and conditions are likely rare or manageable in the short term in the CDU – in keeping with the idea that these are patients who would likely be discharged quickly from a hospital inpatient unit – wrongful discharge from the fast ED may have severe adverse consequences.

Since ED physicians are, when they are in doubt, more likely to admit a patient than discharge them (see discussion in §3.2), the types of patients referred to the CDU are also those who an ED physician would normally admit. We observe that this tendency to admit is particularly pronounced when the ED becomes busy and less time is available to make the disposition decision (refer to Figure 2), which aligns with the findings of other studies (Gorski et al. 2017). The presence of the CDU thus counterbalances this effect by giving physicians a placement option to a unit that emphasizes discharge for those patients who might otherwise be admitted unnecessarily.

**Consultant oversight.** The CDU is typically occupied by patients for whom the gatekeeping decision is more uncertain (i.e., high-risk discharges from the fast ED), whose cases are more complex, and who tend to benefit from the oversight of more experienced physicians (Cooke et al. 2003). This matching of need and experience takes place in the CDU where, as previously mentioned earlier, the CDU consultant makes the final disposition decision for all patients. Patients in the ED, on the other hand, are more or less randomly assigned to ED physicians in a round-robin scheme. This can result in a mismatch between need and experience, with more diagnostically complex cases potentially assigned to more junior physicians. Although a senior physician will review most of these cases, this may not always occur (e.g., when the ED is especially congested). Even when the case is reviewed, it will likely be performed by a senior registrar who is less experienced than a consultant. This suggests that disposition decisions in the CDU will generally be made by more experienced physicians and hence should be of higher quality than those made in the fast ED.

Altogether, the fact that an experienced consultant is always the decision-maker in CDU cases, the increased time available for diagnosis, and the culture of discharge suggest that a patient routed through the CDU should be less likely than a patient routed through the fast ED only to be hospitalized unnecessarily.

Hypothesis 1. *A patient is less likely to be admitted to the hospital unnecessarily if the disposition decision is made in the CDU instead of the fast ED, even after accounting for the* *additional time spent under observation in the CDU.*

While we anticipate a lower average rate of unnecessary hospitalization for patients routed through the CDU, not all patients are likely to benefit equally from being referred into this unit. For some patients, the disposition decision is clear-cut regardless of where it is made or by whom; this is particularly the case for those with a low ex ante likelihood of unnecessary admission (whether because they are very unlikely to be admitted or because they have serious conditions that clearly require hospitalization). On the other hand, other patients arrive at the ED with symptoms and other characteristics (e.g., minor head injuries, undiagnosed chest pain) that make the appropriate disposition decision uncertain and that naturally predisposes them for unnecessary admission. We expect that these patients in particular will benefit from admission to the CDU.

Hypothesis 2. *Compared to patients with a low ex ante probability of unnecessary hospital admission, patients with a higher ex ante probability of unnecessary admission will experience a larger reduction in unnecessary hospitalization rates if the disposition decision is made in the CDU instead of the fast ED.*

**Admission control.** Another critical aspect of the success of the CDU is admission control. Recall that admission to the CDU is entirely controlled by senior ED physicians, who are able to assess whether a patient referred to the CDU has a good chance of being discharged within a short time frame (<24 hours). Yet without a strict admission policy and tight control, the CDU is at risk of being used as a workload buffer by ED physicians, resulting in a backlog of patients who ought to have been admitted or discharged instead. While the CDU might still help prevent unnecessary admissions even if its population were simply a random assortment of ED patients, its effectiveness would be significantly reduced.

Since effective admission control appears to be critical to the success of the two-stage fast-and-slow ED concept, we test for the effect of admission control by determining whether the patients admitted to the CDU are those who stand to benefit from it the most. We anticipate that the effectiveness of the CDU is enhanced by the accurate referral of patients who are more predisposed to unnecessary hospitalization.

Hypothesis 3. *Those patients who are actually referred to the CDU benefit more from it, in terms of reduced probability of unnecessary hospital admission, than a randomly drawn sample of ED patients.*

While we have argued that patients are less likely to be admitted unnecessarily when routed through the CDU, it is unclear what the effect will be on wrongful discharges. On the one hand, if the threshold for admission is higher in the CDU than in the ED, the reduction in the unnecessary hospitalization rate (false positives) may come at the cost of an increase in the wrongful discharge rate (false negatives). Specifically, the culture of discharge in the CDU may lead to an inappropriate discharge for some patients who should be admitted. On the other hand, if the accuracy of gatekeeping decisions as a whole goes up, this puts patients into the hands of experienced CDU consultants who are armed with better information and can make more accurate admission and discharge decisions, so both false positives and negatives may go down. On balance, it is not clear what the overall effect will be, so we leave this as an empirical question that we will answer in the context of our study ED.

EMPIRICAL QUESTION 1. *How does CDU admission in our study hospital affect a patient's likelihood of being wrongfully discharged?*

### 4.2. System effects

The fact that the presence of the CDU improves disposition decisions does not in itself make a case for the CDU, since the CDU binds resources that could otherwise be redeployed in the fast ED. The real question is, therefore, whether the CDU improves disposition decisions more than *"competing alternatives, such as expanding acute care ED space"* (Baugh et al. 2011). Increasing the capacity of the fast ED by repurposing CDU resources, such as staffing and space, would reduce congestion in the fast ED. To evaluate the overall effectiveness of the CDU in reducing gatekeeping errors, it is therefore necessary to understand the impact of congestion on the quality and accuracy of decision making in the fast ED itself.

**ED Congestion.** ED physicians are well aware of the level of congestion in the fast ED, both through direct visual cues and through IT systems that show, for example, the list of waiting patients with their registration details and triage information. In response, physicians exercise a degree of discretion over the time they spend with their patients (Hopp et al. 2007). ED physicians in our study hospital confirm that they are trading off quality and speed: "*When we are crowded we have two competing problems – we know we should not admit patients unnecessarily, yet we have to avoid breaching the ED waiting time target.*" When congestion increases, it is rational for ED physicians to reduce the service time for individual patients, since the opportunity cost of time spent with any given patient increases against the alternative of reducing congestion in the system. When service times are reduced, physicians have less time available to assess a patient and acquire the information necessary to make accurate gatekeeping decisions (Smith et al. 2008, Alizamir et al. 2013). In addition, increased congestion leads to cognitive overload as ED physicians must care for more patients simultaneously (KC 2014). Since the work of ED physicians relies on intuition and heuristics (Croskerry 2002), overload can render these cognitive shortcuts ineffective, resulting in preventable errors (Leape 1994). For example, Graber et al. (2005) found that cognitive factors contributed to 74 out of 100 studied cases of diagnostic error.

Therefore, by reducing the amount of time available for physicians to gather information before making a gatekeeping decision and increasing cognitive overload, congestion is likely to lead to deterioration in decision quality. Moreover, since ED physicians who are in doubt are more likely to admit a patient (potentially unnecessarily) than discharge them (as noted earlier in §3.2), congestion should translate to a significant increase in unnecessary hospital admissions.

Hypothesis 4. *Congestion increases the rate of unnecessary hospital admission.*

This hypothesis implies that there is a trade-off between operating a CDU and a fast-only ED with expanded capacity. In particular, while the CDU may reduce error rates for those patients routed through it, most patients are not admitted to this unit and their gatekeeping decisions are made in the fast ED. These non-CDU patients may actually experience higher error rates in a system with a CDU due to the higher congestion levels that they are exposed to when resources are shifted from the fast ED to establish or enlarge the CDU. Whether CDUs are advisable with respect to their overall impact on the quality of disposition decisions – one of the critical system functions of the ED – is, therefore, an open empirical question.

Empirical Question 2. *How is the rate of unnecessary hospital admission in our study hospital affected when resources are shifted between the fast ED and the CDU?*

## 5. Data Description and Variable Definitions

Our initial study data comprises detailed information relating to 651,028 ED attendances over a period spanning seven years, from December 2006 through December 2013, as well as matching inpatient records for all of those patients admitted from the ED into the hospital during this period. The ED we study is the largest in the region and experienced increasing demand during the study period, with attendances increasing by 4.2% year-on-year, from 215 ED visits per day on average in the first year of our sample to 274 per day in the final year. On average 29.1% of visits result in inpatient admission, with admissions and discharges increasing at approximately the same rate over the sample period (by 4.7% and 4.1% per annum, respectively).

Prior to analysis, the data was preprocessed to ensure, as much as possible, that our results are not affected by various data- or time-related confounds. This included dropping: (i) ∼8.5k observations (obs.) from December 2013, when data entry may have been incomplete; (ii) ∼130k obs. corresponding to children under 16, who cannot be admitted to the CDU; (iii) ∼3.5k obs. with missing or incomplete data; and (iv) ∼18k obs. of patients who left against medical advice, died in the ED, or were transferred to another hospital. The remaining data was then used to generate various variables of interest (to be described later) before excluding: (v) ∼61k obs. from the first 12 months, the warm-up period for generating these variables and (vi) ∼38k obs. from dates associated with public holidays and the Christmas break, when demand and staffing patterns

**Table 2** **Descriptive statistics and correlation table.**

| | | Mean | | | Correlation table | | | |
|---|---|---|---|---|---|---|---|---|
| | N | All | CDU = 0 | CDU = 1 | (1) | (2) | (3) | (4) |
| (1) Total gatekeeping errors (%) | 377,346 | 5.05 | 4.94 | 6.11 | | | | |
| (2) Unnecessary hospitalizations (%) | 377,346 | 4.34 | 4.26 | 5.06 | 0.92*** | | | |
| (3) Wrongful discharges (%) | 377,346 | 0.71 | 0.68 | 1.05 | 0.37*** | −0.02*** | | |
| (4) CDU admission (%) | 377,346 | 9.90 | 0.00 | 100.00 | 0.02*** | 0.01*** | 0.01*** | |
| (5) ED congestion | 377,346 | −0.01 | −0.01 | −0.01 | 0.01*** | 0.01*** | −0.00** | 0.00 |

*Notes:* Columns 'All', 'CDU = 0' and 'CDU = 1' report mean values for the full sample, subsample of patients referred directly from the ED, and subsample referred from the CDU, respectively; Standard deviation of ED congestion equal to 1.01, 1.01 and 1.02 for 'All', 'CDU = 0' and 'CDU = 1', respectively; Pre-standardized mean (standard deviation) of ED congestion equal to 0.70 (0.20) for 'All', 'CDU=0' and 'CDU=1'; Correlation coefficients significant with $^{***}p < 0.001$, $^{**}p < 0.01$, else $p > 0.10$.

vary significantly. Due to a temporary change in coding convention that prevents identification of CDU admissions in December 2009 and January 2010, we also dropped ∼14k obs. from this period. After this, we were left with 377,346 observations to take forward for analysis.[3]

We next describe the main variables used in the analysis. Summary statistics for these variables and correlations between each can be found in Table 2.

### 5.1. Unnecessary admissions and wrongful discharges

The two dependent variables of interest in our analysis capture imprecision in referral (admission) and non-referral (discharge) decisions by ED physicians.

An unnecessary hospitalization occurs when a patient is admitted to an acute hospital bed when admission is unnecessary or excessive for the patient's needs. These patients block beds and use expensive specialist resources and time. We define an unnecessary hospitalization ex post as any patient discharged within 24 hours of admission to an inpatient bed from the ED or CDU without any treatment or procedure performed.[4] We consider a treatment or procedure to have taken place if there is an OPCS-4.6 (HSCIC 2013) intervention or procedure code – the UK equivalent of the American Medical Association's CPT coding system – associated with the post-admission inpatient record. The average unnecessary hospitalization rate for the full sample of 377,346 visits is 4.3% and the rate for the 119,480 visits which resulted in admission is 13.7%.[5]

Wrongful discharges are, if anything, even more concerning. Wrongfully discharged patients often return to the ED in a more serious state, requiring a higher intensity of care than would otherwise

[3] Consistent findings are obtained using an expanded sample in which Christmas and public holidays are reintroduced.

[4] Not all ex post unnecessary hospitalizations are avoidable ex ante, though *avoidable* unnecessary hospitalization does occur (Denman-Johnson et al. 1997). For instance, some patients have intrinsically uncertain conditions that may or may not require hospital intervention and so they are admitted to ensure that the hospital can respond swiftly if and when needed. These patients might be discharged within 24 hours without treatment, but ex ante their admission was necessary. This is akin to the difference between recorded adverse events and unrecorded avoidable adverse events (Brennan et al. 1991). We discuss this further in §EC.2 and §EC.7 of the e-companion.

[5] We use the term "rate" to describe the proportion of patients admitted to the hospital unnecessarily. Specifically, for every patient visit to the ED, the visit can be classified binarily: It results in an unnecessary hospitalization or it does not. Taking the average over these binary outcomes gives the corresponding rate.

have been needed if the patient was initially admitted. Pope et al. (2000), for example, found risk-adjusted mortality among patients with acute myocardial infarction who were inappropriately discharged from the ED to be 1.9 times higher than among hospitalized patients. We define an ED discharge as wrongful if the patient re-attends the ED within seven days after discharge from the ED or CDU, has a diagnosis that is in the same assigned category as their previous ED diagnosis, and is subsequently admitted to the hospital. The wrongful discharge rate is 0.7% for the full sample of 377,346 ED visits and 1.0% for the subset of 257,866 discharged patients.

### 5.2. CDU referral

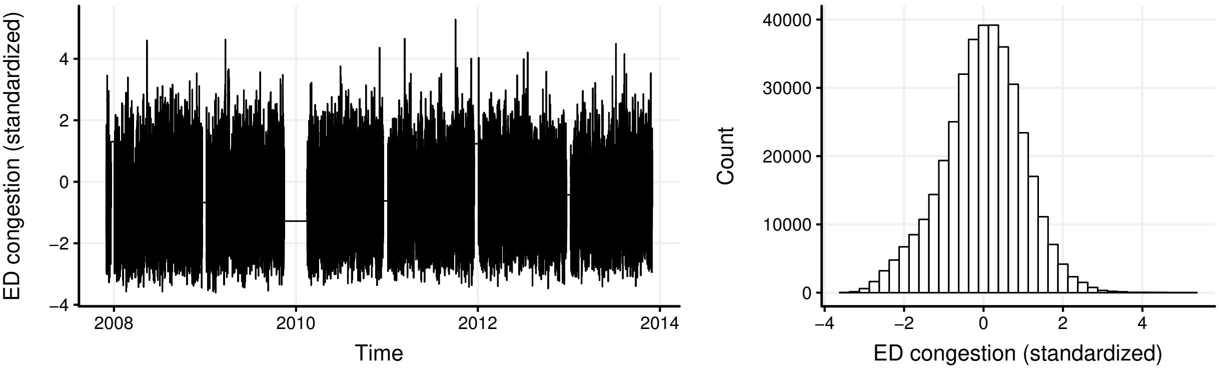
In our data, 37,356 ED patients are sent from the ED to the CDU. Of these patients, (9.9% of the analysis sample), 35.2% are subsequently admitted to an inpatient bed in the main hospital. In the CDU, decisions are made quickly, with a median CDU LOS of 4.5 hours for those who are subsequently admitted and 4.1 hours for those who are subsequently discharged. In contrast, the median LOS in an inpatient hospital bed for a patient classed as an unnecessary hospitalization is 15.5 hours. This suggests that patients in the CDU are processed more quickly than they are in a standard inpatient setting. Confirming this impression, further analysis (documented in §EC.1 of the e-companion) finds that the CDU can process patients at a rate at least 42% faster than hospital inpatient units. Thus, while referral through the CDU does extend the service episode, it is still faster than direct referral to the hospital. This is consistent with findings in the medical literature (e.g. Baugh et al. 2012).

Since the patient population targeted by the CDU comprises those for whom disposition decisions are uncertain and for whom discharge may be risky (see §3.3), we might expect those patients admitted to the CDU to be unnecessarily admitted or wrongfully discharged more frequently. Yet it is also possible that the action of the CDU acts as a countervailing force to keep those rates low, making the net effect unclear a priori. Raw summary statistics suggest that the rates of unnecessary hospitalization for patients admitted from the CDU and from the ED are similar at 14.4% and 13.6%, respectively. This provides initial evidence that the CDU may have a protective effect in reducing gatekeeping errors. In §6, we investigate this formally.

### 5.3. ED congestion

An important variable necessary to evaluate the overall impact of the two-stage gatekeeping approach at the system level (see §4.2) is the level of congestion patients experience when they arrive in the ED. To generate this measure for patient $i$, we first determine which other patients' ED visits overlapped with the period from the arrival of patient $i$ to one hour post-arrival. Taking the sum of those overlapping periods gives us $CensusED_i$. This approximates the number of other patients in the ED (in both the queue and in service) when patient $i$ arrives. Note that the full sample of 651,041 ED visits is used in the calculation of $CensusED_i$.

**Figure 3** **Plot of standardized ED congestion over time (left) with frequency histogram (right).**



Since levels of ED congestion vary throughout the day, across days of the week and seasons, and change over time, and since some of this variation is predictable enough to help determine staffing, we should adjust $CensusED_i$ to account for it. We achieve this by adapting the approach used in Kuntz et al. (2015) and Berry Jaeker and Tucker (2017) to approximate available capacity. Specifically, we estimate capacity using quantile regression to predict the 95th percentile level of occupancy at hour $h$, $CensusED_h^{95th}$. The dependent variable in this regression is the average occupancy level over every hour $h$, starting from midnight on January 1, 2007 and ending at midnight on December 31, 2013. (Note that all dates dropped during the data cleaning process described in §5 are also removed here.) We estimate this model with independent variables: (i) year; (ii) quarter of the year; (iii) time, split into six four-hour windows per day (e.g., midnight to 4 a.m.); (iv) a categorical variable indicating whether it is a Saturday, a Sunday, or a weekday; (v) the interaction between (iii) and (iv); and (vi) the interaction between (v) and a binary variable equal to one if the date is between July 2011 and December 2013 (i.e., the second half of the sample period) and zero otherwise.[6]

The fitted values from the quantile regression model provide us with our estimate of capacity at each hour $h$, $CapacityED_h = \widehat{CensusED}_h^{95th}$. ED congestion, $EDCong_i$, can then be expressed as the ratio of observed occupancy to estimated capacity, i.e., $CensusED_i / CapacityED_{h_i}$, where $h_i$ is the hour patient $i$ arrives. This captures the variation in congestion levels that cannot be explained by predictable and staffable seasonal predictors. Finally, to ease later interpretation of results, we normalize by subtracting the mean, $\mu(EDCong_i)$, and dividing through by the standard deviation, $\sigma(EDCong_i)$, to form $zEDCong_i$. Plots of $zEDCong_i$ are provided in Figure 3.

[6] Note that the interaction term described in (vi) allows for an update to capacity midway through our observation period to capture any changes in the pattern of patient arrivals across hours of the day and/or days of the week.

### 5.4. Predisposition for unnecessary hospitalization

In addition to the aggregate impact of the CDU on unnecessary hospitalization rates, we are also interested in whether the CDU is especially effective for patients who have a higher ex ante likelihood of unnecessary admission. This would give some indication of when a two-stage system might be preferable to a single-stage system and the types of patients who ought to be admitted in the second stage. To determine a patient's risk, we estimate a probit regression on the subset of patients who were not admitted to the CDU (since we are interested in what would have happened to the patient in the absence of a CDU) to predict each patient's likelihood of unnecessary admission, using the definition of an unnecessary hospitalization from §5.1. Control variables included in this regression include all factors known prior to the CDU decision, i.e., all temporal, patient- and diagnosis-related, contextual, and physician-related factors reported in Table 3. After estimating each patient's predicted underlying risk, we classify them into four categories, each comprising 25% of the patients: low, low-medium, medium-high, and high. This forms the variable $PrAdmErr_i$ for each patient arrival $i$.[7]

### 5.5. Control variables

In addition to the primary variables, we also have many control variables, reported in Table 3, that allow us to account for heterogeneity in the patient population and at the hospital. We choose factors correlated with the dependent variables and with the independent variables of interest (§EC.10 of the e-companion provides justification for the controls), and the resulting controls capture patient demographics, temporal factors, differences in diagnosis and condition, contextual factors, and attributes of the assigned physician. Any factors not reported in our data that might be correlated with the variables of interest (and might therefore bias the results through omission) will be accounted for using appropriate empirical methods described in §6.1.

Two controls to be highlighted that become important when discussing our empirical strategy capture the historic unnecessary admission and wrongful discharge rates of the assigned physician. These account for the fact that particular physicians may have a greater propensity for gatekeeping errors, and approximately speaking these controls are calculated as the average case mix adjusted rates of each error type (unnecessary admission and wrongful discharge) made by each physician over the preceding year (see Appendix A for a full description of the calculation of these variables).

Two further controls deserve special attention here. In §3.3 we noted that one benefit of the CDU in our UK context is that it allows ED physicians to increase the time that a patient is under assessment and observation beyond the four-hour target. One might then ask: After a patient is

[7] Due to the low incidence of wrongful discharge (0.7% of the sample), repeating the interaction analysis for these wrongful discharge cases leads to insignificant and unreliable coefficient estimates. These results are reported in §EC.5.2 of the e-companion.

**Table 3 Table of controls.**

| Variable | Type | Description |
|---|---|---|
| Temporal ($\mathbf{T}_i$) | | |
| Year | Categorical (6) | Observation year (offset by one month; e.g., December '07 falls in '08), 2008 through 2013 |
| Daily time trend | Continuous | A variable that takes value one on the first observation date and increases in value by one per day |
| Month | Categorical (12) | Month of the year in which the visit falls, January through December |
| School break | Categorical (7) | If visit occurs during a school break, equals the break type (e.g., Easter, Fall), else set to None |
| Day of week | Categorical (7) | Specifies the day of the week on which the visit occurred, Monday through Sunday |
| Window of arrival x Weekend | Categorical (12) | A four-hourly arrival window (e.g., 4am to 8am) for weekdays, and a separate one for weekends |
| Patient and diagnosis related factors ($\mathbf{D}_i$) | | |
| Age bands | Categorical (16) | The age of the patient, split into 5-year age bands (e.g., 15-20, 20-25,..., 90+) |
| Gender | Binary | A variable equal to one if the patient is male, else zero |
| Triage category | Categorical (6) | The triage level assigned to the patient on ED arrival |
| GP referral | Binary | GP has assessed the patient in the community and referred them directly to the ED |
| Initial severity assessment | Categorical (5) | The nature of the patient's condition (e.g., minor injuries, requires resuscitation) |
| Reason for ED visit | Categorical (30) | The reason for the ED episode (e.g., fall, burn, traffic accident) |
| Diagnosis category | Categorical (22) | The main category of primary diagnosis (e.g. respiratory, cardiovascular) |
| Contextual factors ($\mathbf{C}_i$) | | |
| Mode of arrival | Categorical (8) | The mode of transport used to get to the hospital (e.g., helicopter, ambulance, private transport) |
| ED visits, last year | Continuous | The number of times the patient visited the ED in the previous 12 months |
| ED visits, last month | Continuous | The number of times the patient visited the ED in the previous one month |
| Admissions per ED visit, last year | Continuous | The proportion of hospital admissions to ED visits in the previous 12 months |
| Admissions per ED visit, last month | Continuous | The proportion of hospital admissions to ED visits in the previous one month |
| Zero ED visits, last year | Binary | A variable equal to one if the patient did not attend the ED in the previous 12 months, else zero |
| Zero ED visits, last month | Binary | A variable equal to one if the patient did not attend the ED in the previous month, else zero |
| Physician related factors ($\mathbf{P}_i$) | | |
| Historic unnecessary hospitalization rate | Continuous | The assigned ED physician's unnecessary hospitalization propensity, calculated as in Appendix A |
| Historic wrongful discharge rate | Continuous | The assigned ED physician's wrongful discharge propensity, calculated as in Appendix A |
| New ED physician | Binary | A variable equal to one if we have no data on historic ED physician error rates, else zero |
| Operational/other factors ($\mathbf{O}_i$) | | |
| Hospital congestion | Continuous | The level of congestion of the main hospital inpatient units in to which ED patients are admitted, measured over the one-hour period prior to patient's departure from the ED |
| ED length of stay | Continuous | The length of stay (in minutes) of a patient in the ED. |
| Factors only in the outcome equation ($\mathbf{Y}_i$) | | |
| CDU length of stay (conditional) | Continuous | The length of stay (in minutes) of a patient in the CDU for those patients admitted to the CDU (and zero for those not admitted). |
| CDU congestion (conditional) | Continuous | The congestion level of the CDU for those patients admitted to the CDU (and zero for those not admitted), measured over the one-hour period prior to patient's departure from the ED |

*Notes:* If a variable is categorical, the number in $(\cdot)$ in the " Type" column indicates the number of levels; If a patient did not visit the ED in the previous 12 months (or month) then the "Admission per ED visit, last year" ("last month") variable is set equal to zero; All contextual factors relating to ED visits and admission rates exclude any visits made in the seven days prior to arrival in order to prevent a mechanical relationship with the wrongful discharge variable; The historic ED physician error rates are set equal to zero for those patients who saw a "New ED physician;" CDU length of stay (conditional) and CDU congestion (conditional) are included only in the outcome equation as they perfectly predict the dependent variable in the selection equation.

admitted to the CDU, what drives the change in their likelihood of being a wrongful discharge or unnecessary hospitalization? Is it purely the additional time that these patients spend receiving further diagnostic evaluation and observation? Or are there other mechanisms at work – for example, the culture of discharge and better matching of more experienced staff to more complex cases, as discussed in §4.1? If the former, then one could argue that any benefits from the two-stage system could also be achieved in a single-stage system without time constraints. Therefore, to truly demonstrate the benefit of a two-stage system, we separate out the time effects by controlling for the duration of time that a patient spends (i) in the ED and (ii) in the CDU, if admitted there. This then allows us to isolate the direct impact of admission to the CDU (i.e., the shift in the intercept) while controlling for differences in the time that the patient spends under observation in the ED and the CDU. This point is discussed further in §EC.8 of the e-companion.

## 6. Models and Results: The Two-Stage Gatekeeping System

In this section we describe the method of estimation used to determine the impact of the CDU on both types of gatekeeping error, then present results and robustness.

### 6.1. Econometric specification

Our empirical strategy separates the identification problem into two parts. In the first, we identify those factors that influence whether the patient is admitted to the CDU. In the second, we determine whether a patient is unnecessarily hospitalized or wrongfully discharged, allowing this to depend on whether the patient was admitted to the CDU. More specifically, the first-stage (selection) equation takes the form

$$\begin{aligned} CDU_i^* &= \delta_0 + \mathbf{X}_i \boldsymbol{\delta}_1 + \mathbf{Z}_i \boldsymbol{\delta}_2 + zEDCong_i \delta_3 + \epsilon_i^{\delta}\,, \qquad (1) \\ CDU_i &= \mathbb{1}[CDU_i^* > 0]\,, \qquad (2) \end{aligned}$$

where $CDU_i^*$ is an unobserved latent variable, the vector $\mathbf{X}_i$ contains the set of all controls (reported in Table 3), the vector $\mathbf{Z}_i$ contains the set of instrumental variables (to be described in §6.2), $CDU_i$ is the observed dichotomous variable that indicates whether the patient was sent to the CDU, and $\mathbb{1}[\cdot]$ is the indicator function. The second-stage (outcome) equation takes the form

$$\begin{aligned} AdmErr_i^* &= \beta_0 + \mathbf{X}_i \boldsymbol{\beta}_1 + CDU_i \beta_2 + zEDCong_i \beta_3 + \epsilon_i^{\beta}\,, \qquad (3) \\ AdmErr_i &= \mathbb{1}[AdmErr_i^* > 0]\,, \qquad (4) \end{aligned}$$

where $AdmErr_i^*$ and $AdmErr_i$ are the latent and observed variables, respectively, for unnecessary hospitalizations. The latent variable equation for wrongful discharges ($DischErr_i$) is the same as for unnecessary hospitalizations, with coefficient vector $\boldsymbol{\beta}$ replaced with $\boldsymbol{\alpha}$.[8]

Rather than estimate the first- and second-stage models described above individually, we estimate them jointly with a recursive bivariate probit (biprobit) model using full information maximum likelihood (Maddala 1983). In doing so, we assume that the errors – $(\epsilon_i^{\delta}, \epsilon_i^{\alpha})$ or $(\epsilon_i^{\delta}, \epsilon_i^{\beta})$ – are jointly distributed according to the standard bivariate normal distribution with unit variances and correlation coefficients $\rho^{\alpha}$ or $\rho^{\beta}$, which are estimated as parameters in the models. (More information on the biprobit model is given in §EC.3.2 of the e-companion.) The biprobit model corrects for potential sample selection bias arising from the fact that patients chosen for admission to the CDU might be more (or less) likely to be unnecessarily admitted or wrongfully discharged than those for whom the disposition decision is made in the ED.

We first ask whether there is evidence that a second gatekeeping stage (the CDU), which allows ED physicians who are uncertain about a disposition decision to pass that decision on to another gatekeeper, can help reduce unnecessary hospitalizations. This would be confirmed by coefficient $\beta_2 < 0$ in the outcome equation. We are also interested in any evidence of a change in the wrongful discharge rate, estimated by $\alpha_2$, when patients are routed through the CDU.

[8] Note that ED congestion, $zEDCong_i$, is simply a control. In subsequent analysis in §7 we will obtain reliable estimates for the coefficients of $\beta_3$ and $\alpha_3$, the effect of ED congestion on unnecessary hospitalizations and wrongful discharges, respectively. This will then allow us to identify the overall net effect of the two-stage gatekeeping system.

After establishing the main effects, we explore whether, as hypothesized, patients who are more predisposed to being admitted unnecessarily also benefit more from admission into the CDU. To do this, we add $PrAdmErr_i$ as an additional control into the selection and outcome equations specified in (1) and (3). We then insert into the outcome equation an interaction term between $PrAdmErr_i$ and $CDU_i$, which allows the relative size of the impact of CDU admission on a patient's likelihood of unnecessary hospitalization to differ depending on their ex ante risk of unnecessary admission.[9]

### 6.2. Instrumental variables

While the biprobit model can be estimated without instrumental variables (IVs), estimation is improved and coefficients are more reliable when IVs are provided (Wilde 2000, Maddala 1983). These IVs should affect the CDU admission decision, so they appear in the selection equation (i.e., they are relevant), but they should not affect the unnecessary hospitalization or wrongful discharge rates, so they do not appear in the outcome equation (i.e., they are valid). Our biprobit model uses two IVs, included in the vector $\mathbf{Z}_i$. Summary statistics for these IVs are available in Table 4.

The first IV is the CDU admission propensity of the assigned physician. This is approximately equal to the physician's average rate of CDU referrals over the previous 12 months relative to the rate expected given the case mix of patients they treated (calculation described in Appendix A). A patient assigned to a physician who is predisposed to admit patients to the CDU will be more likely to be sent there as well, satisfying the relevance condition. A potential issue with this IV is that physician rates of CDU referral and error may not be independent. To account for this, we add a control for the physician's historical unnecessary hospitalization or wrongful discharge propensity in the respective selection and outcome equations. After this, the physician's predisposition to admit patients to the CDU should not be correlated with the residuals in the outcome equations, satisfying the validity condition.

Our second IV is CDU congestion, $zCDUCong_i$. This is calculated in the same way as ED congestion in §5.3 except that we time-weight over the one-hour period leading up to the departure of patient $i$ from the ED. If the CDU is congested, beds and other resources are constrained and the CDU then becomes less available to ED physicians as an option. This is similar to other findings in the literature regarding admission to specialist units, e.g., intensive care units (Chan et al. 2017)

[9] One might be concerned about multicollinearity between $PrAdmErr_i$ and the vector $\mathbf{X}_i$ of controls, since most of these controls are also used as predictors for $PrAdmErr_i$ (see §5.4). In practice, however, this does not bias results but may inflate standard errors and make it harder to identify an effect if one does exist. Since all of our interaction effects (the effects of interest) are already highly significant (see Table 7), multicollinearity is likely not an issue here. To validate this, we calculate the variance inflation factors (VIFs). The VIFs for the variable $PrAdmErr_i$ range from 3.3 to 14.8 across the levels, while for the interaction term ($PrAdmErr_i \times CDU_i$) they range from 1.08 to 1.54. As expected, the VIFs for the interaction terms are well within the range where multicollinearity is not a concern. However, as a robustness check, in §EC.5.1 of the e-companion we report on an alternative model specification where all of the covariates that are used as a predictor of both $PrAdmErr_i$ and appear in $\mathbf{X}_i$ are dropped from Equations (1) and (3). Results in this model are nearly identical, further allaying multicollinearity concerns.

**Table 4 Descriptive statistics and correlation table for the instrumental variables.**

| | | Mean | | | Correlation table | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | All | CDU = 0 | CDU = 1 | (1) | (2) | (3) | (4) | (5) |
| (6) Phys. CDU use rate | 377,346 | −0.06 | −0.07 | 0.02 | $0.00^{**}$ | 0.00 | $0.01^{***}$ | $0.15^{***}$ | $-0.05^{***}$ |
| (7) CDU congestion | 377,346 | 0.01 | 0.01 | −0.05 | $0.01^{***}$ | $0.01^{***}$ | −0.00 | $-0.02^{***}$ | $0.17^{***}$ |

*Notes:* Columns 'All', 'CDU = 0' and 'CDU = 1' report mean values for the full sample, subsample where $\text{CDU}_i = 0$ and subsample where $\text{CDU}_i = 1$, respectively; Correlation table column numbers correspond to: (1) Total gatekeeping errors, (2) Unnecessary hospitalization, (3) Wrongful discharge, (4) CDU admission, and (5) ED congestion; Pre-standardized mean (standard deviation) of CDU congestion is equal to 0.65 (0.22), 0.65 (0.22), 0.63 (0.21) for 'All', 'CDU=0' and 'CDU=1', respectively; Correlation coefficients significant with $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

and obstetric operating theaters (Freeman et al. 2017). Thus, when the CDU is busy we expect fewer CDU admissions, satisfying the relevance condition.

For patients who are not admitted to the CDU, CDU congestion should have no direct effect on their likelihood of being an unnecessary hospitalization or wrongful discharge. For patients admitted to the CDU, however, CDU congestion may affect decision-making in the CDU. Therefore, CDU congestion is only a valid IV for the subset of patients who were *not* admitted to the CDU (i.e., 90.1% of the final sample). To account for this, we include in the outcome equation as a control an additional interaction between $CDU_i$ and $zCDUCong_i$. This variable will take a value of zero (and hence have no influence) if patient $i$ is not admitted to the CDU, and it will take a value equal to $zCDUCong_i$ if patient $i$ is admitted to the CDU. (Hence, it controls for the fact that congestion levels in the CDU may directly affect the patient $i$'s likelihood of being a gatekeeping error.) One might also be concerned that CDU congestion could be correlated with busyness in the main hospital, which could influence admission decisions. To account for this, in both the selection and outcome equations we control for the congestion level of the hospital (calculated in the same way as CDU congestion).

Hypothesis testing of the IVs to identify any signs of over-, under-, or weak identification provide strong evidence that the IVs are valid ($p$-values $> 0.10$), relevant ($p$-values $< 0.001$), and achieve a maximal relative bias significantly less than 10%, as desired (see §EC.4 of the e-companion). Our results are also robust to using the IVs individually.

### 6.3. Results

Table 5 compares coefficient estimates from a set of probit models (columns (1p)–(3p)) against the main biprobit models (columns (2b) and (3b)). Column (1p) shows that both instrumental variables are relevant in the selection equation, while columns (2p) and (3p) suggest that in comparison to parents whose gatekeeping decisions are made in the ED, patients referred to the CDU are no less likely (coef. $= -0.008$, $p$-value $= 0.619$) to be admitted unnecessarily but are more likely to be wrongfully discharged (coef. $= 0.130$, $p$-value $< 0.001$). Thus, taken at face value, it appears as if the CDU may be detrimental rather than helpful. However, these probit results should be interpreted with caution, since they fail to account for selection effects.

**Table 5 Coefficient estimates for CDU impact.**

| | Probit | | | Biprobit | |
|---|---|---|---|---|---|
| | (1p) CDU | (2p) AdmErr | (3p) DischErr | (2b) AdmErr | (3b) DischErr |
| CDU referral | – | $-0.008$ | $0.130^{***}$ | $-0.449^{***}$ | $-0.441^{***}$ |
| | | (0.017) | (0.027) | (0.036) | (0.101) |
| CDU length of stay | – | $-0.004^{*}$ | $-0.000$ | $-0.003^{*}$ | $-0.000$ |
| | | (0.002) | (0.002) | (0.002) | (0.002) |
| CDU congestion | $-0.076^{***}$ | – | – | – | – |
| | (0.003) | | | | |
| Phys. CDU use rate | $1.031^{***}$ | – | – | – | – |
| | (0.021) | | | | |
| $\rho$ | – | – | – | $0.247^{***}$ | $0.332^{***}$ |
| | | | | (0.019) | (0.059) |
| N | 377,346 | 377,346 | 377,346 | 377,346 | 377,346 |
| Log-lik | −94,990 | −53,824 | −14,698 | −148,716 | −109,666 |
| Pseudo-R$^2$ | 0.220 | 0.201 | 0.080 | – | – |

*Notes:* See table 3 for control structure; *Robust standard error* in parentheses; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models; $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

Turning our attention fully to the biprobit models in Table 5, we see evidence of positive correlation between the selection and outcome equations, with estimated correlation coefficients $\rho = 0.247$ ($p$-value$< 0.001$) and $\rho = 0.332$ ($p$-value$< 0.001$) in columns (2b) and (3b), respectively. This indicates that there are unobservables that, on average, make a patient more likely to be admitted to the CDU and also more prone to becoming an unnecessary hospitalization or wrongful discharge. This is consistent with our expectation: If CDU admission control is effective, then those patients who are admitted to the CDU ought to have a more uncertain disposition than the average ED arrival (based on both observables and unobservables). Other findings would indicate that the CDU is being used inappropriately.

After accounting for endogeneity, the biprobit models provide strong evidence that patients admitted to the CDU are significantly less likely to (i) be hospitalized unnecessarily (coef.$= -0.449$, $p$-value$< 0.001$ in column (2b)) and (ii) be wrongfully discharged (coef.$= -0.441$, $p$-value$< 0.001$ in column (3b)). This confirms our hypothesis that routing patients with unresolved diagnoses through the CDU can help to significantly reduce the number of unnecessary hospitalizations (Hypothesis 1). Importantly, this reduction in unnecessary admission does not come at the cost of an increase in wrongful discharge. In fact, and to answer Empirical Question 1, CDU admission decreases wrongful discharge rates, suggesting that gatekeeping decisions are generally of higher quality in the CDU. Our results also indicate that a patient's reduction in the chance of unnecessary hospital admission due to CDU admission increases along with the length of time (in hours) that the patient spends in the CDU (coef.$= -0.003$, $p$-value$= 0.048$ in column (2b)). Overall, while time is an important mechanism through which the CDU helps to reduce unnecessary hospitalizations, other important mechanisms, such as the culture of discharge and tighter consultant oversight, also appear to contribute to the CDU's effectiveness.

To see how much better gatekeeping decisions are in the CDU than in the ED, we convert coefficient estimates to average treatment effects (ATEs) and average treatment effects on the

**Table 6 Inferred unnecessary hospitalization and wrongful discharge rates using different CDU allocation strategies.**

| | Existing | Counterfactuals | | | | Effect Sizes | |
|---|---|---|---|---|---|---|---|
| | | | | | | ATE | ATT |
| CDU usage rate | 9.9% | 0% | 9.9% | 25% | 100% | – | – |
| CDU allocation strategy | No Change | None | Random | High Risk Only | All | – | – |
| Unnecessary hospitalizations | 4.34% | 4.96% | 4.68% | 2.66% | 2.17% | −2.79 p.p | -6.00 p.p |
| Wrongful discharges | 0.71% | 0.92% | 0.86% | –‡ | 0.28% | −0.64 p.p | -2.05 p.p |

*Notes:* CDU usage rate reports the % of patients referred to the CDU under different scenarios; CDU allocation strategy details which patients are referred into the CDU; 'Existing' column reports current unnecessary hospitalization and wrongful discharge rates; 'Counterfactuals' columns report rates inferred from results in Tables 5 and 7 under four scenarios: (i) no patients are routed through the CDU; (ii) the same % of patients are referred into the CDU as observed in the data, but referral is random; (iii) the 25% of patients identified as having a high ex ante risk of being admitted unnecessarily are routed into the CDU; and (iv) a hypothetical best case scenario, in which there is sufficient capacity for all patients to be routed through the CDU; ATE and ATT report the average treatment effect and average treatment effect on the treated, respectively, in percentage points (p.p); ‡ Wrongful discharge rate omitted for reasons outlined in Footnote 7.

treated (ATTs). These results, reported in Table 6, show that without the CDU, the rates of unnecessary hospitalization and wrongful discharge would have been 4.96% and 0.92%, respectively, while in our sample these rates were 4.34% and 0.71%. On the other hand, if all ED patients were routed through the CDU, our model predicts that these rates would instead drop by 50% and 60% to 2.17% and 0.28%, respectively. The CDU thus reduces unnecessary hospitalizations (ATE$=-2.79$ percentage points (p.p.)) and wrongful discharges (ATE$=-0.64$ p.p.).

The ATTs are even larger than the ATEs, with values $-6.00$ p.p. and $-2.05$ p.p. for unnecessary hospitalization and wrongful discharge, respectively. This suggests that rather than randomly allocating patients to the CDU, ED physicians are especially good at referring those patients who benefit most from the CDU. This speaks to the success of the CDU's admission control policy, consistent with Hypothesis 3. In fact, if patients were simply allocated to the CDU at random, then we estimate that the rates of unnecessary hospitalization and wrongful discharge would have been 4.68% and 0.86%, respectively (see Table 6). This indicates that approximately 50% ($=\frac{4.68-4.34}{4.96-4.34}$) of the net beneficial effect of the CDU with respect to unnecessary admissions can be attributed to effective admission control.

In Table 7, we report the results of interaction models that test whether patients with a higher ex ante risk of unnecessary admission also experience a larger marginal benefit from referral to the CDU. The low, low-medium, medium-high and high-risk groups contain the same number of patients, and 1.6%, 9.9%, 17.6% and 10.4% of patients in each group, respectively, are admitted to the CDU, while 0.1%, 0.7%, 4.1% and 12.5%, respectively, are classified as unnecessary hospitalizations. Results from the interaction model indicate, interestingly, that admitting a low-risk patient to the CDU can increase their likelihood of unnecessary admission (coef.$=0.637$, $p$-value$<0.001$).[10] For the low-medium risk category, we find that admission to the CDU has no tangible

[10] Note that this finding is based on a relatively small number of observations (only 103 low-risk patients are hospitalized unnecessarily, with 33 of these admitted via the CDU), and CDU admission for this subset of patients occurs rarely (in only 1.6% of cases). Moreover, there may be unexplained factors, unobservable to researchers but available to ED physicians, which have led to us misclassify the risk level of these patients. If this were the case,

**Table 7** **Effect of CDU referral on unnecessary hospitalizations by patient risk category.**

| | Coefficients | | | Effect Size | | | |
|---|---|---|---|---|---|---|---|
| | Base Effect | CDU Interaction | Marginal Difference | CDU = 0 (%) | CDU = 1 (%) | ATE (p.p.) | ATT (p.p.) |
| Low Risk | – | $0.560^{***}$<br>(0.097) | $0.560^{***}$<br>(0.097) | 0.09 | 0.49 | 0.40 | 1.47 |
| Low-Med Risk | $0.315^{***}$<br>(0.043) | 0.043<br>(0.054) | $-0.517^{***}$<br>(0.092) | 0.70 | 0.79 | 0.09 | 0.21 |
| Med-High Risk | $0.679^{***}$<br>(0.048) | $-0.390^{***}$<br>(0.041) | $-0.433^{***}$<br>(0.041) | 5.04 | 2.16 | −2.88 | −4.85 |
| High Risk | $0.752^{***}$<br>(0.053) | $-0.611^{***}$<br>(0.042) | $-0.221^{***}$<br>(0.027) | 14.01 | 4.79 | −9.22 | −15.10 |

*Notes:* See Table 3 for control structure; Estimation made using the biprobit model specification; $\rho = 0.279^{***}$, $N = 377{,}346$, Log-lik $= -150{,}639$; *Robust standard error* in parentheses; Base effect column specifies the difference in unnecessary hospitalization rates across risk categories; CDU interaction column reports the effect of CDU admission on unnecessary hospitalization rates by risk category; Marginal difference column tests for statistically significant differences in the effect of CDU admission across risk categories; CDU = 0 (resp., CDU = 1) column reports the expected % of patients who would have been unnecessarily hospitalized if no (resp., all) patients in that risk category were routed through the CDU; ATE and ATT report the average treatment effect and average treatment effect on the treated, respectively, for patients from each risk category, in percentage points (p.p); Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

impact on a patient's propensity to be hospitalized unnecessarily (coef. $= 0.043$, $p$-value $= 0.420$). Meanwhile, admission to the CDU is beneficial for the medium-high risk patients (coef. $= -0.390$, $p$-value $< 0.001$) and even more so for those in the high-risk group (coef. $= -0.611$, $p$-value $< 0.001$). The marginal difference column in Table 7 tests for statistical differences between the effect of the CDU across the risk groups and finds that all are significant at the 0.1% level. Overall, these results are consistent with Hypothesis 2, indicating that patients who are more naturally predisposed to being admitted unnecessarily also stand to benefit most from admission to the CDU.

We see a further demonstration of the potential for the CDU to facilitate significant reductions in gatekeeping error rates if we return to the counterfactual scenario presented in Table 6. We report findings from a counterfactual scenario in which the CDU is expanded to accommodate the 25% of patients who are classified as having a high ex ante risk of unnecessary admission. Based on the results from our interaction model, we find that if we were able to accurately identify and route all of these patients into the CDU, then the unnecessary hospitalization rate would drop from the current level of 4.34% to just 2.66%. This is close to the 2.17% unnecessary admission rate that would be achieved from admitting 100% of the highest-risk patients to the CDU.

## 6.4. Robustness

In the e-companion we report the results of robustness tests performed to verify the CDU's benefits in reducing gatekeeping errors. In §EC.5.3, we explore a combined measure of gatekeeping errors, total errors, and find that the results are similar to those for unnecessary hospitalizations due to the very low relative frequency of wrongful discharge. In §EC.6, we report results using 1:1 nearest

then admission of these patients to the CDU may act as a proxy for these other unobserved factors, explaining the positive coefficient. We caution against over-interpretation of this particular finding. That being said, if the coefficient is correctly estimated, it suggests that a low-risk patient admitted to the CDU may be more likely to subsequently be admitted unnecessarily.

neighbor matching to better balance the covariate distributions between the treatment and control groups (i.e., the groups of patients admitted or not admitted to the CDU). Results for unnecessary admissions are almost identical, while the effect of the CDU on wrongful discharges becomes insignificant. This is likely due to the rarity of wrongful discharges in the matched sample. In §EC.7, we test the robustness of the results to different definitions of wrongful discharge (where we (i) use a three-day readmission window and (ii) remove the restriction that the diagnosis categories must be the same across the two ED visits) and unnecessary hospitalization (where we (i) use a 12- or 48-hour or two-night time window for discharge after admission, (ii) require that the patient not only does not receive treatment but also that no diagnosis is assigned, and (iii) compare inpatient LOS to median LOS for patients within the same disease category). All results are consistent with those presented in §6.3.

## 7. Counterfactual Analysis

Having established that the CDU does improve the accuracy of the disposition decision for those patients routed through it, the natural question is: What is the aggregate effect of the CDU on all patients who visit the ED? As discussed earlier, CDU resources could be redeployed in the ED, increasing the ED's capacity and thereby reducing congestion. This could improve decision-making in the ED and lower the gatekeeping error rates. Thus, even though we find that the two-stage gatekeeping system reduces gatekeeping errors for those patients fully routed through it, the overall benefits of the second stage, if any, are not obvious. To examine the combined system as a whole, we perform a counterfactual analysis.

In order to calculate the counterfactual, we first need to understand the impact of congestion on decision making by physicians operating in the first gatekeeping stage (i.e., within the fast ED and not the CDU). This will allow us to determine how much more accurate decisions would be in a single-stage system with expanded capacity.

### 7.1. Impact of congestion on decision making in the fast ED

We next identify how ED congestion impacts decisions made by ED physicians in the fast ED. This means we study the upper half of the two-stage gatekeeping process shown in Figure 1. However, as congestion increases, so too might the rate at which ED physicians leverage the CDU option. This could change the composition of the patients for whom the ED physicians are making admission and discharge decisions. While we account partially for these differences with our set of controls (reported in Table 3), there may still be factors unobservable to us but observable to the physician (e.g., patient acuity, medical history) that influence whether the physician leverages the CDU option. Thus, despite only 9.9% of patients being routed into the CDU, it is necessary to ensure that our findings are not confounded by unobserved differences in the patient case mix arising from changes in CDU usage as congestion levels increase.

To account for the selection effect described above, we use a similar estimation strategy to that used in identifying the impact of the CDU, as described in §6.1. However, instead of using a biprobit model, we estimate the selection and outcome equations jointly with a Heckman probit sample selection (heckprob) model using full information maximum likelihood (Maddala 1983; see §EC.3.1 of the e-companion for more information on this model). The heckprob model corrects for potential sample selection bias arising when the outcome (here, whether or not a patient is unnecessarily hospitalized or wrongfully discharged by an ED physician) is only observed when the patient is selected into the sample (which here means that the ED physician chooses not to refer the patient to the CDU). Bias arises from the fact that (a) patients may not be assigned to the CDU at random and (b) the coefficients, in particular the coefficient of interest, $zEDCong$, may vary depending on whether or not the patient was admitted to the CDU. Like the biprobit model, the heckprob model requires us to estimate the selection and outcome equations under the assumption that their errors are jointly distributed according to the standard bivariate normal distribution. However, in addition we must (1) censor the outcome variable, $AdmErr_i$ or $DischErr_i$, whenever $CDU_i = 1$, and (2) set $\alpha_2 = \beta_2 = 0$ in the outcome equation.[11]

Model (1p) in Table 8 indicates that as congestion in the fast ED increases, patients are referred to the CDU more frequently. Since this may lead to differences in the patient mix in the fast ED when it is busy compared to when it is quiet, we must correct with the heckprob models for potential endogeneity (though, for completeness, we also report the results using a probit model specification in models (2p) and (3p)). After correcting for endogenous selection using the heckprob models, we find evidence that as fast ED congestion increases, ED physicians are more likely to admit patients to the hospital unnecessarily (coef. $= 0.036$, mfx. $= 0.30\%$, $p$-value $< 0.001$ in heckprob (2h)). This is consistent with Hypothesis 4. At the same time, ED physicians become less likely to wrongfully discharge patients when the fast ED is congested (coef. $= -0.018$, mfx. $= -0.03\%$, $p$-value $= 0.036$ in heckprob (3h)). This evidence suggests that when the fast ED becomes congested, physicians overcompensate for the increase in clinical uncertainty by increasing the rate at which they admit uncertain cases, which surpasses the rate required to keep the wrongful discharge rate constant.[12]

[11] In addition, we also must drop ED LOS from the vector $\mathbf{X}_i$ of controls. This is because we are interested in the total effect of congestion on errors and in order to capture the total effect, we need to be careful to avoid controlling for any factors that might mediate the relationship between congestion and the error rate. ED LOS is one such mediator. Specifically, when the system is congested, there is a delay in the start of treatment (as shown in Figure 3), increasing the average time that a patient will spend in the ED. As noted in §3.2 and due to the four-hour waiting time target, any delay in the start of treatment also directly reduces the time available for an ED physician to spend with the patient, thus increasing the likelihood of error. Therefore, ED congestion affects the time that the patient spends in the ED, and the amount of time that the patient spends in the ED affects their likelihood of being admitted unnecessarily or wrongfully discharged. This makes ED LOS a bad control in this analysis. Note also that CDU LOS and CDU congestion drop out of the estimation automatically, since the outcome is censored whenever $CDU_i = 1$. This means that the outcome equation only contains observations for which $CDU_i = 0$, in which case CDU LOS and CDU congestion always equal zero (a patient who is not admitted to the CDU has an LOS of zero and is not directly

**Table 8 Coefficient estimates to establish ED physicians' response to increased congestion in the fast ED.**

| | Probit | | | Heckprob | |
|---|---|---|---|---|---|
| | (1p) CDU | (2p) AdmErr | (3p) DischErr | (2h) AdmErr | (3h) DischErr |
| ED congestion | 0.053$^{***}$ | 0.033$^{***}$ | −0.019$^{*}$ | 0.036$^{***}$ | −0.018$^{*}$ |
| | (0.004) | (0.005) | (0.009) | (0.005) | (0.009) |
| $\rho$ | – | – | – | −0.208$^{***}$ | −0.111 |
| | | | | (0.047) | (0.096) |
| N | 377,346 | 339,990 | 339,990 | 339,990 | 339,990 |
| Log-lik | −97,596 | −47,029 | −12,554 | −144,594 | −110,150 |

*Notes:* See Table 3 and Footnote 11 for control structure; CDU estimation made on the full sample; Unnecessary hospitalization and wrongful discharge models estimated on the sample of patients not admitted to the CDU; *Robust standard error* in parentheses; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

To give an idea of the scale of the effects, we compare the expected CDU admission rate and each type of gatekeeping error under two extreme congestion states in the fast ED: one in which the congestion level is very low ($2\sigma$ below the mean) and one in which it is very high ($2\sigma$ above the mean). We find that under the high congestion scenario, the probability of a patient being admitted to the CDU, hospitalized unnecessarily from the fast ED, or wrongfully discharged from the fast ED is 11.49%, 5.56%, 0.67%, respectively. This compares with 8.53%, 4.37%, and 0.80%, respectively, under the low congestion scenario.[13] Moving from the low to high congestion scenario thus represents an approximate increase in a patient?s likelihood of CDU admission by 34.7% and of experiencing unnecessary hospitalization by 27.1%, and it decreases their likelihood of being wrongfully discharged by 17.2%. The congestion state of the fast ED thus has a surprisingly large impact on ED physicians' decision making.

## 7.2. The system-wide effect of a second gatekeeping stage

Having established the effect of congestion on the first gatekeeping stage, we are now ready to determine the total system-wide effect of operating a two-stage gatekeeping system comprising a fast-and-slow ED compared to a single-stage system with an expanded fast-only-ED.

In this analysis we need to account for the fact that any increase in ED capacity would translate into a reduction in ED congestion, as the resources (e.g., physicians, nurses, beds) consumed by the CDU would be available for ED use instead. To adjust for this, we use the same approach we used in

affected by congestion levels).

[12] One alternative explanation for these findings could be that admitting a patient is administratively less time-consuming for ED physicians than discharging them, i.e., busy physicians may err on the side of admission to save time. However, in our particular context the opposite is true: Additional paperwork (a venous thromboembolism assessment and drug chart) must be completed in order to admit a patient, so admission is in fact more time consuming. All else being equal, we would expect fewer and not more unnecessary admissions, ruling out this alternative explanation.

[13] The calculations for unnecessary admission and wrongful discharge effectively assume that the CDU does not exist, so they give the respective rates under the two workload scenarios assuming all gatekeeping decisions were made in the fast ED (i.e., no patients were referred to the CDU). This explains why the average unnecessary admission and wrongful discharge rates reported here are higher than those in Table 2, as error rates are higher on average when there is no CDU.

§5.3 to create a measure of ED capacity, $CapacityED_h$, for every hour $h$, and we now create a measure for CDU capacity, $CapacityCDU_h$. Together (i.e., taking $CapacityED_h + CapacityCDU_h$), these variables capture the expected amount of capacity in the combined system at every hour $h$. Setting $EDCong_i^* = CensusED_i/(CapacityED_{h_i} + CapacityCDU_{h_i})$, where $h_i$ is the hour patient $i$ arrives, gives us our estimate of what the congestion would have been in the combined system when patient $i$ arrived.[14] To ensure that the original and updated measures of congestion in the fast ED are on the same scale, we then standardize using the original mean, $\mu(EDCong_i)$, and standard deviation, $\sigma(EDCong_i)$. This shows that if the resources consumed by the CDU were redeployed in the fast ED, the capacity of the fast ED would increase by approximately 20%, which would reduce average congestion levels in the fast ED by approximately $0.61\sigma$.[15]

Substituting the original values of $zEDCong_i$ for the updated values achieved through pooling ED and CDU capacity into heckprob (1e), we estimate that in the fast-only-ED system the unnecessary hospitalization rate would be reduced by 0.19 p.p. In §6.3 we estimated that the unnecessary hospitalization rate would rise from the current level of 4.34% with the CDU to 4.96% if no patients were routed through the CDU (see Table 6). However, this ignored the possibility of closing the CDU and redistributing its resources to increase capacity in the fast ED. We estimate that if this choice were made, then the unnecessary hospitalization rate would equal 4.77% ($= 4.96\% - 0.19\%$) – still a deterioration relative to the status quo of 4.34%. Moreover, this rate is still higher than the rate that would be obtained from simply allocating patients at random to the CDU, as calculated in §6.3 and reported in Table 6 to equal 4.68%. That said, the absence of effective admission control significantly diminishes the beneficial effect of the two-stage gatekeeping system relative to an enlarged single-stage system.

We next investigate the effect of expanding the CDU to accommodate the 25% of patients who are classified as having a high ex ante risk of experiencing unnecessary admission. As reported in Table 6, the unnecessary hospitalization rate would drop to 2.66%. However, expanding the CDU in this way would require a major reallocation of resources from the fast ED. To identify the impact of this choice on congestion in the fast ED, we must reconstruct $CapacityCDU_h$ from the historic data under the assumption of an expanded CDU. To do so, we assume each patient designated as high-risk was admitted to the CDU at the time they left the fast ED, with their LOS in the CDU drawn randomly from the empirical distribution of CDU LOS. We find that to accommodate this expansion of the CDU by approximately 125%, fast ED capacity would

[14] We take a conservative view and assume that all patients who were treated in the CDU could have instead been relocated elsewhere in the hospital without the need for any additional institutional capacity, meaning that all resources from the CDU can be redeployed to the fast ED. We thus estimate an upper bound on the gains that could be achieved from pooling fast ED and CDU capacity.

[15] If we use number of beds as a proxy for capacity instead, then we arrive at very similar results. Adding the 8 beds in the CDU to the 30 adult cubicles in the fast ED would correspond to a capacity increase of approximately 26%.

have to be reduced by approximately 26%, resulting in an increase in average ED congestion levels by approximately $1.35\sigma$. This increase in congestion would result in a 0.24 p.p. increase in the unnecessary hospitalization rate for those patients not referred to the CDU, meaning that under this scenario, we predict that the rate of unnecessary hospitalizations would equal 2.90% $(= 2.66\% + 0.24\%)$. This is a 33% reduction from the status quo, suggesting that an increase in CDU capacity may be warranted, especially if it is possible to selectively refer these high-risk patients to the CDU.

Overall, we find that the two-stage gatekeeping system that is in place in our study hospital is more effective in reducing gatekeeping errors than an alternative pooled system that combines the resources of the fast ED and the CDU into a single fast-only-ED. This is despite the fact that such a setup causes a negative spillover (by way of increased congestion and a consequent increase in gatekeeping errors) onto those patients treated in the fast ED. We also find that increasing the capacity of the CDU in our study hospital from 9.9% to 25% of ED patients has the potential to reduce the rate of unnecessary hospitalization significantly, by up to 33%.

## 8. Conclusions

This study expands the notion of gatekeeping by providing an in-depth empirical examination of a fast-and-slow ED. Our data reveal a number of key insights as to when, in particular, such a two-stage gatekeeping system is likely to outperform an expanded single-stage system.

First, in order for a two-stage gatekeeping system to be worthwhile, each stage must have a different strategic emphasis. In our study context, the fast ED focuses on stabilization, assessment, and acute treatment. While the disposition decision for each patient is still the ultimate decision that physicians must make, this gatekeeping function is often overshadowed by the acute needs of those patients still waiting to be seen. By contrast, the CDU (i.e., the slow ED) separates the gatekeeping decision from the stabilization, assessment and acute treatment function for those patients for whom the appropriate disposition decision is not immediately clear. Its primary focus is the appropriate placement (either in the hospital or at home) of patients following additional observation and further assessment for diagnosis or exclusion of specific conditions. The CDU therefore prioritizes gatekeeping accuracy and is supported by an alignment of resources (e.g., more experienced decision-making staff) and processes (e.g., patients are expected to spend more time in the unit) with the specificities of this task. By contrast, if the roles and functions of the two units were largely identical, then a two-stage system would likely confer little to no advantage and may instead increase unnecessary friction (e.g., from handoffs).

Second, a two-stage system is appropriate for a patient population with a certain degree of complexity and acuity, as not all patients benefit from admission into the second gatekeeping stage. For some patients the appropriate disposition decision is already clear-cut, a gatekeeping decision made in the first stage would be just as accurate as any second-stage decision, and any time spent in the second stage would be wasted and, moreover, costly for the system. Therefore in order for the fast-and-slow ED concept to make sense, there must be a sufficiently large subset of patients who stand to benefit from being routed through the second stage. At a regional trauma center like our study ED, which caters to highly acute and complex cases, there is a large pool of high-risk admissions who benefit greatly from the CDU. We posit that in other emergency medicine contexts, like urgent treatment centers and walk-in clinics that typically cater to lower acuity patients, the volume of such patients may be insufficient to justify a two-stage gatekeeping system.

Third, we show that admission control is key to the success of the two-stage gatekeeping system. In our study hospital, we find that those patients who are routed through the CDU are also those who stand to benefit more from admission. By contrast, if patients were instead randomly allocated to the CDU then its effectiveness would be significantly diminished. This means that there must be (a) strict criteria for second-stage admission to ensure that the unit is not simply used as a workload buffer, and (b) first-stage gatekeepers must be able to identify with a relatively high degree of accuracy the subset of patients who stand to benefit most from referral to the second stage. This finding also suggests that a prediction model of a patient's likelihood of undergoing unnecessary hospital admission could be a useful decision support tool for ED clinicians deciding which patients to refer to the CDU.

Fourth, from a systems-level perspective, we observe that in a two-stage system there is tension between the first and second stage in terms of resources and performance, and benefits must be weighed with potential adverse effects. The second gatekeeping stage draws resources away from the first stage, increasing congestion and potentially leading to worse initial gatekeeping decisions. In our study context we find that while the CDU benefits those patients routed through it, its presence actually harms those patients (the majority) remaining in the fast ED. On balance, in our context we find that the CDU is a net benefit for the ED overall. However, in other contexts this may or may not be the case depending on the tradeoff between (i) the negative effects of increased congestion in the first stage and (ii) the benefits that come from routing a subset of patients through the second stage.

A natural question following from these findings would be: If two gatekeeping stages are better than one, at least in this context, could the system be extended to three or even more stages? We posit that the criteria discussed above suggest a natural limit to the reasonable number of gatekeeping stages. In order for a third stage to make sense, for example, the roles of the second and third stages must be sufficiently distinct, there must be enough patients who stand to benefit from being placed there, it must be possible to identify these patients in advance, and any new gatekeeping stage must add enough value to the whole system to justify the removal of resources from the other stages. Given the fact that according to the operational policy of our study hospital, the CDU already has a focus on accurate placement of patients in the hospital or at home, as well as the fact that error rates are already low for patients routed through this unit, it is unlikely that a third gatekeeping stage would confer any additional advantage in this context. Instead, we would advocate for an expanded CDU, as proposed in §7.2.

While this study has been able to demonstrate the potential effectiveness of a two-stage gatekeeping approach in reducing gatekeeping errors and identify conditions under which such a system is especially beneficial, our data do not enable us to tease out the exact mechanisms that contribute to the CDU's net beneficial effect in our study context. While it has been possible to test some mechanisms (e.g., the effects of time and admission control) directly, others (e.g., the effect of consultant oversight) remain conjectures, albeit probable conjectures backed by contextual evidence. Furthermore, while this analysis demonstrates the advantage of the CDU over a pooled ED in the study hospital, the question remains of how to distribute resources between the fast ED and CDU to optimize patient flow and minimize gatekeeping errors. Answering this question would require analytical work that goes beyond the scope of this paper, and so it is left for future research.

Although our study focuses on emergency care, the benefits of multi-stage gatekeeping are likely to extend to other industries and health contexts. For example, accurate diagnosis of rare diseases in primary care takes seven years on average in the US and five years in the UK (Shire 2013). Such cases are costly because patients visit their primary care physician multiple times, undergo multiple tests, and see multiple specialists. Our results suggest that a potential solution may be to designate a subset of more experienced primary care physicians (with a track record of identifying rare diseases) as second-stage gatekeepers, allowing primary care physicians to refer patients to them. More generally, our findings demonstrate that two-stage gatekeeping systems could reduce overuse of inappropriate specialist services while improving the accuracy of referrals, a win-win for both the system and the patient.

Finally, in this study we have focused on the benefit of the two-stage gatekeeping system in reducing the rates of gatekeeping errors; however, there may be other benefits. For example, the CDU also appears to act as a workload buffer: As the fast ED becomes congested, more patients are referred to the CDU. If the CDU were not present, then the hospital inpatient units might instead be used for this purpose. If patients must be in one or the other, the CDU may be the better option because admission to the hospital exposes patients to additional risks and is costly (as discussed in §1). When patients are referred unnecessarily to the CDU, however, they spend less time there than they would if they were in an inpatient unit instead (see §EC.1 of the e-companion). While some of the other potential benefits of the multi-stage gatekeeping approach lie outside the scope of this paper, they may be well worth future exploration.

## References

Alizamir S, de Véricourt F, Sun P (2013) Diagnostic accuracy under congestion. *Management Science* 59(1):157–171.

Batt R, Terwiesch C (2015) Waiting patiently: An empirical study of queue abandonment in an emergency department. *Management Science* 61(1):39–59.

Baugh CW, Venkatesh AK, Bohan JS (2011) Emergency department observation units: a clinical and financial benefit for hospitals. *Health Care Management Review* 36(1):28–37.

Baugh CW, Venkatesh AK, Hilton JA, Samuel PA, Schuur JD, Bohan JS (2012) Making greater use of dedicated hospital observation units for many short-stay patients could save $3.1 billion a year. *Health Affairs* 31(10):2314–2323.

Berry Jaeker J, Tucker A (2017) Past the point of speeding up: The negative effects of workload saturation on efficiency and quality. *Management Science* 63(4):1042–1062.

Blatchford O, Capewell S (1997) Emergency medical admissions: taking stock and planning for winter. *British Medical Journal* 315(7119):1322–1323.

Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, Newhouse JP, Weiler PC, Hiatt HH (1991) Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study I. *New England Journal of Medicine* 324(6):370–376.

Brillman J, Mathers-Dunbar L, Graff L, Joseph T, Leikin JB, Schultz C, Severance Jr HW, Werne C (1995) Management of observation units. *Annals of Emergency Medicine* 25(6):823–830.

Bunik M, Glazner JE, Chandramouli V, Emsermann CB, Hegarty T, Kempe A (2007) Pediatric telephone call centers: how do they affect health care use and costs? *Pediatrics* 119(2):e305–e313.

Chan CW, Farias VF, Escobar G (2017) The impact of delays on service times in the intensive care unit. *Management Science* 63(7):2049–2395.

Chan CW, Green LV, Lu Y, Leahy N, Yurt R (2013) Prioritizing burn-injured patients during a disaster. *Manufacturing & Service Operations Management* 15(2):170–190.

Christensen J, Levinson W, Dunn P (1992) The heart of darkness: The impact of perceived mistakes on physicians. *Journal of General Internal Medicine* 7(4):424–431.

Conley J, OBrien CW, Leff BA, Bolen S, Zulman D (2016) Alternative strategies to inpatient hospitalization for acute medical conditions: a systematic review. *JAMA Internal Medicine* 176(11):1693–1702.

Cooke M, Higgins J, Kidd P (2003) Use of emergency observation and assessment wards: A systematic literature review. *Emergency Medicine Journal* 20(2):138–142.

Cosby KS, Roberts R, Palivos L, Ross C, Schaider J, Sherman S, Nasr I, Couture E, Lee M, Schabowski S, Ahmad I, Scott RD (2008) Characteristics of patient care management problems identified in emergency department morbidity and mortality investigations during 15 years. *Annals of Emergency Medicine* 51(3):251–261.

Croskerry P (2002) Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Academic Emergency Medicine* 9(11):1184–1204.

Denman-Johnson M, Bingham P, George S (1997) A confidential enquiry into emergency hospital admissions on the Isle of Wight, UK. *Journal of Epidemiology and Community Health* 51(4):386–390.

FitzGerald G, Jelinek GA, Scott D, Gerdtz MF (2010) Emergency department triage revisited. *Emergency Medicine Journal* 27(2):86–92, URL `http://dx.doi.org/10.1136/emj.2009.077081`.

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

Freeman M, Savva N, Scholtes S (2019) Economies of scale and scope in hospitals: An empirical study of volume spillovers. *Management Science (Forthcoming)* .

Galipeau J, Pussegoda K, Stevens A, Brehaut JC, Curran J, Forster AJ, Tierney M, Kwok ES, Worthington JR, Campbell SG, et al. (2015) Effectiveness and safety of short-stay units in the emergency department: A systematic review. *Academic Emergency Medicine* 22(8):893–907.

Gawande A (2015) Overkill. *New Yorker* URL `http://www.newyorker.com/magazine/2015/05/11/overkill-atul-gawande`.

Gorski J, Batt R, Otles E, Shah M, Hamedani A, Patterson B (2017) The impact of emergency department census on the decision to admit. *Academic Emergency Medicine* 24(1):13–21.

Graber ML, Franklin N, Gordon R (2005) Diagnostic error in internal medicine. *Archives of Internal Medicine* 165(13):1493–1499.

Hasija S, Pinker E, Shumsky R (2005) Staffing and routing in a two-tier call centre. *International Journal of Operational Research* 1(1/2):8–29.

Hopp W, Iravani S, Yuen G (2007) Operations systems with discretionary task completion. *Management Science* 53(1):61–77.

HSCIC (2013) OPCS-4 classification. Technical report, Health & Social Care Information Centre, URL `http://systems.hscic.gov.uk/data/clinicalcoding/codingstandards/opcs4`.

Inouye SK, Zhang Y, Jones RN, Shi P, Cupples LA, Calderon HN, Marcantonio ER (2008) Risk factors for hospitalization among community-dwelling primary care older patients: development and validation of a predictive model. *Medical Care* 46(7):727–731.

Kachalia A, Gandhi TK, Puopolo AL, Yoon C, Thomas EJ, Griffey R, Brennan TA, Studdert DM (2007) Missed and delayed diagnoses in the emergency department: A study of closed malpractice claims from 4 liability insurers. *Annals of Emergency Medicine* 49(2):196–205.

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

KC D, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* 55(9):1486–1498.

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

Kim S, Chan CW, Olivares M, Escobar G (2014) ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* 61(1):19–38.

Kuntz L, Mennicken R, Scholtes S (2015) Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science* 61(4):754–771.

Leape LL (1994) Error in medicine. *JAMA* 272(23):1851–1857.

Lee H, Pinker E, Shumsky R (2012) Outsourcing a two-level service process. *Management Science* 58(8):1569–1584.

Lo SM, Choi KTY, Wong EML, Lee LLY, Yeung RSD, Chan JTS, Chair SY (2014) Effectiveness of emergency medicine wards in reducing length of stay and overcrowding in emergency departments. *International Emergency Nursing* 22(2):116–120.

Maddala G (1983) *Limited Dependent and Qualitative Variables in Econometrics* (New York: Cambridge University Press).

NHS (2013) 2014/15 NHS standard contract. Technical report, NHS England, URL `https://www.england.nhs.uk/nhs-standard-contract/14-15/`, Last accessed: 2016-09-15.

Pope JH, Aufderheide TP, Ruthazer R, Woolard RH, Feldman JA, Beshansky JR, Griffith JL, Selker HP (2000) Missed diagnoses of acute cardiac ischemia in the emergency department. *New England Journal of Medicine* 342(16):1163–1170, URL `http://dx.doi.org/10.1056/NEJM200004203421603`.

Roberts MV, Baird W, Kerr P, O'Reilly S (2010) Can an emergency department-based clinical decision unit successfully utilize alternatives to emergency hospitalization? *European Journal of Emergency Medicine* 17(2):89–96.

Saghafian S, Hopp WJ, Iravani SMR, Cheng Y, Diermeier D (2018) Workload management in telemedical physician triage and other knowledge-based service systems. *Management Science* 64(11):5180–5197.

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.

Saghafian S, Hopp WJ, Van Oyen MP, Desmond JS, Kronick SL (2014) Complexity-augmented triage: A tool for improving patient safety and operational efficiency. *Manufacturing & Service Operations Management* 16(3):329–345.

Schull MJ, Vermeulen MJ, Stukel TA, Guttmann A, Leaver CA, Rowe BH, Sales A (2012) Evaluating the effect of clinical decision units on patient flow in seven canadian emergency departments. *Academic Emergency Medicine* 19(7):828–836.

Shire (2013) Rare disease impact report: Insights from patients and the medical community. Technical report, Shire, URL `https://globalgenes.org/wp-content/uploads/2013/04/ShireReport-1.pdf`, Last accessed: 2016-10-02.

Shumsky R, Pinker E (2003) Gatekeepers and referrals in services. *Management Science* 49(7):839–856.

Shurtz I (2013) The impact of medical errors on physician behavior: Evidence from malpractice litigation. *Journal of Health Economics* 32(2):331–340.

Smith M, Higgs J, Ellis E (2008) Factors influencing clinical decision making. *Clinical Reasoning in the Health Professions*, 89–100 (Butterworth Heinemann Elsevier), 3rd edition.

Song H, Tucker A, Graue R, Moravick S, Yang J (2019) Capacity pooling in hospitals: The hidden consequences of off-service placement. *Management Science (Articles in Advance)* .

Studdert DM, Mello MM, Gawande AA, Gandhi TK, Kachalia A, Yoon C, Puopolo AL, Brennan TA (2006) Claims, errors, and compensation payments in medical malpractice litigation. *New England Journal of Medicine* 354(19):2024–2033.

Wilde J (2000) Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters* 69(3):309–312.

Zhang Z, Luh H, Wang C (2011) Modeling security-check queues. *Management Science* 57(11):1979–1995.

## Appendix A: Physician-level Controls

In §6.2 we introduce the history of CDU use by the physician assigned to a patient as an instrumental variable. Here we elaborate on how this IV is calculated.

We wish to identify the propensity of a physician to admit patients to the CDU after controlling for observable differences in patient characteristics. To do this, we first estimate a probit model of the form

$$CDU_i^* = \delta_0 + \mathbf{T}_i\boldsymbol{\delta}_1 + \mathbf{D}_i\boldsymbol{\delta}_2 + \mathbf{C}_i\boldsymbol{\delta}_3 + \epsilon_i^\delta\,, \tag{5}$$

$$CDU_i = \mathbb{1}[CDU_i^* > 0]\,, \tag{6}$$

where $\mathbf{T}_i$, $\mathbf{D}_i$ and $\mathbf{C}_i$ specify the temporal, patient- and diagnosis-related and contextual controls outlined in Table 3, and where $\epsilon_i^\delta \sim \mathcal{N}(0,1)$, $CDU_i^*$ is a latent variable and $CDU_i$ is the observed dichotomous variable that indicates whether the patient was sent to the CDU. This model gives the patient's baseline risk of being admitted to the CDU if treated by an " average" physician. We then take the fitted values from the auxiliary equation, $\widehat{CDU_i^*}$, and estimate a random effects probit model of the form

$$CDU_{i_{pm}}^* = \delta_{pm} + \widehat{CDU_{i_{pm}}^*} + \epsilon_{i_{pm}}^\delta\,, \tag{7}$$

$$CDU_{i_{pm}} = \mathbb{1}[CDU_{i_{pm}}^* > 0]\,, \tag{8}$$

where $\epsilon_{i_{pm}}^\delta$, $CDU_{i_{pm}}^*$ and $CDU_{i_{pm}}$ are as previously defined but for the subset of observations $i$ assigned to physician $p$ in the 12 month period $[m-12, m-1]$, indexed $i_{pm}$. The random intercept $\delta_{pm}$ then captures variation in CDU admission rates across physicians and within physicians over time. The value of the IV for a patient who arrives in month $m$ and is assigned to physician $p$ is then set equal $\delta_{pm}$.

The controls in $\mathbf{P}_i$ of Table 3, which capture a physician's historic unnecessary hospitalization and wrongful discharge rates, are calculated in the same way as for CDU admission propensity.

# e-companion to
# "Gatekeeping Under Congestion: An Empirical Study of Referral Errors in the Emergency Department"

## Appendix EC.1: Comparison of Inpatient (Specialist) and CDU LOS

The results in our main paper suggest that implementing an intermediate unit between the ED and hospital inpatient units to serve patients for whom considerable diagnostic uncertainty exists (in our case, the CDU) can help reduce the number of unnecessary hospital admissions. However, unless this intermediate unit operates more efficiently than a standard inpatient unit, it offers little benefit and all patients in the CDU should simply be admitted to the hospital instead. Here we compare these two alternatives.

Ignoring wrongful discharges (regarding which the CDU may offer an additional advantage), our sample of admitted patients includes five patient categories. There are patients (1) admitted necessarily from the ED to an inpatient bed or (2) admitted unnecessarily from the ED to an inpatient bed. There are also patients who are (3) admitted from the ED to the CDU and then discharged, (4) admitted from the ED to the CDU and subsequently not deemed to be an unnecessary hospitalization, or (5) admitted from the ED to the CDU and then classed as an unnecessary hospitalization. We assume, conservatively, that the CDU was a waste of time for every patient who was admitted to it (i.e., class (4) or (5)), i.e., their LOS is not reduced at all despite the additional tests, better routing, etc. that the CDU may provide. For all 13,156 patients in our sample who enter the hospital via the CDU, this amounts to 93,077 "wasted" hours. For the CDU to break even, each of the 24,200 patients discharged from the CDU (i.e., those in class (3)) must have an average CDU stay that is more than 3.85 hours shorter than their stay would have been as a hospital inpatient.

To determine whether this condition is satisfied, we again take a conservative approach and assume that if those patients who were discharged from the CDU had instead been admitted to the hospital, then *all* of them would have been identified and discharged within 24 hours (with no treatment performed), i.e., they would have all been unnecessary hospitalizations, in class (2). Thus we compare the length of stay associated with patients of classes (2) and (3). In doing so we account for differences in the characteristics of those patients admitted and subsequently discharged from the hospital directly rather than through the CDU (e.g., since the former may be inherently riskier, they may also be likely to stay longer). To this end, we construct an ordinary least squares (OLS) model that takes the form

$$LOS_i = \lambda_0 + \mathbf{W}_i \boldsymbol{\lambda}_1 + CDU_i \lambda_2 + \epsilon_i^\lambda \,, \tag{EC.1}$$

where $\epsilon_i^\lambda \sim \mathcal{N}(0, \sigma_\lambda^2)$ and $\mathbf{W}_i$ is a control vector that contains all of the temporal, patient- and diagnosis-related, and contextual controls from Table 3. This model indicates that on average, a patient treated in the CDU would have spent 8.78 additional hours in the hospital had they been admitted directly. In other words, the hospital "saves" 199,937 hours of time by giving ED physicians the option to refer patients to the CDU rather than admit them directly to the hospital. This longer patient processing time in hospital inpatient units is not surprising: General wards have higher patient heterogeneity than the CDU, which is specifically set up to route patients towards either hospitalization or discharge. This difference in processing time is also consistent with findings in the medical literature (e.g. Baugh et al. 2012).

Combining the "wasted" and "saved" hours, we find that relative to hospital use, the CDU saves 106,860 hours over 1,840 days, reducing required capacity at our study hospital by approximately 2.4 beds (assuming 100% bed utilization). Stated another way, over the sample period, the CDU consumed 267,748 hours (and the equivalent resources), however, had the CDU not been in place, we conservatively estimate that 467,685 hours ($= 267,748 + 199,937$) would have been required. This implies an efficiency gain of approximately 42.8% ($= 1 - \frac{267{,}748}{467{,}685}$).

## Appendix EC.2: Definition of Unnecessary Hospitalizations

We define an admission as unnecessary if the patient is discharged within 24 hours of admission to an inpatient hospital bed without a recorded treatment in their discharge record. As we mention in Footnote 4 of the main paper, this is an ex post assessment and not all ex post unnecessary hospitalizations are avoidable ex ante. While this remains a limitation of our study, it turns out that under fairly natural assumptions, the estimated effect of congestion on the unnecessary hospitalization rate is a conservative estimate of the effect of congestion on the rate of *avoidable* unnecessary hospitalization. To reach this conclusion, we let:

- $N(c)$ be the expected number of patients admitted to the hospital if the ED congestion level is $c$;
- $N(c) = N_n(c) + N_u(c)$, where $N_n(c)$ and $N_u(c)$ are the expected number of necessary and unnecessary admissions, respectively, as observed ex post after discharge of the patient from the hospital;
- $N_u(c) = N_{ua}(c) + N_{uu}(c)$, where $N_a(c)$ and $N_{na}(c)$ are the expected number of unnecessary admissions that are, respectively, ex ante avoidable and ex ante unavoidable; and
- $r_u(c) = \frac{N_u(c)}{N(c)}$ and $r_{ua}(c) = \frac{N_{ua}(c)}{N(c)}$ be the rates of unnecessary and of avoidable admissions, respectively.

The quantities of interest are the slopes of the regression lines of the rates of unnecessary admission rate and of the rate of avoidable unnecessary admissions as a function of congestion $c$, i.e., $r_u(c)$ and $r_{ua}(c)$. We make three assumptions:

1. The expected numbers of necessary admissions and of unavoidable unnecessary admissions do not change with congestion $c$, i.e., $N_n'(c) = N_{uu}'(c) = 0$.
2. The unnecessary admission rate is an non-decreasing function of congestion ($r_u'(c) \geq 0$).
3. There is a positive number of necessary admissions, i.e., $N = N_n + N_u > N_u$.

These assumptions imply that $0 \leq r_u'(c) \leq r_{ua}'(c)$, i.e., that the slope of the unnecessary admissions rate underestimates the slope of the avoidable unnecessary admissions rate.

Proof. Since $N = N_n + N_u$ and $N_u = N_{uu} + N_{ua}$, assumption (1) implies that $N' = N_u' = N_{ua}'$. Hence

$$r_u' = \frac{N_u' N - N' N_u}{N^2} = \frac{N_{ua}'(N - N_u)}{N^2}$$

and therefore assumptions (2) and (3) imply $N_{ua}' \geq 0$. Hence

$$r_{ua}' = \frac{N_{ua}' N - N' N_{ua}}{N^2} = \frac{N_{ua}'(N - N_{ua})}{N^2} = r_u' + \frac{N_{ua}' N_{un}}{N^2} \geq r_u'.$$

The key assumption for our analysis is that the expected numbers of patients who need to be admitted to the hospital (i.e., those who are (i) ex post necessary and (ii) ex post unnecessary but ex ante unavoidable) are uncorrelated with our measure of congestion, $c$. As $c$ is adjusted for systematic seasonal variation (using a method described in §5.3), we assume that after accounting for seasonal variation, serious acute events or illnesses that require hospitalization occur randomly and independently in the community. This means that elevated congestion levels in the ED are largely caused by patients who are less seriously ill but concerned enough to come to the hospital – patients sometimes referred to as the "worried well."

## Appendix EC.3: Model Specification – Further Details

### EC.3.1. Heckman probit sample selection model

In §7.1 of the paper, we employ a Heckman probit sample selection (heckprob) model in order to identify the effect of congestion on admission and discharge errors for patients for whom the disposition decision (admit or discharge) was made in the ED. The heckprob model assumes an existing underlying relationship (StataCorp 2013)

$$AdmErr_i^* = \beta_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + zEDCong_i\beta_3 + \epsilon_i^{\beta}\,, \tag{EC.2}$$

in the case of admission error (with identical formulation for discharge errors), but it also assumes that we only observe the binary outcome

$$AdmErr_i = \mathbb{1}[AdmErr_i^* > 0] = \begin{cases} 1 & \text{if } AdmErr_i^* > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{EC.3}$$

This model further assumes that the dependent variable (admission or discharge error) is not always observed. Specifically, to estimate the effect of congestion on decisions made in the ED, we assume that the dependent variable is only observed in the case where $ED_i = 1 - CDU_i = \mathbb{1}[CDU_i^* < 0] = 1$, where

$$CDU_i^* = \delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3 + \epsilon_i^{\delta}\,, \tag{EC.4}$$

and where $\epsilon^{\beta} \sim N(0,1)$, $\epsilon^{\delta} \sim N(0,1)$, and $\text{corr}(\epsilon^{\beta}, \epsilon^{\delta}) = \rho$. The log likelihood to be maximized in the case where $CDU_i = \mathbb{1}[CDU_i^* > 0] = 1$ is (Van de Ven and Van Praag 1981)

$$\begin{aligned} lnL = & \sum_{\substack{i\in S \\ AdmErr_i \neq 0}} ln\left[\Phi_2\{\beta_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + zEDCong_i\beta_3, \delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3, \rho\}\right] \\ & + \sum_{\substack{i\in S \\ AdmErr_i = 0}} ln\left[\Phi_2\{-(\beta_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + zEDCong_i\beta_3), \delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3, -\rho\}\right] \\ & + \sum_{i\notin S} ln\left[1 - \Phi\{\delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3\}\right] \end{aligned} \tag{EC.5}$$

where $S$ is the set of observations for which the dependent variable ($AdmErr_i$) is observed, $\Phi_2(\cdot)$ is the cumulative distribution function (CDF) of a bivariate normal distribution with mean vector $(0,0)^T$ and unit variances, and $\Phi(\cdot)$ is the CDF of a standard normal distribution. To be specific, the bivariate normal cdf is

$$\text{Prob}(X_1 < x_1, X_2 < x_2) = \int_{-\infty}^{x_2}\int_{-\infty}^{x_1} \phi_2(z_1, z_2, \rho)dz_1 dz_2$$

which we denote $\Phi_2(x_1, x_2, \rho)$, with corresponding density

$$\phi_2(z_1, z_2, \rho) = \frac{e^{-(1/2)(x_1^2+x_2^2-2\rho x_1 x_2)/(1-\rho^2)}}{2\pi(1-\rho^2)^{1/2}} .$$

Note that in the equation above we assume that the dependent variable is only observed in the case where $ED_i = 1 - CDU_i = \mathbb{1}[CDU_i^* < 0] = 1$. This is similar to traditional Heckman sample selection models, which are used when the outcome is not observed in the case of non-selection (e.g., if we had no further information about those patients admitted to the CDU). In our case, however, we observe the outcomes of gatekeeping decisions made both in the ED and in the CDU. It is therefore possible for us to estimate the coefficients of the outcome equation under both regimes (i.e., with the ED physician and with the CDU physician as the decision-maker). This estimation can be made jointly using an endogenous switching regression model, or both sides of the equation can be estimated separately by "tricking" the Heckman selection model, as described in Lee (1978). Our formulation above employs this trick.

It is also possible, however, to estimate the full endogenous switching regression model (switch). The switch model differs from the heckprob model in that we jointly estimate the two structural equations

$$AdmErr_{0i}^* = \beta_{00} + \mathbf{X}_i\boldsymbol{\beta}_{01} + zEDCong_i\beta_{03} + \epsilon_{0i}^{\beta} \qquad \text{if } CDU_i = 0\,, \tag{EC.6}$$

and

$$AdmErr_{1i}^* = \beta_{10} + \mathbf{X}_i\boldsymbol{\beta}_{11} + zEDCong_i\beta_{13} + \epsilon_{1i}^{\beta} \qquad \text{if } CDU_i = 1\,, \tag{EC.7}$$

where

$$CDU_i^* = \delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3 + \epsilon_i^{\delta}\,. \tag{EC.8}$$

Instead of estimating the two regimes given in Equations (EC.6) and (EC.7) separately, as we did using the heckprob model approach, under the switch model we estimate them jointly. The log likelihood to be maximized is given by

$$\begin{aligned} lnL = \sum_i \Bigg[ &(1-CDU_i)\Big\{AdmErr_{0i} * ln\big(\Phi_2\{\beta_{00} + \mathbf{X}_i\boldsymbol{\beta}_{01} + zEDCong_i\beta_{03}, -(\delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3), -\rho_0\}\big) \\ &\quad + (1-AdmErr_{0i}) * ln\big(\Phi_2\{-(\beta_{00} + \mathbf{X}_i\boldsymbol{\beta}_{01} + zEDCong_i\beta_{03}), -(\delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3), \rho_0\}\big)\Big\} \\ &+ CDU_i\Big\{AdmErr_{1i} * ln\big(\Phi_2\{\beta_{10} + \mathbf{X}_i\boldsymbol{\beta}_{11} + zEDCong_i\beta_{13}, \delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3, \rho_1\}\big) \\ &\quad + (1-AdmErr_{1i}) * ln\big(\Phi_2\{-(\beta_{10} + \mathbf{X}_i\boldsymbol{\beta}_{11} + zEDCong_i\beta_{13}), \delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3, -\rho_1\}\big)\Big\}\Bigg] \end{aligned}$$

where $\Phi_2(\cdot)$ is the CDF of a bivariate normal distribution with mean vector $(0,0)^T$ and unit variances.

In the log likelihood function, this method requires the simultaneous estimation of significantly more parameters than the heckprob model, though in practice estimating both sides of the equation jointly (as is the case here) versus separately (as is the case using the heckprob model) should result in very similar coefficient estimates. We have estimated the model jointly using the switch model described here and find nearly identical results (not reported).

### EC.3.2. Bivariate probit model

In §6.1 of the paper, we employ a recursive bivariate probit (biprobit) model to identify the effect of admission to the CDU on admission and discharge errors for patients. Like the heckprob model, the biprobit model assumes an existing underlying relationship (Greene 2012, pp.738–752)

$$AdmErr_i^* = \beta_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + CDU_i\beta_2 + zEDCong_i\beta_3 + \epsilon_i^{\beta}, \tag{EC.9}$$

in the case of admission error (with identical formulation for discharge error), but that we only observe the binary outcome

$$AdmErr_i = \mathbb{1}[AdmErr_i^* > 0] = \begin{cases} 1 & \text{if } AdmErr_i^* > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{EC.10}$$

Note the first difference between this and the heckprob model is the additional term $CDU_i\beta_2$ in Equation (EC.9). The second difference is that the dependent variable is observed both in the case where $CDU_i = \mathbb{1}[CDU_i^* > 0] = 1$ and where $CDU_i = \mathbb{1}[CDU_i^* > 0] = 0$, where $CDU_i^*$ is as given in Equation (EC.4). To construct the log likelihood let

$$q_{1i} = \begin{cases} 1 & \text{if } AdmErr_i = 1 \\ -1 & \text{otherwise.} \end{cases} \tag{EC.11}$$

and

$$q_{2i} = \begin{cases} 1 & \text{if } CDU_i = 1 \\ -1 & \text{otherwise.} \end{cases} \tag{EC.12}$$

The log likelihood, $lnL$, can then be written

$$\begin{aligned} lnL = \textstyle\sum_i ln[\Phi_2\{\ & q_{1i} * (\beta_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + CDU_i\beta_2 + zEDCong_i\beta_3 + \text{offset}_i^{\beta}), \\ & q_{2i} * (\delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3 + \text{offset}_i^{\delta}), q_{1i}q_{2i}\rho)] \end{aligned} \tag{EC.13}$$

where $\Phi_2(\cdot)$ is the cumulative distributive function (CDF) of a bivariate normal distribution with mean vector $(0,0)^T$ and unit variances, as specified in §EC.3.1.

## Appendix EC.4: Relevance and Validity of the Instruments

In this section, we perform formal testing to assess the relevance and validity of the two instrumental variables (IVs) employed in the paper.

### EC.4.1. Tests of under- and weak identification

The underidentification test is a Lagrange multiplier test to determine whether the equation is identified. Specifically, the test determines whether the excluded instruments are correlated with the potential endogenous regressor, i.e., that the excluded instruments are relevant in the selection (first-stage) equation. Weak identification, on the other hand, occurs when the excluded instruments are correlated with the endogenous regressors, but only weakly. When this happens, estimators can perform poorly: Estimates may be inconsistent, tests for the significance of coefficients may lead to the wrong conclusions, and confidence intervals are likely to be incorrect. We test for both under- and weak identification as follows.

First, we note that most tests are based on a linear IV regression model where the dependent variable in the outcome equation and the endogenous variable are continuous. In order to perform formal testing we therefore follow convention and treat the binary unnecessary hospitalization, wrongful discharge, and CDU admission variables as continuous. While this means that the true critical values of the tests and significance levels may differ from those reported here, we note that differences in estimated parameters that arise from using a continuous rather than binary model specification are often small, and that the estimated coefficients using these models (not shown) are consistent with those reported in the main paper.

In testing for both underidentification and weak identification we use the method developed by Sanderson and Windmeijer (2016), implemented in and reported by the `ivreg2` command in Stata 12.1 (Baum et. al. 2010). The Sanderson-Windmeijer (SW) first-stage chi-squared Wald statistic is distributed as chi-squared with $(I_E - N_{EN} + 1)$ degrees of freedom under the null that the particular endogenous regressor of interest is underidentified, where $I_E$ is the number of excluded instruments ($= 2$ here) and $N_{EN}$ is the number of endogenous regressors ($= 1$ here). For the unnecessary hospitalization model, the SW Chi-sq statistic is calculated to take a value of 430.53 with 2 d.f., which has corresponding $p$-value$< 0.0001$. For the wrongful discharge model, the SW Chi-sq statistic takes value 435.42 with 2 d.f. and corresponding $p$-value$< 0.0001$. This means that there is strong evidence in favor of rejecting the null hypothesis of underidentification in both cases at, e.g., the 0.1% significance level, and so we conclude that the excluded instruments are relevant.

Turning next to the issue of weak identification, the SW first-stage $F$-statistic is the $F$ form of the SW chi-squared test statistic and can be used as a diagnostic for whether a particular endogenous regressor is weakly identified. In particular, the $F$-statistic can be compared against the critical values for the Cragg-Donald $F$-statistic reported in Stock and Yogo (2005) to determine whether the instruments perform poorly. The test has the null hypothesis that the maximum bias of the IV estimator relative to the bias of ordinary least squares, i.e., $\left|\frac{\mathbb{E}[\hat{\beta}_{IV}]-\beta}{\mathbb{E}[\hat{\beta}_{OLS}]-\beta}\right|$, is $b$, where $b$ is some specified value such as 10%. For a single endogenous regressor, assuming the model to be estimated under limited information maximum likelihood, the critical $F$-values are 8.68, 5.33, and 4.42 for maximum biases of $b = 10\%$, 15%, and 20%, respectively. If the estimated $F$-statistic is less than a particular critical value, then we conclude that the instruments are weak for that level of bias. Here, the estimated SW $F$-statistic is equal to 215.19 for the unnecessary hospitalization model and 217.63 for the wrongful discharge model, indicating that the maximal bias is likely to be tiny. Thus we are not concerned that our models are affected by the problem of weak instruments.

### EC.4.2. Testing for overidentification

In addition to making sure the excluded instruments are relevant, we also check that they are valid, i.e., (1) uncorrelated with the error term (i.e., orthogonal to epsilon) and (2) correctly excluded from the outcome equation (i.e., only indirectly influencing dependent variable $y$). The test for overidentification for the biprobit model uses the $\chi^2$ statistic in a test of the joint significance of the instruments in the outcome equation. In particular, we include the instruments in both the selection and outcome equations and rely on identification based on the nonlinear functional form alone. The null hypothesis is that the instruments are not jointly significant in the outcome equation (Guilkey and Lance 2014, footnote 8, p. 31). For the unnecessary hospitalization biprobit model $\chi^2 = 0.47$, $p$-value$= 0.790 > 0.10$, and for the wrongful discharge model $\chi^2 = 0.75$, $p$-value$= 0.385 > 0.10$. Together these results indicate no evidence that the instruments are jointly significant, hence we have no reason to suspect that they are not valid.

## Appendix EC.5: Additional Analyses

In this study we performed a number of analyses that are not reported in the main paper. Here we report results from an interaction model that addresses multicollinearity concerns, an interaction model for wrongful discharges, and an alternative measure of error that we call total gatekeeping errors, which is the sum of wrongful discharges and unnecessary hospitalizations.

### EC.5.1. Mitigating multicollinearity concerns in the interaction model

In Footnote 9 of the paper we report a potential concern with the interaction model, namely that $PrAdmErr_i$ and the vector $\mathbf{X}_i$ of controls may be collinear (since most of these controls are used as predictors for $PrAdmErr_i$; see §5.4). To address this concern, we re-estimate the interaction model but drop from $\mathbf{X}_i$ all controls that are used in producing $PrAdmErr_i$.

One problem with this approach, however, is that these controls are contained in both the selection equation (which predicts CDU admission) and the outcome equation (which predicts admission errors). These are important controls in predicting CDU admission and so we would like to include them in some way, yet removing the controls from only the outcome equation is not an option since this would in effect force the model to treat these variables as instrumental variables. Since these variables likely do not satisfy the exogeneity condition, this would be inappropriate.

To get around this problem, we repeat the process reported in §5.4 for producing $PrAdmErr_i$, and for each patient we estimate a probit model to determine their likelihood of being admitted to the CDU. Control variables included in this regression are identical to those used in producing $PrAdmErr_i$ and include all factors known prior to the CDU decision, i.e., all temporal, patient and diagnosis, contextual- and physician-related factors reported in Figure 3. We then take the fitted values from this probit regression, $PrCDUAdmit_i$.

This new variable $PrCDUAdmit_i$ is added as a new control variable in both the selection and outcome equations in the bivariate model, allowing us to drop from $\mathbf{X}_i$ all controls that are used in producing $PrAdmErr_i$ in both equations. We can then recalculate the VIFs and find that for the $PrAdmErr_i$ control they range from 1.70 to 2.62, while for the interaction between $PrAdmErr_i$ and $CDU_i$ they range from 1.08 to 1.50, and all VIFs are significantly lower than the threshold of 10, beyond which multicollinearity starts becoming a bigger concern. Therefore, this new model specification is very unlikely to suffer from multicollinearity concerns.

Reproducing the results from Table 7 also indicates that multicollinearity is not an issue. Table EC.1 reports the updated results, showing that the coefficients of the interaction terms are nearly identical, as are the estimated effect size columns. Overall, this demonstrates the robustness of the results reported in the paper.

**Table EC.1 Effect of CDU referral on unnecessary hospitalizations by patient risk category after addressing multicollinearity concerns.**

| | Coefficients | | | Effect Size | | | |
|---|---|---|---|---|---|---|---|
| | Base Effect | CDU Interaction | Marginal Difference | CDU = 0 (%) | CDU = 1 (%) | ATE (p.p.) | ATT (p.p.) |
| Low Risk | – | $0.415^{***}$<br>(0.092) | $0.415^{***}$<br>(0.097) | 0.09 | 0.32 | 0.24 | 1.24 |
| Low-Med Risk | $0.583^{***}$<br>(0.038) | −0.036<br>(0.050) | $-0.451^{***}$<br>(0.088) | 0.73 | 0.67 | −0.07 | −0.19 |
| Med-High Risk | $1.257^{***}$<br>(0.037) | $-0.456^{***}$<br>(0.038) | $-0.420^{***}$<br>(0.039) | 5.24 | 1.91 | −3.33 | −6.12 |
| High Risk | $1.791^{***}$<br>(0.037) | $-0.642^{***}$<br>(0.040) | $-0.185^{***}$<br>(0.027) | 14.19 | 4.39 | −9.80 | −16.67 |

*Notes:* See Table 3 for control structure; Estimation made using the biprobit model specification; $\rho = 0.279^{***}$, $N =$377,346, Log-lik $= -$150,639; *Robust standard error* in parentheses; Base effect column specifies the difference in admission errors across risk categories; CDU interaction column reports the effect of CDU admission on admission errors by risk category; Marginal difference column tests for statistically significant differences in the effect of CDU admission across risk categories; CDU = 0 (resp., CDU = 1) column reports the expected % of patients who would have been unnecessarily hospitalized if no (resp., all) patients in that risk category were routed through the CDU; ATE and ATT report the average treatment effect and average treatment effect on the treated, respectively, for patients from each risk category, in percentage points (p.p); Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

**Table EC.2 Effect of CDU referral on wrongful discharges by patient risk category.**

| | Coefficients | | | Effect Size | | | |
|---|---|---|---|---|---|---|---|
| | Base Effect | CDU Interaction | Marginal Difference | CDU = 0 (%) | CDU = 1 (%) | ATE (p.p.) | ATT (p.p.) |
| Low Risk | – | 0.231<br>(0.195) | 0.231<br>(0.195) | 0.12 | 0.25 | 0.13 | 0.14 |
| Low-Med Risk | 0.059<br>(0.040) | 0.162<br>(0.135) | −0.068<br>(0.162) | 0.32 | 0.52 | 0.20 | 0.21 |
| Med-High Risk | 0.022<br>(0.051) | 0.047<br>(0.123) | −0.116<br>(0.073) | 0.63 | 0.72 | 0.09 | 0.09 |
| High Risk | 0.037<br>(0.066) | −0.058<br>(0.118) | $-0.104^{*}$<br>(0.049) | 1.79 | 1.55 | −0.23 | −0.26 |
| $\rho$ | | 0.080<br>(0.063) | | | | | |
| N | | 377,346 | | | | | |
| Log-lik | | −112,406 | | | | | |

*Notes:* Estimation made using the biprobit model specification; *Robust standard error* in parentheses; Base effect column specifies the difference in discharge errors across risk categories; CDU interaction column reports the effect of CDU admission on discharge errors by patient risk category; Marginal difference column tests for statistically significant differences in the effect of CDU admission across risk categories; CDU = 0 (resp., CDU = 1) column reports the expected % of patients who would have been wrongfully discharged if no (resp., all) patients in that risk category were routed through the CDU; ATE and ATT report the average treatment effect and average treatment effect on the treated, respectively, for patients from each risk category, in percentage points (p.p); Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

### EC.5.2. Wrongful discharge interactions

In Sections 5.4, 6.1, and 6.3 we discuss and report results from an interaction model that tests whether patients with higher ex ante risk of unnecessary hospitalization also benefit more from being admitted to the CDU. As mentioned in Footnote 7, we do not report the results of interaction models for wrongful discharges in the paper due to their low incidence (0.7% of the sample). This means that repeating the interaction analysis for discharge errors is likely to lead to insignificant and unreliable coefficient estimates. However, in the interests of completeness we report the results from this interaction model here.

To run this analysis, we generate a patient's predicted underlying risk of being wrongfully discharged using the same method reported in §5.4, then allocate patients into four categories depending on their risk level: low, low-medium, medium-high, and high. This forms the variable $PrDischErr_i$. We then add $PrDischErr_i$ as an additional control into the selection and outcome equations specified in (1) and (3) in the paper. Next,

**Table EC.3 Coefficient estimates for CDU impact on total gatekeeping errors.**

| | Total GK Error | |
|---|---|---|
| | (1) Probit | (2) Biprobit |
| CDU referral | 0.004 | $-0.472^{***}$ |
| | (0.015) | (0.037) |
| CDU length of stay | $-0.003^{\dagger}$ | $-0.002^{\dagger}$ |
| | (0.002) | (0.001) |
| $\rho$ | – | $0.267^{***}$ |
| | | (0.020) |
| N | 377,346 | 377,346 |
| Log-lik | −62,920 | −157,800 |
| Pseudo-R$^2$ | 0.166 | – |

*Notes: Robust standard error* in parentheses; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

we add into the outcome equation an interaction term between $PrDischErr_i$ and $CDU_i$, which allows the relative size of the impact of CDU admission on a patient's likelihood of being wrongfully discharged to differ depending on their ex ante risk of being discharged in error. Results are reported in Table EC.2. As anticipated, none of the interaction effects are significant in this table.

### EC.5.3. Total gatekeeping errors

While we perform separate analysis for both types of error (unnecessary hospitalization and wrongful discharge), it is also interesting to consider what happens to total gatekeeping errors (i.e., the sum of these two types of error) as patients are routed through the CDU and exposed to congestion. This is especially interesting for the effect of congestion, which works in opposite directions for these two error types. In Table EC.3 we report results using a probit and biprobit specification to examine the impact of CDU admission on a patient's likelihood of experiencing either type of gatekeeping error. Consistent with the main results, we find that patients routed through the CDU are significantly less likely to incur gatekeeping errors than those admitted or discharged directly from the ED (coef. $= -0.472$, $p$-value $< 0.001$).

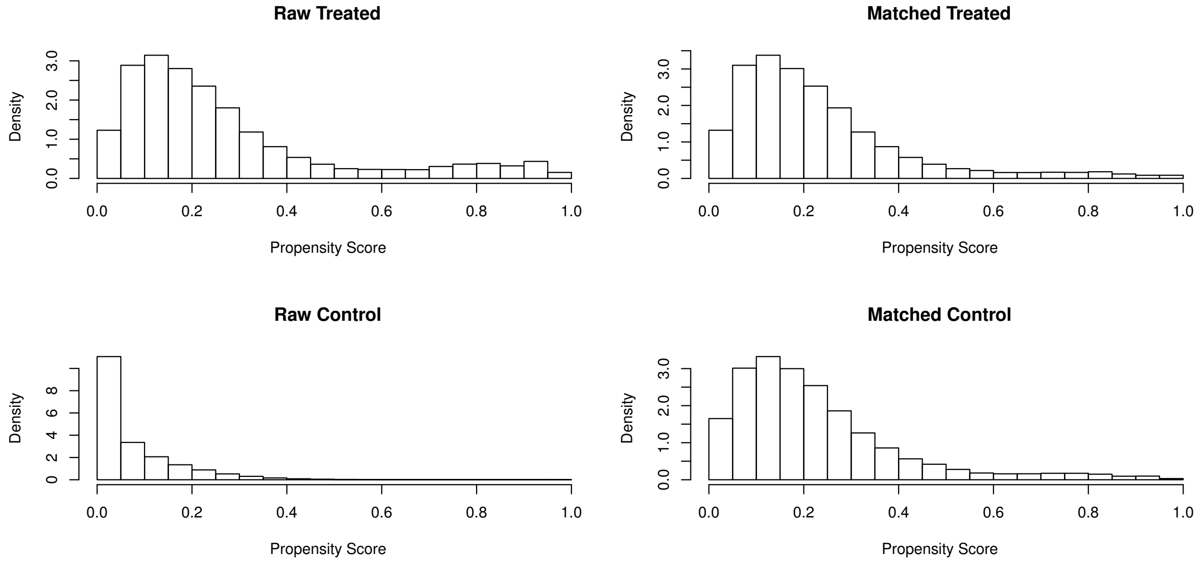
Turning to the effect of congestion, we find that a one standard deviation increase leads to a 0.17 ($p$-value $< 0.001$) percentage point increase in a patients likelihood of experiencing either type of gatekeeping error. This is smaller than the effect on admission errors reported in the paper, likely due to the fact that discharge errors become less likely with an increase in ED congestion. Overall, though, this finding indicates that the total number of errors made by physicians in the ED increases with congestion. (Table omitted for brevity.)

## Appendix EC.6: Propensity Score Matching

### EC.6.1. Background

Matching is a method for reducing dependence on statistical modeling assumptions when making causal inferences. This is especially valuable when working with observational data, where the treatment effect (in our case assignment to the CDU) is not randomly assigned. The goal of preprocessing using matching methods is to reduce the strength of the relationship between the treatment effect ($CDU_i$) and the control variables ($X_i$). Most matching methods work by retaining all of the treated observations in the dataset

**Figure EC.1 Density of propensity scores before (left column) and after (right column) matching for the treated (top row) and control (bottom row) groups.**



and selecting a set of non-treated observations that are similar (where similarity is defined by the matching method of choice) to the treated units based on the controls $X_i$. One of the main benefits of matching is that it can increase efficiency by removing observations outside of an area where the model can reasonably extrapolate.

The simplest type of matching occurs when there exist two observations, one treated and one untreated, and an exact match can be made (meaning that the two observations are identical based on controls $X_i$). This is known as one-to-one exact matching. In practice, when there are many control variables exact matching is not possible and the goal of matching methods becomes balancing the covariate distributions across the two groups (treated and untreated).

When one-to-one exact matching is not possible, various matching methods can be used. Below, we report results using 1:1 nearest neighbor matching, though other methods yield similar results. Balance is achieved using a logit model to estimate a propensity score – the probability of an individual receiving the treatment condition – and then selecting control observations that are similar in their propensity. We also impose the condition that the closest propensity score can be no greater than 0.2 standard deviations away from the switchers propensity score. This condition is a conservative distance that has been shown to reduce more than 90% of bias due to observable differences between treatment and control groups (Gu and Rosenbaum 1993). We are unable to find a match for 2,218 of the patients in the treatment group, thus we discard them from the matched dataset. The matching is performed using the `MatchIt` package in `R` (Ho et. al. 2011). Figure EC.1 shows the distribution of propensity scores before (left column) and after (right column) matching.

Before matching, the average rates of unnecessary hospitalization and wrongful discharge in the treatment group (those admitted to the CDU) and the control group (those not admitted to the CDU) are 5.06% (resp.,

**Table EC.4 Coefficient estimates for CDU impact – Matched sample.**

| | AdmErr | DischErr |
|---|---|---|
| CDU referral | $-0.416^{**}$ | 0.009 |
| | (0.127) | (0.224) |
| CDU length of stay | $-0.003^{*}$ | $-0.000$ |
| | (0.002) | (0.002) |
| $\rho$ | $0.222^{**}$ | 0.089 |
| | (0.080) | (0.140) |
| N | 70,276 | 70,276 |
| Log-lik | $-62{,}658$ | $-51{,}251$ |

*Notes:* All estimations made using a biprobit model specification; *Robust standard error* in parentheses; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

1.05%) and 4.26% (resp., 0.68%), respectively. After matching, the rate of unnecessary hospitalization (resp., wrongful discharge)for the control group changes to 7.20% (resp., 1.09%), and for the treatment group it changes to 5.34% (resp., 0.80%). Therefore, based purely on the matched summary statistics, admission to the CDU does appear to reduce the unnecessary hospitalization rate, as discussed in the paper, but there is some question as to whether CDU admission may actually increase the wrongful discharge rate. We therefore investigate this further in §EC.6.2.

### EC.6.2. Matching with Engogeneity Correction

One limitation with the matching method is that it still does not account for the fact that patients may differ based on unobservables (although they are more similar based on observables, so the extent to which they differ based on unobservables is also likely to be reduced). Therefore, we have also re-estimated the two-stage models from the paper to account for endogeneity, if any, in the matched sample of 70,276 patients. Results from the biprobit models are reported in Table EC.4.

Matched results for unnecessary hospitalizations are almost identical to those reported in the paper on the full sample. On the other hand, when using the matched sample the significance of the CDU in reducing wrongful discharges reduces and becomes insignificant (coef. 0.009, $p$-value$= 0.97$). The lack of an effect is likely due to the low prevalence of discharge errors in the matched sample (only 662 out of 70,276 observations) relative to the approximately 150 coefficients to be estimated in the models. This may be leading to issues from overfitting. However, since the effect becomes insignificant, in the counterfactual analysis in §7 we take a conservative approach and discuss only unnecessary hospitalizations.

## Appendix EC.7: Alternative Measures for Dependent and Independent Variables

In this section we discuss alternative measures for: (i) unnecessary hospitalization, (ii) wrongful discharge, and (iii) ED congestion.

### EC.7.1. Unnecessary Hospitalization

In §EC.2 we discuss how patient discharge within 24 hours of hospitalization with no procedure performed is an ex post observation that does not mean that a hospitalization is avoidable ex ante. What we are really interested in is the subset of unnecessary hospitalizations that were avoidable. Specifically, if we let

- $N(c)$ be the expected number of patients admitted to the hospital if the ED congestion level is $c$,
- $N(c) = N_n(c) + N_u(c)$, where $N_n(c)$ and $N_u(c)$ are the expected number of necessary and unnecessary admissions, respectively, as observed ex post after discharge of the patient from the hospital, and
- $N_u(c) = N_{ua}(c) + N_{uu}(c)$, where $N_{ua}(c)$ and $N_{uu}(c)$ are the expected number of unnecessary admissions that are, respectively, ex ante avoidable and ex ante unavoidable,

then the patients of interest are given by $N_{ua}$, i.e., they are ex post unnecessary and ex ante avoidable. Instead of observing $N_{ua}$ directly, which would require a clinical team to review every medical record and determine whether they believe the admission (discharge) error was ex ante avoidable, we instead create a measure for $N_u$ – the number of ex post unnecessary admissions. We show in §EC.2 that under mild conditions the effect size we observe using $N_u$ will, if anything, be an underestimate of the effect of congestion on $N_{ua}$.

A natural question, then, is whether patient discharge within 24 hours of hospitalization with no procedure performed (denote this $AdmErr$) is a good measure for $N_u$, the number of ex post unnecessary admissions. Specifically, suppose that $AdmErr = \alpha N_n + \beta N_u$, where $\alpha$ is the proportion of ex post necessary admissions that we incorrectly identify as unnecessary, and $\beta$ is the proportion of ex post unnecessary admissions that we correctly identify. In an ideal world $\alpha = 0$ and $\beta = 1$, but again, without clinical review of every medical record to determine whether admission was unnecessary ex post, there is going to be some measurement error. However, so long as $AdmErr$ and $N_u$ are highly correlated and there is no systematic bias in our estimate, this will not be overly problematic. This will certainly be the case when $\alpha$ is close to 0 and $\beta$ is close to 1.

To test the robustness of the results to the above, we employ various alternative definitions of an admission error. Changing our definition is equivalent to changing the $\alpha$ and the $\beta$ discussed above. Below we describe four alternative definitions of an unnecessary admission and discuss the expected effect on $\alpha$ and $\beta$.

1. Patient discharge within **12 hours** of hospitalization with no procedure performed: We shorten the time window over which we record a patient as an unnecessary admission. This is likely to reduce the values of both $\alpha$ and $\beta$, since with a shorter time window we are likely to capture fewer patients who actually needed to be admitted ex post but are also likely to leave out some patients whose admission was ex post avoidable.

2. Patient discharge within **48** hours of hospitalization with no procedure performed: We lengthen the time window over which we record a patient as an unnecessary admission. This is likely to increase the value of both $\alpha$ and $\beta$, since with a longer time window we are likely to capture more patients who actually needed to be admitted ex post and also to capture some additional patients whose admission was ex post avoidable.

3. Patient discharge less than **two nights** (i.e., passing midnight fewer than two times) after hospitalization with no procedure performed: We account for the fact that discharge from the hospital typically takes place during daytime hours and within a particular time window, so a fixed time window measured in hours might not capture this dynamic associated with discrete daily discharges. Since using nights instead of hours extends the time window (it now varies between 24 and 48 hours), it is likely to

increase the value of both $\alpha$ and $\beta$, as again we are likely to capture more patients who actually needed to be admitted ex post and also some additional patients whose admission was ex post avoidable.

4. Patient discharge within 24 hours of hospitalization with no procedure performed **and with no ICD-10 diagnosis code assigned other than the code for "signs and symptoms"**: This means that the patient not only had no procedure performed, but also that hospital staff could make no clear diagnosis of an actual condition. We would expect this definition to significantly reduce $\alpha$. However, at the same time it may lead to a non-trivial decrease in $\beta$, since while the patient is in the hospital they may be diagnosed with a problem that did not actually require hospitalization ex post. For example, the ICD-10 code "S41.XX" corresponds to patients with an "Open wound of shoulder and upper arm." In practice this patient could have had the wound treated and stitched in the ED and then be discharged, and so their admission may have been unnecessary. However, they would not be included as an unnecessary hospitalization using this new definition.
5. Patient **discharge is significantly faster ($< 0.1\times$ the median length of stay) than other patients admitted in the same diagnosis group**: This measure differs from the others used here in that it does not depend on whether or not a procedure was performed or a diagnosis was assigned, and it is not measured over a fixed time period constant for all patients. Instead, we look at patients with similar diagnoses and determine whether any were discharged significantly faster than others. The effect on the $\alpha$ and $\beta$ is unclear and depends on whether significantly faster discharge is a sign of reduced necessity or of lower severity.

The correlations between each of these measures and the measure of unnecessary hospitalization employed in the paper are 0.556, 0.778, 0.848, 0.623, and 0.305 respectively.

Using these alternative definitions of an admission error, we re-run the models from the paper. In the first five columns of Table EC.5 we report coefficient estimates for the effect of CDU admission on the likelihood of a patient being admitted unnecessarily, while in the first five columns of Table EC.6 we report coefficient estimates for the effect of congestion. As Table EC.5 shows, admission to the CDU reduces all form of admission error, regardless of how this variable is defined. Furthermore, the coefficient of congestion is consistent and positive in all cases in Table EC.6, which strongly corroborates the direction of the congestion effect on unnecessary hospitalizations reported in the main paper. Note that the evidence is weaker and significant at only the 10% level when using the "discharge within 12 hours" definition, and insignificant using the "$< 0.1\times$ the median length of stay" definition. This lack of significance is likely due to the rarity with which unnecessary hospitalizations occur when using these definitions: only 1.4% and 1.0% of the time, respectively.

In conclusion, we have demonstrated that the results in the paper are robust to different definitions of an unnecessary hospitalization.

### EC.7.2. Wrongful Discharge

In the paper we define a wrongfully discharged patient as one who is discharged from the ED or CDU, re-attends the ED within seven days, receives a diagnosis in the same category as the diagnosis made during

**Table EC.5 Coefficient estimates to establish effect of CDU on unnecessary hospitalizations and wrongful discharges – Using alternative definitions of the DVs.**

| | Admission Errors | | | | | Discharge Errors | |
|---|---|---|---|---|---|---|---|
| | (1) 12hrs | (2) 48hrs | (3) 2 nights | (4) 24hrs, no diagnosis | (5) Short stay | (1) 3 days | (2) 7 days, any diagnosis |
| CDU referral | $-0.446^{***}$ | $-0.535^{***}$ | $-0.529^{***}$ | $-0.438^{***}$ | $-0.666^{***}$ | $-0.522^{***}$ | $-0.301^{***}$ |
| | (0.057) | (0.030) | (0.031) | (0.070) | (0.071) | (0.108) | (0.074) |
| CDU length of stay | −0.001 | 0.000 | −0.001 | $-0.005^{\dagger}$ | $0.004^{**}$ | 0.002 | 0.001 |
| | (0.002) | (0.001) | (0.001) | (0.003) | (0.001) | (0.002) | (0.001) |
| $\rho$ | $0.256^{***}$ | $0.266^{***}$ | $0.267^{***}$ | $0.214^{***}$ | $0.372^{***}$ | $0.357^{***}$ | $0.265^{***}$ |
| | (0.032) | (0.016) | (0.016) | (0.038) | (0.043) | (0.065) | (0.042) |
| N | 377,346 | 377,346 | 377,346 | 377,346 | 377,346 | 377,346 | 377,346 |
| Log-lik | −117,638 | −169,423 | −161,380 | −119,574 | −114,891 | −106,495 | −115,774 |

*Notes:* All estimations made using the biprobit model specification; *Robust standard error* in parentheses; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

**Table EC.6 Coefficient estimates to establish ED physicians' response to increased congestion – Using alternative definitions of the DVs.**

| | Admission Errors | | | | | Discharge Errors | |
|---|---|---|---|---|---|---|---|
| | (1) 12hrs | (2) 48hrs | (3) 2 nights | (4) 24hrs, no diagnosis | (5) Short stay | (1) 3 days | (2) 7 days, any diagnosis |
| ED congestion | $0.012^{\dagger}$ | $0.030^{***}$ | $0.024^{***}$ | $0.035^{***}$ | 0.008 | $-0.016^{\dagger}$ | $-0.017^{*}$ |
| | (0.007) | (0.004) | (0.005) | (0.007) | (0.007) | (0.010) | (0.007) |
| $\rho$ | $-0.305^{***}$ | $-0.204^{***}$ | $-0.212^{***}$ | $-0.152^{*}$ | $-0.335^{***}$ | −0.156 | −0.090 |
| | (0.064) | (0.044) | (0.046) | (0.069) | (0.069) | (0.110) | (0.067) |
| N | 339,990 | 339,990 | 339,990 | 339,990 | 339,990 | 339,990 | 339,990 |
| Log-lik | −117,216 | −162,808 | −155,632 | −119,349 | −115,032 | −107,540 | −115,031 |

*Notes:* All estimations made using the heckprob model specification; *Robust standard error* in parentheses; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

their previous visit, and is subsequently admitted to the hospital. While we use a seven-day window, other time windows (e.g., 24 hours, 48 hours, 72 hours) are used in the medical literature. Given that wrongful discharge is rare even when calculated using a seven-day window, shortening the window too much will make wrongful discharge such a rare event that our analysis is likely to lack power to identify an effect if one exists. However, since we do want to test the robustness of the results to different time periods, we re-run the analysis using a 72 hour time window for re-attendance.

Shortening the re-attendance time window from seven days to three days (i.e., 72 hours) drops the percentage of cases identified as wrongful discharges in the full sample from 0.71% to 0.54%. Note that even though we shorten the window by approximately 57%, we only lose approximately 25% of the cases previously identified as wrongful discharges. This suggests that most patients who we flag in the paper as wrongfully discharged re-attend the ED within a short time window after discharge (approx. 75% return within three days). The correlation between the wrongful discharge rate for re-attendance over seven days versus three days is 0.87.

To explore another alternative definition, we re-run the analysis and remove the restriction that the patients diagnosis category must be the same on the first and second visits. This allows for the possibility of an incorrect initial diagnosis when the patient first visits the ED and a correct second diagnosis when they re-attend. However, we note that this expanded definition increases the risk of incorrectly flagging as wrongful discharges those patients who were in fact correctly discharged but who returned within a week for an unrelated condition. We find that 1.06% of patients satisfy this alternative definition of a wrongful

**Table EC.7 Coefficient estimates to establish ED physicians' response to increased congestion – using alternative definitions of congestion.**

| | 2hr from arrival | | 4hr from arrival | | 1hr from discharge | |
|---|---|---|---|---|---|---|
| | (1) AdmErr | (2) DischErr | (1) AdmErr | (2) DischErr | (1) AdmErr | (2) DischErr |
| ED congestion | 0.033$^{***}$ | −0.021$^{*}$ | 0.024$^{***}$ | −0.019$^{*}$ | 0.010$^{*}$ | −0.000 |
| | (0.005) | (0.009) | (0.005) | (0.010) | (0.005) | (0.009) |
| $\rho$ | −0.208$^{***}$ | −0.116 | −0.214$^{***}$ | −0.122 | −0.222$^{***}$ | −0.123 |
| | (0.047) | (0.095) | (0.047) | (0.093) | (0.046) | (0.091) |
| N | 339,990 | 339,990 | 339,990 | 339,990 | 339,990 | 339,990 |
| Log-lik | −144,576 | −110,128 | −144,557 | −110,100 | −144,600 | −110,138 |

*Notes:* All estimations made using the heckprob model specification; *Robust standard error* in parentheses; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

discharge, compared to 0.71% using the definition in the paper, with correlation between the two measures taking value 0.82.

Given the high correlation between each of the proposed alternative measures and the original measure used in the paper, it is unlikely that the results will differ significantly on this subsample. However, we re-run the analysis from the paper using alternative definitions of wrongful discharge and report the results for the effect of CDU admission on wrongful discharges in the last two columns of Table EC.5 and the results for the effects of congestion on wrongful discharges in the last two columns of Table EC.6. These results are entirely consistent with the main results in terms of size, direction and significance.

### EC.7.3. Alternative congestion measures

Recall that our congestion measure is designed to capture the fact that when the ED is more crowded physicians have less time to spend with any individual patient, meaning that they must make decisions under increased uncertainty. Indications of this phenomenon can be seen in Figure 2, where we plot ED congestion against the time between a patients arrival and when they are first seen by an ED physician. Clearly, as congestion increases, patients spend more time waiting to be seen. Due to the four-hour waiting time target, a longer wait translates directly into shorter service time. Another possible effect is that when the ED is more crowded, physicians become more generally error-prone due to cognitive overload.

One problem with our current workload measure is that while it is likely to catch the former effect (increased diagnostic uncertainty), it is less likely to capture the latter (physicians becoming more error-prone) because we only measure workload over the first hour after a patient is first admitted to the ED. Since only 8.2% of patients are discharged within one hour, this means that we are not measuring workload at the time when ED physicians are likely to be making the patients disposition decision (i.e., closer to the time of discharge). To the extent that ED congestion in the first hour after admission is correlated with ED congestion in the subsequent hours, this is not so much of a concern. Yet this point deserves further investigation.

We perform sensitivity analysis to determine what happens as we change the time window over which we measure ED congestion, investigating three variations. First, we measure congestion over the first (i) two hours and (ii) four hours after arrival to the ED (rather than one hour). While only 8.2% of patients are discharged within one hour of arriving in the ED, 29.5% are discharged within two hours and 94.4% within four hours. Thus as we extend this time horizon, we are increasingly likely to capture the level of congestion closer to when the gatekeeping decision is made. The correlation between congestion over the first hour and over the first two (resp., four) hours is 0.94 (resp., 0.79), thus we should expect similar results. Results are reported in the first four columns of Table EC.7 and show that the findings in the paper are not sensitive to the choice to measure congestion over the first hour instead of over the first two or four hours after patient arrival.

To address the possibility of physicians becoming more error-prone when making disposition decisions in a busy ED, we measure congestion over the one hour prior to a patients departure from the ED, which is when the disposition decision is generally being made. However, this alternative definition is less likely to capture the effect of congestion on increased waiting time (and hence may also not capture the effects of increased diagnostic uncertainty) because it is measured after the patient is already in service, rather than before they start. That this is the case can be demonstrated visually: Specifically, Figure EC.2 (left) reproduces the left-hand plot of Figure 2 in the paper, while Figure EC.2 (right) is the same except that on the $x$-axis we have *pre-departure* congestion, rather than *post-arrival* congestion. Figure EC.2 then shows that the correlation between time to be seen and ED congestion, when measured over the one hour pre-departure from the ED, is much weaker than it is with the measure used in the paper (i.e., the correlation seen in the left side is much stronger than in the right side of Figure EC.2). Thus, we are less likely to be capturing the effect of shortening service times. Since this is one of the main effects we are interested in, this inability to capture it is unfortunate.

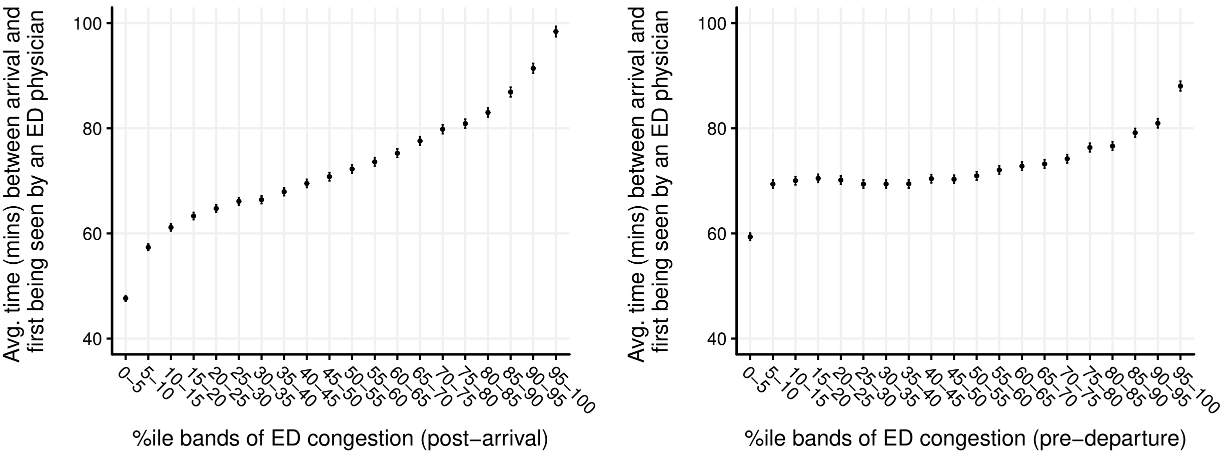
However, we report the results in the last two columns of Table EC.7. The fact that the congestion effect significantly weakens indicates that the increase in error rates with congestion is best explained by the shortening of service times when the ED is congested, leading to increased diagnostic uncertainty when decisions are being made. This is as we hypothesized, and so the paper uses congestion measures from time of arrival rather than from time of departure.

## Appendix EC.8: Benefit of the CDU Beyond the Time Component

As previously mentioned in the paper, one advantage of the CDU in the NHS context of this study is that that a CDU patient is considered “off the clock.” That is to say, even if the patient stay exceeds four hours, the patient no longer contributes to breaches of the four-hour target for an ED stay. We consider what might happen if this four-hour target did not exist and, as an alternative to a CDU, patients were allowed to stay in the ED longer. In this case, the additional testing and assessment provided in the CDU might instead be performed in the ED.

The central question we will resolve is whether, above and beyond the benefits conferred by keeping patients under observation for a longer period, the CDU offers additional benefits for regulating admission and discharge error rates. In other words: Is the “secret sauce” of the CDU simply that its patients are allowed to remain longer? And in a system like the NHS where ED stays are time-limited, is the CDU simply a mechanism that buys more time, and would allowing longer ED stays obtain the same quality results? While our discussion leading up to Hypotheses 1 and 3 lays out the theoretical arguments for the two-stage

**Figure EC.2 Mean time between patient ED arrival and being seen by an ED physician as a function of ED congestion, with 95% confidence bands, where ED congestion is measured over the one-hour period post-arrival in the ED (left) and one hour pre-departure from the ED (right).**



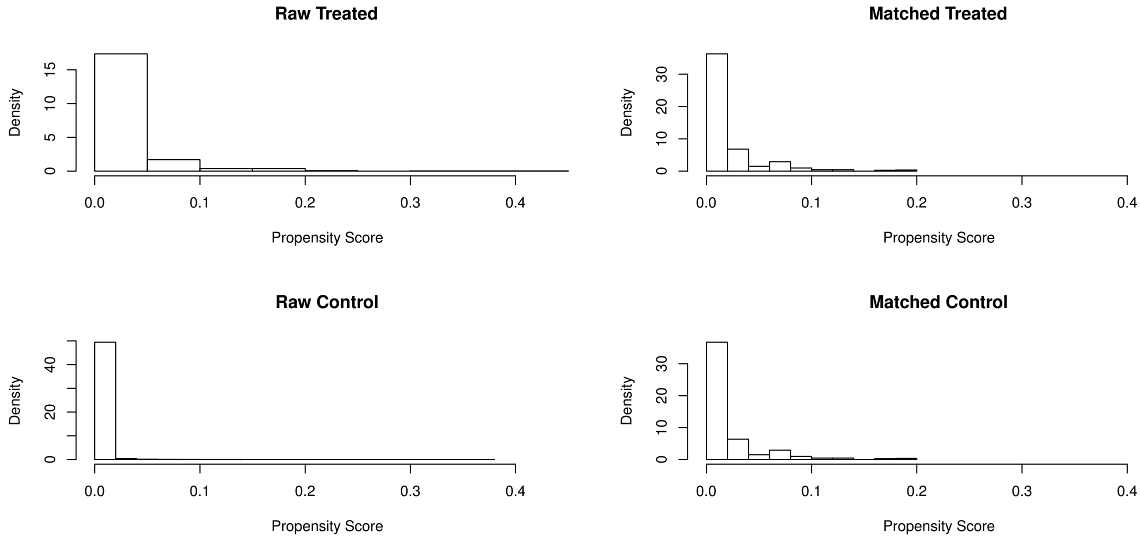
system offering additional advantages beyond the additional time it provides for assessment, in this section we will demonstrate that this is the case.

In discussing this point, we make the argument using unnecessary hospitalization, but it can also be extended to wrongful discharge. Suppose that the longer a patient stays in the ED and (if applicable) the CDU, the less likely they are to be an admitted unnecessarily (because, e.g., more time is spent on diagnosis or more uncertainty is resolved). If the entire benefit of the CDU is related to time, then after we control for the time component, the effect of CDU admission on a patient's unnecessary hospitalization propensity should be effectively zero. If, on the other hand, the CDU has an effect above and beyond that of time, then admission to the CDU should result in a step change effect (i.e., shift the intercept) on a patients likelihood of unnecessary admission. The purpose of the model as specified in the paper is to identify that step change.

Before we discuss how we estimate the step change in the full sample, we use a reduced sample to demonstrate that the CDU does have a benefit above and beyond the time component. Specifically, we subset the data to only those patients who spent less than four hours total in *both* the ED and (if applicable) the CDU. Only 576 patients are admitted to the CDU from the ED and subsequently depart from the CDU within four hours of arrival to the ED, and we take these as our "treated" sample. We can use this dataset to find a matched group of patients who spent the same amount of time in *just* the ED as patients in the treated group spent in *both* the ED and CDU, and who also match based on other observable factors. This is our "control" group. Figure EC.3 shows the density of propensity scores before and after matching.

Since the patients in Figure EC.3 are matched on time spent in the system (as well as other observables), there should in theory be no difference in the unnecessary hospitalization rates of the two groups *unless* (i) admission to the CDU leads to a step change in this propensity or (ii) there are unobservables that make the rates differ across these two groups. If (ii) were true then we would expect that patients admitted to the CDU would be *more* likely to be admission errors based on unobservables. However, comparing our control and treated groups we find precisely the opposite: Patients admitted to the CDU have a 1.22% probability

**Figure EC.3 Density of propensity scores before (left column) and after (right column) matching for the treated (top row) and control (bottom row) groups.**



of being an admission error, versus 3.30% for those patients not admitted to the CDU. (A two-sample t-test for equality of proportions with continuity correction has $p$-value of 0.0291, indicating rejection of the null hypothesis that the two proportions are equal.) This means that patients admitted to the CDU are 63% less likely to be admitted unnecessarily, despite the fact that, if anything, we should expect a higher unnecessary hospitalization rate in this treated group (based on unobservables). This is strong evidence that there is an additional step change benefit to CDU admission above and beyond the time component.

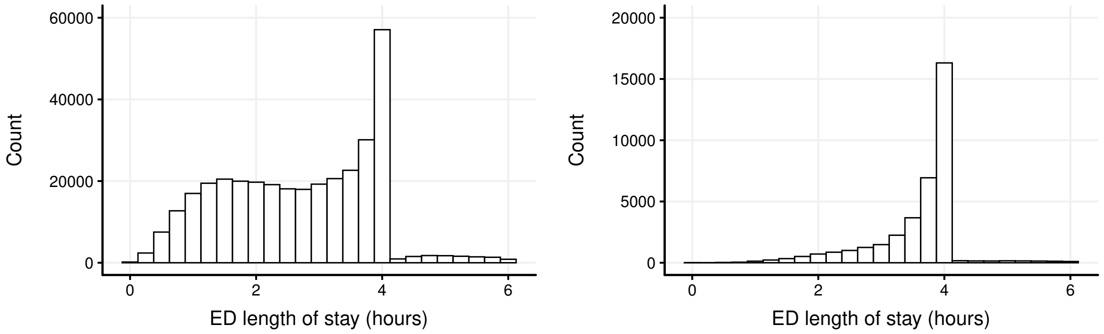
Returning to the step change effect in the full sample, note that in §5.5 we discuss two important controls: (i) the amount of time the patient spends in the ED and (ii) the amount of time the patient spends in the CDU. The longer a patient spends in the ED and/or CDU, the more time they have to be observed by a physician, undergo testing, etc., and this works similarly for patients in the CDU, which may reduce their likelihood of being admitted in error, and these controls account for that fact. The CDU dummy variable, the coefficient of which we are estimating using endogeneity correction techniques, then captures the step change benefit that arises simply from being admitted to the CDU (regardless of how long the patient spends there).

Thus, we believe there is sufficient evidence to prove that the two-stage system has advantages over the equivalent single-stage system without time targets.

## Appendix EC.9: ED Length of Stay

In Figure 2 of the paper, we plot a histogram of ED length of stay for all patients. We reproduce this in Figure EC.4, where we separate ED patients based on whether they were admitted to the CDU (right) or not (left). As shown (and as discussed in Footnote 1 of the paper), a large proportion of patients are admitted to the CDU close to the four-hour target threshold.

**Figure EC.4** **(Left) Histogram of ED length of stay for patients not admitted to the CDU; (Right) Histogram of ED length of stay for patients admitted to the CDU.**



## Appendix EC.10: Full Model Output

In the table that spans the next few pages, we report full model results from linear regression models for each of the main outcome variables as a function of the controls. In addition, column 1 shows results of regressing ED congestion against the set of controls. Note that in our models we want to control for any factors that appear to be correlated with both the dependent variable *and* the independent variable of interest (ED congestion). Given that this is the case for almost all the factors reported in the table, we argue that it is important to include all of these factors in the models.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | ED congestion | CDU referral | AdmErr | DischErr |
| Year2008 | – | – | – | – |
| Year2009 | −0.035 | −0.025 | −0.013 | 0.004 |
| Year2010 | −0.074 | −0.063 | −0.022 | 0.010 |
| Year2011 | −0.178 | $-0.105^{*}$ | −0.037 | 0.016 |
| Year2012 | −0.253 | $-0.144^{*}$ | −0.053 | 0.021 |
| Year2013 | −0.266 | $-0.177^{*}$ | −0.066 | 0.027 |
| Trend | 0.0002 | $0.0001^{*}$ | 0.00004 | −0.00002 |
| Month01 | – | – | – | – |
| Month02 | $0.329^{***}$ | −0.002 | −0.0001 | 0.001 |
| Month03 | $0.456^{***}$ | $-0.009^{**}$ | −0.001 | 0.001 |
| Month04 | $0.332^{***}$ | −0.008 | −0.003 | 0.001 |
| Month05 | $0.306^{***}$ | $-0.013^{*}$ | 0.00004 | 0.004 |
| Month06 | $0.353^{***}$ | $-0.017^{*}$ | −0.003 | 0.004 |
| Month07 | $0.348^{***}$ | $-0.019^{*}$ | −0.005 | 0.004 |
| Month08 | $0.203^{***}$ | $-0.022^{*}$ | −0.002 | 0.003 |
| Month09 | $0.226^{***}$ | $-0.024^{*}$ | −0.006 | 0.004 |
| Month10 | $0.355^{***}$ | $-0.028^{*}$ | −0.010 | 0.005 |
| Month11 | $0.322^{***}$ | $-0.032^{*}$ | −0.012 | 0.006 |
| Month12 | $0.298^{***}$ | 0.003 | 0.0003 | 0.001 |
| SchoolHol None | – | – | – | – |
| SchoolHol Autumn half term | $-0.043^{***}$ | $-0.006^{*}$ | $0.005^{*}$ | −0.0005 |
| SchoolHol Easter | $-0.040^{***}$ | 0.001 | 0.002 | $0.002^{*}$ |
| SchoolHol Spring half term | $-0.113^{***}$ | −0.001 | 0.004 | 0.0002 |
| SchoolHol Summer | $-0.039^{***}$ | 0.0002 | 0.001 | 0.0005 |
| SchoolHol Summer Half term | $-0.122^{***}$ | −0.001 | −0.001 | −0.002 |
| DoW0-Sun | – | – | – | – |
| DoW1-Mon | $0.104^{***}$ | −0.0005 | 0.005 | 0.0001 |
| DoW2-Tue | $-0.301^{***}$ | −0.001 | $0.005^{*}$ | 0.0004 |
| DoW3-Wed | $-0.414^{***}$ | −0.0003 | 0.005 | 0.001 |

| | | | | |
|---|---|---|---|---|
| DoW4-Thu | $-0.338^{***}$ | $-0.0001$ | $0.009^{***}$ | $0.0005$ |
| DoW5-Fri | $-0.251^{***}$ | $0.003$ | $0.008^{**}$ | $0.001$ |
| DoW6-Sat | $0.036^{***}$ | $0.005^{***}$ | $0.002$ | $0.0002$ |
| ArrivedWkday04-08 | – | – | – | – |
| Arrived00-04 | $0.534^{***}$ | $-0.041^{***}$ | $0.024^{***}$ | $-0.0004$ |
| ArrivedWkday08-12 | $0.931^{***}$ | $-0.005^{*}$ | $-0.005^{**}$ | $-0.002^{*}$ |
| ArrivedWkday12-16 | $1.475^{***}$ | $-0.016^{***}$ | $0.004^{*}$ | $-0.003^{***}$ |
| ArrivedWkday16-20 | $1.517^{***}$ | $-0.033^{***}$ | $0.016^{***}$ | $-0.003^{**}$ |
| ArrivedWkday20-24 | $1.142^{***}$ | $-0.044^{***}$ | $0.024^{***}$ | $-0.002^{*}$ |
| ArrivedWknd04-08 | $-0.041^{*}$ | $-0.010^{*}$ | $0.002$ | $-0.001$ |
| ArrivedWknd08-12 | $0.747^{***}$ | $-0.013^{***}$ | $-0.001$ | $-0.003$ |
| ArrivedWknd12-16 | $1.262^{***}$ | $-0.028^{***}$ | $0.007^{*}$ | $-0.002$ |
| ArrivedWknd16-20 | $1.100^{***}$ | $-0.034^{***}$ | $0.020^{***}$ | $-0.002$ |
| ArrivedWknd20-24 | $0.937^{***}$ | $-0.046^{***}$ | $0.020^{***}$ | $0.002$ |
| Age_Bands16-20 | – | – | – | – |
| Age_Bands20-25 | $-0.035^{***}$ | $0.003$ | $-0.002$ | $0.0004$ |
| Age_Bands25-30 | $-0.036^{***}$ | $0.004^{*}$ | $-0.003^{*}$ | $0.0003$ |
| Age_Bands30-35 | $-0.043^{***}$ | $0.009^{***}$ | $-0.002$ | $0.001$ |
| Age_Bands35-40 | $-0.053^{***}$ | $0.010^{***}$ | $-0.003$ | $0.001$ |
| Age_Bands40-45 | $-0.072^{***}$ | $0.011^{***}$ | $0.003^{*}$ | $0.0002$ |
| Age_Bands45-50 | $-0.077^{***}$ | $0.008^{***}$ | $0.002$ | $0.001$ |
| Age_Bands50-55 | $-0.080^{***}$ | $0.007^{***}$ | $0.003$ | $0.001$ |
| Age_Bands55-60 | $-0.083^{***}$ | $0.005^{*}$ | $0.001$ | $0.001$ |
| Age_Bands60-65 | $-0.081^{***}$ | $-0.004$ | $0.002$ | $0.0003$ |
| Age_Bands65-70 | $-0.077^{***}$ | $-0.007^{**}$ | $-0.003$ | $0.00000$ |
| Age_Bands70-75 | $-0.090^{***}$ | $-0.010^{***}$ | $-0.003$ | $0.001$ |
| Age_Bands75-80 | $-0.090^{***}$ | $-0.014^{***}$ | $-0.018^{***}$ | $-0.0002$ |
| Age_Bands80-85 | $-0.116^{***}$ | $-0.013^{***}$ | $-0.028^{***}$ | $-0.001$ |
| Age_Bands85-90 | $-0.123^{***}$ | $-0.015^{***}$ | $-0.035^{***}$ | $-0.001$ |
| Age_Bands90+ | $-0.129^{***}$ | $-0.012^{***}$ | $-0.037^{***}$ | $0.0005$ |
| SexF | – | – | – | – |
| SexM | $-0.010^{***}$ | $0.001$ | $-0.003^{***}$ | $0.0001$ |
| Triage_Cat1 | – | – | – | – |
| Triage_Cat2 | $-0.005$ | $0.008$ | $0.011^{*}$ | $-0.001$ |
| Triage_Cat3 | $0.012$ | $0.035^{***}$ | $0.015^{**}$ | $0.003$ |
| Triage_Cat4 | $0.102^{***}$ | $-0.020^{***}$ | $-0.008$ | $0.003$ |
| Triage_Cat5 | $-0.010$ | $-0.026^{**}$ | $-0.012$ | $0.006$ |
| Triage_CatMT | $0.056$ | $0.012$ | $-0.028^{***}$ | $0.001$ |
| GP_Referral | $0.035$ | $0.032^{***}$ | $0.017^{**}$ | $-0.002$ |
| Init_SeverityEA | – | – | – | – |
| Init_SeverityHMA | $-0.127^{***}$ | $0.017^{***}$ | $0.017^{***}$ | $-0.001$ |
| Init_SeverityLMA | $0.028^{*}$ | $0.027^{***}$ | $0.008^{*}$ | $-0.0005$ |
| Init_SeverityMI | $0.061^{***}$ | $0.015^{***}$ | $0.005$ | $-0.007^{***}$ |
| Init_SeverityRE | $0.016$ | $-0.018^{***}$ | $0.014^{***}$ | $-0.005^{***}$ |
| Visit_ReasonALCO | – | – | – | – |
| Visit_ReasonASSLT | $0.036$ | $0.037^{***}$ | $0.015^{*}$ | $0.002$ |
| Visit_ReasonBITES | $0.048$ | $0.027^{**}$ | $0.027^{***}$ | $0.004$ |
| Visit_ReasonBURN | $0.132^{***}$ | $0.021^{*}$ | $0.030^{***}$ | $0.002$ |
| Visit_ReasonCHEM | $0.249$ | $-0.027$ | $0.014$ | $0.0001$ |
| Visit_ReasonCHILD | $0.102$ | $0.045$ | $0.013$ | $0.025$ |
| Visit_ReasonCYCLE | $0.027$ | $0.019^{*}$ | $0.023^{***}$ | $0.003$ |
| Visit_ReasonDSH | $0.038$ | $0.106^{***}$ | $0.021^{**}$ | $0.001$ |
| Visit_ReasonDV | $0.127$ | $0.026$ | $0.020$ | $-0.002$ |
| Visit_ReasonENT | $0.208^{***}$ | $0.018$ | $0.074^{***}$ | $0.008^{*}$ |
| Visit_ReasonFALL | $0.066^{*}$ | $0.030^{***}$ | $0.021^{***}$ | $0.004$ |
| Visit_ReasonFLU | $0.132$ | $-0.026$ | $-0.003$ | $-0.005$ |
| Visit_ReasonFWORK | $0.319$ | $-0.004$ | $0.008$ | $0.002$ |
| Visit_ReasonGMED | $0.054$ | $0.022^{**}$ | $0.018^{**}$ | $0.007^{*}$ |

| | | | | |
|---|---|---|---|---|
| Visit_ReasonGSUR | $0.109^{***}$ | 0.009 | $0.045^{***}$ | 0.005 |
| Visit_ReasonGYNA | $0.125^{**}$ | 0.014 | $0.058^{***}$ | $-0.0001$ |
| Visit_ReasonHORSE | 0.065 | $0.028^{**}$ | $0.022^{**}$ | 0.003 |
| Visit_ReasonICE | $-0.120$ | 0.018 | 0.018 | 0.001 |
| Visit_ReasonMAJI | $0.795^{**}$ | $-0.134$ | 0.056 | 0.003 |
| Visit_ReasonNEURO | 0.058 | $0.085^{***}$ | $0.019^{*}$ | 0.004 |
| Visit_ReasonORTHO | 0.073 | $0.089^{***}$ | $0.024^{**}$ | 0.003 |
| Visit_ReasonOTHER | $0.070^{**}$ | $0.018^{**}$ | $0.029^{***}$ | $0.005^{*}$ |
| Visit_ReasonOV | $-0.025$ | 0.026 | $0.030^{*}$ | 0.004 |
| Visit_ReasonPC4 | 0.153 | 0.035 | $-0.010$ | $-0.007$ |
| Visit_ReasonPLAS | 0.055 | 0.009 | 0.008 | 0.003 |
| Visit_ReasonRTA | $0.109^{***}$ | 0.012 | $0.014^{*}$ | 0.002 |
| Visit_ReasonSEX | $-0.372$ | $-0.056$ | $-0.006$ | $-0.007$ |
| Visit_ReasonSPORT | $0.105^{***}$ | $0.017^{*}$ | $0.025^{***}$ | 0.003 |
| Visit_ReasonTHROM | $0.790^{*}$ | $-0.073$ | $-0.066$ | $-0.0005$ |
| Visit_ReasonUROL | $0.150^{***}$ | $0.022^{*}$ | $0.041^{***}$ | 0.006 |
| DiagCatCV | – | – | – | – |
| DiagCatDE | $-0.014$ | $0.016^{***}$ | $-0.113^{***}$ | $0.002^{*}$ |
| DiagCatDM | $-0.092$ | $0.041^{*}$ | $-0.119^{***}$ | $-0.004$ |
| DiagCatDV | 0.058 | 0.017 | $-0.134^{***}$ | $-0.002$ |
| DiagCatEN | $-0.040^{*}$ | $-0.046^{***}$ | $-0.060^{***}$ | $-0.001$ |
| DiagCatET | $-0.028^{*}$ | $-0.033^{***}$ | $-0.084^{***}$ | $0.006^{***}$ |
| DiagCatEX | $0.147^{**}$ | $-0.013$ | $-0.124^{***}$ | $-0.003$ |
| DiagCatEY | $0.038^{***}$ | $-0.025^{***}$ | $-0.121^{***}$ | $-0.0002$ |
| DiagCatGI | $-0.084^{***}$ | $-0.006^{***}$ | $-0.096^{***}$ | $0.008^{***}$ |
| DiagCatGU | $-0.051^{***}$ | $-0.007^{**}$ | $-0.116^{***}$ | $0.007^{***}$ |
| DiagCatHM | $-0.042^{*}$ | $-0.034^{***}$ | $-0.071^{***}$ | $-0.001$ |
| DiagCatID | $-0.074^{***}$ | $-0.047^{***}$ | $-0.103^{***}$ | $-0.002$ |
| DiagCatNO | 0.021 | $0.011^{**}$ | $-0.114^{***}$ | $-0.002$ |
| DiagCatNS | $-0.001$ | $0.027^{***}$ | $-0.114^{***}$ | $0.003^{***}$ |
| DiagCatOG | 0.002 | $-0.055^{***}$ | $-0.044^{***}$ | $0.005^{***}$ |
| DiagCatPD | $-0.037$ | 0.005 | $-0.099^{***}$ | 0.017 |
| DiagCatPS | $-0.001$ | $0.178^{***}$ | $-0.138^{***}$ | $-0.002^{*}$ |
| DiagCatRH | $-0.049^{***}$ | $0.016^{***}$ | $-0.121^{***}$ | 0.001 |
| DiagCatRS | $-0.042^{***}$ | $-0.025^{***}$ | $-0.110^{***}$ | $0.002^{**}$ |
| DiagCatSH | $0.049^{**}$ | $0.164^{***}$ | $-0.131^{***}$ | $-0.003^{*}$ |
| DiagCatSK | $0.052^{***}$ | $-0.025^{***}$ | $-0.120^{***}$ | $0.002^{***}$ |
| DiagCatUNKN | $0.114^{***}$ | $-0.004^{*}$ | $-0.111^{***}$ | $0.017^{***}$ |
| Arrival_ModeAM | – | – | – | – |
| Arrival_ModeFO | $-0.015$ | 0.002 | $-0.008^{***}$ | $-0.0004$ |
| Arrival_ModeHE | 0.028 | $-0.035^{***}$ | $-0.024^{**}$ | $-0.001$ |
| Arrival_ModeOT | $-0.005$ | $-0.008$ | $-0.012^{**}$ | $0.004^{*}$ |
| Arrival_ModePO | 0.012 | $0.030^{***}$ | $-0.002$ | 0.001 |
| Arrival_ModePR | $0.021^{***}$ | $0.007^{***}$ | $-0.003^{**}$ | $0.002^{***}$ |
| Arrival_ModePU | $0.056^{***}$ | 0.008 | $-0.006$ | 0.001 |
| Arrival_ModeTA | $-0.025^{*}$ | 0.0001 | $-0.007^{**}$ | 0.002 |
| VisitsLY | 0.0002 | $-0.00005$ | 0.0003 | $0.0002^{**}$ |
| AvgAdmitsLY | $-0.017^{**}$ | $-0.017^{***}$ | $-0.001$ | 0.0004 |
| ZeroVisitsLY | 0.006 | $-0.007^{***}$ | $-0.002^{*}$ | $-0.002^{***}$ |
| VisitsLM | $-0.001$ | 0.002 | 0.001 | $0.002^{**}$ |
| AvgAdmitsLM | $-0.004$ | $-0.008^{**}$ | $-0.002$ | $0.002^{*}$ |
| ZeroVisitsLM | $-0.001$ | $-0.001$ | $0.007^{**}$ | $-0.001$ |
| New_Doctor | $-0.025^{**}$ | 0.004 | $-0.012^{***}$ | $0.002^{**}$ |
| DocCDURate | $-0.179^{***}$ | $0.067^{***}$ | $-0.031^{***}$ | $-0.001$ |
| DocDischErrRate | $0.232^{***}$ | $-0.007$ | $-0.032^{***}$ | $0.022^{***}$ |
| DocAdmErrRate | $0.043^{***}$ | $-0.048^{***}$ | $0.095^{***}$ | $-0.003^{*}$ |
| Hosp_Congestion | $0.050^{***}$ | 0.001 | 0.00001 | 0.0001 |
| ED_Congestion | – | 0.001 | $-0.0003$ | $-0.0004^{**}$ |

| | | | | |
|---|---|---|---|---|
| CDU_Congestion | $0.150^{***}$ | $-0.004^{***}$ | $0.001^{*}$ | $-0.0001$ |
| CDU_Congestion_Conditional | $0.040^{***}$ | $-0.021^{***}$ | $-0.002$ | 0.0001 |
| ED_LOS | $0.198^{***}$ | $0.024^{***}$ | $0.010^{***}$ | $0.0004^{**}$ |
| CDU_LOS_Conditional | $-0.001$ | $0.045^{***}$ | $-0.0002^{*}$ | $-0.00002$ |
| CDU_Decision | 0.011 | – | $-0.008^{***}$ | $0.003^{***}$ |
| Constant | $-2.055^{***}$ | $-0.048^{*}$ | $0.040^{*}$ | 0.007 |
| Observations | 377,346 | 377,346 | 377,346 | 377,346 |
| R$^2$ | 0.270 | 0.420 | 0.076 | 0.008 |
| Adjusted R$^2$ | 0.270 | 0.420 | 0.076 | 0.007 |

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

## References

Baum CF, Schaffer ME and Stillman S (2010) ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. URL http://ideas.repec.org/c/boc/bocode/s425401.html, Accessed: 2016-01-06.

Baugh CW, Venkatesh AK, Hilton JA, Samuel PA, Schuur JD, Bohan JS (2012) Making greater use of dedicated hospital observation units for many short-stay patients could save $3.1 billion a year. *Health Affairs* 31(10):2314–2323.

Greene WH (2012) *Econometric Analysis* (Upper Saddle River, NJ: Prentice Hall), 7th edition.

Gu XS and Rosenbaum PR (1993) Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms. *J Comput Graph Stat* 2(4):405–420.

Guilkey DK and Lance PM (2014) Program impact estimation with binary outcome variables: Monte Carlo results for alternative estimators and empirical examples. Sickles R, Horrace W, eds., *In Festschrift in Honor of Peter Schmidt*, 5–46 (New York: Springer).

Hansen L (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029–1054.

Ho DE, Imai K, King G and Stuart EA (2011) MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software* 42(8):1–28.

Lee LF (1978) Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables. *International Economic Review* 19(2):415–433.

Sanderson E and Windmeijer F (2016) A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics* 190(2):212–221.

Sargan J (1958) The estimation of economic relationships using instrumental variables. *Econometrica* 26:393–415.

StataCorp (2013) Stata 13 Base Reference Manual. College Station, TX: Stata Press. URL https://www.stata.com/manuals13/rheckprobit.pdf, Accessed: 2019-01-23.

Stock J, Yogo M (2005) Testing for weak instruments in linear IV regression. Andrews D, Stock J, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80–108 (Cambridge University Press).

Tan T, Netessine S (2014) When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science* 60(6):1574–1593.

Van de Ven, WPMM, Van Praag, BMS (1981) The demand for deductibles in private health insurance: A probit model with sample selection. *Journal of Econometrics* 17(2):229–252.