# Gatekeeping under Congestion: 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 effect of congestion on the accuracy of gatekeeping decisions (admission to the hospital or discharge home) and the effectiveness of a second gatekeeping stage (a clinical decision unit (CDU) in our context) at reducing gatekeeping errors. While the total error rate increases with congestion, ED physicians prevent an increase in the potentially more harmful wrongful discharges by lowering the threshold for hospital admission. This leads to an increase in unnecessary hospitalizations precisely at times when the gatekeeping system should protect the scarce specialist resource from the surge of demand in the ED. We show that the introduction of a second gatekeeping stage, to which ED physicians can pass those patients for whom they are unable to make confident referral decisions, can mitigate this effect. When used as a second gatekeeping stage, we find evidence that the CDU reduces both unnecessary hospitalization and wrongful discharges, by 11.6% and 11.3%, respectively. We also demonstrate that the two-stage gatekeeping system performs better than a single-stage system that pools the capacity of both stages.

*Key words*: gatekeeping; congestion; referral error; health care: hospitals; service operations; econometrics
*History*: October 21, 2018

## 1. Introduction

As first-line clinicians in a hospital's emergency department (ED), emergency physicians 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 to hospital beds, deciding whether a patient needs to be admitted to the hospital for further diagnosis and specialist treatment, or else can be safely discharged home after receiving treatment in the ED.

Getting the gatekeeping decision wrong puts patients at risk and is costly for the system. Patients who are wrongfully discharged will revisit the ED with the same complaint but in poorer health. If then admitted to the hospital, they will often have a worse prognosis and need more resource-intensive interventions. Patients who are unnecessarily admitted are put at risk of hospital-acquired infections, adverse events such as falls or medication errors, and may experience general physical and mental deterioration due to reduced mobility and an unfamiliar environment. Inouye et al. (2008) point out that an unnecessary hospitalization "*can initiate the terminal downward spiral for an older person,*" resulting in "*delirium, falls, functional decline, institutionalization, and death.*"

Unnecessary admissions will not only put the admitted patients at risk but will also 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, if beds are unnecessarily blocked, then this reduces the hospital's flexibility to treat future arrivals of patients appropriately (Song et al. 2018). Third, when a surge in emergency admissions pushes the hospital beyond its emergency bed capacity, pre-booked elective beds will have to be used as a buffer, leading to cancellations of elective patients and idling of expensive surgical resources (Freeman et al. 2018).

The decision as to whether or not a patient requires hospitalization can be challenging at the best of times, and even more so when demand surges and the ED becomes congested. During busy periods ED physicians are forced to trade off the best treatment for the patient at hand with the need for making fast gatekeeping decisions so as to increase throughput and reduce waiting times for the patients still to be seen. As a consequence, when busy, gatekeeping decisions are often taken with less information and under increased cognitive load, which makes them more prone to error.

In this paper, we present an empirical study of the relationship between ED congestion levels and ED gatekeeping errors. The study is performed on a multi-year dataset of over 300,000 ED attendances to a large UK teaching hospital. As expected, we find that the gatekeeping error rate increases with increasing ED congestion. However, the rates of the two types of error – wrongful discharges and unnecessary hospitalizations – respond differently to congestion. In the study ED, moving from low to high levels of ED congestion (from minus two to plus two standard deviations from the mean) is associated with a 20.9% increase in the rate of unnecessary hospitalizations, and a 16.9% *drop* in the rate of wrongful discharges. We hypothesize that the reason for this asymmetry is a "safety first" principle: ED physicians assume, in case of doubt, that a patient is safer in the hospital than in the community. When congestion puts physicians under time pressure, meaning that gatekeeping decisions are taken sooner with less information, more of these decisions are uncertain. Calling on the “safety first” principle leads ED physicians to admit more of those patients with residual uncertainty to protect them from the risk of being discharged to a less safe out-of-hospital environment.

While this asymmetric response of gatekeeping errors to congestion may be in the best interest of the patient at hand (given the reduced time and increased uncertainty when making the gatekeeping decision), it has severe negative consequences for the system as a whole: Demand surges in the ED lead to proportionately larger demand surges downstream in the hospital inpatient units. This bullwhip-type effect is precisely the opposite of what the hospital expects from its gatekeeping system. Instead of rationing access to specialist resources when demand surges, so as to protect these scarce and expensive resources for those patients who stand to benefit from them the most, the ED gatekeeping system opens the floodgates to the hospital.

After providing evidence of this ED bullwhip phenomenon with our data, we turn our attention to a design feature that can tame this undesirable effect: Adding a second gatekeeping stage in the ED. After stabilizing a patient, ED physicians will now first decide whether they have enough information to make an accurate gatekeeping decision, or else whether additional investigation is needed to improve their confidence when making this decision. Gatekeeping decisions for the former patients are taken immediately, while the latter patients are instead transferred to a bedded clinical decision unit (CDU) in the ED. Patients transferred to the CDU then stay for an extended period of observation and may have further diagnostic tests performed before the gatekeeping decision is taken. Unlike a hospital admission, transferring patients to the CDU is seamless because the CDU is part of the ED; no formal hand-over or bureaucratic admissions procedure is necessary as the emergency medicine team remains entirely responsible for these patients. However, like a hospital admission, transferring a patient to the CDU does move this patient “off the clock” and they no longer directly contribute to congestion in the ED.

Using the context of our study ED, which includes a CDU, we show that the presence of the CDU reduces both rates of unnecessary hospitalization and wrongful discharge. Specifically, after accounting for non-random assignment of patients to the CDU using appropriate sample selection methods, we estimate that the presence of the CDU in the study ED prevents 6 unnecessary hospital admissions for every 1,000 ED visits, and 57 unnecessary hospital admissions for every 1,000 patients referred into the CDU. Finally, we show that the redeployment of resources from the CDU to the ED in the study hospital, which would reduce congestion in the ED, is significantly less effective at reducing unnecessary admissions than the CDU itself. Implementing a two-stage gatekeeping system is, therefore, an effective way to tame the ED bullwhip effect and to safeguard scarce hospital resources for the most vulnerable patients when ED demand surges.

## 2. Contribution to the Literature

This paper contributes first and foremost to the healthcare management literature, as published in operations management (OM) and medical journals. In addition, some of the paper's insights are relevant for several areas within the general OM literature. In this section, we outline how the paper's contributions are positioned within these literature streams.

### 2.1. Healthcare management literature

The paper is, to our best knowledge, the first study of how congestion affects gatekeeping error rates in patient flow systems and provides the first evidence that a specific design feature of a gatekeeping system – a second gatekeeping stage – mitigates this negative congestion effect.

**2.1.1. Empirical healthcare operations.** The paper contributes to a growing body of research within the OM literature that studies the impact of non-clinical variables, such as system congestion, 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), or queue abandonment (Batt and Terwiesch 2015). Of specific 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 does not avoid deterioration in system performance, as evidenced by increased ICU readmission rates, it does ration access to congested services to the neediest patients. In contrast to these studies, we focus on upstream congestion faced by gatekeepers who refer patients to specialist services (acute hospital beds in our case). We find that the rationing pattern observed in KC and Terwiesch (2012) and Kim et al. (2014) is *reversed* upstream: When gatekeepers become busy, they refer *more* patients than necessary to the specialist unit, thus increasing congestion downstream. Similar behavior has been observed elsewhere. 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. Gorski et al. (2017) show that hospital admission rates from the ED increase with congestion. Building on these studies, we provide evidence that error-avoidance behavior specific to healthcare – an emphasis on avoiding missed referrals in the interest of a "safety first" principle – will naturally lead to increased unnecessary referrals under congestion, and then show that a second gatekeeping stage can reduce both forms of error, and thereby mitigate the over-admission phenomenon.

This paper is also closely related to a series of recent analytical papers on ED triage in the OM literature. While triaging has traditionally prioritized patients based on levels 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 duration of stay and comorbidity profile. Most relevant to our work are two modeling papers that study the ED triage process (Saghafian et al. 2012, 2014). These propose augmenting triage by segmenting ED patients based not only on severity but also by their (i) likelihood of being admitted, and (ii) 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 to a two-stage process. While our paper complements these studies with an empirical examination, our context differs in two important ways. First, a two-stage gatekeeping process streams patients into the second stage during service itself, while triaging puts patients into a specific queue before the start of service. We therefore study the effect of congestion on ED physicians' admission decisions rather than on the typically much faster triage decision made by triage nurses (Saghafian et al. 2018). Second, our outcomes of interest differ from the prevailing average cost and waiting time concerns and focus on admission and discharge errors.

**2.1.2. Crowding in emergency departments.** In addition to the healthcare operations literature, our paper also contributes to the medical literature on ED crowding. ED congestion has worsened over the past decade as capacity has failed to keep pace with the growth in attendance due to aging populations (Pines et al. 2011). In the US, for example, ED visits between 1997 and 2007 grew at almost twice the rate of population growth (Tang et al. 2010), while ED attendances in England grew by 47% between 1997 and 2012, compared to population growth of 10% over this period (NAO 2013). In the US, the ED is now the primary point of entry to the hospital, admitting more than half of non-obstetric cases (Greenwald et al. 2016). Mitigating ED congestion is a significant policy concern and countries have adopted a wide range of interventions designed to manage the problem (Boyle et al. 2012). Examples include telephone advice centers, fast tracks, increases in capacity and staffing, changes in boarding practices, and, most relevant to our study, the use of observation units such as CDUs (Pines et al. 2011). Yet these approaches have met with only limited success in mitigating congestion. In February 2016, only 87.8% of patients attending EDs in England were admitted, transferred or discharged within four hours of their arrival – significantly below the target of 95% and the lowest rate since records began (NHE 2016). In contrast to the prevalent focus on throughput or direct patient outcomes, our paper focuses on the effect that congestion has on admission and discharge errors and highlights the importance of CDUs in containing gatekeeping error deterioration when EDs become crowded.

### 2.2. Operations management literature

This paper also contributes to the general OM literature by providing an empirical study that complements the predominantly analytical research on general gatekeeping systems and on the speed-quality trade-off in queuing systems with discretionary service completion.

**2.2.1. Gatekeeping.** Gatekeeping systems are customer flow systems comprised of multiple service tiers, with the progression from a lower to a higher tier controlled by gatekeepers who have a dual role. They can (1) provide a range of services themselves and (2) also have the option to refer a more complex customer on to the next service tier, which consists of more highly skilled and more costly providers (Shumsky and Pinker 2003). Early studies in the OM literature have focused on two-tier systems with a single gatekeeper and have studied 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.

Insights from this economic modeling literature are not readily transferable to contexts in which gatekeeping decisions are not economically motivated but may instead follow professional or social norms, as is likely the case for salaried ED physicians. In such a context, empirical or experimental studies are likely to provide better insights into the behavior of gatekeeping systems. To date, such studies are rare (though exceptions exist, e.g. Freeman et al. 2017, Gorski et al. 2017) and the question of how system states, such as the level of congestion in the system, affect the accuracy of gatekeeping decisions has not been addressed. As highlighted earlier, though, these effects are important as gatekeeper referral errors are both costly and may worsen individual outcomes as well as system performance. Our paper is, to our best knowledge, the first empirical study of the effect of congestion on gatekeeping errors and is particularly relevant for contexts where gatekeepers are professionally rather than financially incentivized.

**2.2.2. Speed-quality tradeoff.** This paper also contributes to the OM literature on the speed-quality trade-off in queuing systems when servers have discretion over task completion (e.g. Hopp et al. 2007, Anand et al. 2011, Kostami and Rajagopalan 2013). As with the gatekeeping literature, most of this work is analytical and empirical studies are rare (e.g. Tan and Netessine 2014). Discretionary service completion in queuing systems can lead to surprising results, in particular in relation to overtreatment, which is where our study interacts most closely with this stream of literature. For example, Hopp et al. (2007) find that, in contrast to standard queuing systems, increasing capacity when service completion is discretionary may, in fact, increase congestion as a result of additional service components being added when servers are under light load. Freeman et al. (2017) offer some empirical support for such behavior. Wang et al. (2010) extend these results to a decentralized context where servers in diagnostic centers trade-off diagnostic accuracy and congestion, and explore the effects of asymmetric error costs, and Alizamir et al. (2013) characterizes the optimal policy for the diagnosis of customer types (e.g. patients requiring hospitalization or not) when servers can decide to perform additional diagnostic testing to resolve type uncertainty. Focusing on services for which customers cannot themselves ascertain their needs (as is often the case in healthcare), Debo et al. (2008) demonstrate analytically that queuing dynamics can create heterogeneity in the customer base that can be exploited to induce additional service when arrival rates are low. However, Paç and Veeraraghavan (2015) show that congestion may act as a deterrent to such overtreatment. In contrast to these studies, which are all concerned with single-tier queuing systems and provide analytical insights, we study a two-tier gatekeeping system and offer empirical observations. Specifically, we show that in contrast to the analytical results for the single-tier case, upstream congestion in the two-tier case *induces* overtreatment when gatekeepers weigh a missed referral as a more serious error than an unnecessary referral.

# 3. Setting Description

## 3.1. The emergency department

The ED at our study hospital is visited by 250 patients per day on average, and operates in a manner similar to the majority of hospitals in the US, UK and worldwide. Patients self-present or arrive by ambulance with a variety of complaints and symptoms, some of which can be easily managed in the ED (e.g., wound suturing, casting, splinting), while others are complex and require admission to the hospital 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 be the sign of a more serious or even life-threatening condition (e.g. chest or abdominal pain). These patients require careful diagnosis before an admission or discharge decision can be taken.

After a patient arrives, the degree of urgency is assessed by a triage nurse. Unless the patient needs immediate attention, they register and wait to be seen for further assessment by an ED physician. The physician may order diagnostic tests (e.g. blood tests, imaging) and may consult a specialist in the hospital. If, after assessment, the physician determines that the patient requires a level of care beyond that which can be provided in the ED, she can admit the patient to an acute bed in the hospital. Otherwise, after treating the symptoms, the patient will be discharged and may be advised to arrange an outpatient or primary care follow-up appointment. ED physicians thus act as gatekeepers to expensive hospital inpatient beds and ration access by admitting only those patients whose needs cannot be met in the ED setting (Blatchford and Capewell 1997).

### 3.2. Gatekeeping challenges in the emergency department

While playing a crucial role as gatekeepers to expensive inpatient beds, ED physicians must make decisions under time-pressure and often there exists significant uncertainty in the medical diagnosis. Consequently, on occasion an error may be made. Graber (2013), for example, estimates that one in ten medical diagnoses made in EDs are inaccurate, with errors in the diagnostic process being the leading cause of internal investigations and claims of malpractice (Kachalia et al. 2007, Cosby et al. 2008). When physicians are exposed to elevated levels of congestion, even less time is available for diagnosis, increasing diagnostic uncertainty and the likelihood that an error will be made. It thus follows intuitively that gatekeeping errors (i.e. unnecessary hospitalizations or wrongful discharges) will be affected by the congestion level in the ED through the diagnostic process.

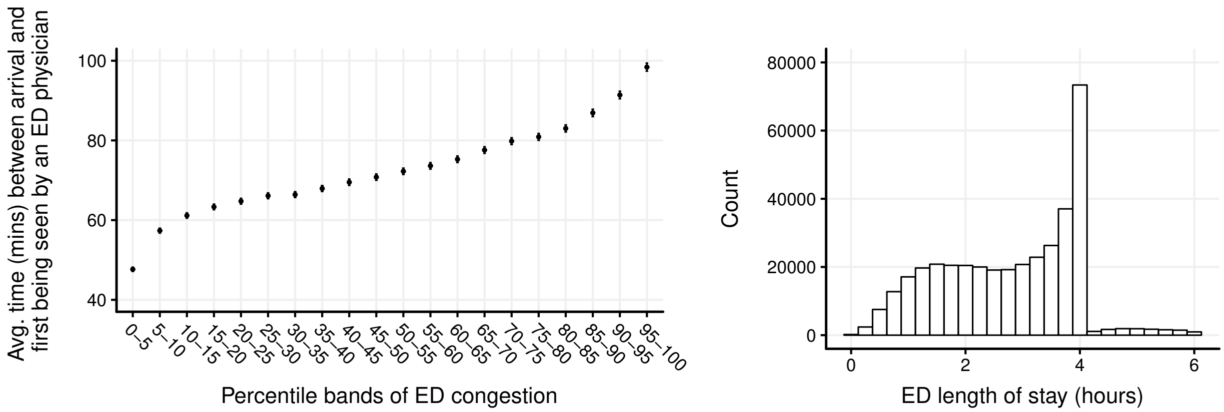
To demonstrate how ED congestion translates into a reduced time available for diagnosis, we plot in Figure 1 (left) the mean time between ED arrival and the patient's first contact with a physician as a function of ED congestion. Each point in the plot corresponds to one of 20 percentile bands of ED congestion of width 5%. (Note that ED congestion is adjusted for differences across hours of the day and various other time-related factors using a method described in Section 5.1.) As congestion in the ED increases, the average time between a patient's arrival and their first contact with a physician increases from under 50 minutes to over 95 minutes.

In England, the time pressure caused by congestion is further exacerbated by the government's 4-hour waiting-time target. This requires that 95% of patients must leave the ED within four hours of arrival. Failure to meet this target in any month attracted a fine of £200 per breach (NHS 2013), which could results in costs to the study hospital of between £75,000 (5% breaches) and £300,000 (20% breaches) per month. As a consequence, breaches of the 4-hour target were taken seriously, as can be seen in Figure 1 (right). This meant that any delay in the start of treatment effectively translated into a direct reduction in the time available to spend with the patient.

### 3.3. The clinical decision unit

The clinical decision unit (CDU), also known as an observation unit, is a dedicated bedded area that is separate from the main ED but is organizationally integrated with the ED and staffed by emergency physicians and nurses. The unit is designed to provide services such as further diagnostic evaluation, additional testing, and continuation of therapy for patients who require care beyond the initial level that can be provided in the ED (Ross et al. 2012). Patients admitted to the CDU are expected to have symptom complexes that can be resolved within 6-24 hours, with further

**Figure 1** **(Left) Mean time between patient arrival at 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.**




assessment determining whether inpatient admission is required at the end of their CDU stay (Hassan 2003). Various clinical and operational advantages of CDUs have been documented in the literature, including improved patient satisfaction, safety, and length of stay (see Cooke et al. 2003, for an excellent survey), as well as considerable cost savings, estimated by Baugh et al. (2012) at \$3.1 billion per year in the US. However, the benefit of a CDU to regulate admission and discharge error rates in the presence of congestion has, to our knowledge, not yet been examined.

In the NHS context of this study, one additional advantage of the CDU is that the patient is considered "off the clock." That is to say, the patient no longer contributes to breaches of the 4-hour target even if they stay for longer than four hours. This is important, since if the 4-hour target were not imposed then an alternative to having a CDU could be to allow patients to instead spend a longer period of time in the ED. In this case, the additional testing and assessment provided in the CDU might instead be performed in the ED. A central question that we will resolve is whether there are additional benefits of the CDU in regulating admission and discharge error rates above and beyond those that come from keeping these patients under observation for a longer period.

## 4. ED Gatekeeping under Congestion

### 4.1. Trading off quality and speed

ED physicians are well aware of the level of congestion in their ED, both through direct visual cues and from information provided by IT systems that show, for example, the list of waiting patients with their registration details and triage information. In response, they will exercise a degree of discretion over the time they spend with their patients (Hopp et al. 2007). ED physicians in the 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 with individual patients, since the opportunity cost of time spent with the current patient increases against the alternative of reducing congestion in the system.

When service times are reduced in response to increased congestion, 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). The work of ED physicians relies on intuition and heuristics (Croskerry 2002), and cognitive overload can render these cognitive shortcuts ineffective, resulting in preventable errors (Leape 1994). For example, in a study of 100 cases of diagnostic errors. Graber et al. (2005) found that cognitive factors contributed in 74% of cases. In summary, as congestion reduces the amount of information at the time of a gatekeeping decision and increases cognitive overload, decision quality will deteriorate.

Hypothesis 1. *As system congestion increases, the rate of gatekeeping errors in the ED increases.*

### 4.2. Trading off admission and discharge errors

Medical errors have been shown to have a negative emotional impact on physicians (Christensen et al. 1992), can result in malpractice investigations and/or litigation (Studdert et al. 2006), and also 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 hospitalization) and false negatives (wrongful discharge) in their gatekeeping decision. The overtreatment phenomenon in healthcare suggests that medical professionals, when faced with uncertainty, will more often choose to do more rather than less (Gawande 2015). For example, unnecessary referral to specialists occurs more frequently 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). A physician in our study hospital put it starkly: "*No-one has ever been sued for admitting a patient to the hospital.*" We can therefore assume that ED physicians have a higher disutility for a discharge error than for an admission error.

Gatekeepers who have asymmetric disutilities for the two types of errors can trade them off against one another. If they reduce the thresholds for one type of decision, say for hospital admission, the rate of unnecessary admissions will increase. At the same time, however, the rate of wrongful discharges will be reduced because in cases of doubt they are now more likely to admit. Per Hypothesis 1 we expect that as congestion increases, the overall error propensity increases. Thus as congestion increases, ED physicians have a choice between accepting more unnecessary admissions or more wrongful discharges. Since an unnecessary admission is less severe in this context, it is rational for the ED physician to reduce the decision threshold for admission to protect against an increase in the rate of the more severe discharge error.

Hypothesis 2. *As system congestion increases, the proportion of unnecessary hospital admissions as a percentage of total gatekeeping errors will increase.*

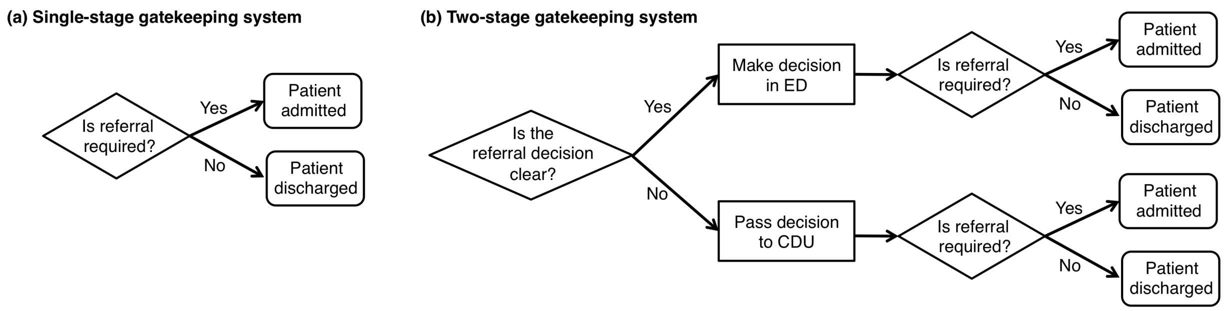
### 4.3. The two-stage gatekeeping system

Hypotheses 1 and 2 imply an increase in unnecessary hospitalizations when congestion increases. Instead of protecting scarce specialist resources in times of demand surges, the incentives in the gatekeeping system lead to a flooding of these resources. Adding a second gatekeeping stage to the system can reduce this undesirable effect. First, it allows for a more formal decoupling of the two tasks of the ED physician, the clinical task of stabilizing and the task of diagnosing the patient and making the gatekeeping decision. Second, it improves the match between patients with heterogeneous needs with gatekeepers with heterogeneous experience and resources. Third, when the gatekeepers in the first stage are subject to time targets (as is the case for the ED in this study – see Section 3.3 for more), a second stage can allow additional time for assessment and diagnosis.

Figure 2 illustrates the difference between a single-stage and a two-stage gatekeeping system. In a single-stage system, the ED physician is required to take a binary gatekeeping decision, to either admit the patient to the hospital or discharge her home. In a two-stage gatekeeping system, the ED physician has an additional decision node. If the gatekeeping decision is not clear, then she can classify a patient as "requiring more work" and refer them to the second stage of the gatekeeping system. This second-stage will see patients for whom the gatekeeping decision is more difficult, and so can be equipped with skills and resources that are more attuned to these patients. Examples include specialized testing equipment or specially trained or experienced gatekeepers. Processes in the second stage can also be better aligned with patient needs: instead of being designed for quickly and effectively stabilizing and treating patients with acute needs, the focus is instead on accurately assessing whether the patient requires hospitalization. The second stage gatekeeper in the CDU can then make a better-informed gatekeeping decision at a later stage, after more information has been gathered. Since a two-stage gatekeeping system decouples the clinical and gatekeeping tasks and allows EDs to better match resources, processes and gatekeeper experience with the complexity of the gatekeeping decision, we expect it to reduce gatekeeping errors.

Hypothesis 3. *A two-stage gatekeeping system reduces both types of ED gatekeeping errors.*

**Figure 2 Flow charts of the traditional single-stage gatekeeping process (left) and the proposed two-stage gatekeeping process (right).**




While the two-stage gatekeeping system reduces errors, it does come at a cost. Specifically, resources must be allocated to the second stage that might otherwise have been deployed in the first stage. It is, therefore, necessary to ascertain that the reduction in gatekeeping errors achieved by the two-stage gatekeeping system is greater than what would otherwise be achieved if the same resources were redeployed to increase the capacity of the first stage.

Note, though, that any capacity diverted from the second-stage to the first-stage would only be useful when the first-stage is congested, and will be idle at times of low demand. In fact, when there is discretion in task completion time, adding capacity may even increase congestion and reduce the benefits of pooling capacity (Hopp et al. 2007). Moreover, since patients in the ED are more-or-less randomly allocated to ED physicians in a round-robin scheme, there is no guarantee that in a pure single-stage system the specialized capability diverted from the second stage would be matched with the patients who stand to benefit most from it. As triage in EDs is based on urgency and not on diagnostic complexity, the matching of, e.g., more capable gatekeepers with more complex gatekeeping patients would not occur systematically, as it does in the two-stage system.

By contrast, the second stage in a two-stage system adds value independently of congestion levels in the first stage. In the context of our study, for example, patients are referred to the CDU even when demand is low in the ED. Therefore, the two-stage gatekeeping system can help to reduce gatekeeping errors at all times, while the additional single-stage capacity will only reduce these errors when congested. Furthermore, the two-stage system ensures a direct match of patients for whom the gatekeeping decision is more difficult with the specialized capability to make these decisions accurately, rather than the round-robin approach to allocating these resources in a single-stage system. Overall, we expect that this should make the two-stage gatekeeping system more efficient at reducing gatekeeping errors than a single-stage capacity increase, i.e. an equal marginal investment in both systems will lead to a larger reduction of errors in the two-stage system.

Hypothesis 4. *A suitably sized second stage in a two-stage gatekeeping system reduces ED gatekeeping errors more than the redeployment of the second stage resources to increase the capacity of the single-stage gatekeeping system.*

Note that this is a marginal effect hypothesis. If the second stage becomes oversized, it will become beneficial to move capacity from the second to the first stage. Determining the optimal size of the two stages is an interesting challenge that is beyond the scope of this paper.

## 5. Data Description and Variable Definitions

The data for our study is comprised of 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 has experienced increasing demand pressure over recent years, with attendances up 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 patients who arrive at the hospital are admitted to an inpatient bed, with admissions and discharges increasing at approximately the same rate over the sample period (by 4.7% and 4.1% per annum, respectively).

To prepare the data for analysis, it was pre-processed to ensure, as far as is possible, that our results are not affected by various data or time-related confounds. This included dropping: (i) ∼8.5k obs. from December 2013, when data entry may not have been completed; (ii) ∼130k obs. corresponding to children under the age of 16, who cannot be admitted to the CDU; (iii) ∼3.5k obs. with missing or incomplete data; and (iv) ∼18k obs. for patients who left against medical advice, died in the ED, or were transferred to another hospital. We then use this data set to generate various variables of interest (to be described later), before: (v) excluding ∼61k obs. from the first 12 months, the warm-up period for generating these variables; and (vi) removing ∼38k obs. from dates close to public holidays and the Christmas break when demand and staffing patterns vary significantly. Due to a temporary change in coding convention that prevents identification of admissions to the CDU in December 2009 and January 2010, we also drop ∼14k obs. corresponding to this period. After this, we were left with 377,331 observations to take forward for analysis.

We next describe the main variables used in the analysis. Summary statistics for these variables and correlations between each can be found in Table 1.

### 5.1. ED congestion

Our main independent variable of interest is the level of congestion that patients experience when they arrive in the ED. To generate this measure for patient $i$, we first determine which patients'

**Table 1 Descriptive statistics and correlation table.**

| | N | Mean | | | Correlation table | | | |
|---|---|---|---|---|---|---|---|---|
| | | All | CDU = 0 | CDU = 1 | (1) | (2) | (3) | (4) |
| (1) Total gatekeeping errors (%) | 377,331 | 5.05 | 4.94 | 6.11 | | | | |
| (2) Unnecessary admissions (%) | 377,331 | 4.34 | 4.26 | 5.06 | 0.92*** | | | |
| (3) Wrongful discharges (%) | 377,331 | 0.71 | 0.68 | 1.05 | 0.37*** | −0.02*** | | |
| (4) CDU admission (%) | 377,331 | 9.90 | 0.00 | 100.00 | 0.02*** | 0.01*** | 0.01*** | |
| (5) ED congestion | 377,331 | −0.01 | −0.01 | −0.01 | 0.01*** | 0.01*** | −0.00** | 0.00 |

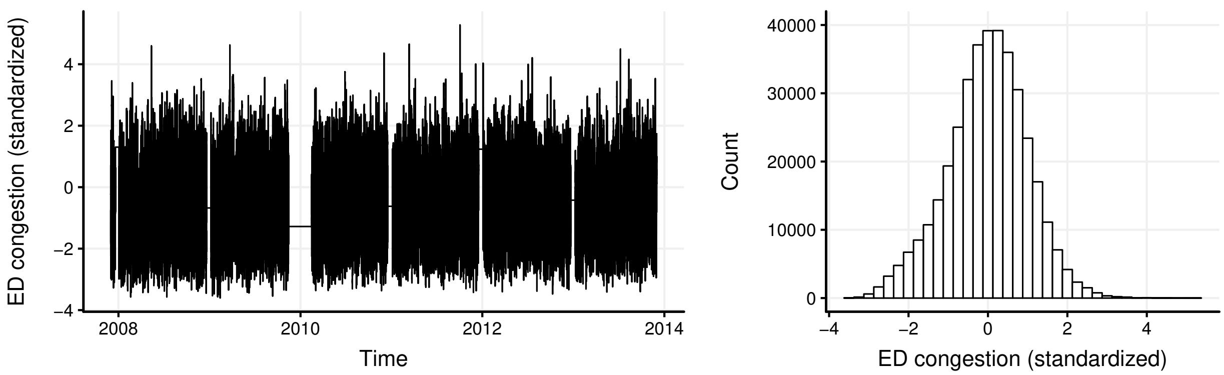
*Notes:* Columns 'All', 'CDU = 0' and 'CDU = 1' report mean values for the full sample, subsample of patients referred directly from the ED, and subsample referred from the CDU, respectively; Standard deviation of ED congestion equal to 1.01, 1.01 and 1.02 for 'All', 'CDU = 0' and 'CDU = 1', respectively;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

ED visits overlapped with the period from arrival to one hour post-arrival of patient $i$. 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. As hypothesized, we expect the more crowded the ED, the less time a physician has to spend with any individual patient, reducing the amount of information available when the gatekeeping decision is being made. Note that the full sample of 651,041 ED visits contribute to the calculation of $CensusED_i$.

It is well known that levels of congestion in EDs vary throughout the day, across days of the week and seasons, and change over time. Since some of this is predictable and staffing can be partially set to meet demand, we should adjust $CensusED_i$ to account for these differences. We achieve this by employing a variation on 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 1st January 2007 and ending at midnight on 31st December 2013. (Note that all dates dropped during the data cleaning process described at the start of Section 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 4a.m., etc.), (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.

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 of arrival of patient $i$. This captures the variation in congestion levels that cannot be explained by predictable and staffable seasonal predictors. Finally, to ease later interpretation of

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




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.

### 5.2. Admission and discharge errors

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

An admission error (or 'unnecessary hospitalization') occurs when a patient is admitted to an acute hospital bed despite that admission being unnecessary or excessive to their needs. These patients block beds and use expensive specialist resources and time. We define an unnecessary hospitalization ex-post as patients who are discharged within 24 hours of being admitted to an inpatient bed in the hospital from the ED or CDU without treatment or procedure performed. The second of these conditions is met if there is no 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 rate of unnecessary hospitalizations for the full sample of 373,663 visits is 4.3% and for the 119,474 visits which resulted in admission is 13.7%.

Discharge errors (or 'wrongful discharges') are, if anything, even more concerning. These patients often come back to the ED in a more serious state, requiring a higher intensity of care than would otherwise have been needed if correctly 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 record ED patients as a wrongful discharge if, after discharge from the ED or CDU, they re-attend the ED within 7 days, have a diagnosis assigned within the same category as during their previous visit, and are subsequently admitted to the hospital. The rate of wrongful discharges in the full sample is 0.7% and is 1.0% for the subset of 257,857 discharged patients.

### 5.3. Unnecessary versus avoidable hospitalizations

In this paper, we define an admission as unnecessary if the patient is discharged within 24 hours of admission to an inpatient bed in the hospital without a recorded treatment in their discharge record. This is an ex-post assessment and not all ex-post unnecessary hospitalizations are avoidable ex-ante, though *avoidable* unnecessary hospitalization does occur (Denman-Johnson et al. 1997). 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). While this remains a limitation of this study, it turns out that under fairly natural assumptions the estimated effect of congestion on the rate of unnecessary hospitalization is a conservative estimate of the effect of congestion on the rate of *avoidable* unnecessary hospitalization (see Appendix B for assumptions and proof).[1]

The key assumption from Appendix B is that the expected numbers of patients who need admission to the hospital (those who are (i) ex-post necessary and (ii) ex-post unnecessary but ex-ante unavoidable) is uncorrelated with our measure of congestion, $c$. As $c$ is adjusted for systematic seasonal variation, the assumption is that: after accounting for seasonal variation, events of serious acute illness 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 worried enough to visit. Sometimes these patients are referred to as the "worried well." In Section EC.5 of the e-companion we show that elevated congestion levels are likely to be associated with the worried well, as well as discussing a relaxation of these assumptions where we allow for unobserved heterogeneity in the case-mix of patients as the congestion level changes.

### 5.4. CDU referral

In our data we observe whether or not a patient is referred into the CDU by a physician from the ED. Of the 37,356 ED patients that are sent to the CDU (9.9% of the analysis sample), 35.2% are subsequently admitted, and the rest are discharged. Once a patient is in the CDU, decisions are made quickly, with a median CDU length of stay (LOS) of 4.5 hours for those who are subsequently admitted, and 4.1 hours for those who are subsequently discharged. This compares with a median LOS in an inpatient hospital bed of 15.5 hours for a patient classed as an unnecessary

[1] 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 classed as resulting in either an unnecessarily hospitalization of not (i.e. it is binary). Taking the average over these binary outcomes gives the correspond 'rate'.

hospitalization. This suggests that the CDU is able to more quickly process patients than can be achieved in a standard inpatient setting. Further analysis (documented in Section EC.1 of the e-companion) finds that the CDU is at least 42% faster at processing patients routed through it than if they had been admitted to a hospital inpatient unit. Thus, while referral through the CDU does extend the service episode, it does so less than if all patients were instead referred directly to the hospital. This is consistent with findings in the medical literature (e.g. Baugh et al. 2012).

Moreover, of those patients admitted from the CDU, 14.4% are then identified to be unnecessary hospitalizations, similar to the 13.6% error rate for those admitted directly from the ED. This is despite the fact that patients admitted from the CDU are subject to considerably more diagnostic uncertainty and thus inherently more likely to be admitted in error.

### 5.5. Control variables

In addition to the primary variables, we also have available a large number of control variables, reported in Table 2, that allow us to account for heterogeneity in the patient population and at the hospital. Important controls are factors that are correlated with the dependent variables and with the independent variables of interest (Section EC.8 of the e-companion provides justification for the controls). The 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 so through omission may bias the results) will be accounted for using appropriate empirical methods described in Section 6.1.

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

## 6. Models and Results I: Response to Congestion

We focus first on testing Hypotheses 1 and 2 (the effect of congestion on gatekeeping errors) and return to Hypotheses 3 and 4 (concerned with estimating the effect of the CDU) in Sections 7 and 8, respectively. In testing the first two hypotheses the presence of the CDU is a confounding factor in the econometric analysis. Specifically, we would like to identify how ED congestion impacts decisions made directly by ED physicians, i.e. corresponding to only those patients not passed to the CDU. This means we want to study the upper half of the two-stage gatekeeping process shown in Figure 2. However, as congestion increases, so too might the rate at which ED physicians leverage

**Table 2 Table of controls.**

| | Type | Description |
|---|---|---|
| Temporal ($\mathbf{T}_i$) | | |
| Year | Categorical (6) | Observation year (offset by one month so 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 (24) | A two-hourly arrival window (e.g., 2am to 4am) for weekdays, and a separate one for weekends |
| Patient and diagnosis related factors ($\mathbf{D}_i$) | | |
| Age bands | Categorical (17) | The age of the patient, split into 5-year age bands (e.g., 15-20, 20-25,..., 95+) |
| Gender | Binary | A variable equal to one if the patient is male, else zero |
| Triage category | Categorical (7) | The triage level assigned to the patient on ED arrival |
| Initial severity assessment | Categorical (7) | The nature of the patient's condition (e.g., minor injuries, requires resuscitation, etc.) |
| Reason for ED visit | Categorical (32) | The reason for the ED episode (e.g., fall, burn, road traffic accident, etc.) |
| Diagnosis category | Categorical (22) | The main category in which the primary diagnosis falls (e.g. respiratory, cardiovascular, etc.) |
| Contexual factors ($\mathbf{C}_i$) | | |
| Mode of arrival | Categorical (8) | The mode of transport used to get to the hospital (e.g., helicopter, ambulance) |
| 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 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 ED physician error rate | Continuous | The unnecessary hospitalization, wrongful discharge, or total gatekeeping errors propensity of the assigned ED physician, 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 ($\mathbf{O}_i$) | | |
| CDU congestion (conditional) | Continuous | The congestion level of the CDU for those patients admitted to the CDU (and 0 for those not), measured over the one hour period prior to departure of the patient from the ED |
| 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 departure of the patient from the ED |

*Notes:* 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 of the contextual factors relating to ED visits and admissions per ED visit exclude any visits made in the 7 days prior to arrival, to prevent a mechanical relationship with the wrongful discharge variable; The historical ED physician error rate is set equal to 0 for those patients who saw a "New physician."

the CDU option. This could change the composition of the patients for whom the ED physicians are taking the admission or discharge decision. While we account partially for these differences with our set of controls (reported in Table 2), there may still exist factors unobservable to us, but observable to the physician (e.g., patient complexion, medical history) that influence whether or not the physician leverages the CDU option. Thus, despite only 9.9% of patients being passed to the CDU, it will be 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. In this section, we describe the empirical approach used to resolve this endogeneity concern.

### 6.1. Econometric specification

Our empirical strategy separates the identification problem into two parts. The first looks to identify those factors that influence whether or not the patient is admitted to the CDU. The second determines whether or not a patient is unnecessarily hospitalized and/or wrongfully discharged, allowing this to depend on whether or not the patient was admitted to the CDU. More specifically, the first stage (selection) equation takes the form

$$CDU_i^* = \delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3 + \epsilon_i^{\delta}, \tag{1}$$

$$CDU_i = \mathbb{1}[CDU_i^* > 0]\,, \tag{2}$$

where $CDU_i^*$ is a latent variable, the vector $\mathbf{X}_i$ contains the set of all controls (reported in Table 2), the vector $\mathbf{Z}_i$ contains the set of instrumental variables (to be described in Section 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

$$AdmErr_i^* = \beta_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + CDU_i\beta_2 + zEDCong_i\beta_3 + \epsilon_i^\beta\,, \tag{3}$$

$$AdmErr_i = \mathbb{1}[AdmErr_i^* > 0]\,, \tag{4}$$

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

Rather than estimate the first and second stage models described above individually, we estimate them jointly with a Heckman probit sample selection (heckprob) model using full information maximum likelihood (Maddala 1983). The heckprob model corrects for potential sample selection bias arising 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. This modeling approach is necessary as we are interested in the effect of ED congestion on errors for those patients for whom the ED physician made the gatekeeping decision (rather than a physician in the CDU). To estimate the heckprob model we must: (1) censor the outcome variable $AdmErr_i$, $DischErr_i$ or $TotErr_i$ whenever $CDU_i = 1$, (2) set $\alpha_2, \beta_2, \gamma_2 = 0$ in the outcome equation, and (3) estimate the selection and outcome equations simultaneously under the assumption that their errors – $(\epsilon_i^\delta, \epsilon_i^\alpha)$, $(\epsilon_i^\delta, \epsilon_i^\beta)$ or $(\epsilon_i^\delta, \epsilon_i^\gamma)$ – are jointly distributed according to the standard bivariate normal distribution with unit variances and correlation coefficients $\rho^\alpha$, $\rho^\beta$ or $\rho^\gamma$ which are estimated as parameters in the models.[2]

[2] Traditionally, Heckman sample selection models 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 outcome when the gatekeeping decision is made both in the ED and the CDU. It is possible, therefore, for us to estimate the coefficients under both regimes (i.e., when the gatekeeping decision is made by either the ED or a CDU physician). This estimation can be made jointly using an endogenous switching regression model, or instead by estimating both sides of the equation separately by "tricking" the Heckman selection model to do so, as described in Lee (1978). We employ this trick by censoring the dependent variable in the outcome equation depending on whether $CDU_i$ takes the value zero or one. Censoring when $CDU_i = 1$ allows us to estimate the effect of ED congestion on error rates made by ED physicians while censoring when $CDU_i = 0$ allows us to estimate the effect on decisions made in the CDU instead. Joint estimation (not reported) results in nearly identical estimates of the coefficients and $\rho$.

### 6.2. Instrumental variables

While the heckprob 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, and so appear in the selection equation (i.e., be relevant), but not affect the rate of unnecessary hospitalizations, wrongful discharges, or total gatekeeping errors, and so do not appear in the outcome equation (i.e., be valid). We use two IVs, included in the vector $\mathbf{Z}_i$. Summary statistics for these IVs are available in Table 3.

The first IV is the CDU admission propensity of the assigned physician. This is calculated in the same way as the physician's historic error propensity (as mentioned in Section 5.5 and described in Appendix A), and 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. A patient assigned to a physician who is predisposed to admit patients to the CDU will be more likely to be sent there, satisfying the relevance condition. A potential issue with this IV is that being assigned to a physician who is more likely to admit to the CDU may also affect the likelihood of that patient being admitted or discharged in error, since 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, wrongful discharge or total gatekeeping error propensity in the respective 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 the busyness of the CDU. Congestion in the CDU, $zCDUCong_i$, is calculated in the same way as ED congestion in Section 5.1, except that we time-weight instead over the one hour period leading up to the departure of patient $i$ from the ED. If the CDU is congested then it becomes less available to ED physicians as an option, since beds and other resources are constrained. This is similar to findings in the literature relating to e.g. admission to the intensive care unit (Chan et al. 2017) and obstetric operating theaters (Freeman et al. 2017). Thus when the CDU is busy we expect there to be fewer CDU admissions, satisfying the relevance condition. For patients who are not admitted to the CDU, the CDU congestion level should have no direct effect on their likelihood of being hospitalized unnecessarily or wrongfully discharged. However, to the extent that CDU congestion may be correlated with busyness in the main hospital, we control for this using the congestion level of the hospital (calculated in the same way as CDU congestion).

Hypothesis testing of the IVs to identify whether there are 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 Section EC.2 of the e-companion). Our results are also robust to using the IVs individually.

**Table 3** **Descriptive statistics and correlation table for the instrumental variables.**

| | N | Mean | | | Correlation table | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | All | CDU = 0 | CDU = 1 | (1) | (2) | (3) | (4) | (5) |
| (6) Phys. CDU use rate | 377,331 | −0.07 | −0.08 | 0.01 | 0.01*** | 0.00** | 0.00** | 0.15*** | −0.05*** |
| (7) CDU congestion | 377,331 | 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 $CDU_i = 0$ and subsample where $CDU_i = 1$, respectively; Correlation table column numbers correspond to: (1) Total gatekeeping errors, (2) Unnecessary hospitalization, (3) Wrongful discharge, (4) CDU admission, (5) ED congestion; Pre-standardized mean (standard deviation) of CDU congestion 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 $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

### 6.3. Results

Before presenting the full set of results, we start by reporting in Table 4 coefficient (coef.) estimates with robust standard errors using a standard probit estimation for each of the four dependent variables in the selection and outcome equations. Examining the model coefficients, we find evidence that as ED physicians become busier, they (1) increase the rate at which they refer patients to the CDU (coef. = 0.059, $p$-value < 0.001), (2) make more total gatekeeping errors (coef. = 0.017, $p$-value < 0.001), (3) admit more patients unnecessarily (coef. = 0.025, $p$-value < 0.001), and (4) make fewer wrongful discharges (coef. = −0.016, $p$-value = 0.027). These responses to increasing levels of diagnostic uncertainty are consistent with Hypotheses 1 and 2, i.e. that physicians simultaneously become more error-prone and more cautious, admitting significantly more patients to the hospital unnecessarily, thus reducing the relative rate of wrongful discharges.[3] In the remainder of this section we investigate our hypotheses using the empirical strategy outlined in Section 6.2.

Given that ED congestion is significant in the selection equation (model (1) of Table 4) we must correct with the heckprob models for potential endogeneity to ensure that the coefficients of ED congestion in the outcome equations are not biased. Heckprob model coefficients are reported in Table 5. In heckprob (1e), (2e) and (3e) we identify the effect of ED congestion for only the subset of patients for whom the gatekeeping decision is made directly by an ED physician, i.e. censoring when $CDU_i = 1$. For completeness, in heckprob (1c), (2c) and (3c) we report this instead for only those patients admitted to the CDU, i.e. censoring when $CDU_i = 0$.

After correcting for endogenous selection, we find evidence consistent with that of probits (2), (3) and (4) in Table 4. In particular, evidence from Table 5 show that when the ED is more congested, ED physicians are likely to make more gatekeeping errors (coef. = 0.020, $p$-value < 0.001 in heckprob (1e)) and are more likely to admit patients to the hospital unnecessarily (coef. = 0.027, $p$-value < 0.001 in heckprob (2e)). At the same time, ED physicians become less likely to discharge patients in error when the ED is congested (coef. = −0.016, $p$-value = 0.048 in heckprob (3e)).

[3] While we report the coefficients corresponding to CDU referral, we caution against interpreting the effects at this stage since this variable is endogeous – full details of how to estimate these coefficients is provided in Section 7.

**Table 4** **Base coefficient estimates using probit model specification.**

| | (1) CDU | (2) TotErr | (3) AdmErr | (4) DischErr |
|---|---|---|---|---|
| ED congestion | $0.059^{***}$ | $0.017^{***}$ | $0.025^{***}$ | $-0.016^{*}$ |
| | (0.003) | (0.004) | (0.004) | (0.007) |
| CDU referral | – | $0.024^{\dagger}$ | 0.009 | $0.135^{***}$ |
| | | (0.013) | (0.014) | (0.023) |
| CDU congestion | $-0.059^{***}$ | – | – | – |
| | (0.003) | | | |
| Phys. CDU use rate | $1.062^{***}$ | – | – | – |
| | (0.019) | | | |
| N | 377,331 | 377,331 | 377,331 | 119,474 |
| Log-lik | −97,302 | −63,134 | −54,168 | −14,626 |
| Pseudo-$R^2$ | 0.201 | 0.164 | 0.196 | 0.085 |

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

All of this evidence is consistent with ED physicians overcompensating for the increased clinical uncertainty when congestion increases, by increasing the rate at which they admit uncertain cases beyond the level required to keep the wrongful discharge rate constant.[4]

While not the primary effect of interest, the impact of demand pressure in the ED on the likelihood of a patient who is in the CDU being admitted to the hospital in error is interesting (coef. $= 0.031$, $p$-value$= 0.011$ in heckprob (1c)). This suggests that the ED exerts downstream pressure on the CDU when busy to release additional buffer capacity. In particular, as the ED becomes more congested, more patients are referred to the CDU (coef. $= 0.059$, $p$-value$< 0.001$ in probit (1) from Table 4), and from the CDU more patients are admitted to the hospital, which frees up capacity to accept additional incoming patients.[5] This effect is interesting from an incentive perspective. In particular, ED physicians may have an incentive to divert patients to the CDU during periods of peak physician shortage in the ED, as this can help to prevent critical gatekeeping errors. This forces the CDU to make faster decisions, which reduces the benefit of CDU during peak demand. It is important, therefore, that ED physicians only admit those patients to the CDU that will benefit most from being there. How can they be incentivized to do this? This is an interesting problem beyond the scope of this paper that might be taken up in future research.

To give an idea of the scale of the effects, we convert coefficient estimates into average partial

[4] If admitting a patient to the hospital were administratively less time consuming for ED physicians than discharging them home, then this could provide an alternative explanation for these findings, i.e., physicians may err towards admitting patients when busy to save time. However, in our particular context the opposite is true: additional paperwork in the form of a venous thromboembolism assessment and drug chart must be completed if a patient is to be admitted. Consequently, admission is in fact more time consuming, meaning that all else being equal we would expect fewer and not more unnecessary admissions, ruling out this alternative explanation.

[5] Note that while the coefficients appear to suggest that physicians in the CDU are more affected by congestion than those in the ED (0.031 versus 0.027), converting to average partial effects we find that the effect on ED physicians is about twice as large as on their CDU counterparts.

**Table 5 Coefficient estimates to establish ED physicians' response to increased congestion, using heckprob model specification.**

| | Decision made by ED physicians | | | Decision made in the CDU | | |
|---|---|---|---|---|---|---|
| | (1e) TotErr | (2e) AdmErr | (3e) DischErr | (1c) TotErr | (2c) AdmErr | (3c) DischErr |
| ED congestion | 0.020$^{***}$ | 0.027$^{***}$ | −0.016$^{*}$ | 0.026$^{*}$ | 0.031$^{*}$ | −0.010 |
| | (0.004) | (0.005) | (0.008) | (0.012) | (0.012) | (0.022) |
| $\rho$ | −0.056 | 0.005 | −0.178$^{\dagger}$ | 0.195$^{*}$ | 0.209$^{**}$ | −0.078 |
| | (0.059) | (0.066) | (0.090) | (0.076) | (0.076) | (0.141) |
| N | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 |
| N uncensored | 339,975 | 339,975 | 339,975 | 37,356 | 37,356 | 37,356 |
| Log-lik | −152,082 | −144,216 | −109,902 | −105,420 | −104,287 | −99,502 |

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

(marginal) effects (APEs) with 95% confidence intervals ($CI_{95}$s). This shows that moving from a low ($2\sigma$ below the mean) to high ($2\sigma$ above the mean) congestion state in the ED will: increases in absolute terms the probability of admission to the CDU by 3.29%, $CI_{95} = (2.94\%, 3.64\%)$, total gatekeeping errors by 0.70%, $CI_{95} = (0.39\%, 1.02\%)$, being hospitalized unnecessarily by 0.80%, $CI_{95} = (0.51\%, 1.10\%)$, and decreases the wrongfully discharge probability by $-0.13\%$, $CI_{95} =$ (-0.25%,-0.00%). Compared with the average rate of CDU use and error rates reported in Table 1, this represents a relative increase (decrease) of approximately 39.8%, 15.3%, 20.9% and −16.9%, respectively. The congestion state of the ED thus has a surprisingly large impact on the decision taken by physicians in the ED. In fact, if we (conservatively) assume a cost of £1000 per unnecessary hospitalization, then if all patients had been treated in the ED during a period of high congestion, over-referrals by ED physicians would have cost the hospital approximately £5m extra.

### 6.4. Robustness

We have argued that the increase in unnecessary hospitalization and decrease in wrongful discharges is an indication of ED physicians becoming more cautious when faced with increasing levels of diagnostic uncertainty. An alternative explanation could be that as the ED becomes more congested the risk profile of the patients changes. We address this concern in Section EC.5 of the e-companion by using an instrumental variable approach to demonstrate that after accounting for potential correlation between ED congestion and the error terms the results are, if anything, conservative.

We have also tested the robustness of the results to different definitions of wrongful discharge – using a 3-day time window for readmission – and unnecessary hospitalization – using (i) a 12- or 48- hour time window for discharge after admission, (ii) requiring that the patient not only does not receive treatment but also that no diagnosis is assigned, and (iii) comparing inpatient length of stay to median length of stay for patients within the same disease category. We also test sensitivity to different time periods for measuring congestion, time-averaging over the first two hours and four hours post-admission, and the one hour prior to discharge. All results (reported in Section EC.4 of the e-companion) are consistent, other than when measuring congestion over the one hour prior to discharge. As we show, this is because congestion close to discharge does not capture the delay in the time to been seen by an ED physician (refer to Figure 1) that congestion on arrival causes.

Finally, in Section EC.7 of the e-companion, we test an alternative model set-up where we explore the effect of congestion on admission (discharge) decisions, and subsequently on unnecessary hospitalizations (wrongful discharges) *conditional* on being admitted (discharged). Conceptually, this corresponds to a model in which we map the patient health status into one dimension, e.g., severity. Those more severe (above some upper threshold $U$, say) will be admitted, those less severe (below threshold $L$, say) discharged, while those in the middle interval are candidates for the CDU. This model allows us to see how the admission and discharge thresholds ($U$ and $L$) change with congestion, and the subsequent effect on error rates. We find consistent results using this approach.

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

Having established the response of physicians in the ED to increased congestion and diagnostic uncertainty, we next look at what action might be taken to mitigate this effect. We would like to know whether the two-stage gatekeeping process reduces the high rate of gatekeeping errors in referrals of patients from the ED into acute inpatient beds. In this section we describe the method of estimation and present results.

### 7.1. Empirical specification

The empirical approach that we adopt is similar to that described in Section 6.1. Rather than use a heckprob model we estimate the models in Equations(1)–(4) with a recursive bivariate probit (biprobit) model with full information maximum likelihood (Maddala 1983). These models have the same error structure as the heckprob model but differ in that censoring is not performed and $\alpha_2, \beta_2, \gamma_2$ (the coefficient that captures the effect of CDU admission on a patients likelihood of being a wrongful discharge, unnecessary hospitalization, or total gatekeeping error, respectively) are left as free parameters to be estimated in the models. We first ask whether there is evidence that decoupling the gatekeeping decision and allowing ED physicians to, when uncertain, pass on the gatekeeping decision to a second gatekeeping stage can help to reduce total gatekeeping errors and unnecessary hospitalizations. This would be confirmed by coefficients $\beta_2 < 0$ and $\gamma_2 < 0$ in the respective outcome equations. We are also interested in if there is any evidence of a change in the rate of wrongful discharges, estimated by $\alpha_2$, when patients are routed through the CDU.

Two controls deserve special attention here. In Section 3.3 we noted that one of the benefits of the CDU in the NHS context of our study is that it allows ED physicians to increase the time that a patient is under assessment and observation beyond the 4-hour target. After admission to the CDU, one might ask, then, what drives the change in a patient's likelihood of being a wrongful discharge or unnecessary hospitalization? Is it the additional time that these patients spend receiving further diagnostic evaluation and observation? Or is it due to the co-location of these patients with more complex diagnostic needs and the ability to match more specialized, e.g., staff, equipment, and processes to these patients. If the former, then one could argue that any benefits from the two-stage system could be achieved in a single-stage system without time constraints. Therefore, to demonstrate the benefit of the two-stage system we are interested in isolating the second of these effects. To achieve this, we control for the duration of time that a patient spends (i) in the ED and (ii) in the CDU, if appropriate. 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 patients spent under observation in the ED and/or the CDU. For more on this point as well as additional robustness checks, see Section EC.6 of the e-companion.

### 7.2. Results

Table 6 shows evidence of positive correlation between the selection and outcome equations in all of the biprobit models, with estimated correlation coefficients $\rho = 0.193$ ($p$-value$< 0.001$), $\rho = 0.173$ ($p$-value$< 0.001$), and $\rho = 0.187$ ($p$-value$< 0.001$). 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 being a total gatekeeping error, unnecessary hospitalization or wrongful discharge. This is consistent with expectation: patients admitted to the CDU should be more complicated than the average ED arrival, else this more expensive service would be being used inappropriately.

These biprobit models provide strong evidence that patients admitted to the CDU are significantly less likely to (i) be hospitalized unnecessarily (coef.$= -0.412$, $p$-value$< 0.001$ in column (2o)) and (ii) be wrongfully discharged (coef.$= -0.201$, $p$-value$= 0.026$ in column (3o)). Unsurprisingly, total gatekeeping errors also decrease (coef.$= -0.408$, $p$-value$< 0.001$ in column (1o)). This confirms our hypothesis that routing customers with unresolved diagnosis through a two-stage gatekeeping system can help to significantly reduce the number of gatekeeping errors made.

To see how much better admission decisions are when made in the CDU rather than by a physician in the ED, we convert coefficient estimates to average treatment effects (ATEs) and average treatment effects on the treated (ATTs). This shows that if no patients were referred through the CDU the rate of unnecessary hospitalizations and wrongful discharges would have been 4.91% and 0.79%, respectively. These change to 2.25% and 0.46%, respectively, if all patients are instead routed through the CDU. Thus the CDU acts to significantly reduce unnecessary

**Table 6 Coefficient estimates for CDU impact.**

| | (1) TotErr | | (2) AdmErr | | (3) DischErr | |
|---|---|---|---|---|---|---|
| | (1s) CDU | (1o) TotErr | (2s) CDU | (2o) AdmErr | (3s) CDU | (3o) DischErr |
| CDU referral | – | $-0.408^{***}$ | – | $-0.412^{***}$ | – | $-0.201^{*}$ |
| | | (0.043) | | (0.044) | | (0.090) |
| CDU length of stay | – | $-0.002$ | – | 0.002 | – | $-0.002$ |
| | | (0.003) | | (0.005) | | (0.003) |
| CDU congestion | $-0.079^{***}$ | – | $-0.080^{***}$ | – | $-0.078^{***}$ | – |
| | (0.003) | | (0.003) | | (0.003) | |
| Phys. CDU use rate | $1.009^{***}$ | – | $1.017^{***}$ | – | $1.104^{***}$ | – |
| | (0.021) | | (0.021) | | (0.019) | |
| $\rho$ | $0.193^{***}$ | | $0.173^{***}$ | | $0.187^{***}$ | |
| | (0.023) | | (0.022) | | (0.051) | |
| N | 377,331 | | 377,331 | | 377,331 | |
| Log-lik | −158,845 | | −149,987 | | −109,287 | |

*Notes:* All estimations made using a biprobit model specification; *Robust standard error* in parentheses; Columns (1s), (2s) and (3s) report coefficient estimates for the first-stage (selection) equation, while columns (1o), (2o) and (3o) report coefficients for the second-stage (outcome) equation; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

hospitalizations (ATE$= -2.67\%$) and wrongful discharges (ATE$= -0.33\%$). The ATTs are even larger than the ATEs, taking values $-5.7\%$ and $-0.8\%$ for unnecessary hospitalization and wrongful discharge, respectively. This suggests that ED physicians are especially good at routing into the CDU patients who would have otherwise been a gatekeeping error.

### 7.3. Robustness

In the e-companion we report results from robustness tests that were performed to verify the benefits of the CDU in reducing errors. In Section EC.3, we report results using 1:1 nearest neighbour 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, respectively). Results are nearly identical to those presented in Section 7.2. Additionally, in Sections EC.4.1 and EC.4.2 we report consistent results using different definitions of unnecessary hospitalization and wrongful discharge, respectively (see Section 6.4 for details on the alternative definitions).

## 8. Counterfactual Analysis

Having established the usefulness of the CDU, the natural question is: how large should the CDU be? After all, the resources in the CDU 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 rates of unnecessary hospitalization. Thus, even though we find that the two-stage gatekeeping system reduces gatekeeping errors, the effect of the second gatekeeping stage in the combined system is not obvious. To examine the combined system, we perform a counterfactual analysis.

In this analysis we need to take account of two factors. First, any increase in ED capacity would translate into a reduction in ED congestion, as the resources (e.g., physicians, nurses, treatment rooms) consumed by the CDU would have been available for use by the ED instead. Second, that physicians in the CDU may be more experienced/skilled than those in the ED, and so redeploying them from the CDU to the ED would increase the average skill-level of physicians in the ED.

To account for the former, recall from Section 5.1 that we have a measure of ED capacity, $CapacityED_h$, for every hour $h$. Using the same approach, we can 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 of arrival of patient $i$, gives us our estimate of what congestion would have been in the combined system when patient $i$ arrived.[6] To ensure that the original and updated measures of ED congestion 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 could have been redeployed in the ED then this would have had the effect of reducing average congestion levels in the ED by approximately $0.59\sigma$.

To adjust for the latter, recall that we control in our models (see Appendix A) for the historic unnecessary hospitalization rate of the ED physician to whom a patient is assigned. A physician who is particularly error-prone will have a higher value for this variable, and one who is more experienced and skilled a lower value. We can use this information to 'upgrade' the skill level of physicians operating in the ED after pooling. Suppose that the physicians in the CDU make more accurate gatekeeping decisions than 90% of their ED counterparts. Then $P_{10}(PhysAdmErr)$, the 10th percentile of observed rates of unnecessary hospitalization by ED physicians, gives the estimated accuracy of CDU physicians when making admission decisions. Now let $\mu_i(PhysAdmErr)$ give the average accuracy of those physicians assigned to patients who arrive at a similar time as patient $i$, e.g. over the one hour prior and one hour post $i$'s arrival. If we were to combine ED and CDU capacity, then we suppose that this average would have instead been $\mu_i^*(PhysAdmErr) = \frac{\mu_i(PhysAdmErr)CapacityED_{h_i} + P_{10}(PhysAdmErr)CapacityCDU_{h_i}}{CapacityED_{h_i} + CapacityCDU_{h_i}}$. If patient $i$ had arrived under these conditions then the chance of them being assigned to a more skilled physician would be higher. To capture this, we suppose that the accuracy of the physician assigned to patient $i$ would be upgraded from the observed value in the two-stage system, $PhysAdmErr_i$, to the expected value in a single stage system, $PhysAdmErr_i^* = PhysAdmErr_i - (\mu_i(PhysAdmErr) - \mu_i^*(PhysAdmErr))$.

Substituting (i) the original values of $zEDCong_i$ for the updated values achieved through pooling ED and CDU capacity, $zEDCong_i^*$, and (ii) the original values of $PhysAdmErr_i$ for the updated

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

values, $PhysAdmErr_i^*$, into heckprob (1e), we estimate that in the pooled system the rate of unnecessary hospitalizations would be reduced by 0.26 percentage points. (Approximately 50% of the reduction comes from the decrease in congestion, the rest coming from the improvement in average physician ability). In Section 7.2 we estimated that the rate of unnecessary hospitalizations would rise from the current level of 4.34% with the CDU to 4.91% if the CDU were closed. However, this ignored the possibility of redistributing resources from the CDU, if it were closed, to increase capacity and physician skill levels in the ED. Accounting for this, we estimate instead that if the CDU were closed and resources could be redistributed to the ED, then the unnecessary hospitalization rate would equal 4.65% (= 4.91% − 0.26%) − still a deterioration relative to the status quo of 4.34%. Note that this represents a lower bound on the benefit of the CDU, assuming all patients admitted to the CDU can be redistributed in the hospital at no cost (see Footnote 6).

While this analysis demonstrates the advantage of the CDU over a pooled ED resource in the study hospital, the question of how resources should be distributed between the ED and CDU to optimize patient flow and minimize gatekeeping errors remains. This requires analytical work that goes beyond the scope of this empirical paper and is left for future research.

## 9. Conclusions

This empirical study of the effect of system congestion on gatekeeping errors in the context of an ED provides two related insights. First, our data shows that when ED congestion increases, the ED physicians in our study hospital change the trade-off point on the speed-quality curve in the interest of speed, as expected. This leads to a deterioration of the quality of their gatekeeping decisions. However, the rates of missed referrals (wrongful discharges) and unnecessary referrals (unnecessary hospital admissions) do not increase proportionally. When gatekeepers regard a missed referral as a more severe error than an unnecessary referral, as is the case in the context of an emergency department, they protect the rate of the more severe error by lowering the threshold for specialist referrals when congestion increases. This increases the rate of unnecessary referrals disproportionately and therefore causes excess false demand for downstream specialist services precisely at a time when the gatekeeper should ration access to specialists more stringently to protect the specialist unit from the upstream demand surge. The result is a demand amplification effect (ED bullwhip), where an upstream demand surge in the gatekeeping system causes a relatively larger downstream demand surge in the specialist unit. In the ED context, this has repercussions for the safety and efficiency of the hospital as a whole (Kuntz et al. 2015, Eriksson et al. 2017).

To alleviate this problem, operations managers would benefit from system changes that are designed to adjust the speed-quality trade-off without a significant increase in cost. The second insight of this paper is that a two-stage gatekeeping system can achieve this when there is considerable heterogeneity in the difficulty of the gatekeeping decision. The two-stage system, implemented in our study ED as a clinical decision unit (CDU), allows first-stage gatekeepers to better match the heterogeneous patient characteristics that become apparent during the first gatekeeping stage with more appropriate resources and gatekeeper characteristics at the second stage, leading to overall improved decision quality. This study provides empirical evidence that a two-stage gatekeeping system can significantly reduce rates of both missed and unnecessary specialist referrals.

While 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, on average, seven years in the US and five years in the UK (Shire 2013). Such cases are costly as patients visit their primary care physician (PCP) 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 PCPs (with a track-record of identifying rare diseases) as second-stage gatekeepers, allowing PCPs 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 ED becomes congested, more patients are referred into the CDU. If the CDU were not present, then the hospital inpatient units might instead be used for this purpose. Given the choice between a patient being referred into the CDU versus into the inpatient units, we may prefer them to spend time in the CDU since, e.g., admission to the hospital exposes patients to additional risks and is costly (see the discussion in Section 1), and when patients are referred unnecessarily to the CDU they spend less time there than if they had instead been admitted (see Section EC.1 of the e-companion). While outside the scope of this paper, future work might explore some of the other benefits of multi-stage gatekeeping.

## References

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

Anand K, Paç M, Veeraraghavan S (2011) Quality-speed conundrum: Trade-offs in customer-intensive services. *Management Science* 57(1):40–56.

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, 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.

Boyle A, Beniuk K, Higginson I, Atkinson P (2012) Emergency department crowding: Time for interventions and policy evaluations. *Emergency Medicine International* .

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.

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 C, Farias V, 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.

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.

Debo L, Toktay L, Van Wassenhove L (2008) Queuing for expert services. *Management Science* 54(8):1497–1512.

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.

Eriksson C, Stoner R, Eden K, Newgard C, Guise J (2017) The association between hospital capacity strain and inpatient outcomes in highly developed countries: A systematic review. *Journal of General Internal Medicine* 32(6):686–696.

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 (2018) Economies of scale and scope in hospitals: An empirical study of volume spillovers, London Business School, Working paper.

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 (2013) The incidence of diagnostic error in medicine. *BMJ Quality & Safety* 22(Suppl 2):ii21–ii27.

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

Greenwald PW, Estevez RM, Clark S, Stern ME, Rosen T, Flomenbaum N (2016) The ED as the primary source of hospital admission for older (but not younger) adults. *The American Journal of Emergency Medicine* 34(6):943–947.

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.

Hassan T (2003) Clinical decision units in the emergency department: Old concepts, new paradigms, and refined gate keeping. *Emergency Medicine Journal* 20(2):123–125.

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 C, 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.

Kostami V, Rajagopalan S (2013) Speed-quality trade-offs in a dynamic model. *Manufacturing & Service Operations Management* 16(1):104–118.

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.

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

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

NAO (2013) Emergency admissions to hospital: managing the demand. Technical report, National Audit Office, URL `https://www.nao.org.uk/report/emergency-admissions-hospitals-managing-demand/`, Last accessed: 2016-09-21.

NHE (2016) Worst NHS figures ever triggered by 'unprecedented funding slowdown'. Technical report, National Health Executive, URL `http://www.nationalhealthexecutive.com/Health-Care-News/worst-nhs-performance-figures-ever-triggered-by-unprecedented-funding-slowdown`, Last updated: 2016-04-16, Last accessed: 2016-09-15.

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.

Paç M, Veeraraghavan S (2015) False diagnosis and overtreatment in services, the Wharton School, Working paper.

Pines JM, Hilton JA, Weber EJ, Alkemade AJ, Al Shabanah H, Anderson PD, Bernhard M, Bertini A, Gries A, Ferrandiz S, Kumar VA, Harjola VP, Hogan B, Madsen B, Mason S, Öhlén G, Rainer T, Rathlev N, Revue E, Richardson D, Sattarian M, Schull MJ (2011) International perspectives on emergency department crowding. *Academic Emergency Medicine* 18(12):1358–1370.

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`.

Ross MA, Aurora T, Graff L, Suri P, O'Malley R, Ojo A, Bohan S, Clark C (2012) State of the art: Emergency department observation units. *Critical Pathways in Cardiology* 11(3):128–138.

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 (Articles in Advance)*

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.

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 (2018) Capacity pooling in hospitals: The hidden consequences of off-service placement, The Wharton School, Working paper.

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.

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.

Tang N, Stein J, Hsia RY, Maselli JH, Gonzales R (2010) Trends and characteristics of US emergency department visits, 1997-2007. *JAMA* 304(6):664–670.

Wang X, Debo L, Scheller-Wolf A, Smith S (2010) Design and analysis of diagnostic service centers. *Management Science* 56(11):1873–1890.

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 Section 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/diagnosis related and contextual controls outlined in Table 2, 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 baseline risk of a patient 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 defined before 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 2, which capture a physician's historic unnecessary hospitalization, wrongful discharge and total gatekeeping error rates, are calculated in the same way as for CDU admission propensity.

## Appendix B: Comparing Unnecessary and Avoidable Unnecessary Admission Rates

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 admissions and 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 rate of unnecessary admissions 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

$$r'_{ua} = \frac{N'_{ua} N - N' N_{ua}}{N^2} = \frac{N'_{ua}(N - N_{ua})}{N^2} = r'_u + \frac{N'_{ua} N_{un}}{N^2} \geq r'_u.$$

# 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 Efficiency

The results in our paper suggest that implementing an intermediate unit that exists between the ED and hospital inpatient units (the CDU in our case) where patients for whom there exists considerable diagnostic uncertainty can be admitted can help to reduce the number of unnecessary hospital admissions. However, we must also show that this intermediate unit can operate more efficiently than a standard inpatient unit else it offers little benefit (instead all patients who are currently referred in to the CDU could simply be admitted to the hospital). Here we compare these two alternatives.

Ignoring wrongful discharges, for which our analysis shows there is also an additional advantage of the CDU, in our sample of admitted patients there exist five classes of patient: those admitted from the ED to an inpatient bed who are (1) not unnecessary hospitalizations or (2) are unnecessary hospitalizations, and in addition those instead admitted to the CDU who are (3) discharged or are (4) admitted and subsequently not deemed to be an unnecessary hospitalization or are (5) admitted and then classed as an unnecessary hospitalization. Assume, conservatively, that for every patient who was admitted from the CDU (i.e., those of class (4) or (5)) all of the time they spent in the CDU was wasted, i.e., their LOS is not reduced at all despite the additional tests, better routing, etc. of patients after assessment in the CDU. For all 13,156 patients in our sample who enter the hospital via the CDU this thus adds up to 93,077 'wasted' hours. For the CDU to break-even, therefore, each of the 24,200 patients who are discharged from the CDU (i.e., those in class (3)) must have an average stay that is more than 3.85 hours shorter than it would have been if they had instead been admitted directly to the hospital.

To determine whether the condition above is satisfied, again we take a conservative approach and assume that if those patients who were discharged from the CDU had been admitted to the hospital instead then *all* of them would have been identified and discharged within 24h (with no treatment performed), i.e., they would instead have been unnecessary hospitalizations (i.e., of class (2)). Thus we need to compare the length of stay associated with patients of classes (2) and (3). In doing so we should account for differences in the characteristics of those patients admitted and subsequently discharged from the hospital directly rather than through the CDU, since e.g. the former may be inherently more risky and hence more likely to stay longer. To do this, 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, diagnosis related and contextual controls from Table 2. This model indicates that a patient treated in the CDU would have spent 8.78 hours more in the hospital if they had instead been admitted directly, meaning that the hospital 'saves'

199,937 hours of time as a consequence of ED physicians referring these patients to the CDU rather than admitting them directly to the hospital. The longer processing time of patients in the hospital than in the CDU is not surprising, since once admitted to a general inpatient ward heterogeneity of the patient pool increases, while the CDU is specifically set up to route patients in to the hospital who require hospitalization and discharge those who do not.

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

## Appendix EC.2: Relevance and Validity of the Instruments

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

### EC.2.1. Tests of Under- and Weak Identification

The underidentification test is a Lagrange multiplier (last month) 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, arises when the excluded instruments are correlated with the endogenous regressors, but only weakly. Estimators can perform poorly when instruments are weak: estimates may be inconsistent, tests for the significance of coefficients may lead to the wrong conclusions, and confidence intervals are likely to be be incorrect. Here we describe how we test for both of these properties.

First it is important to note that the majority of tests are based on a linear IV regression model where the dependent variable in the outcome equation and the endogeneous variable are continuous. In order to perform formal testing we therefore follow convention and treat the binary short-stay observational admission, 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 that are 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 of 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 2239.20 with 2 d.f., which has corresponding $p$-value $< 0.0001$. For the wrongful discharge model, the SW Chi-sq statistic takes value 2170.26 with 2 d.f. and corresponding $p$-value $< 0.0001$. This means that there is strong evidence to reject the null hypothesis of underidentification in both cases at e.g. the 0.1% significance level, and so it is possible to 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 or not the instruments perform poorly. The relevant test has 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 the conclusion is that the instruments are weak for that level of bias. Here, the estimated SW $F$-statistic is equal to 1119.23 for the unnecessary hospitalization model, and equal to 1084.78 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.

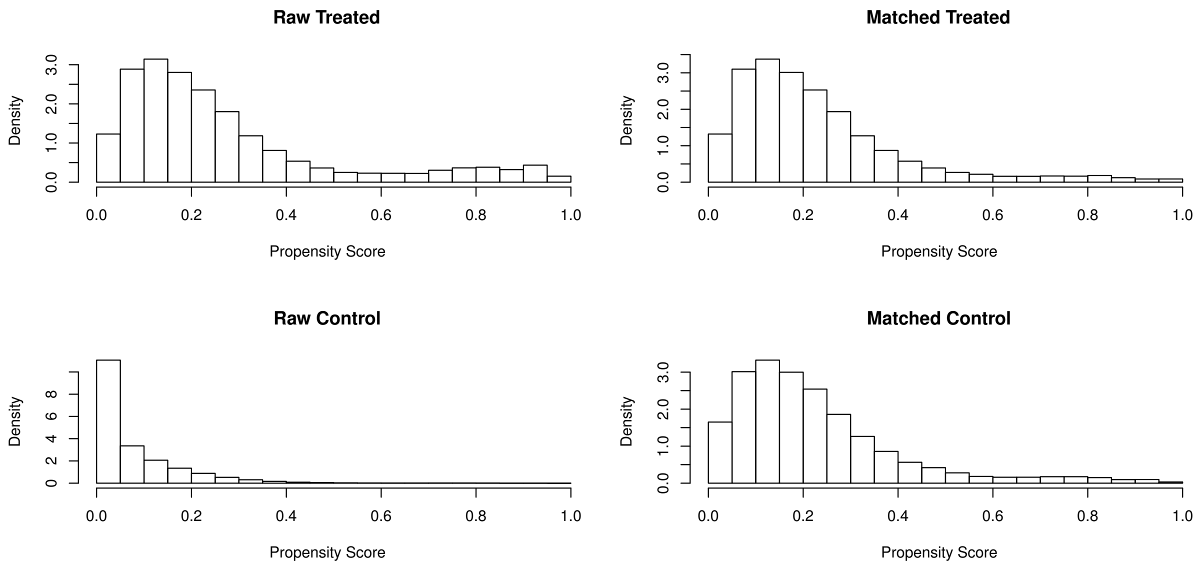
### EC.2.2. Testing for Overidentification

In addition to the excluded instruments being "relevant", it is also important to 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 influence 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 $\chi^2 = 4.21$, $p$-value $= 0.122 > 0.10$, for the wrongful discharge model $\chi^2 = 3.56$, $p$-value $= 0.169 > 0.10$. Together these results indicate no evidence of joint significance of the instruments, and hence we have no reason to suspect that they are not valid.

## Appendix EC.3: Propensity Score Matching

Matching is a method for reducing dependence on statistical modelling 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 idea of preprocessing using matching methods is to reduce the strength of the relationship between the treatment effect ($CDU_i$) and the control variables ($X_i$). The way in which most matching methods work is to retain all of the treated observations in the data set and to select 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.

**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.**




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 a large number of control variables then exact matching is not possible and instead the goal of matching methods is to balance the covariate distributions across the two groups (treated and untreated).

When one-to-one exact matching is not possible, there are various matching methods that can be used. In the discussion below we report results using nearest neighbour matching, with other methods giving 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. 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 rate of unnecessary hospitalization (resp., wrongful discharge) in the treatment group (i.e., those admitted to the CDU) and the control group (i.e., those not admitted to the CDU) was 5.06% (resp., 1.05%) and 4.25% (resp., 0.68%), respectively. After matching, the rate for the control group changes to 7.36% (resp., 0.80%) for unnecessary hospitalization (resp., wrongful discharge). Therefore, admission to the CDU does appear to reduce the unnecessary hospitalization rate, as found in the paper, but there is some question as to whether admission to the CDU may actually increase the wrongful discharge rate. We investigate this further below.

One limitation with the matching method is that it still does not account for the fact that patients may differ based on unobservables (though because they are more similar based on observables 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 72,134 patients. Results from the biprobit models are reported in Table EC.1.

**Table EC.1 Coefficient estimates for CDU impact – Matched sample.**

| | (1) TotErr | | (2) AdmErr | | (3) DischErr | |
|---|---|---|---|---|---|---|
| | (1s) CDU | (1o) TotErr | (2s) CDU | (2o) AdmErr | (3s) CDU | (3o) DischErr |
| CDU referral | – | $-0.419^{***}$ | – | $-0.462^{***}$ | – | $-0.092$ |
| | | (0.098) | | (0.098) | | (0.220) |
| CDU length of stay | – | $0.007^{\dagger}$ | – | $0.014^{*}$ | – | $-0.001$ |
| | | (0.004) | | (0.006) | | (0.003) |
| CDU congestion | $-0.088^{***}$ | – | $-0.089^{***}$ | – | $-0.092^{***}$ | – |
| | (0.005) | | (0.005) | | (0.005) | |
| Phys. CDU use rate | $1.206^{***}$ | – | $1.213^{***}$ | – | $1.261^{***}$ | – |
| | (0.036) | | (0.036) | | (0.036) | |
| $\rho$ | $0.203^{**}$ | | $0.194^{**}$ | | 0.150 | |
| | (0.060) | | (0.060) | | (0.134) | |
| N | 72,134 | | 72,134 | | 72,134 | |
| Log-lik | −64,990 | | −63,042 | | −51,828 | |

*Notes:* All estimations made using a biprobit model specification; *Robust standard error* in parentheses; Columns (1s), (2s) and (3s) report coefficient estimates for the first-stage (selection) equation, while columns (1o), (2o) and (3o) report coefficients for the second-stage (outcome) equation; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

We find results consistent with those in Section 7.2 of the paper. Similar to the paper, we convert coefficient estimates to average treatment effects (ATEs) and average treatment effects on the treated (ATTs). This shows that if none of the patients in our matched sample were referred through the CDU, then the rate of unnecessary hospitalizations and wrongful discharges would have been 9.00% and 1.08%, respectively. These change to 4.08% and 0.85%, respectively, if all patients in the matched sample are instead routed through the CDU. Thus the CDU acts to significantly reduce unnecessary hospitalizations (ATE$= -4.92\%$) and wrongful discharges (ATE$= -0.23\%$). The ATTs are more similar to the ATEs in the matched sample than in the full sample used in the paper, taking values $-5.4\%$ and $-0.3\%$ for unnecessary hospitalization and wrongful discharge, respectively.

One difference between the matched and unmatched results is that, when using the matched sample, the significance of the CDU in reducing wrongful discharges reduces and becomes insignificant (coef. $-0.092$, $p$-value$> 0.10$). This is due to the fact that the sample size is significantly reduced and we do not have enough power to achieve a tight confidence interval. In fact, only 277 out of 34,501 observations in the control group in the matched sample are wrongfully discharged, as compared to 2,299 in the control group in the unmatched sample. Thus, it is not entirely surprising that the standard errors are inflated and the significance of the CDU coefficient reduces. However, since the effect becomes insignificant, in the counterfactual analysis in Section 8 of the paper we take a conservative approach and discuss only unnecessary hospitalizations.

## Appendix EC.4: 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.4.1. Unnecessary Hospitalization

In Section 5.3 of the paper we discuss how being discharged 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;
- $N_u(c) = N_{ua}(c) + N_{uu}(c)$, 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;

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 team to go through every medical record and determine whether they believe the admission (discharge) error could have been avoided ex-ante, we instead create a measure for $N_u$ – the number of ex-post unnecessary admissions. We show in Appendix C of the paper that under mild conditions the effect size we observe using $N_u$ will, if anything, be an underestimate of the effect of congestion on $N_{ua}$.

Another question that naturally arises following this is whether being discharged 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 being 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 a team going through every medical record to determine whether they believe the 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 then 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 note that it is possible to define an admission error in various ways. Changing our definition is equivalent to changing the $\alpha$ and the $\beta$ discussed above. Below we describe four alternative definitions of an unnecessary admission, as well as a discussion of the expected effect on $\alpha$ and $\beta$:

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

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

3. Being discharged within 24 hours of hospitalization with no procedure performed **and with no ICD-10 diagnosis code assigned other than one relating to "Signs and Symptoms"**: This means

**Table EC.2 Coefficient estimates to establish ED physicians' response to increased congestion, using heckprob model specification – Using alternative definitions of unnecessary hospitalization.**

| | Decision made by ED physicians | | | |
|---|---|---|---|---|
| | (1) 12 hours | (2) 48 hours | (3) 24 hour w/ no diagnosis | (4) Short-stay |
| ED congestion | 0.021** | 0.024*** | 0.025*** | 0.022*** |
| | (0.007) | (0.004) | (0.007) | (0.005) |
| $\rho$ | −0.076 | 0.054 | −0.120† | −0.354*** |
| | (0.160) | (0.053) | (0.067) | (0.048) |
| N | 377,331 | 377,331 | 377,331 | 377,331 |
| N uncensored | 339,975 | 339,975 | 339,975 | 339,975 |
| Log-lik | −116,759 | −162,323 | −118,801 | −136,166 |

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

that the patient not only had no procedure performed, but also no clear diagnosis of an actual condition could be made while the patient was in the hospital. Our expectation is that doing so should significantly reduce $\alpha$. However, at the same time this may lead to a non-trivial decrease in $\beta$, since a patient may be diagnosed with a problem while in hospital 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". Now, 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.

4. Being discharged significantly faster ($< 0.1\times$ the median length of stay) than other patients admitted in the same diagnosis group: This measure is different in that it does not depend on whether or not a procedure was performed or a diagnosis was assigned, and is not measured over a fixed time period constant for all patients. Instead, it looks at patients with similar diagnoses, and decides whether or not they were discharged significantly faster than patients with the same diagnosis. The effect on the $\alpha$ and $\beta$ is unclear, and depends on whether being discharged significantly faster than predicted is a sign of reduced necessity or of lower severity.

The correlation between each of these measures and the measure of unnecessary hospitalization employed in the paper is 0.556, 0.778, 0.623, and 0.572, respectively.

In Table EC.2, we report coefficient estimates for the effect of congestion on the likelihood of a patient being classed as an unnecessary hospitalization in each of these cases. As can be seen, the coefficient of congestion is consistent, positive in sign, and significant in all cases, strong evidence that the direction of the congestion effect on unnecessary hospitalizations is as reported in the main paper.

In Table EC.3, we report the effect of CDU admission on the likelihood of a patient being classed as an unnecessary hospitalization, after controlling for endogeneity, as in Section 7 of the paper. Again, all results hold, suggesting the CDU effect is also robust against different definitions of an unnecessary hospitalization.

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

**Table EC.3 Coefficient estimates for CDU impact – Using alternative definitions of unnecessary hospitalization.**

| | (1) 12 hours | | (2) 48 hours | | (3) 24 hour w/ no diagnosis | | (4) Short-stay | |
|---|---|---|---|---|---|---|---|---|
| | (1s) CDU | (1o) AdmErr | (2s) CDU | (2o) AdmErr | (3s) CDU | (3o) AdmErr | (4s) CDU | (4o) AdmErr |
| CDU referral | – | $-0.471^{***}$ | – | $-0.411^{***}$ | – | $-0.186^{*}$ | – | $-0.619^{***}$ |
| | | (0.062) | | (0.036) | | (0.077) | | (0.047) |
| CDU length of stay | – | 0.001 | – | 0.000 | – | 0.000 | – | 0.008 |
| | | (0.006) | | (0.002) | | (0.009) | | (0.006) |
| CDU congestion | $-0.079^{***}$ | – | $-0.079^{***}$ | – | $-0.075^{***}$ | – | $-0.079^{***}$ | – |
| | (0.003) | | (0.003) | | (0.003) | | (0.003) | |
| Phys. CDU use rate | $0.934^{***}$ | – | $0.944^{***}$ | – | $0.957^{***}$ | – | $1.046^{***}$ | – |
| | (0.021) | | (0.020) | | (0.020) | | (0.020) | |
| $\rho$ | $0.230^{***}$ | | $0.147^{***}$ | | 0.024 | | $0.363^{***}$ | |
| | (0.033) | | (0.019) | | (0.040) | | (0.026) | |
| N | 377,331 | | 377,331 | | 377,331 | | 377,331 | |
| Log-lik | −117,615 | | −171,515 | | −121,194 | | −140,269 | |

*Notes:* All estimations made using a biprobit model specification; *Robust standard error* in parentheses; Columns (1s), (2s), (3s) and (4s) report coefficient estimates for the first-stage (selection) equation, while columns (1o), (2o), (3o) and (4o) report coefficients for the second-stage (outcome) equation; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models.
$^{***}p<0.001$, $^{**}p<0.01$, $^{*}p<0.05$, $^{\dagger}p<0.10$.

## EC.4.2. Wrongful Discharge

In the paper we define a wrongful discharge as a patient who is discharged from the ED or CDU, who then re-attends the ED within 7 days, has a diagnosis assigned within the same category as during their previous visit, and is subsequently admitted to the hospital. While we use a 7 days window, other time windows are used in the medical literature, e.g. 24 hours, 48 hours, 72 hours, etc. Given that wrongful discharge is a rare event even when calculated using a 7 day time window, shortening the time window by too much will make it such a rare event that our analysis is likely to lack power to identify an effect if one exists. However, we do want to test the robustness of the results to different time periods. As such, we have re-run the analysis from the paper where we use a 72 hour time window over which to observe re-attendances.

If we shorten the re-attendance time window from 7 days to 3 days (i.e., 72 hours), then the percentage of cases identified as wrongful discharges in the full sample drops to 0.54% from 0.71%. Interestingly, even though we shorten the observation window by approx. 57%, we only lose approx. 25% of the cases previously identified as wrongful discharges. This suggests that most of those patients who we flag as a wrongful discharge in the paper actually re-attend the ED within a short time-window of being discharged (approx. 75% return within 3 days). The correlation between the wrongful discharge rate for re-attendances over 7 days versus 3 days is 0.87. Given the high correlation, it is unlikely that the results will differ significantly on this subsample. Despite this, we have re-run the analysis from the paper with the shorter time window for re-attendance, and report the results for the effects of congestion on wrongful discharges in Table EC.4, and for the effect of CDU admission on wrongful discharges in Table EC.5.

Tables EC.4 and EC.5 show that our results are robust to the time window over which we measure re-attendances. In particular, looking at Column (3e) from Table EC.4, we see that at higher levels of congestion, fewer discharge errors occur ($p$-value $= 0.084$). Similarly, Column (3o) of Table EC.5 shows that patients who are admitted to the CDU are less likely to be wrongful discharges, after accounting for endogeneity ($p$-value $= 0.014$). Together this provides evidence that our results are robust to a shortening in the time window for re-attendance.

**Table EC.4 Coefficient estimates to establish ED physicians' response to increased congestion, using heckprob model specification – Wrongful discharge measured over a 3 day time window.**

| | Decision made by ED physicians | | | Decision made in the CDU | | |
|---|---|---|---|---|---|---|
| | (1e) TotErr | (2e) AdmErr | (3e) DischErr | (1c) TotErr | (2c) AdmErr | (3c) DischErr |
| ED congestion | 0.020$^{***}$ | 0.026$^{***}$ | −0.016$^{\dagger}$ | 0.032$^{**}$ | 0.032$^{**}$ | 0.007 |
| | (0.004) | (0.005) | (0.009) | (0.012) | (0.012) | (0.025) |
| $\rho$ | 0.001 | 0.037 | −0.188$^{\dagger}$ | 0.216$^{**}$ | 0.238$^{**}$ | −0.100 |
| | (0.070) | (0.067) | (0.104) | (0.073) | (0.072) | (0.158) |
| N | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 |
| N uncensored | 339,975 | 339,975 | 339,975 | 37,356 | 37,356 | 37,356 |
| Log-lik | −150,191 | −144,114 | −107,190 | −104,968 | −104,187 | −98,828 |

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

**Table EC.5 Coefficient estimates for CDU impact – Wrongful discharge measured over a 3 day time window.**

| | (1) TotErr | | (2) AdmErr | | (3) DischErr | |
|---|---|---|---|---|---|---|
| | (1s) CDU | (1o) TotErr | (2s) CDU | (2o) AdmErr | (3s) CDU | (3o) DischErr |
| CDU referral | – | −0.406$^{***}$ | – | −0.412$^{***}$ | – | −0.256$^{*}$ |
| | | (0.044) | | (0.044) | | (0.105) |
| CDU length of stay | – | −0.002 | – | 0.002 | – | 0.000 |
| | | (0.003) | | (0.005) | | (0.003) |
| CDU congestion | −0.079$^{***}$ | – | −0.080$^{***}$ | – | −0.078$^{***}$ | – |
| | (0.003) | | (0.003) | | (0.003) | |
| Phys. CDU use rate | 1.009$^{***}$ | – | 1.017$^{***}$ | – | 1.108$^{***}$ | – |
| | (0.021) | | (0.021) | | (0.019) | |
| $\rho$ | 0.182$^{***}$ | | 0.173$^{***}$ | | 0.189$^{**}$ | |
| | (0.024) | | (0.022) | | (0.060) | |
| N | 72,134 | | 72,134 | | 72,134 | |
| Log-lik | −156,678 | | −149,987 | | −106,077 | |

*Notes:* All estimations made using a biprobit model specification; *Robust standard error* in parentheses; Columns (1s), (2s) and (3s) report coefficient estimates for the first-stage (selection) equation, while columns (1o), (2o) and (3o) report coefficients for the second-stage (outcome) equation; Likelihood ratio (Pr $> \chi^2$) $< 0.0001$ in all models.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

### EC.4.3. 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, which means that they must make decisions under increased uncertainty. That this phenomena occurs is shown in Figure 1 of the paper, where we plot ED congestion against the time between a patient arriving and first being seen by an ED physician. Clearly as congestion increases, patients spend more time waiting to be seen. Due to the 4-hour waiting time target, waiting longer to be seen translates directly into shorter service times. Another effect that might be occurring is that when the ED is more crowded the physicians become more generally error prone, e.g., due to cognitive overload they are more likely to make an error.

One problem with our current workload measure is that while it is likely to catch the former effect (i.e., 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 admission of a patient to the ED. Since only 8.2% of patients are discharged within 1 hour, this means that we are not measuring workload at the time at which ED physicians are likely to be taking the disposition decision (i.e., closer to the time of discharge).

**Table EC.6 Coefficient estimates to establish ED physicians' response to increased congestion, using heckprob model specification – ED congestion measured over 2 hours post-arrival.**

| | Decision made by ED physicians | | | Decision made in the CDU | | |
|---|---|---|---|---|---|---|
| | (1e) TotErr | (2e) AdmErr | (3e) DischErr | (1c) TotErr | (2c) AdmErr | (3c) DischErr |
| ED congestion | 0.015*** | 0.022*** | −0.020* | 0.031** | 0.038** | −0.013 |
| | (0.004) | (0.005) | (0.008) | (0.011) | (0.012) | (0.022) |
| $\rho$ | −0.032 | 0.028 | −0.137 | 0.220** | 0.237** | −0.060 |
| | (0.060) | (0.067) | (0.093) | (0.072) | (0.072) | (0.132) |
| N | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 |
| N uncensored | 339,975 | 339,975 | 339,975 | 37,356 | 37,356 | 37,356 |
| Log-lik | −151,986 | −144,115 | −109,787 | −105,326 | −104,184 | −99,393 |

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

**Table EC.7 Coefficient estimates to establish ED physicians' response to increased congestion, using heckprob model specification – ED congestion measured over 4 hours post-arrival.**

| | Decision made by ED physicians | | | Decision made in the CDU | | |
|---|---|---|---|---|---|---|
| | (1e) TotErr | (2e) AdmErr | (3e) DischErr | (1c) TotErr | (2c) AdmErr | (3c) DischErr |
| ED congestion | 0.012** | 0.018*** | −0.020* | 0.040*** | 0.050*** | −0.018 |
| | (0.004) | (0.005) | (0.008) | (0.011) | (0.012) | (0.021) |
| $\rho$ | −0.039 | 0.025 | −0.141 | 0.220** | 0.238** | −0.061 |
| | (0.060) | (0.068) | (0.092) | (0.072) | (0.072) | (0.133) |
| N | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 |
| N uncensored | 339,975 | 339,975 | 339,975 | 37,356 | 37,356 | 37,356 |
| Log-lik | −151,981 | −144,108 | −109,780 | −105,317 | −104,174 | −99,386 |

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

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.

Given the above concern, we can perform some sensitivity analysis to determine what happens as we change the time windows over which we measure ED congestion. We investigate three variations. First, we measure congestion over the first (i) 2 hours and (ii) 4 hours after arrival to the ED (rather than 1 hour). While only 8.2% of patients are discharged within 1 hour of arriving in the ED, 29.5% are discharged within 2 hours, and 94.4% within 4 hours. Thus as we extend this time horizon out, 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), and thus we should expect similar results. Results are reported in Tables EC.6 and EC.7, and show that the findings in the paper are not sensitive to the choice of measuring congestion over the first hour versus over the first two or four hours after arrival of a patient.

In addition to the above, we have also measured congestion over the 1 hour prior to a patient departing from the ED. This measure should isolate the effect of physicians becoming more error prone when making disposition decisions while busy, because it is measured close to the time when the disposition decision is being made. However, it is less likely to capture the effect of congestion on increased waiting time (and hence also not capture the effects of increased diagnostic uncertainty), as this measure is taken after the patient is

**Figure EC.2 Mean time between patient arrival at the ED 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 (left) post-arrival in the ED versus (right) pre-departure from the ED.**

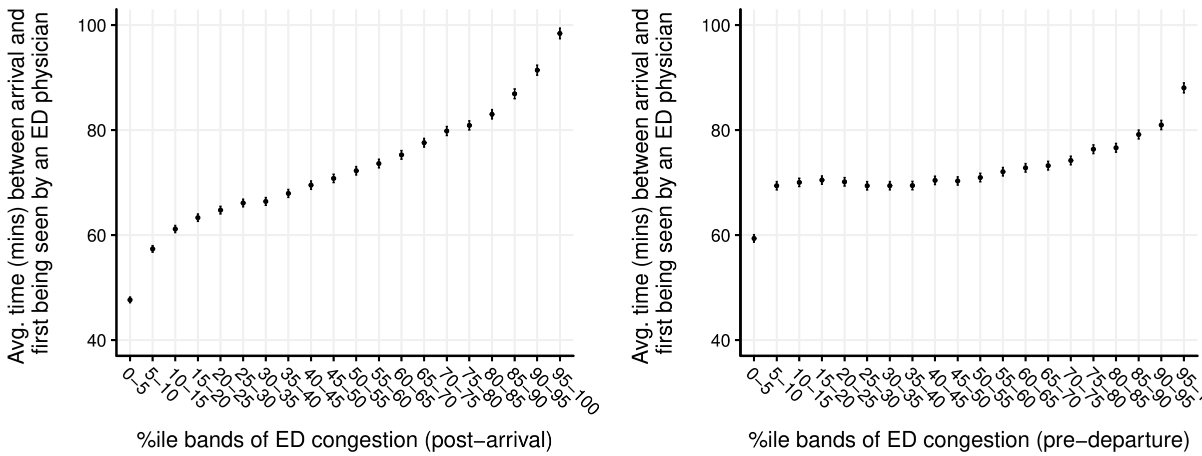



**Table EC.8 Coefficient estimates to establish ED physicians' response to increased congestion, using heckprob model specification – ED congestion measured over 1 hour pre-discharge.**

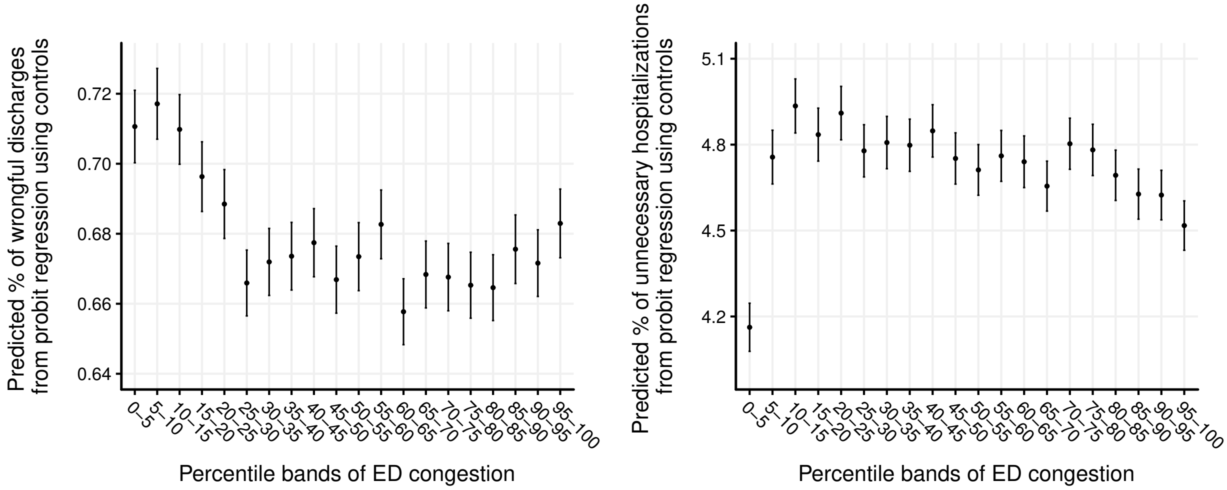
| | Decision made by ED physicians | | | Decision made in the CDU | | |
|---|---|---|---|---|---|---|
| | (1e) TotErr | (2e) AdmErr | (3e) DischErr | (1c) TotErr | (2c) AdmErr | (3c) DischErr |
| ED congestion | 0.005 | $0.008^{\dagger}$ | −0.004 | $0.034^{**}$ | $0.048^{***}$ | −0.028 |
| | (0.004) | (0.005) | (0.008) | (0.011) | (0.012) | (0.021) |
| $\rho$ | −0.040 | 0.023 | −0.146 | $0.220^{**}$ | $0.237^{**}$ | −0.055 |
| | (0.060) | (0.069) | (0.090) | (0.073) | (0.073) | (0.133) |
| N | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 |
| N uncensored | 339,975 | 339,975 | 339,975 | 37,356 | 37,356 | 37,356 |
| Log-lik | −152,006 | −144,136 | −109,813 | −105,349 | −104,201 | −99,412 |

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

already in service, rather than before they start service. That this is the case can be demonstrates visually. Specifically, Figure EC.2 (left) re-produces the left-hand plot of Figure 1 from 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. As can be seen in Figure EC.2, 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 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. This is a problem for us, since this is one of the main effects that we are interested in.

Despite this, the results are reported in Table EC.8. The fact that most of the coefficients become insignificant indicates that the main explanation for why error rates increase with congestion is that service times are shortened, leading to increased diagnostic uncertainty when decisions are being made, as hypothesized. Thus, we prefer to use in the paper congestion measures taken from time of arrival rather than from time of departure.

**Figure EC.3** **Mean likelihood of a patient being a wrongful discharge (left) or unnecessary admission (right) as a function of ED congestion, with 95% confidence bands.**




## Appendix EC.5: Endogeneity of ED Congestion

In the paper we claim that the increase in the rate of unnecessary hospitalizations (and reduction in the rate of wrongful discharges) at higher levels of ED congestion can be explained as a consequence of physicians adjusting their admission threshold when busy, and adopting a preference for ‘safety-first’ in the face of increased diagnostic uncertainty. An alternative explanation that we must consider is that when the ED is more congested, it may be because of a surge in arrivals of patients who are both more likely to be admitted in error and less likely to be discharged wrongfully.

In order to check whether this is the case, we adopt a two-step approach. In the first step, we look to see whether based on *observables*, there is any evidence that patients arriving when the ED is busier are more likely to be unnecessary hospitalizations or wrongful discharges. If based on observables this seems to be the case, then we can adjust for this in our analysis by controlling for these factors. However, this leaves the possibility that these patients are also more likely to be errors based on *unobservables*. We discuss our investigations into these effects in this section.

### EC.5.1. Observables

First, it is important to note that in our models we adjust for temporal variation with a range of controls (e.g., hourly dummies, day of the week, month, school holidays, trend, year) and also remove time periods where the patient population may significantly differ (e.g., public holidays, the Christmas period). Thus, the effects we observe are unlikely be a consequence of systematic time-related correlation between patient error propensity and ED congestion. Despite this, we look to see whether there is any evidence that patients differ in their error propensities at different levels of ED congestion. Figure EC.3 shows the fitted values from a probit regression of (left) wrongful discharges and (right) unnecessary hospitalizations against the set of control variables given in Table 2 of the paper.

Figure EC.3 shows two interesting features. First, when the ED is relatively quiet (i.e., ED congestion is in the lower quantile (i.e., $< 25\%$)) patients appear to be more likely to be wrongful discharges based on observables. After the 25th percentile, there appears to be little correlation between wrongful discharge propensity and congestion based on observables. Second, when the ED is extremely quiet (i.e., ED congestion is less than the 5th percentile), the chance of a patient being an unnecessary hospitalization is very low based on observables. This jumps after the 5th percentile, before gradually decreasing as ED congestion increases beyond that level. (Note that the vast majority of those patients who experience extremely low levels of congestion arrive to the ED during the early hours of the morning (between midnight and 6am) when few other patients are in the ED.) Based on observables, then, there is some evidence that a patient's likelihood of being a wrongful discharge or unnecessarily hospitalization varies with congestion levels.

Now, since we control for the set of variables used in creating Figure EC.3 in our models, the relationship shown in Figure EC.3 will be taken care of (i.e., this effectively becomes a flat line in the final regression). However, it is still possible that there are also unobservable factors that affect the likelihood of a patient being wrongfully discharged or unnecessarily admitted that also vary as congestion in the ED varies. The rest of this section is devoted to addressing these concerns.

### EC.5.2. Unobservables

To start, let us assume that the extent to which patients differ in their error propensity based on unobservables is identical to the extent to which they differed based on observables. If this were the case, then how much would the effects that we report in the paper be an overestimate of the true effects? We can test this by regressing the predicted values used to generate Figure EC.3 against congestion and then comparing to the effects reported in the paper. Doing this, we find that a one standard deviation increase in congestion would result in a -0.0057% decrease and 0.012% increase in wrongful discharges and unnecessary hospitalizations, respectively, based on unobservables. This compares with a -0.031% decrease and 0.20% increase in wrongful discharges and unnecessary hospitalizations, respectively, found in the paper. In other words, if the magnitude to which patients differed in their error propensity based on unobservables is the same as the magnitude to which they differ based on observables, then this would explain only 18% $(= -0.031/-0.0057)$ of the effect of congestion on wrongful discharges that we find, and 6% $(= 0.012/0.20)$ of the effect on unnecessary hospitalizations. This means that the difference in a patient's error propensity with respect to congestion based on unobservables would have to be significantly greater than the difference based on observables. Since this is unlikely, this provides initial evidence that unobservables do not drive the effects.

Another test indicated by Figure EC.3 is to drop all observations where congestion levels are low (below the 25th percentile) and then re-estimate the congestion effects. Why do this? Because Figure EC.3 suggests that before the 25th percentile, there may be significant differences in patients' wrongful discharge propensity based on observables, while after this point there appears little difference in their error propensities. For unnecessary hospitalization, after the 25th percentile we see that based on observables we would expect congestion and error rates to be negatively correlated, which would work in the opposite direction to the effect that we find in the paper (in the paper we find that as congestion increases, the rate of unnecessary hospitalizations increases). Thus, if we assume similarly that patients differ based on unobservables below the 25th percentile but are approximately the same after that point, then re-estimating on this subsample should

**Table EC.9 Coefficient estimates to establish ED physicians' response to increased congestion, using heckprob model specification – Dropping observations here congestion less than 25th %ile.**

| | Decision made by ED physicians | | | Decision made in the CDU | | |
|---|---|---|---|---|---|---|
| | (1e) TotErr | (2e) AdmErr | (3e) DischErr | (1c) TotErr | (2c) AdmErr | (3c) DischErr |
| ED congestion | 0.025*** | 0.034*** | −0.021 | 0.024 | 0.018 | 0.032 |
| | (0.007) | (0.007) | (0.013) | (0.018) | (0.019) | (0.032) |
| $\rho$ | −0.057 | −0.009 | −0.209* | 0.179* | 0.191* | −0.003 |
| | (0.064) | (0.067) | (0.097) | (0.082) | (0.084) | (0.153) |
| N | 282,998 | 282,998 | 282,998 | 282,998 | 282,998 | 282,998 |
| N uncensored | 255,038 | 255,038 | 255,038 | 27,960 | 27,960 | 27,960 |
| Log-lik | −114,916 | −109,306 | −82,627 | −79,762 | −78,954 | −75,082 |

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

give us some sense as to how robust our results are. The results from the HeckProb models are reported in Table EC.9.

Table EC.9 shows that, if anything, the results reported in our paper are conservative. If we ignore those cases where congestion levels are low (below the 25th percentile), we find that every one standard deviation increase in congestion decreases wrongful discharges by -0.042% and increases unnecessary hospitalizations by 0.26% (as compared to a -0.031% decrease and 0.20% increase reported in the paper). Note that the effect of congestion on wrongful discharges (coef. −0.021) is statistically insignificant at the 10% level, though is borderline with a $p$-value of 0.104. This may be a power issue, however, since the coefficient becomes more negative but the standard error is inflated from 0.008 in Table 5 of the paper to 0.013 here. However, since the effect becomes insignificant, in the counterfactual analysis in Section 8 of the paper we take a conservative approach and discuss only unnecessary hospitalizations.

To add further credibility to our findings, we perform a further robustness check and use an instrumental variable approach to correct for any potential endogeneity that results from unobservable factors being correlated with both ED congestion and error propensity. To do this, we need an instrumental variable (IV) that is correlated with ED congestion but uncorrelated with a patient's likelihood of being admitted unnecessarily or discharged wrongfully. We choose for this IV the congestion of the ED at the same time and on the same day one week prior to the patient's arrival in the ED (this is similar to the IV used in, e.g., Tan and Netessine 2014). Due to the time lag, this should have no direct effect on the patient's likelihood of being admitted or discharged in error. However, to the extent that busy periods in hospitals tend to cluster, how congested the ED was in the previous week is expected to be correlated with ED congestion one week later. There is a strong positive association between the two variables, with correlation equal to $\rho = 0.458$, $p$-value$< 0.0001$. This provides initial evidence that the instrument satisfies the relevance condition. Thus, we take a two-step approach and first regress ED congestion against all of the exogenous covariates plus our IV using OLS, i.e.,

$$zEDCong_i = \omega_0 + \mathbf{X}_i\boldsymbol{\omega}_1 + \mathbf{Z}_i\boldsymbol{\omega}_2 + zEDCongLW_i\omega_3 + \epsilon_i^\omega\,, \qquad \text{(EC.2)}$$

where $\epsilon_i^\omega \sim \mathcal{N}(0, \sigma_\omega^2)$, and then substitute $zEDCong_i$ in the selection and outcome equations (e.g. in Equations (1) and (3) in the paper) with the fitted values from the regression specified in Equation (EC.2), i.e.,

**Table EC.10 Coefficient estimates to establish ED physicians' response to increased congestion, using heckprob model specification – Using instrumental variable approach for congestion.**

| | Decision made by ED physicians | | | Decision made in the CDU | | |
|---|---|---|---|---|---|---|
| | (1e) TotErr | (2e) AdmErr | (3e) DischErr | (1c) TotErr | (2c) AdmErr | (3c) DischErr |
| ED congestion | 0.024* | 0.039*** | −0.040* | 0.018 | 0.028 | −0.031 |
| | (0.009) | (0.010) | (0.018) | (0.025) | (0.027) | (0.046) |
| $\rho$ | −0.042 | 0.026 | −0.168 | 0.202* | 0.210* | −0.049 |
| | (0.063) | (0.073) | (0.102) | (0.078) | (0.080) | (0.143) |
| N | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 | 377,331 |
| N uncensored | 339,975 | 339,975 | 339,975 | 37,356 | 37,356 | 37,356 |
| Log-lik | −142,451 | −135,170 | −102,807 | −98,651 | −97,581 | −93,165 |

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

$z\widehat{EDCong}_i$.[7] Re-estimating the heckprob models from the paper using this approach gives us the results presented in Table EC.10.

As can be seen, in Table EC.10 we estimate a coefficient (coef.) of 0.039, $p$-value$< 0.001$ (versus coef. of 0.027 originally) in the heckprob model for unnecessary hospitalization when the ED physicians make the disposition decision, and coef. of $-0.040$, $p$-value$= 0.026$ (versus coef. of $-0.016$ originally) in the equivalent model for wrongful discharges. Thus, if anything, based on unobservables as the ED becomes busy, these findings suggest that patients become less, rather than more, likely to be admission errors and more, rather than less, likely to be discharge errors. This is not too surprising, because if we assume all patients with 'serious' conditions attend the ED anyway, then a surge in ED admissions is likely a result of an increase in the 'worried well', who all else being equal we would expect to be less likely to be admitted in error. The worried well may also be more likely to re-attend the ED within a short period of time if discharged, and ED physicians may then be forced to admit them as a precautionary measure, explaining why there may also be an increase in wrongful discharges based on unobservables.

Overall, the congestion effects on wrongful discharges and unnecessarily hospitalizations are robust against endogeneity concerns, as we have shown by demonstrating:

1. That if patients differ based on unobservables to the same extent as they differ based on observables, then this would explain only a small proportion of the congestion effect that we report in the paper;
2. That if we drop from our data set observations when workload levels are low – which is where patients seem to differ in their error propensity based on observables and hence possibly also differ based on unobservables – then we find that our results hold up and, if anything, are conservative;
3. That if we account for endogeneity using a two-stage approach where we use an instrument for congestion (where that instrument is the congestion level one week prior), then again the results reported in the paper appear to be an underestimate of the true effects.

[7] The estimated coefficient for $\omega_3$ is highly significant in Equation (EC.2) with $p$-value$< 0.0001$, suggesting that $zEDCongLW_i$ is a strong IV.

## Appendix EC.6: Additional benefit of the CDU beyond the time component

As mentioned in the paper, in the NHS context of this study, one additional advantage of the CDU is that the patient is considered “off the clock.” That is to say, the patient no longer contributes to breaches of the 4-hour target even if they stay for longer than four hours. This is important, since if the 4-hour target were not imposed then an alternative to having a CDU could be to allow patients to instead spend a longer period of time in the ED. In this case, the additional testing and assessment provided in the CDU might instead be performed in the ED. A central question that we will resolve is whether there are additional benefits of the CDU in regulating admission and discharge error rates above and beyond those that come from keeping these patients under observation for a longer period

In essence, the question is: Is the “secret sauce” of the CDU simply that the patients are allowed to remain longer. In effect, in the UK system with a 4hr time limit for a normal ED stay, is the CDU simply a mechanism that buys more time. Perhaps exactly the same quality results could be achieved by simply letting the patient stay in the ED longer? While the build up to Hypotheses 3 and 4 lays out the theoretical arguments as to why there are additional advantages to the two-stage system above and beyond the additional time it provides for assessment, in this section we will provide evidence from the data to show that this is the case.

In discussing this point below we will make the argument using unnecessary hospitalization, but the same arguments can be extended to wrongful discharges. Suppose that the longer a patient stays in the ED and (if applicable) CDU, the less likely they are to be an admitted unnecessarily (because, e.g., more time is spent on diagnosis, more uncertainty is resolved, etc.). If the entire benefit of the CDU were related to time, then after we control for the time component the effect of CDU admission on a patient’s unnecessarily 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 being admitted unnecessarily. The purpose of the model as specified in the paper is to identify that step-change.

Shortly we will come to discussing how we estimate the step-change in the full sample, but let us start with a reduced sample which can be used to demonstrate that there does exist a benefit of the CDU above and beyond the time component. Specifically, we subset the data to only those patients who spent less than 4 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 4 hours of arrival to the ED. These we take as our “treated” sample. We can use this data set to find a matched group of patients who spent the same amount of time in *just* the ED as did these other patients in *both* the ED and CDU, as well as matching based on other observable factors. These are our “control” group. Figure EC.4 shows the density of propensity scores before and after the matching.

Since the patients in Figure EC.4 are matched on time spent in the system (as well as other observables), there should in theory be no difference between the two groups in their rates of unnecessary hospitalization *unless* (i) admission to the CDU leads to a step-change in this propensity, or (ii) there are unobservables that make the rates across these two groups different. If (ii) were true then we would expect that patients admitted

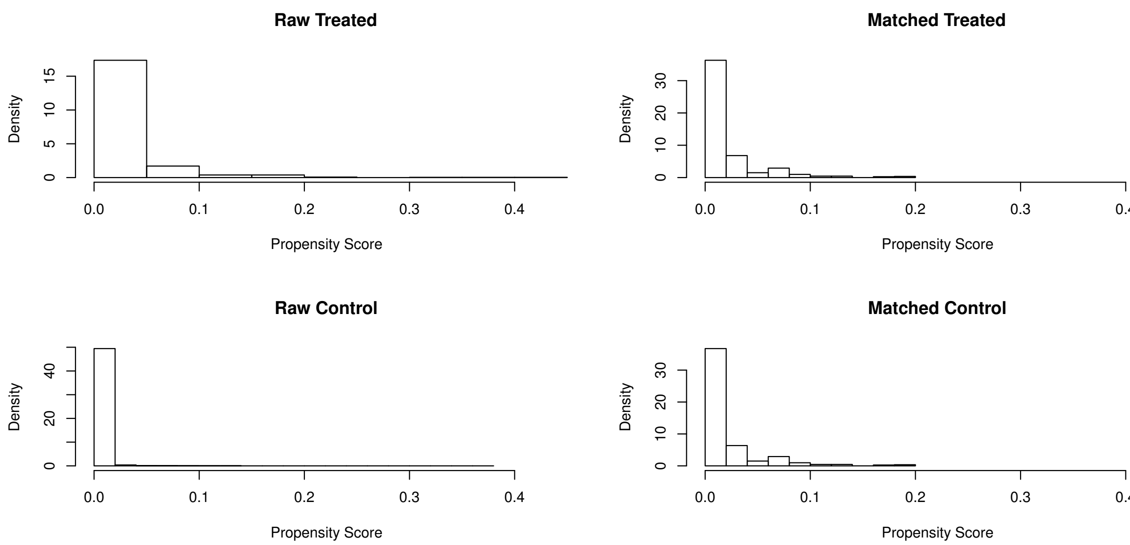



**Figure EC.4** **Density of propensity scores before (left column) and after (right column) matching for the treated (top row) and control (bottom row) groups.**

to the CDU should 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 of being an admission error, versus 3.30% for those patients not admitted to the CDU. (A 2-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 rate of unnecessary hospitalization 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.

Now, returning to the step-change effect in the full sample for which we report results in the paper, note that in Section 7.1 of the paper we discuss two important controls: (i) the amount of time that the patient spends in the ED, and (ii) the amount of time that the patient spends in the CDU. The longer a patient spends in the ED, the more time they have for observation. Similarly for the time they spend in the CDU. These controls therefore account for the fact that a longer period spent in the ED and (if appropriate) the CDU give more time for observation, testing, etc., which may reduce their likelihood of being admitted in error. 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.7: Alternative model specification

In portraying the two-stage gatekeeping process, we model the decision as follows: First, the ED physician makes a decision as to whether they have sufficient information to make the disposition (admit or discharge) decision. If not, then the patient is sent to the CDU. For patients that do not get sent to the CDU, the ED physician subsequently makes a separate admit/discharge decision. We do this because in the paper we focus on the question of how ED physicians trade-off admission and discharge errors as the hospital becomes more congested. To do this we retain the full sample of patients for whom ED physicians are making this decisions, and see how the relative rates of each error change with congestion conditioning on the fact that the physician in the ultimate decision maker.

An alternative representation of the ED physician's decision making process is to suppose that they are making a single disposition decision from among three options: (i) admit to the hospital, (ii) admit to the CDU, (iii) discharge from the ED. To see why this might make sense, suppose that we can represent a patient's condition by a latent variable $P$, ranging, say, from 1-100 (with 100 being most severe). If the condition is severe enough (i.e., $P$ is large enough) then the patient should be admitted, otherwise they should be discharged. The job of the ED physician is to determine the $P$ of a patient, and use this to make the admit/discharge decision. However, this is done imprecisely, or with noise. If the patient is sick enough or healthy enough, the physician has an easy task. It is the patients "in the middle" and/or is a patient for whom $P$ is most difficult to assess who are the most challenging to decide on the correct disposition. The CDU is designed precisely to deal with patients "in the middle."

In this conceptualization, the physician effectively has two cut-points in mind. For example, if $P$ is above $P_{high}$, admit the patient. If $P$ is below $P_{low}$ , discharge the patient. If the patient is in the middle, send the patient to the CDU. An interesting question is how these values, $P_{high}$ and $P_{low}$, vary with congestion. We can answer this question by estimating a probit model of the form:

$$ADM_i^* = \delta_0 + \mathbf{X}_i\boldsymbol{\delta}_1 + \mathbf{Z}_i\boldsymbol{\delta}_2 + zEDCong_i\delta_3 + \epsilon_i^\delta\,, \tag{EC.3}$$

$$ADM_i = \mathbb{1}[ADM_i^* > 0]\,, \tag{EC.4}$$

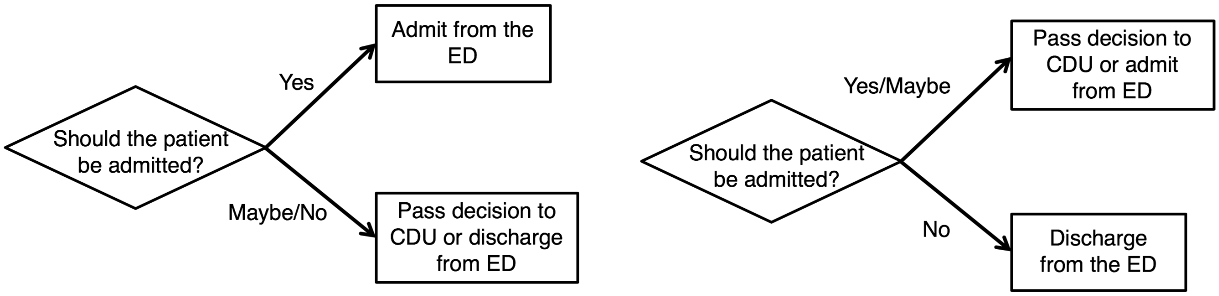
where $\epsilon_i^\delta \sim \mathcal{N}(0,1)$, $ADM_i^*$ is a latent variable, the vector $\mathbf{X}_i$ contains the set of all controls (reported in Table 2), the vector $\mathbf{Z}_i$ contains the set of instrumental variables to be described later, $ADM_i$ is the observed dichotomous variable that indicates whether the patient was admitted directly from the ED, and $\mathbb{1}[\cdot]$ is the indicator function. A similar model can be specified for $DIS_i$, the observed dichotomous variable that indicates whether the patient was discharged directly from the ED. These two models are shown in Figure EC.5

Equation (EC.3) amounts to estimating whether the ED physician either (a) admits the patient to the hospital, or else (b) either sends the patient to the CDU or discharges them home from the ED. In other words, this estimates the cut-point for admission, $P_{high}$. Replacing $ADM_i$ with $DIS_i$ in Equation (EC.3) allows us to estimate instead the cut-point for discharge. The coefficient of congestion, $\delta_3$, in these models would then allow us to determine how these cut-points change as the ED becomes more or less crowded.

It turns out that we can combine the model specified above with the equations for unnecessary hospitalization or wrongful discharge and estimate them simultaneously using a Heckman probit sample selection (heckprob) model. Specifically, let the second stage (outcome) equation takes the form

$$AdmErr_i^* = \beta_0 + \mathbf{X}_i\boldsymbol{\beta}_1 + zEDCong_i\beta_3 + \epsilon_i^\beta\,, \tag{EC.5}$$

**Figure EC.5 Alternative models where either the ED physician admits the patient or does not (left) or discharges the patient or does not (right).**




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

where $\epsilon_i^\beta \sim \mathcal{N}(0,1)$, and where $AdmErr_i^*$ and $AdmErr_i$ are the latent and observed variables for unnecessary hospitalizations, respectively. The latent variable equation for wrongful discharges ($DischErr_i$) is the same as for unnecessary hospitalizations, with coefficient vector $\boldsymbol{\beta}$ replaced with $\boldsymbol{\alpha}$. Similar to the paper, rather than estimate the first and second stage models described above individually, we can estimate them simultaneously with a heckprob model using full information maximum likelihood. The heckprob model allows us to correct for potential sample selection bias arising from the fact that patients may not be admitted or discharged at random. In order to estimate the heckprob model we must: (1) censor the outcome variable $AdmErr_i$ (resp., $DischErr_i$) whenever $ADM_i = 0$ (resp., $DIS_i = 0$), and (2) then estimate the selection and outcome equations simultaneously under the assumption that their 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.

In the heckprob models described above we would be able to see how congestion affects the likelihood of a patient being either admitted or discharged directly from the ED (i.e., estimate the cut-points), while in the outcome equation we could identify how congestion affects the likelihood of the patient being an unnecessary admission (or wrongful discharge) *conditional* on them being either admitted (or discharged) by an ED physician. To improve estimation we need instrumental variables. Similar to the paper, the first instrumental variable we use is the historic propensity of the assigned physician to admit patients directly (or discharge directly when estimating the model for discharge), and the second is the congestion level of the CDU. When a patient is assigned to a physician who historically admits or discharges a high proportion of patients, the patient at hand is also more likely to be admitted or discharged. We would also expect that when the CDU is busy ED physicians are more likely to have to make the admission or discharge by themselves. However, after controlling for the historic rate at which the ED physician makes admission or discharge errors, we should not expect these instrumental variables to be significant in the outcome equation. The results are reported in Table EC.11.

Table EC.11 shows a number of interesting results. First, when the ED becomes more congested, physicians in the ED approximately maintain (slightly decreasing) the rate at which they admit patients to the hospital (each standard deviation increase reduces the likelihood of admission by just 0.23%). Thus, the rate at which

**Table EC.11 Coefficient estimates to establish ED physicians' response to increased congestion, using heckprob model specification – Using alternative modeling approach.**

| | Admission Decisions | | Discharge Decisions | |
|---|---|---|---|---|
| | (1a) Admit | (2a) AdmErr | (1d) Discharge | (2d) DischErr |
| ED congestion | $-0.013^{***}$ | $0.025^{***}$ | $-0.030^{***}$ | $-0.011$ |
| | (0.003) | (0.005) | (0.003) | (0.009) |
| $\rho$ | 0.063 | | $-0.099^{*}$ | |
| | (0.054) | | (0.042) | |
| N | 377,331 | – | 377,331 | – |
| N uncensored | – | 106,318 | – | 233,657 |
| Log-lik | −162,105 | | −144,176 | |

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

hospital capacity is being utilized by patients being admitted from the ED is significantly increased when congestion in the ED increases (if there are 10 patients in the ED and the admission rate is 20% then 2 patients are admitted, while if there are 50 patients and the admission rate is still 20% then 10 patients will be admitted): downstream capacity is stretched. Interestingly, conditional on being admitted, a patient is much more likely to be an unnecessary hospitalization when the ED is busy. This indicates that ED physicians are becoming less discerning in who they admit when busy: the types of patients that they admit to the hospital are those who are more likely to be an unnecessary hospitalization. In summary, while we should expect ED physicians to be reserving hospital beds for only those most needy patients when the ED is busy, instead significantly more patients are admitted who appear to be ex-post unnecessary.

At the same time, patients become significantly less likely to be discharged when the ED is more crowded: a one standard deviation increase in congestion reduces the likelihood of discharge by 0.60%. Conditional on being discharged, however, there patients are no more or less likely to be wrongful discharges (i.e., there is no evidence that physicians are becoming more or less accurate in deciding who to discharge). This indicates that the main reason why we see a decrease in wrongful discharges when the ED is more congested is that physicians become less likely to discharge those more borderline cases. Since the rate of both discharges and admissions goes down when the ED is congested, this must mean that the CDU is being increasingly utilized. This is exactly what we find in the paper.

The above findings are consistent with those that we report in the paper: an increase in congestion results in more unnecessary hopsitalizations and fewer wrongful discharges. They do, though, add an interesting additional dimension to the story: the increase in admission errors is driven predominantly by ED physicians becoming less discerning in who they admit, while the decrease in wrongful discharges is driven mainly by the fact that fewer patients are discharged overall and these patients are instead sent to the CDU.

## Appendix EC.8: Full model output

In the table that spans the next few pages we report full model results from the probit regressions reported in Table 4 of the paper. We also, in Column 1, show results where ED congestion is regressed 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 of those factors reported in the below Table (with the exception of Year, which appears insignificant in all models), we argue that it is important to include all of these factors in the models.

| | (1)<br>ED congestion | (2)<br>CDU referral | (3)<br>TotErr | (4)<br>AdmErr | (5)<br>DischErr |
|---|---|---|---|---|---|
| **Year** | | | | | |
| 2008 | 0 | 0 | 0 | 0 | 0 |
| | (.) | (.) | (.) | (.) | (.) |
| 2009 | -0.0351 | -0.179 | 0.00713 | -0.0343 | 0.113 |
| | (-0.61) | (-1.30) | (0.04) | (-0.19) | (0.35) |
| 2010 | -0.0793 | -0.416 | 0.0557 | -0.0571 | 0.324 |
| | (-0.69) | (-1.52) | (0.17) | (-0.16) | (0.50) |
| 2011 | -0.176 | -0.670 | 0.0636 | -0.144 | 0.513 |
| | (-1.02) | (-1.63) | (0.13) | (-0.27) | (0.53) |
| 2012 | -0.236 | -0.925 | 0.0525 | -0.253 | 0.718 |
| | (-1.02) | (-1.69) | (0.08) | (-0.36) | (0.56) |
| 2013 | -0.254 | -1.138 | 0.0998 | -0.305 | 0.939 |
| | (-0.88) | (-1.66) | (0.12) | (-0.34) | (0.59) |
| **Month** | | | | | |
| Jan | 0 | 0 | 0 | 0 | 0 |
| | (.) | (.) | (.) | (.) | (.) |
| Feb | $0.342^{***}$ | -0.0221 | 0.00635 | -0.00740 | 0.0562 |
| | (40.44) | (-1.09) | (0.25) | (-0.28) | (1.15) |
| Mar | $0.490^{***}$ | $-0.0611^{*}$ | 0.00708 | -0.00416 | 0.0495 |
| | (44.52) | (-2.33) | (0.22) | (-0.12) | (0.79) |
| April | $0.350^{***}$ | -0.0702 | -0.0109 | -0.0258 | 0.0310 |
| | (21.52) | (-1.82) | (-0.23) | (-0.51) | (0.34) |
| May | $0.316^{***}$ | $-0.111^{*}$ | 0.0448 | 0.00242 | 0.175 |
| | (15.86) | (-2.36) | (0.78) | (0.04) | (1.58) |
| June | $0.370^{***}$ | $-0.155^{**}$ | 0.0137 | -0.0348 | 0.161 |
| | (15.23) | (-2.67) | (0.20) | (-0.46) | (1.18) |
| July | $0.368^{***}$ | $-0.164^{*}$ | 0.00782 | -0.0374 | 0.142 |
| | (12.84) | (-2.41) | (0.09) | (-0.42) | (0.89) |
| Aug | $0.209^{***}$ | $-0.183^{*}$ | 0.0279 | -0.00397 | 0.113 |
| | (6.27) | (-2.33) | (0.29) | (-0.04) | (0.61) |
| Sep | $0.232^{***}$ | $-0.185^{*}$ | 0.0116 | -0.0382 | 0.152 |
| | (6.06) | (-2.03) | (0.10) | (-0.32) | (0.71) |
| Oct | $0.381^{***}$ | $-0.206^{*}$ | 0.000764 | -0.0746 | 0.220 |
| | (8.88) | (-2.03) | (0.01) | (-0.57) | (0.92) |
| Nov | $0.342^{***}$ | $-0.239^{*}$ | -0.000210 | -0.0875 | 0.243 |
| | (7.18) | (-2.11) | (-0.00) | (-0.60) | (0.91) |
| Dec | $0.315^{***}$ | 0.00724 | -0.0171 | -0.0278 | 0.0552 |
| | (31.53) | (0.31) | (-0.59) | (-0.90) | (0.98) |
| **School Break** | | | | | |
| 0.None | 0 | 0 | 0 | 0 | 0 |
| | (.) | (.) | (.) | (.) | (.) |
| Autumn half term | $-0.0490^{***}$ | $-0.0642^{**}$ | 0.0237 | 0.0342 | -0.0237 |
| | (-5.69) | (-3.08) | (0.96) | (1.29) | (-0.50) |
| Easter | $-0.0278^{**}$ | 0.0251 | $0.0514^{*}$ | 0.0273 | $0.128^{*}$ |
| | (-3.12) | (1.20) | (2.03) | (1.01) | (2.57) |
| Spring half term | $-0.115^{***}$ | -0.0215 | 0.0401 | 0.0452 | 0.0151 |
| | (-11.88) | (-0.95) | (1.46) | (1.54) | (0.28) |

| | | | | | |
|---|---|---|---|---|---|
| Summer | -0.0535*** | -0.0280 | 0.00537 | -0.00470 | 0.0325 |
| | (-6.90) | (-1.51) | (0.24) | (-0.20) | (0.78) |
| Summer half term | -0.128*** | 0.00864 | -0.0261 | -0.0118 | -0.0812 |
| | (-12.91) | (0.38) | (-0.93) | (-0.39) | (-1.52) |
| **Day of Week** | | | | | |
| 0-Sun | 0 | 0 | 0 | 0 | 0 |
| | (.) | (.) | (.) | (.) | (.) |
| 1-Mon | 0.197*** | 0.00832 | 0.0697 | 0.0476 | 0.160* |
| | (11.58) | (0.27) | (1.94) | (1.26) | (2.00) |
| 2-Tue | -0.196*** | 0.0233 | 0.0842* | 0.0732 | 0.133 |
| | (-11.52) | (0.76) | (2.33) | (1.93) | (1.65) |
| 3-Wed | -0.314*** | 0.0131 | 0.0755* | 0.0609 | 0.149 |
| | (-18.47) | (0.43) | (2.10) | (1.61) | (1.86) |
| 4-Thu | -0.237*** | 0.0134 | 0.107** | 0.0982** | 0.138 |
| | (-13.90) | (0.44) | (2.97) | (2.60) | (1.71) |
| 5-Fri | -0.135*** | 0.0389 | 0.108** | 0.0913* | 0.172* |
| | (-7.95) | (1.28) | (2.99) | (2.41) | (2.15) |
| 6-Sat | 0.0380*** | 0.0274* | 0.0302* | 0.0365* | 0.00997 |
| | (7.47) | (2.31) | (2.08) | (2.29) | (0.39) |
| **Hour of Arrival** | | | | | |
| Weekday 02-04 | 0 | 0 | 0 | 0 | 0 |
| | (.) | (.) | (.) | (.) | (.) |
| Weekday 04-06 | 0.485*** | 0.111*** | -0.0724* | -0.0983** | 0.0492 |
| | (25.76) | (3.78) | (-2.03) | (-2.61) | (0.72) |
| Weekday 06-08 | -0.296*** | 0.167*** | -0.272*** | -0.328*** | -0.00938 |
| | (-19.29) | (6.01) | (-7.38) | (-8.18) | (-0.14) |
| Weekday 08-10 | -0.0277* | 0.0872*** | -0.331*** | -0.373*** | -0.0795 |
| | (-2.33) | (3.66) | (-11.05) | (-11.53) | (-1.41) |
| Weekday 10-12 | 1.502*** | 0.0631** | -0.297*** | -0.332*** | -0.0439 |
| | (131.69) | (2.75) | (-10.74) | (-11.30) | (-0.81) |
| Weekday 12-14 | 1.470*** | 0.0669** | -0.211*** | -0.216*** | -0.0959 |
| | (132.18) | (2.93) | (-7.75) | (-7.51) | (-1.74) |
| Weekday 14-16 | 1.516*** | -0.0607** | -0.164*** | -0.164*** | -0.0809 |
| | (136.22) | (-2.60) | (-5.99) | (-5.67) | (-1.45) |
| Weekday 16-18 | 1.394*** | -0.0672** | -0.0979*** | -0.0918** | -0.0776 |
| | (125.16) | (-2.88) | (-3.61) | (-3.21) | (-1.39) |
| Weekday 18-20 | 1.685*** | -0.128*** | -0.0806** | -0.0716* | -0.0660 |
| | (150.38) | (-5.42) | (-2.96) | (-2.49) | (-1.19) |
| Weekday 20-22 | 1.495*** | -0.147*** | -0.0157 | -0.00644 | -0.0293 |
| | (129.21) | (-6.17) | (-0.57) | (-0.22) | (-0.53) |
| Weekday 22-24 | 0.740*** | -0.0889*** | 0.0158 | 0.0294 | -0.0468 |
| | (61.12) | (-3.67) | (0.57) | (1.00) | (-0.82) |
| 00-02 | 0.997*** | -0.111*** | -0.00144 | 0.00324 | -0.00603 |
| | (67.90) | (-4.22) | (-0.05) | (0.10) | (-0.10) |
| Weekend 02-04 | 0.368*** | 0.0759 | 0.0103 | -0.0303 | 0.208 |
| | (15.17) | (1.73) | (0.19) | (-0.54) | (1.92) |
| Weekend 04-06 | 0.572*** | 0.0585 | -0.0857 | -0.101 | 0.0134 |
| | (19.93) | (1.26) | (-1.50) | (-1.68) | (0.11) |
| Weekend 06-08 | -0.293*** | 0.102* | -0.247*** | -0.368*** | 0.192 |
| | (-11.70) | (2.18) | (-4.04) | (-5.48) | (1.72) |
| Weekend 08-10 | -0.0123 | 0.0343 | -0.313*** | -0.442*** | 0.137 |
| | (-0.57) | (0.83) | (-6.03) | (-7.73) | (1.36) |

| | | | | | |
|---|---|---|---|---|---|
| Weekend 10-12 | 1.332*** | -0.0416 | -0.248*** | -0.332*** | 0.130 |
| | (63.64) | (-1.05) | (-5.18) | (-6.47) | (1.33) |
| Weekend 12-14 | 1.299*** | -0.0916* | -0.194*** | -0.248*** | 0.106 |
| | (63.12) | (-2.32) | (-4.12) | (-4.93) | (1.08) |
| Weekend 14-16 | 1.419*** | -0.163*** | -0.118* | -0.177*** | 0.175 |
| | (68.54) | (-4.07) | (-2.52) | (-3.55) | (1.80) |
| Weekend 16-18 | 1.252*** | -0.111** | -0.0387 | -0.0676 | 0.141 |
| | (60.07) | (-2.78) | (-0.83) | (-1.37) | (1.44) |
| Weekend 18-20 | 1.125*** | -0.171*** | -0.0287 | -0.0362 | 0.0827 |
| | (53.67) | (-4.25) | (-0.61) | (-0.74) | (0.83) |
| Weekend 20-22 | 1.259*** | -0.158*** | -0.0170 | -0.0528 | 0.181 |
| | (58.13) | (-3.88) | (-0.36) | (-1.06) | (1.81) |
| Weekend 22-24 | 0.837*** | -0.125** | 0.0399 | -0.0111 | 0.268** |
| | (38.01) | (-3.04) | (0.83) | (-0.22) | (2.68) |
| **Age Bands** | | | | | |
| 16-20 | 0 | 0 | 0 | 0 | 0 |
| | (.) | (.) | (.) | (.) | (.) |
| 20-25 | -0.0209*** | -0.00235 | 0.0148 | -0.00193 | 0.0563 |
| | (-3.50) | (-0.16) | (0.77) | (-0.09) | (1.64) |
| 25-30 | -0.0144* | 0.0170 | 0.0322 | 0.0130 | 0.0765* |
| | (-2.30) | (1.08) | (1.60) | (0.59) | (2.14) |
| 30-35 | -0.0171** | 0.0956*** | 0.0702*** | 0.0542* | 0.0972** |
| | (-2.63) | (5.99) | (3.43) | (2.42) | (2.66) |
| 35-40 | -0.0172* | 0.127*** | 0.0640** | 0.0362 | 0.123*** |
| | (-2.56) | (7.83) | (3.05) | (1.58) | (3.30) |
| 40-45 | -0.0246*** | 0.162*** | 0.118*** | 0.118*** | 0.0683 |
| | (-3.62) | (10.03) | (5.73) | (5.27) | (1.76) |
| 45-50 | -0.0246*** | 0.141*** | 0.123*** | 0.118*** | 0.0922* |
| | (-3.57) | (8.55) | (5.91) | (5.23) | (2.37) |
| 50-55 | -0.0230** | 0.121*** | 0.150*** | 0.143*** | 0.106** |
| | (-3.25) | (7.19) | (7.20) | (6.35) | (2.68) |
| 55-60 | -0.0181* | 0.0987*** | 0.137*** | 0.127*** | 0.122** |
| | (-2.45) | (5.55) | (6.30) | (5.41) | (2.96) |
| 60-65 | -0.0241*** | -0.00333 | 0.158*** | 0.155*** | 0.0930* |
| | (-3.29) | (-0.18) | (7.48) | (6.82) | (2.22) |
| 65-70 | -0.0193* | -0.0119 | 0.113*** | 0.104*** | 0.0866* |
| | (-2.57) | (-0.65) | (5.22) | (4.50) | (2.00) |
| 70-75 | -0.0268*** | -0.0486** | 0.138*** | 0.119*** | 0.154*** |
| | (-3.42) | (-2.60) | (6.37) | (5.14) | (3.61) |
| 75-80 | -0.0227** | -0.0648*** | 0.0130 | -0.00748 | 0.0816 |
| | (-2.96) | (-3.54) | (0.59) | (-0.32) | (1.85) |
| 80-85 | -0.0329*** | -0.0164 | -0.0714** | -0.0939*** | 0.0416 |
| | (-4.31) | (-0.92) | (-3.20) | (-3.95) | (0.92) |
| 85-90 | -0.0405*** | -0.0192 | -0.135*** | -0.163*** | 0.0468 |
| | (-5.06) | (-1.05) | (-5.75) | (-6.55) | (0.99) |
| 90+ | -0.0327*** | 0.0135 | -0.143*** | -0.188*** | 0.110* |
| | (-3.68) | (0.68) | (-5.63) | (-6.92) | (2.21) |
| **Gender** | | | | | |
| F | 0 | 0 | 0 | 0 | 0 |
| | (.) | (.) | (.) | (.) | (.) |
| M | -0.00238 | -0.0315*** | -0.0649*** | -0.0705*** | -0.0296* |
| | (-0.91) | (-5.17) | (-8.81) | (-8.93) | (-2.08) |

| | | | | | |
|---|---|---|---|---|---|
| **Triage Category** | | | | | |
| 1 | 0 | 0 | 0 | 0 | 0 |
| | (.) | (.) | (.) | (.) | (.) |
| 2 | 0.0393 | 0.204*** | 0.114* | 0.111* | 0.0411 |
| | (1.87) | (4.69) | (2.22) | (2.12) | (0.27) |
| 3 | 0.0113 | 0.475*** | 0.142** | 0.100* | 0.390** |
| | (0.60) | (12.14) | (3.05) | (2.10) | (2.85) |
| 4 | 0.00538 | -0.366*** | -0.589*** | -0.765*** | 0.180 |
| | (0.29) | (-9.20) | (-12.39) | (-15.54) | (1.31) |
| 5 | -0.161*** | -0.588*** | -0.603*** | -1.491*** | 0.403* |
| | (-5.23) | (-5.77) | (-5.21) | (-4.94) | (2.31) |
| 6 | -0.226*** | 0.336*** | 0.0557 | 0.0172 | 0.338* |
| | (-11.89) | (8.44) | (1.18) | (0.35) | (2.44) |
| Major trauma | 0.00343 | 0.165** | -1.019*** | -1.189*** | -0.00191 |
| | (0.12) | (2.70) | (-7.42) | (-7.22) | (-0.01) |
| **Mode of Arrival** | | | | | |
| Ambulance | 0 | 0 | 0 | 0 | 0 |
| | (.) | (.) | (.) | (.) | (.) |
| Foot | -0.107*** | -0.356*** | -0.176*** | -0.174*** | -0.0599 |
| | (-11.46) | (-14.64) | (-5.50) | (-4.93) | (-1.01) |
| Helicopter | -0.0358 | -0.442*** | -1.006*** | -0.988*** | 0 |
| | (-1.08) | (-4.96) | (-3.97) | (-3.87) | (.) |
| Other | -0.0910*** | -0.360*** | -0.0336 | -0.0889 | 0.183* |
| | (-5.23) | (-7.77) | (-0.66) | (-1.58) | (2.13) |
| Police/prison transport | -0.00970 | 0.413*** | -0.276*** | -0.293*** | -0.0713 |
| | (-0.39) | (10.42) | (-3.72) | (-3.53) | (-0.58) |
| Private transport | -0.0709*** | -0.241*** | -0.0790*** | -0.110*** | 0.0930*** |
| | (-20.13) | (-31.48) | (-8.64) | (-11.42) | (4.67) |
| Public transport | -0.0680*** | -0.260*** | -0.210*** | -0.270*** | 0.0316 |
| | (-5.17) | (-6.95) | (-3.87) | (-4.16) | (0.39) |
| Taxi | -0.0800*** | -0.279*** | -0.137*** | -0.184*** | 0.0770 |
| | (-7.88) | (-10.68) | (-4.11) | (-4.94) | (1.39) |
| **Non-Categorical Variables** | | | | | |
| Daily trend | 0.000177 | 0.000660 | -0.0000173 | 0.000247 | -0.000534 |
| | (1.12) | (1.76) | (-0.04) | (0.51) | (-0.61) |
| Visits last year | 0.000516 | 0.0150*** | 0.000914 | -0.000931 | 0.00531* |
| | (0.68) | (11.86) | (0.55) | (-0.51) | (2.09) |
| Avg admits last year | 0.0112 | -0.212*** | 0.0214 | 0.0236 | 0.0475 |
| | (1.92) | (-16.53) | (1.40) | (1.45) | (1.62) |
| Zero visits last year | 0.00242 | -0.140*** | -0.0540*** | -0.0334** | -0.109*** |
| | (0.61) | (-16.06) | (-4.85) | (-2.76) | (-5.51) |
| Visits last month | -0.00382 | 0.00993 | 0.0188 | 0.0109 | 0.0285 |
| | (-0.65) | (0.95) | (1.42) | (0.75) | (1.41) |
| Avg admits last month | 0.00710 | -0.199*** | 0.0309 | 0.0368 | 0.0783 |
| | (0.65) | (-8.72) | (1.10) | (1.21) | (1.63) |
| Zero visits last month | -0.00261 | -0.115*** | 0.0569* | 0.0950** | -0.0756 |
| | (-0.25) | (-5.75) | (2.10) | (3.19) | (-1.76) |
| No doctor history | 0.0154 | 0.166*** | -0.0579* | -0.106*** | 0.0929* |
| | (1.84) | (9.85) | (-2.34) | (-3.85) | (2.29) |
| Hospital congestion | 0.0572*** | 0.00759* | 0.00149 | -0.0000557 | 0.00847 |
| | (38.03) | (2.17) | (0.35) | (-0.01) | (1.04) |
| CDU congestion | 0.181*** | -0.0596*** | | | |

| | | | | | |
|---|---|---|---|---|---|
| | (135.86) | (-18.78) | | | |
| Conditional CDU congestion | | | 0.00117 | 0.000954 | 0.00292 |
| | | | (0.11) | (0.08) | (0.15) |
| Physician CDU usage | 0.00716 | 1.278*** | | | |
| | (0.89) | (70.20) | | | |
| Physician admission error rate | 0.0486 | | | 1.407*** | |
| | (1.32) | | | (43.72) | |
| Physician discharge error rate | -0.0977*** | | | | 2.046*** |
| | (-3.55) | | | | (18.49) |
| Physician total error rate | 0.000526 | | 1.635*** | | |
| | (0.01) | | (42.01) | | |
| CDU admission | 0.0725*** | | 0.0244 | 0.00867 | 0.135*** |
| | (15.97) | | (1.90) | (0.62) | (5.92) |
| ED congestion | | 0.0588*** | 0.0174*** | 0.0245*** | -0.0162* |
| | | (18.38) | (4.33) | (5.63) | (-2.21) |
| Constant | -1.417*** | -1.352*** | -1.614*** | -1.809*** | -2.648*** |
| | (-19.71) | (-8.16) | (-8.07) | (-8.47) | (-6.56) |
| Observations | 377331 | 377331 | 377331 | 377331 | 377331 |

Note: We have also suppressed the output from the factors relating to Initial severity assessment, reason for ED visit, and diagnosis category, due to the large number of levels in each of these factors.
$t$ statistics in parentheses
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

## 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.

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.

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.

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.