Does Hospital Crowding Matter?

Evidence from Trauma and Orthopedics in England*

Thomas P. Hoe†

June 10, 2019

Abstract

This paper estimates and documents the impact of hospital crowding on patient health outcomes and medical treatment decisions. Exploiting pseudo-random variation in emergency admits, I find that a one standard deviation admission shock increases the unplanned readmission rate by 4.1%. Non-parametric and heterogeneity analyses indicates these effects are caused by 'quicker and sicker' discharges due to bed constraints. The crowding impacts are larger at hospital departments with fewer beds, higher severity patients, and stronger incentives to admit additional patients.

*I am extremely grateful to Richard Blundell for advice and suggestions throughout the development of this paper, and to Orazio Attanasio, David Chan, Amitabh Chandra, David Cutler, Alejandro Estefan, Jonathan Gruber, Amanda Kowalski, Neale Mahoney, Robert McMillan, Magne Mogstad, Lars Nesheim, David Oliver, Aureo de Paula, Imran Rasul, John Van Reenen, Pasquale Schiraldi, Marcos Vera-Hernández, and Heidi Williams for helpful discussions. I thank seminar participants at Arizona, Cornell PAM, The Department of Health, EMCON 2016, The Institute for Fiscal Studies, Manchester CHE, Mannheim, McGill, NASMES 2018, NHS Improvement, Oxford HERC, RES 2016 and 2019, Stanford CHP/PCOR, TADC 2017, Toulouse, and University College London. I also thank many healthcare professionals from the English NHS for providing me with insights on the industry. This work was supported by the Economic and Social Research Council grant number ES/J500185/1. The data was made available by NHS Digital under data sharing agreement NIC-369748-C1Y0V.

†Cornell University, Department of Policy Analysis and Management. Email: thomas.hoe@cornell.edu.

1 Introduction

The last 20 years have seen policymakers devote major efforts to improve the quality of health care provision (Institute of Medicine, 2001). Recent efforts have focused on the use of economic incentives, with pay-for-performance schemes are now widespread across both public and private health care systems. These efforts aim to better align provider incentives with those of the patient population and bridge some of the conflicts of interest inherent in health care provision (Arrow, 1963). Setting these incentives requires an understanding of the production function of health care providers, and a growing number of empirical papers taken steps in this direction.[1]

One important aspect of hospital production is commonly referred to as 'hospital crowding', which has been defined as 'a situation in which the defined need for [...] services outstrips available resources [...] and wait times exceed a reasonable period' (American College of Emergency Physicians). Research has shown that these situations, where demand is high relative to supply, are associated with poor health outcomes for patients and long wait times (Hoot and Aronsky, 2008). Concerns have been especially prevalent in public hospital systems such as Canada, the U.K., and the Veteran's Administration in the U.S. While several studies have examined policies which implicitly regulate crowding – such as readmission penalties (Gupta, 2017) and ED wait time targets (Gruber et al., 2018) – there are few studies that directly examine hospital crowding itself.[2]

This paper uses linked administrative data to estimate and carefully document the effects of hospital crowding. Using quasi-random shocks to hospital admissions, I document negative impacts on patient health outcomes, variation in medical treatment decisions, and show that these effects are correlated with provider incentives.

I first set up a simple framework that defines hospital crowding shows how this relates to provider incentives. I define crowding as occurring when production inputs are rationed because capacity constraints are reached. The resulting reduction in inputs per patient raises the prospect that crowding can have adverse effects on patient health outcomes. The provider incentives that govern these outcomes, taking capacity as given, are the payoffs from admitting patients (payments and any non-pecuniary incentives such as wait time reductions) relative to

[1] Examples include: Bartel et al. (2014), Chan (2016), Chan (2018), Chan (2019), Friedrich and Hackmann (2017), Gowrisankaran et al. (2017) and Gruber and Kleiner (2012).

[2] Exceptions to this include Joskow (1980) and Freedman (2016).

the penalties linked to crowding (outcome-based penalties and any direct valuation of patient health outcomes). Situations where admission incentives are high relative to penalties will typically result in more crowding.

I investigate these issues empirically using data from the England National Health Service (NHS), a prominent public health care system with perceived crowding problems.[3] I focus on inpatient care, an area that has been associated with crowding pressures and accounts for the large majority of secondary care spending. Within inpatient care, I look specifically at trauma and orthopedic departments which treat patients with injuries such as broken bones and arthritis. An advantage of this setting is that the causes of emergency admissions are plausibly random (which facilities identification) and the data records relevant health outcomes. Trauma and orthopedic departments are also an important in their own right, being the third largest department measured by admissions (7.2% in 2013).

I identify the crowding effects using variation in emergency admissions. A clear concern with this type of variation is that hospitals may react to admissions, adjusting their production inputs, constraints or patient population. But I show that, conditional on hospital-specific seasonality controls, these admissions are quasi-random. This property mitigates several identification concerns because it implies hospitals cannot respond ahead of time to variations in emergency admissions. It does not guarantee exogeneity, however, since hospitals may adjust their production inputs at short notice in response to the shocks (e.g. by selectively admitting different patients as they become busier). I therefore test for these 'within-day' responses across several margins using the linked administrative data and a rich set of patient characteristics. I show there is no meaningful selection or labor supply response, indicating that emergency admissions are plausibly exogenous. This allows me to estimate the effects of crowding by comparing patient health outcomes and hospital treatment decisions across high- and low-admission days.

I find that crowding increases the rate of unplanned readmissions but has no impact on patient mortality. In response to a one standard deviation shock (around 2 emergency admits, relative to an average day at a 65-bed department), 7-day unplanned readmissions increase by 0.114 percentage points (4.1% relative to baseline). The additional readmissions occur at trauma and orthopedic departments (rather than other departments), suggesting they are symptoms of poor care in the initial episode.

[3]See, for example: https://www.nytimes.com/2018/01/03/world/europe/uk-national-health-service.html

Exploring treatment decisions, I find that the salient impacts of crowding are to cause delays prior to surgery and for patients to be discharged sooner. A one standard deviation shock increases the likelihood that patients wait at least a day before receiving surgery by 0.374 percentage points (1.1% relative to baseline) and reduces the average length of stay by 0.060 days (1.4%).

A comparison of the crowding impacts across elective and emergency patients shows that the majority of the additional readmissions – around 80% – affect emergency patients discharged on the day of the shock. These are the same patients that I also find have shorter hospital stays. I therefore focus in on these patients and attempt to establish whether the causal mechanism is that patients are discharged 'quicker and sicker' (Morrisey et al., 1988).

I present three pieces of further evidence consistent with the quicker and sicker hypothesis. First, I show that the length of stay and readmission impacts are strongly correlated across diagnosis groups: it is primarily low severity patients that are discharged earlier in response to crowding, and same patients are those that experience increases in readmissions. Second, I use a non-parametric specification and compare the functional form of the crowding impacts on length of stay and readmission. These estimated functions closely resemble one another, each showing that emergency admissions have a near-linear impact on the respective outcome. Finally, I look at how crowding affects the prevalence of medical conditions among readmitted patients using the granular diagnosis information for each episode of care. This reveals a series of conditions that could plausibly have gone unnoticed or discounted if hospitals were lowing their discharge threshold in response to crowding (e.g. infections, hemorrahges, pain, urinary disorders). While these findings do not definitively prove causality, they provide strongly suggestive evidence that quicker and sicker discharges is the primary consequence of hospital crowding.

Finally, I turn to the issue of capacity constraints and policy responses. I address this empirically by exploring the heterogeneity in the crowding estimates between hospitals. I again find evidence consistent with the quicker and sicker mechanism, along with three other important findings: the crowding impacts are larger at departments with less bed capacity, higher severity patients, and those facing more pressure to meet elective patient wait time targets. There is no correlation with physician capacity, suggesting that it is bed capacity that is the binding constraint.

I conclude with a discussion of the potential policy responses to crowding. While the empirical results suggest that increasing bed capacity may mitigate the crowding impacts, the

key incentive question is the balance of admission and outcomes-based incentives. In England, hospitals face higher payoffs on the margin from elective admissions because it helps them meet tight 'Referral-to-Treatment' wait time targets, and the results show this is indeed correlated with bigger crowding impacts. So one clear option to reduce hospital crowding is to weaken these targets. But more generally, settings where elective admissions attract high rewards (e.g. strong wait time targets or financial incentives) relative to outcome-based penalties and incentives (e.g. readmission penalties, intrinsic incentives) are expected to experience more severe crowding pressures. Policymakers must trade-off the consequences of these crowding pressures with any impact of lowering hospital utilization. This paper provides provides important and novel evidence to help assess this trade-off.

I contribute to two strands of literature. The first is on hospital crowding itself, which has an established medical literature (for a review, see Hoot and Aronsky (2008)) and has received some attention in economics (Joskow, 1980; Freedman, 2016). The former literature has paid less attention to causality, while the latter has focused mostly on admission decisions and capacity. My contribution is to identify the causal effects of crowding on health outcomes and link this carefully to the quicker and sicker mechanism. By highlighting the relevant incentives and policy issues with hospital crowding, this paper brings together the issues of crowding and hospital access (Lindsay and Feigenbaum, 1984; Propper et al., 2008).

The second literature concerns the use of economic incentives with health care providers. There has been extensive research in both economics (Mullen, Frank and Rosenthal, 2010; Gupta, 2017; Gruber, Hoe and Stoye, 2018) and the medical field (Lindenauer et al., 2007; Jha et al., 2012; Sutton et al., 2012). These papers often use policy-evaluation tools to study the impact of a specific policy. The evidence to date is mixed, with Jha et al. (2012) noting "we still have not identified the right mix of incentives and targets to ensure that pay for performance will drive improvements in patient outcomes". My study emphasizes the importance of in-depth studies of hospital production, such as crowding, for identifying the specific mechanisms and trade-offs that are critical for effective policy design.

The remainder of the paper is organized as follows. Section 2 sets out an economic framework for the crowding impacts and relevant incentives. Section 3 describes the institutional setting and describes inpatient departments and the English NHS. Section 4 describes the HES data. Section 5 presents the empirical specification and discusses identification. Section 6 describes the baseline results. Section 7 goes into more detail about the crowding mechanism. Section 8

discusses capacity constraints, incentives and welfare. Section 9 concludes.

2 Economic framework

I start by setting out a simple economic model that defines the mechanism through which crowding can impact patients and illustrates how incentives can influence the likelihood and extent of these impacts.

Consider a single health care provider with a single production input that must be shared across multiple patients q. Patients are homogeneous. The production input has fixed capacity constraint X and inputs per patient are denoted x. The model is static and the provider must decide how many patients to admit in the single period. The provider receives a fixed payoff p for each admission. Unit costs per input are c. The provider gains a payoff from the health production it generates in patients $h(x)$ which it values in units w. The health production function is assumed to be continuous and increasing concave in x. I use linear functions for p and w for simplicity.

The payoff p could reflect be contracted payments from the government (or an insurer or individual) or other non-pecuniary incentives associated with admitting patients (e.g. meeting wait time targets). The valuation of health production w could reflect pay-for-performance incentives such as readmission penalties (Gupta, 2017) or physicians' intrinsic incentives (Kolstad, 2013). The single production input can be thought of as a composite measure of health care inputs and I discuss how to generalize the model to distinct production inputs later.

The hospital production problem is

$$\begin{aligned} \max_{\{q,x\}} \quad & q\big(p - cx + wh(x)\big) \\ \text{s.t.} \quad & qx \leq X. \end{aligned} \tag{1}$$

Taking q as given and solving for x gives

$$x^*(q) = \begin{cases} \overline{x} & \text{when } qx \leq X \quad \text{(slack input constraint)} \\ X/q & \text{when } qx > X \quad \text{(binding input constraint),} \end{cases} \tag{2}$$

where $\overline{x}$ satisfies the equation $wh'(x) = c$. These are the optimal allocations, which are constant at *overlinex* when there is no constraint on inputs, and are rationed once the constraint binds.

Now consider the hospital problem of setting q once $x^*(q)$ has been substituted in. When the constraint is slack, the first order condition is always positive when $p > cx$, which will typically be the case, and therefore indicates that q should be set such that the constraint at least binds. In the binding case, the first order condition becomes

$$p + wh\left(\frac{X}{q}\right) = wh'\left(\frac{X}{q}\right)\frac{X}{q}. \tag{3}$$

Equation (3) shows that admission decision has a trade-off between payoff gains (first term) plus health gains for the marginal patients (second term) against the health losses for inframarginal patients (third term). At low levels of q, when the constraint is slack, increasing admissions purely generates gains as input allocations do not change. When the constraint binds, crowding impacts begin to bite, reducing inputs per patient and in turn reducing the marginal gain from admissions. Eventually the crowding impacts will be sufficiently large in magnitude to drive the marginal gains from admission to zero, at which point no more admissions will be made.

The crowding effects that I study in this paper are the health losses and input reductions for inframarginal patients which occur when the capacity constraints bind. These responses are represented by the right hand term in Equation (3).

The incentives that determine the extent of crowding are the payoff incentives p, the valuation of health production w, and the marginal (health) product of the production input $h'(x)$. Figure 1 illustrates these incentives graphically. In Figure 1a I show production inputs x as a function of q. When q is low and the constraint does not bind then allocations are constant at $\overline{x}$. Once q is such that the constraint binds then crowding causes the allocations to decline linearly with q. The impact on health outcomes is shown in Figure 1b which plots $-h(x)$ against q. In both figures I mark two levels of payoff incentives: $q(p')$ and $q(p'')$ where $p'' > p'$. A shift from p' to p'' increases crowding, in turn lowering inputs per patient and worsening health outcomes. Similar outcomes occur if p is held fixed and the valuation of health production w, or the marginal health production $h'(x)$, are lower.

The same intuition carries over if the model is extended, either by incorporating multiple inputs, heterogeneous patients, or uncertainty. Multiple inputs can be included, each with its own capacity constraint and the crowding mechanisms will work in a similar fashion along each dimension. With heterogeneous patients, the $h(x)$ function would vary and this would lead hospitals to allocate inputs across these patients to equalize the marginal returns to health

production. Finally, uncertainty can be modeled by replacing q with $q + \varepsilon$, where ε is an unknown admission shock with a known distribution. Hospitals will then set q prior to the realization of ε. This uncertainty can be thought of as the hospital having control over certain types of admissions q (elective patients) but not others ε (emergency patients). In this setup, setting q will determine the likelihood that the constraint binds and there will be a distribution of outcomes depending on the ε realizations. Irrespective of these details, crowding can be defined in the same manner, and the incentives play a similar role.

The empirical contribution of this paper is to evaluate crowding impacts by exploiting exogenous variation in admissions. Using shocks to q, I estimate the derivatives $\partial x/\partial q$ and dh/dq across a range of health outcomes and production inputs.

3 Institutional setting

The empirical setting for this paper is the English NHS and specifically hospital inpatient departments. Below I provide a general description of the English NHS and the relevant incentives of this system, followed by a description of how hospital inpatient departments operate.

3.1 The English NHS

The English NHS is a single-payer health care system funded through general taxation. Hospital care is provided to all English residents free of charge. The large majority of hospitals that provide publicly-funded care are publicly run, and subject to centrally-set government regulations.

The most salient policies relevant to crowding in an inpatient setting are operational targets for elective patient wait times. The time spent waiting by elective patients is not in the hospital, but rather at home prior to the hospital appointment. The 'Referral-To-Treatment' (RTT) target specifies that patients should wait no longer than 3 months between the time of their referral from primary care to their inpatient surgery. Unless specified otherwise, it is this definition of wait time that I refer to throughout the paper. The target was introduced in 2005. It has been shown to have dramatically reduced wait times, and is enforced through a combination of financial penalties and senior management incentives (Propper et al., 2008). Hospitals are fined £300 for every patient that is not treated in under 3 months, and senior hospital managers can be fired for failing to meet the target.

Another set of relevant incentives are financial. Public hospitals have an obligation to break even financially and are also subject to rewards and penalties implicitly linked to crowding. During the period I study, hospitals were subject to a prospective payment reimbursement model and other pay-for-performance schemes including, for example, information disclosure, open hospital competition (Cooper et al., 2011; Gaynor et al., 2013; Bloom et al., 2015), readmission penalties, and best practice payments. The extent to which these financial incentives bind is unclear and, as in other settings (Duggan, 2000), public hospitals often running a financial deficit but do not facing closure. There are no direct volume incentives for individual physicians, with all doctors and nurses being paid fixed salaries.

Parallel to the NHS, there is a private health care system that provides elective inpatient care (but not emergency care). Private hospitals are excluded from this study, but accounted for 14% of hip replacements (a common orthopedic procedure) conducted over the period 2008-2013 were privately funded (Kelly and Stoye, 2015).

3.2 Hospital inpatient departments

Inpatient departments are where the majority of hospital care for serious injuries and illnesses is provided. These departments are organised by medical specialty, which group together related diagnoses and medical procedures. Examples include cardiology (diagnoses relating to the heart), neurology (nervous system), and general surgery. Inpatient departments in England have been associated with

Patients in inpatient departments are classified as either elective or emergency cases. Elective patients are those that require treatment but it is not urgent. A common example is a hip replacement. Elective patients obtain an inpatient appointment after first seeking a referral from a primary care physician and then having an initial assessment at an outpatient consultation with a secondary care physician. If treatment is required, the patient will join a wait list and be given an inpatient appointment at a pre-specified time in the future, which may be several weeks or months later. Emergency patients in contrast often have severe conditions that require immediate treatment. Common examples include broken bones. These patients first attend the emergency department (ED), arriving by their own means or via an ambulance. The ED provides triage and initial treatment and then a decision is made about whether further treatment is required. The majority of ED cases are discharged without further treatment, but those that do require treatment are admitted to an appropriate inpatient department.

Hospitals have various ways to control the flow of patients in and out of inpatient departments. The inflow of elective patients is primarily controlled through appointments. These are set in advance but can be canceled or rescheduled at short notice (even on the day of surgery). There is far less control over the inflow of emergency patients, and there is often no option but to accept urgent and severe cases patients. For less urgent or severe cases, hospitals can potentially divert ambulances to alternative hospitals or adjust the threshold for inpatient admissions from the ED. The outflow of elective and emergency patients is controlled by discharge decisions. Patients are evaluated daily and discharged once they are deemed medically fit and are able to leave the hospital. Upon discharge, patients may be sent home or transferred to another hospital or care facility.

Inpatient departments are staffed by a combination of physicians, nurses and other support staff (e.g. physiotherapists). Hospitals will typically schedule a fixed number of staff for each hour of the day. At certain times and depending on the hospital, there will be one senior physician 'on call' who may be off-site but available remotely, and can come on-site if required.

4 Data

I use administrative data on medical records for inpatient and ED visits from the Hospital Episode Statistics (HES). This data provides a complete picture of secondary care use at public hospitals in England. It allows me to observe each patient's care history and track each episode of care from initiation through to discharge via any transfers. Rich information is available for each episode, including the hospital site, admission and discharge dates, a complete listing of diagnoses (5-digit ICD-10 codes) and procedures (OPCS codes), and a standard set of demographic information. I have inpatient records available for the period 2006 to 2013 and ED records for the period 2010 to 2013.[4]

4.1 Sample selection

The empirical application focuses on trauma and orthopedic patients in adult wards (16+ years of age). These patients present with musculoskeletal conditions such as broken bones and arthritis, and require procedures such as hip or knee replacements or an 'open reduction and internal fixation' (a surgical procedure used to reassemble and fix broken bones into place).

[4]The dates refer to financial years beginning in April and ending in March the following year. This convention is used throughout the paper.

This setting is well suited to the empirical analysis. Inpatient care is an area that has been associated with crowding pressures (British Medical Association, 2017), accounts for the large majority of secondary care spending, and has received comparatively less attention than emergency care. There are also specific advantages to trauma and orthopedic departments: the nature of the emergency injuries (often broken bones) is plausibly random and the data records relevant health outcomes. Trauma and orthopedics are also an important department in their own right, being the third largest department measured by admissions (7.2% in 2013).

4.2 Sample construction

I define the inpatient sample as all trauma and orthopedic patients (elective and emergency) attending general acute hospitals with an active ED. General acute hospitals are full service public hospitals that account for the majority of secondary care, with approximately 91% of trauma and orthopedic patients treated in these hospitals. Of the patients treated at this type of hospital, 86% attend a hospital with an active ED. The sample definition excludes private hospitals, including those conducting some publicly-funded elective care.

To implement the sample definition, I identify trauma and orthopedic patients in the data by the specialty that the physician is categorized as working under. I classify a hospital-year as having an active ED if it admits at least one trauma and orthopaedic patient per week on average across the year. The results are robust to variations in this definition of an active ED. The HES data does not provide information on the bed capacity of the departments included in the sample, but aggregate data from NHS England indicates that the average trauma and orthopedic department in England has approximately 65 beds.[5]

In addition to the inpatient sample, I use data on ED visits. The ED sample is defined as all patients that visit an ED at hospital-years that appear in the inpatient sample. These visits include those that result in an inpatient admission to the trauma and orthopaedic department, as well as all other visits (e.g. those admitted to other inpatient departments, and those not admitted).

Together the sample datasets provides information on 177 trauma and orthopaedic departments, 4.3 million inpatient visits (2006-2013), and 20.7 million ED visits (2010-2013).

[5]This aggregate data is available from 2010. This reports an average trauma and orthopedic bed capacity of 65, compared to 61 by the end of 2013.

4.3 Descriptive statistics

Table 1 presents descriptive statistics for the three datasets I use in the analysis. Panel A shows the hospital-day panel, which is an aggregated version of the inpatient data that records the number of elective and emergency admissions. There are 7.8 elective admissions and 3.8 emergency admissions each day on average across hospitals. In Appendix B I study the time-series properties of the emergency admissions and show that, conditional on hospital-specific seasonality controls, these admissions are serially uncorrelated. This pseudo-random property is useful because it justifies the static empirical framework that I later adopt, and it mitigates a series of endogeneity concerns.

Panel B shows the patient characteristics and several outcomes for inpatient visits. Patients are on average 56 years old, slightly more likely to be female than male, and predominately white. Elective patients wait on average 86 days for an inpatient appointment. The average length of stay is around 4 days. The data also records a number of other measures of patient health (e.g. diagnoses, a co-morbidity index) and treatment decisions (e.g. proportion that face delays, number of procedures).

I use two measures of health outcomes: 7-day unplanned readmission and 30-day in-hospital mortality. These two measures have relative merits. Mortality is an unambiguous outcome with clear and very severe negative consequences. In contrast, the consequences of a readmission are more ambiguous and in many cases far less severe than death. In recent years, the acceptance of readmission as a measure of health care quality has grown, with it being adopted by regulators in pay-for-performance schemes (Gupta, 2017) and it is also commonly used in medical research to evaluate orthopaedic surgery (Kehlet, 2013). I use a readmission definition with a 7-day window to capture the subsequent admissions that are more likely to be associated with the index hospital, and include only unplanned emergency visits to exclude any planned follow-up care. The results are robust to variations in the time window.

The mortality rate is 1.1% and the readmission rate is 2.8%. Examining readmissions in the data illustrates these events are typically associated with undesirable events and common diagnoses including complications with internal devices (e.g. mechanical components of a hip replacement), infections, inflammation, and bleeding. The average length of stay for a readmission is 7.3 days. These facts illustrate that the readmissions events are likely to be associated with negative outcomes, at the very least because of the opportunity cost of time that they involve.

Panel C of Table 1 presents descriptive statistics for ED visits. The average patient that visits the ED is younger and more likely to be male than a trauma and orthopaedic inpatient. Around a third of patients arrive at the ED in an ambulance and two-thirds of patients attend their nearest ED. The average patient spends around 2.5 hours in the ED and around a quarter are admitted to an inpatient department.

5 Empirical specification and identification

I define the degree of crowding as the volume of patients at a hospital given a fixed level of capacity (beds) and other inputs (labor, machinery). Identifying the effects of crowding on outcomes requires exogenous variation in either patient volumes or production inputs. Rather than focus on inputs, which are either difficult to observe at the department-level or have limited (exogenous) variation, I exploit variation in patient volumes that is caused by emergency admissions. These admissions are pseudo-random and, as I argue below, plausibly exogenous which provides ideal variation for identifying the effects of crowding.

5.1 Baseline specification

I use the following baseline specification

$$y_{iht} = \alpha_{d(i)} + \gamma_{h,s(t)} + \beta q_{h,s(t)} + u_{iht} \tag{4}$$

where y_{iht} is an outcome for patient i at hospital h in cohort t, $\alpha_{d(i)}$ is a series of fully interacted individual-level diagnosis, age category, and emergency status fixed effects (giving over 45,000 patient types), $\gamma_{h,s(t)}$ is a series of hospital-specific year, weekly-seasonal and day-of-the-week fixed effects, $q_{h,s(t)}$ is the number of emergency admissions at hospital h on a day defined by $s(t)$, and u_{iht} is an error term. The inclusion of hospital-specific fixed effects in Equation (4) restricts the identifying variation in $q_{h,s(t)}$ to daily-deviations from seasonal averages – which I refer to as 'emergency shocks'.

Since shocks can impact patients at different points during their stay, I estimate Equation (4) for two mutually exclusive patient groups: 'discharge cohorts' are the patients discharged on the day of a shock, and 'admission cohorts' are the patients that arrive the day after a shock.[6] Crowding may impact admission cohorts during the initial part of their hospital stay

[6] I leave a day between the two cohorts so that there is no overlap between the groups from patients that are

(e.g. surgery) and discharge cohorts during the end of their stay (e.g. the discharge process). [7] For admission cohorts I define t as the admission date and $q_{h,s(t)} = q_{h,t-1}$, and for discharge cohorts I define t as the discharge date and $q_{h,s(t)} = q_{h,t}$.

The parameter of interest is β. This reflects the causal impact of emergency admissions on patient outcome y_{iht}. I estimate the baseline specification using an algorithm designed for high-dimensional fixed effects by Correia (2016).[8] Standard errors are clustered at the hospital-level (177 clusters).

5.2 Identification

I make the following identification assumption: $\mathbb{E}[q_{h,s(t)}u_{iht}] = 0$. This states that emergency admissions, after conditioning on patient type and hospital-specific seasonality, are uncorrelated with unobservable factors affecting outcomes.

Under this assumption, the OLS estimates can be interpreted in two ways. First, they can be seen as a linear approximation to the underlying causal relationship, and the estimates should then be interpreted as the causal impact of a shock relative to an average day (i.e. with a zero shock). An alternative interpretation, following Angrist and Krueger (1999), is that OLS estimates of β identify a weighted-average of heterogeneous treatment effects where the treatment effects are measured across outcomes conditional on different discrete shocks. I refer to the parameter estimates obtained using OLS as the 'average crowding effect' (ACE).

Three sources of variation threaten the identification assumption. The first is any supply-side response from hospitals, whereby production inputs (e.g. labor) adjust to changes in emergency admissions. For example, if hospitals seek to maintain service quality throughout busy periods then they will schedule greater resources to be available in these periods. These responses are expected to attenuate the ACE estimates towards zero because increases in admissions would be met with increases in resources.

The second threat is from hospital decisions that affect the selection of patients into the hospital. In a similar way to resourcing decisions, hospitals might seek to schedule or accept less severe (resource-intensive) patients when they expect to be busy with emergency patients.

admitted and discharged within the same day.

[7] In principle, one could think about analyzing the impact of crowding during a patient's hospital stay. The data does not allow for this possibility as it does not record daily treatment activity (e.g. nursing care) or interim health outcomes (e.g. vital signs). There is also a selection problem for assessing mid-tenure shocks: since shocks cause patients to be discharged earlier (a result shown later) then the selection of patients that experience shocks prior to discharge is endogenous.

[8] OLS produces near-identical estimates.

An alternative response might be to accept only the most severe patients to minimize any health impact on patients that are delayed. The selection response could therefore bias the ACE estimates in either direction.

A third threat is present if patient composition differs on high and low admission days, simply because of the types of injuries sustained by the population. This could happen irrespective of hospital behavior and be driven by changes in the weather or patient behaviors. I refer to this threat as 'natural endogeneity'.

5.3 Testing for within-day hospital responses

To test the identification assumption, I use the baseline specification to evaluate whether hospitals respond to emergency admissions by adjusting their production inputs or the selection of admitted patients. Since identification comes from emergency shocks, rather than any expected variation in admissions, these tests focus on within-day hospital responses. The prospect of hospital responses occurring in anticipation of a shocks is mitigated by the fact that the shocks are pseudo-random (see Appendix B).

Table 2 presents ACE estimates for hospital responses. Panel A shows the supply-side responses. This analysis is limited by the data availability, which only records the number of senior physicians present in the hospital. The estimate indicates that there is around 1 additional senior physician present for every 6 unexpected emergency admissions. This effect is small in the sense that a shock of this magnitude corresponds to a 3 standard deviation event and, in comparison, if the same analysis is repeated but for *expected* admissions – so that it measures the scheduling responses that the baseline specification partials out – the physician response is approximately twice as strong. This result indicates that there is a supply-side response, at least in terms of senior physicians, but it is limited in magnitude and, if anything, would bias my crowding estimates towards zero.[9]

Panel B presents the ACE estimates for admitted elective patients. In the first row, I show the impact of emergency admissions on the volume of elective admissions, which hospitals might adjust at short notice by canceling appointments.[10] Emergency shocks have a negative but small and statistically insignificant impact on elective admissions. I then look at whether

[9]This finding is consistent with my discussions with hospital managers, which revealed that hospitals typically have one senior physician on call but there are few other short-term resources that can be called upon. I do not have any data available on the presence of nursing or other support staff.

[10]The data does not record cancellations but any such responses should be reflected in the volumes of admitted patients.

this small impact feeds through to any selection on the characteristics of the admitted elective patients. I use the baseline specification with the mean characteristics of admitted patients as the dependent variable and remove the patient-level fixed effects. I find very limited evidence of selection across this rich set of observables. The only characteristic with statistical significance at the 1% level is the number of days an elective patient has been waiting. This effect is very small: a one standard deviation shock increases the average wait time of admitted patients by less than one day (0.1% of baseline). In the final two rows, I show that linear predictors of mortality or readmission (based on the preceding variables in Panel B plus diagnosis information) reveal very little evidence of selection. There is no impact on predicted mortality and only a very small *negative* impact on predicted readmission (0.05% of baseline). The latter would only have the effect of downward biasing my later crowding estimates. These tests indicate that there is little to no selection on patient characteristics. I reinforce this finding later, particularly with regard to wait times, by testing the robustness of my results to the inclusion of wait times and other variables as controls.

Panel C presents the ACE estimates for admitted emergency patients. Unlike for elective admissions, where I did not observe hospital responses such as cancellations directly, I do observe the relevant responses for emergency patients and so evaluate these directly. I find that emergency shocks have no impact on the likelihood of inpatient admission, and no convincing evidence of effects on hospital choice (which includes ambulance diversion) nor ED wait times. The only marginally statistically significant result is the estimate on the likelihood of attending the nearest ED, although only at the 5% level, which is not especially demanding given the sample size of the ED dataset, and the sign of the estimate is counter-intuitive suggesting that, if anything, more patients arrive at busy inpatient departments. These results show that the selection of emergency patients reaching the trauma and orthopedic departments it not influenced by within-day hospital responses.[11]

While the previous results show that hospitals do not select on emergency admissions, it is still possible that the characteristics of emergency patients arriving on high-shock days differ from those on low-shock days. This corresponds to the natural endogeneity threat. It is important to note that the baseline specification allows for this possibility through the patient-level fixed effects, such that there may be more broken legs (or other injuries) on high- relative to

[11]These results use the full sample of ED visits and thus average across many patients that are not likely to end up a trauma and orthopedic department. Restricting the analysis to patients that are predicted to be trauma and orthopedic patients produces similar results.

low-shock days. But the concern is that, within patient type, there may still be unobservable differences that vary with the emergency shock. For example, broken legs on high-shock days may be more severe than those on low-shock days. Major correlated events – such as terrorist attacks, major road traffic accidents, or epidemics – are potential causes of within-day natural endogeneity.

To illustrate that correlated events are not driving the variation in emergency admissions, I compare the observed data to a simulated Poisson process. If the data is approximately Poisson it suggests the variation is from a series of independent events, rather than from a major event that could generate atypical admissions.[12] Figure A1 in Online Appendix A presents the results of this exercise, showing that the simulated Poisson and the observed data are near-identical.[13] The implication is that natural endogeneity is not a concern.

To summarize, these results support the assertion that emergency shocks are plausibly exogenous. There is very little evidence that hospitals make within-day responses that influence patient selection and, conditional on patient-level fixed effects, the patients arriving on high and low shock days are likely to be similar. In the case of very extreme shocks, there is a small within-day response from senior physicians and this may attenuate the crowding estimates that follow.

6 Baseline results

I first present the ACE estimates for health outcomes, followed by treatment decisions, and then a sub-group analysis that compares elective and emergency patients.

6.1 Health outcomes

Table 3 presents the ACE estimates for health outcomes. Panel A presents the results for admission cohorts and Panel B for discharge cohorts. Both cohorts include all patient types.

I find no impact of crowding on mortality for admission or discharge cohorts, indicating that crowding pressures do not kill patients. This is perhaps not surprising, as one would expect hospitals (or regulators) to prevent such stark outcomes. In contrast, I find positive and statistically significant impacts on unplanned readmission, indicating that more patients return

[12] This follows from the fact that the Poisson distribution can be derived from a large number of independent and identically distributed Bernoulli trials.

[13] Notably, the same analysis for elective admissions, where hospitals can control the level of admissions, shows that the observed and simulated data strongly diverge.

to hospital and are admitted as an emergency inpatient in the week after a positive emergency shock. The impacts for discharge cohorts are approximately five times larger than those for admission cohorts: a one standard deviation increase in emergency admissions (2 patients) is estimated to increase the readmission rate by 0.018 percentage points (0.6% relative to the baseline) for admission cohorts and 0.096 percentage points (3.4%) for discharge cohorts. The sum of the impacts is 0.114 percentage points (4.1%).

To check that these readmissions are indeed related to the original episode of care, I estimate the ACE using indicator variables for the department where the patient was readmitted. Table A3 presents these estimates and shows that almost all of the readmissions are attributable to trauma and orthopedic departments. It is thus clear that these readmission effects are related to the medical care given when and where the crowding occurs.

In summary, crowding does not impact patient mortality but it does lead to increases in unplanned readmissions. Around 84% of the additional readmissions affect patients discharged on the day of the crowding shock.

6.2 Treatment decisions

Table 4 presents the ACE estimates for treatment decisions in the inpatient department. For admission cohorts, I evaluate: whether a patient has a minor operation, defined as an operation with a median length of stay of zero nights ('daycase operation'); whether the patient has to wait at least one night for their primary operation ('delayed operation'); and, the total number of medical procedures they receive ('number of procedures'). I find that crowding leads to more daycase operations, more delays and fewer procedures. The most substantive impact is on delays: a one standard deviation shock implies admission cohorts are 0.374 percentage points (1.1% relative to baseline) more likely to wait at least a day before receiving their primary procedure. The impacts on the number and type of procedures are very small in magnitude (-0.3% and 0.2% of baseline).

Turning to the discharge cohorts, I evaluate the length of stay in days, and whether the patients were transferred to another hospital ('transfers out') or a location other than their usual residence ('discharges home'). I find that crowding leads to reductions in length of stay (patients are being discharged sooner) but has no impact on the likelihood of transfer to another hospital or being sent home rather than to another care provider. A one standard deviation shock reduces average length of stay for discharge cohorts by 0.060 days (1.4% of baseline),

which corresponds to around 1 in 15 patients being discharged a day early.

Tables A1 and A2 shows that these results, and the earlier estimates for health outcomes, are robust to the choice of readmission window and set of control variables.

These results suggest some potential causes of the earlier readmission effects. For admission cohorts, delays prior to surgery may increase readmissions, while for discharge cohorts it may be that they being discharged 'quicker and sicker' (Morrisey et al., 1988).

6.3 Elective and emergency patients

To further explore the link between health outcomes and treatment decisions further, I compare the estimates for elective and emergency patients.

Table 5 presents the ACE estimates for health outcomes and mechanisms separately for each patient type. There are some parallels to the aggregate results: there are no mortality impacts for any of subgroups, and the readmission impacts are very close to zero for admission cohorts of either patient type. In contrast to the aggregate results, the readmission ACE estimates for discharge cohort differ by patient type: they are close to zero for elective patients, but statistically significant and large in magnitude for emergency patients. Overall, emergency discharge cohort account for around 80% of the additional readmissions.

Turning to the treatment decisions in Table 5, the ACE estimates for delays and procedures of elective admission cohorts are statistically significant but much smaller in magnitude than the equivalent estimates for emergency patients. There is no length of stay impact for elective patients, whereas it is statistically significant and large in magnitude for emergency patients.

These results largely emphasize the aggregate estimates from before: the most striking impacts of crowding are shorter hospital stays and increased readmissions of discharge cohorts. But is now clear that these impacts are felt primarily by emergency patients.[14] Both emergency and elective admission cohorts also experience delays, but these effects are not associated with substantive readmission impacts.

[14]The reason for the sharp distinction between elective and emergency patients, particularly with respect to length of stay, is unclear. The same distinction is evident across hospitals and regions suggesting that it is unlikely that the cause is related to the differences in demand elasticity between elective and emergency patients. One potential explanation is physicians do not change treatment decisions for elective patients because the impact on their health outcomes would be greater. Another argument, offered by physicians in discussions, is that the elective pathway is more standardized while the emergency pathway is much heterogeneous, and and thus physicians have more discretion with the treatment of emergency patients. Without any variation in elective patient treatment it is not possible to test between these two hypotheses.

7 Quicker and sicker mechanism

The results to now suggest 'quicker and sicker' discharges may be the the primary consequence of crowding. Detecting causal evidence of the link between length of stay and readmission, however, is more challenging because the variation I am exploiting (crowding) has impacts across several margins. To evaluate the plausibility of the quicker and sicker mechanism, rather than conducting a direct casual analysis, I instead exploit the rich properties of the data and test several implications of the mechanism. This section presents three tests that each focus on the sample of emergency discharge cohorts.

7.1 Patient severity

The first approach I take is to disaggregate the emergency patients by ICD-10 diagnosis codes. If the quicker and sicker mechanism holds then I should expect to find a correlation between the length of stay and the readmission impacts at this more granular level. I therefore estimate the ACEs along these two margins for the 100 most common diagnosis groups.

Figure 3 shows the presents. In Figure 3a the ACE for length of stay is plotted against predicted mortality as a measure of severity. Figure 3b repeats this exercise for the readmission ACE. In both cases, the ACE is increasing in magnitude with patient severity, showing that hospitals discharge low severity emergency patients in response to crowding and it is these same patients that are more likely to be subsequently readmitted. Figure 3c then compares the length of stay and readmission ACEs directly, confirming that the magnitude of the length of stay and readmission impacts are positively associated.[15] This correlation between the effects is consistent with a quicker and sicker mechanism.

7.2 Non-parametric estimates

Another way to test the quicker and sicker hypothesis is through non-parametric versions of the baseline specification. This can be set up by replacing the emergency admission variable in Equation (4) with a series of indicator variables for each discrete value of emergency admissions. If length of stay is indeed driving the readmission responses, then I should observe a similar functional form in the crowding impacts on each variable.

[15] A weighted-regression of the ACE for length of stay or readmission on predicted mortality produces statistically significant estimates in both cases, as does regressing the ACE for readmission on the ACE for length of stay.

Figure 4 presents non-parametric estimates for length of stay (Figure 4a) and readmission (Figure 4b). Over the emergency admit distribution the length of stay and readmission impacts are broadly linear. High shocks reduce length of stay and increase readmissions, while low shocks have the opposite effects. The magnitude of these estimates illustrates that the daily shock, depending on its magnitude, causes the readmission rate for emergency discharge cohorts to vary between 4.8 and 7.4 (49% of the baseline). Most importantly, the strong parallel between the functional form of each relationship is what would occur if earlier discharges caused more readmissions.

These functional forms can be contrasted with similar non-parametric estimates for the emergency admission cohorts. Figure A2 shows that the functional form of the readmission impacts (small, non-linear) and delays (broadly linear) do not correspond to one another.

7.3 Medical causes of readmission

As a final probe of the quicker and sicker mechanism, I examine the diagnoses of the readmissions caused by crowding. The quicker and sicker mechanism would imply that hospitals are lowering their threshold for discharge when they reach capacity constraints, and this would lead to marginally less-than-healthy patients being discharged. If this is happening, the additional readmissions should be associated with relatively minor health conditions that could have been dealt with, or avoided, if the patient had remained in a hospital setting. I therefore test for ACE responses using indicators for each readmission diagnosis.

Table 6 shows the ACE estimates for indicator variables relating to the top-20 readmission diagnosis codes. The first point to note is that the baseline mean of these diagnosis indicators is very low, on average 0.45%, which limits statistical precision of this exercise. Second, in almost all cases (16 of 20 diagnosis, and all of the top 10) the estimated coefficient is positive, suggesting that crowding increases many of these types of readmission. Finally and most importantly, there are several minor conditions which, when under pressure to discharge patients, could plausibly have gone unnoticed, or the severity or likelihood of occurrence been discounted. Specific examples that fit with the quicker and sicker hypothesis include infections, hemorrhage, pain, and urinary disorders. However, these examples are only indicative since the only statistically significant estimates relate to very broad or vague categories, such as fractures codes which are likely to relate to the original diagnosis rather than the specific readmission event.

The preceding analyses presents three different tests of the hypothesis that it is the earlier

discharge which causes the additional readmissions. While no individual test is decisive, there is a body of suggestive evidence that this mechanism is indeed the primary consequence of hospital crowding.

8 Capacity constraints, incentives and welfare

At this point it is still unclear which capacity or resource constraint may be leading to the quicker and sicker discharges, and how, if at all, policymakers should respond to crowding pressures. This section provides some evidence and discussion of these issues.

8.1 Hospital-level heterogeneity

I explore these issues empirically by using between-hospital variation in the ACE. To do this I estimate Equation (4) for each hospital separately – again focusing on the length of stay and readmission impacts for emergency discharge cohorts – and then regress the resulting ACE estimates on hospital characteristics. I use the following characteristics: the mean number emergency admissions (which will also approximate its standard deviation given the Poisson property); the mean overnight census of trauma and orthopedic patients in the hospital (as a proxy for bed capacity); the mean predicted mortality of elective and emergency patients (as proxies for patient severity); the mean number of senior physicians present each day (labor capacity); the mean performance relative to the RTT wait time target; and, a measure of how responsive the hospital is to predicted emergency admissions. The responsiveness measure is defined as the parameter estimate from regressing the daily number of elective admissions on the expected number of emergency admissions (constructed using the hospital-specific seasonality fixed effects in Equation (4)); intuitively, this gives a measure of whether a hospital is rationally scheduling its elective admissions around seasonal expectations of emergency admissions and, more generally, the organizational skill of the department.

Table 7 shows the results of this exercise where all variables have been normalized to the absolute value of their z-score. Column (1) uses the ACE estimates for length of stay as the dependent variable. The length of stay impacts are negatively associated with bed capacity (at the 5% significance level) and wait time target performance (10%), and positively associated with emergency patient severity (1%). Column (2) shows the results using ACE estimates for readmission as the dependent variable. This regression includes the same covariates as

column (1) with the addition of the ACE for length of stay. The readmission impacts are positively associated with emergency patient severity and the length of stay ACE (both at the 1% significance level).

There are four points to note from these regressions. First, the positive association between the readmission and length of stay ACE estimates in column (2) is again consistent with the quicker and sicker mechanism described earlier. Second, the length of stay regression in column (1) indicates that it is bed capacity, rather than physician capacity, that is the relevant capacity constraint associated with the quicker discharges. Third, the association between the ACE estimates and emergency patient severity likely reflect two causal channels: for the length of stay effects, high severity patients (typically elderly) may be blocking bed capacity as they have longer hospital stays, sometimes being medically fit but difficult to discharge (British Medical Association, 2017); and for the readmission effects, high severity patients may experience larger health impacts when length of stay is shortened because they have a different health production function.

Finally, the positive association between the length of stay impacts and the RTT performance in column (1) suggests that stronger incentives to admit elective patients may be contributing to crowding pressures. That is, those hospitals which are performing worse relative to the RTT target – and therefore have a higher payoff on the margin from admitting higher volumes of elective patients – are those that hit the bed capacity constraint and experience crowding outcomes more often. This fits with predictions from the model in Section 2.[16]

8.2 Policy discussion

These results inform how policymakers seeking to maximize consumer welfare should respond to crowding pressures. One potential response is to increase capacity, which in this case would mean more beds. The preceding results suggest this would reduce crowding pressures. It would also come with significant fixed costs in many cases.

A more pertinent question, however, is how hospitals are incentivized to utilize any given level of capacity. Hospitals can control their utilization through the scheduling of elective admissions. Restricting volumes of elective admissions will necessarily reduce utilization and

[16] An alternative interpretation of the correlation with RTT performance is that these are hospitals that are less well managed in general, and so perform worse relative to the RTT target and perform worse in terms of crowding outcomes. However, the fact that there is no correlation with the responsiveness variable suggests that the correlation with RTT performance is not a reflection of organization and management and rather more focused on the RTT performance itself.

crowding pressures, but in turn makes access and wait times more difficult for these patients. Strong admission incentives (wait time targets in this setting, or more generally payoffs that would include financial payments) will therefore mean capacity utilization decisions favor elective patients by lowering wait times, while weak incentives will favor emergency patients through reduced crowding pressures. With this trade-off, it is important to consider whether the policy incentives maximize consumer welfare. This is an empirically challenging question: it involves not only a trade-off between outcomes with very different welfare implications (wait times and readmissions), but also across very different states of the world (requiring elective vs. emergency care).

To provide an indication of the potential for policy change, it is instructive to consider how patients trade-off these outcomes. A ballpark calculation based one existing evidence suggests preferences that imply consumers would be willing to wait around 3 months to avoid a readmission.[17]

This marginal rate of substitution can be compared to an estimate for the marginal rate of technical substitution between readmissions and wait times. If the MRS exceeds the MRTS then it would indicate that reductions in crowding, and any subsequent increase in elective wait times, would improve consumer welfare. In such a scenario, lowering capacity utilization by admitting fewer elective patients would generate welfare gains (from the reduced likelihood of readmission for emergency patients) that exceed the welfare losses (from the wait time increases for elective patients).

Making this comparison requires estimates of the MRTS, which is defined as the ratio of the marginal impact of elective admits on equilibrium wait times and health outcomes. Assuming that the ACE estimates reflect what would happen when elective (rather than emergency) admissions increase, and comparing this to the MRS, indicates that the marginal impact on equilibrium wait times would have to be extremely small for the MRS to exceed the MRTS. This seems unlikely given standard predictions from queuing theory. The implication is that,

[17]Propper (1990) ran a stated preference survey to obtain estimates of the willingness-to-pay (WTP) to reduce wait times in the English NHS. The survey framed the WTP question in the context of an unspecified diagnosis with a 'near-zero decay rate', a characteristic shared by many conditions requiring orthopaedic treatment.Propper (1990) estimates a WTP of approximately £86 (2017 prices) to reduce wait times by one month (just below £3 per day). To obtain estimates of the value placed on readmission, I assume that the cost of a readmission is limited to its opportunity cost, and estimate this as the average length of stay of a readmission event multiplied by average net wages. This gives a valuation of approximately £692 per readmission. Combining these preference estimates with estimates of the relative likelihood of requiring elective or emergency care gives an estimated ex-ante marginal rate of substitution of -101. Similar estimates of preferences can be obtained from Beckert and Kelly (2017), who estimate a demand model for hip replacement patients in England.

at least in the context of a single hospital, less crowding may even decrease consumer welfare.

In practice, this partial welfare calculation, and any full welfare analysis, requires a richer model to make strong policy predictions. In the English context, this would need to incorporate an analysis of demand for public and private sector hospitals, as well as an assessment of crowding at private hospitals. Both of these tasks are beyond the scope of this paper.

Notwithstanding the welfare issues, the policy implications from this paper are twofold. First and foremost, evidence of crowding pressures in isolation is not indicative of inefficient capacity utilization and in fact the opposite can be true. Second, the relative admission incentives and outcomes-based incentives, arguably more so than capacity levels, matter for crowding outcomes. In the English setting, the RTT wait time target, which implicitly rewards hospitals for running at high levels of utilization, may be particularly influential. More generally, settings where elective admissions attract high rewards (e.g. wait time targets, strong financial incentives) relative to the outcome-based penalties and incentives (e.g. readmission penalties, intrinsic incentives) are expected to experience more severe crowding pressures.

9 Conclusion

There is growing interest in understanding the production function of health care providers, and it is an important step for setting provider incentive policies. A salient feature of hospital production is the persistent problem of hospital crowding, which is a major policy issue and has been described as an international crisis (Hoot and Aronsky, 2008).

This paper documents the causal effects of hospital crowding in one of the most well-known public health care systems, the English NHS. Using quasi-random variation in emergency admissions, I show that the primary impact of hospital crowding is that patients are discharged quicker and sicker. These responses lead to worse health outcomes for patients, specifically through additional unplanned readmissions. The impacts are more pronounced at hospital departments with fewer beds, higher severity patient populations, and stronger incentives to admit additional patients.

These results raise important questions about how welfare would be impacted if policymakers pursue an objective of reducing hospital crowding. One possibility is to reduce levels of capacity utilization by admitting fewer elective patients. A first pass at this question suggests that this type of policy is not necessarily desirable because of the adverse welfare impacts of increasing

wait times for elective patients. But there is a need to evaluate these welfare impacts more carefully, taking into account the richness of consumer preferences and the institutional setting. More generally, further research is needed that makes the link between provider incentives, health outcomes and consumer welfare if we are to properly evaluate the efficacy of health care regulations.

References

Angrist, Joshua D. and Krueger, Alan B. (1999). 'Empirical Strategies in Labor Economics', *Handbook of Labor Economics* 3(23), 1277–1366.

Arrow, Kenneth J. (1963). 'Uncertainty and the Welfare Economics of Medical Care', *American Economic Review* 53(5), 941–973.

Bartel, Ann P., Beaulieu, Nancy D., Phibbs, Ciaran S. and Stone, Patricia W. (2014). 'Human Capital and Productivity in a Team Environment: Evidence from the Healthcare Sector', *American Economic Journal: Applied Economics* 6(2), 231–259.

Beckert, Walter and Kelly, Elaine. (2017). 'Divided by Choice? Private Providers, Patient Choice and Hospital Sorting in the English National Health Service', *IFS Working Paper W17/15* .

Bloom, Nicholas, Propper, Carol, Seiler, Stephan and Reenen, John Van. (2015). 'The Impact of Competition on Management Quality: Evidence from Public Hospitals', *Review of Economic Studies* 82(2), 457–489.

British Medical Association. (2017), 'State of the health system: Beds in the NHS'.

Chan, David. (2016). 'Teamwork and Moral Hazard: Evidence from the Emergency Department', *Journal of Political Economy* 124(3).

Chan, David. (2018). 'The Efficiency of Slacking Off: Evidence from the Emergency Department', *Econometrica* 86(3).

Chan, David. (2019), Influence and Information in Team Decisions: Evidence from Medical Residency. Working Paper.

Cooper, Zack, Gibbons, Stephen, Jones, Simon and McGuire, Alistair. (2011). 'Does Hospital Competition Save Lives? Evidence from the English NHS Patient Choice Reforms', *Economic Journal* 121(554), 228–260.

Correia, Sergio. (2016), Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator. Working Paper.

Duggan, Mark G. (2000). 'Hospital Ownership and Public Medical Spending', *Quarterly Journal of Economics* 115(4), 1343–1373.

Freedman, Seth. (2016). 'Capacity and Utilization in Health Care: The Effect of Empty Beds on Neonatal Intensive Care Admission', *American Economic Journal: Economic Policy* 8(2), 154–185.

Friedrich, Benjamin U. and Hackmann, Martin B. (2017). 'The Returns to Nursing: Evidence from a Parental Leave Program', *NBER Working Paper No. 23174* .

Gaynor, Martin, Moreno-Serra, Rodrigo and Propper, Carol. (2013). 'Death by Market Power: Reform, Competition and Patient Outcomes in the National Health Service', *American Economic Journal: Economic Policy* 5(4), 134–166.

Gowrisankaran, Gautam, Joiner, Keith A. and Léger, Pierre-Thomas. (2017). 'Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments', *NBER Working Paper No. 24155* .

Gruber, Jonathan, Hoe, Thomas P. and Stoye, George. (2018). 'Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers', *NBER Working Paper No. 24445* .

Gruber, Jonathan and Kleiner, Samuel A. (2012). 'Do Strikes Kill? Evidence from New York State', *American Economic Journal: Economic Policy* 4(1), 127–57.

Gupta, Atul. (2017). 'Impacts of performance pay for hospitals: The Readmissions Reduction Program', *BFI Working Paper No.2017-07* .

Hoot, Nathan R. and Aronsky, Dominik. (2008). 'Systematic Review of Emergency Department Crowding: Causes, Effects and Solutions', *Annals of Emergency Medicine* 52(2), 126–137.

Jha, Ashish K., Joynt, Karen E., Orav, E. John and Epstein, Arnold M. (2012). 'The Long-Term Effect of Premier Pay for Performance on Patient Outcomes', *New England Journal of Medicine* 366(17), 1606–1615.

Joskow, Paul L. (1980). 'The Effects of Competition and Regulation on Hospital Bed Supply and the Reservation Quality of the Hospital', *Bell Journal of Economics* 11(2), 421–447.

Kehlet, Henrik. (2013). 'Fast-track hip and knee arthroplasty', *The Lancet* 381(9878), 1600–1602.

Kelly, Elaine and Stoye, George. (2015). 'New joints: private providers and rising demand in the English National Health Service', *IFS Working Paper W15/22* .

Kolstad, Jonathan T. (2013). 'Information and Quality when Motivation is Intrinsic: Evidence from Surgeon Report Cards', *American Economic Review* 103(7), 2875–2910.

Lindenauer, Peter K., Remus, Denise, Roman, Sheila, Rothberg, Michael B., Benjamin, Evan M., Ma, Allen and Bratzler, Dale W. (2007). 'Public Reporting and Pay for Performance in Hospital Quality Improvement', *New England Journal of Medicine* 356(5), 486–496.

Lindsay, Cotton M. and Feigenbaum, Bernard. (1984). 'Rationing by Waiting Lists', *American Economic Review* 74(3), 404–417.

Morrisey, M., Sloan, F. and Valvona, J. (1988). 'Shifting Medicare patients out of the hospital', *Health Affairs* 7(5), 52–64.

Mullen, Kathleen J., Frank, Richard G. and Rosenthal, Meredith B. (2010). 'Can you get what you pay for? Pay-for-performance and the quality of healthcare providers', *RAND Journal of Economics* 41(1), 64–91.

Propper, Carol. (1990). 'Contingent Valuation of Time Spent on NHS Waiting Lists', *Economic Journal* 100(400).

Propper, Carol, Sutton, Matt, Whitnall, Carolyn and Windmeijer, Frank. (2008). 'Did 'Targets and Terror' Reduce Waiting Times in England for Hospital Care', *B.E. Journal of Economic Analysis & Policy* 8(2).

Sutton, Matt, Nikolova, Silviya, Boaden, Ruth, Lester, Helen, McDonald, Ruth and Roland, Martin. (2012). 'Reduced Mortality with Hospital Pay for Performance in England', *New England Journal of Medicine* 367(19), 1821–1828.

Tables and figures

Table 1: Descriptive statistics

	Mean	St. dev.	N
Panel A: Hospital-day panel			
Daily elective admissions	7.8	8.4	473,496
Daily emergency admissions	3.8	2.7	473,496
Panel B: Inpatient visits			
Age	56.5	19.7	4,279,803
Male, %	47.3	49.9	4,279,803
White, %	85.2	35.5	4,279,803
Diagnosis count	3.5	2.5	4,279,803
Co-morbidity index	1.8	4.4	4,279,803
Past ED visits	0.8	1.5	4,279,803
Elective wait time, days	86.0	74.0	4,279,803
Daycase treatment, %	26.9	44.3	4,279,803
Delayed treatment, %	33.7	47.3	4,279,803
Number of procedures	1.1	0.9	4,279,803
Length of stay, days	4.3	9.4	4,279,803
Transfers out, %	2.9	16.8	4,279,803
Home discharge, %	93.6	24.5	4,279,803
7-day unplanned readmission, %	2.8	16.4	4,279,803
30-day in-hospital mortality, %	1.1	10.5	4,279,803
Panel C: ED visits			
Age	39.5	26.0	20,671,785
Male, %	50.3	50.0	20,671,785
Ambulance arrival, %	28.9	45.3	20,671,785
Attended nearest ED, %	67.1	47.0	20,671,785
Wait time in ED, minutes	154.2	102.6	20,671,785
Inpatient admission, %	24.7	43.1	20,671,785

Notes: (1) 'Co-morbidity index' is defined according to the Charlson index, on a scale between 0 and 50; (2) 'Past ED visit' is a count over the 12 months prior to the index admission; (3) 'Daycase operation' is equal to 1 if a patient receives a procedure that, for that hospital and year, results in no overnight stay for the median patient, and equal to 0 otherwise; (4) 'Delayed treatment' is equal to 1 if a patient receives their primary procedure the day after admission, and equal to 0 otherwise; (5) 'Attended nearest ED' is defined using straight line distances between patient zip codes and all nearby hospitals.

Table 2: Testing for within-day hospital responses to emergency admissions

	ACE	Std. error	ACE % of baseline	N
Panel A: Supply-side responses				
Senior physician count	0.164***	(0.006)	4.15%	391,359
Panel B: Elective patient admissions				
Volume of admissions	−0.003	(0.009)	−0.04%	463,515
Age	0.002	(0.006)	0.00%	2,765,842
Male, %	0.012	(0.015)	0.02%	2,765,842
White, %	0.021*	(0.012)	0.02%	2,765,842
Diagnosis count	−0.001*	(0.001)	−0.03%	2,765,842
Co-morbidity index	0.000*	(0.000)	−0.03%	2,765,842
Past ED visits	0.000	(0.001)	0.03%	2,765,842
Elective wait time, days	0.075***	(0.027)	0.09%	2,765,842
Predicted mortality, %	0.000	(0.000)	−0.02%	2,765,842
Predicted readmission, %	−0.001***	(0.001)	−0.05%	2,765,842
Panel C: Emergency patient admissions				
Attended nearest ED, %	0.934**	(0.457)	1.39%	13,049,202
Wait time in ED, minutes	0.072*	(0.042)	0.05%	13,049,202
Inpatient admission, %	0.007	(0.007)	0.03%	13,049,202

Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (4); (2) Estimates in Panel A use the hospital-day panel, those in Panel B use the hospital-day panel (volume of admissions) and the inpatient dataset (patient characteristics), and those in Panel C use the ED dataset; (3) The sample used in Panel A excludes days when hospitals make no admissions, since on those days it is not possible to observe the number of active physicians at the hospital; (4) The specification in Panel A includes the number of elective admissions; (5) Specifications for volumes in Panel B use lagged admissions hospital-specific year, weekly-seasonal, and day-of-week fixed effects, while specifications for patient characteristics in Panel B include hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (6) Specifications in Panel C include a fully interacted set of diagnosis, age category, and ambulance arrival fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (7) Standard errors clustered at the hospital-level (177 clusters); (8) ***/**/* indicates statistical significance at the 1/5/10% level.

Table 3: OLS estimates of the effect of emergency admissions on health outcomes

	ACE	Std. err.	ACE % of baseline
Panel A: Admission cohorts			
7-day unplanned readmission, %	0.009**	(0.004)	0.31%
30-day in-hospital mortality, %	0.001	(0.002)	0.13%
Panel B: Discharge cohorts			
7-day unplanned readmission, %	0.048***	(0.007)	1.72%
30-day in-hospital mortality, %	−0.001	(0.002)	−0.05%

Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (4); (2) All specifications include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (3) N = 4,279,803 in all regressions; (4) Standard errors clustered at the hospital-level (177 clusters); (5) ***/**/* indicates statistical significance at the 1/5/10% level.

Table 4: OLS estimates of the effect of emergency admissions on treatment decisions

	ACE	Std. err.	ACE % of baseline
Panel A: Admission cohorts			
Daycase operation, %	0.022^{**}	(0.010)	0.08%
Delayed treatment, %	0.187^{***}	(0.015)	0.55%
Number of procedures	-0.002^{***}	(0.000)	−0.15%
Panel B: Discharge cohorts			
Length of stay, days	-0.030^{***}	(0.002)	−0.70%
Transfers out, %	0.000	(0.005)	0.00%
Home discharge, %	0.003	(0.009)	0.00%

Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (4); (2) All specifications include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (3) N = 4,279,803 in all regressions; (4) Standard errors clustered at the hospital-level (177 clusters); (5) $^{***}/^{**}/^{*}$ indicates statistical significance at the 1/5/10% level.

Table 5: OLS estimates of the effect of emergency admissions on health outcomes and treatment decisions for elective and emergency patients

	Elective		Emergency	
	ACE	SE	ACE	SE
Panel A: Admission cohorts				
7-day unplanned readmission, %	0.003	(0.003)	0.017**	(0.008)
30-day in-hospital mortality, %	0.001	(0.001)	0.002	(0.006)
Daycase operation, %	−0.002	(0.011)	0.005	(0.006)
Delayed operation, %	0.053***	(0.016)	0.398***	(0.022)
Number of procedures	−0.001***	(0.000)	−0.003***	(0.000)
Panel B: Discharge cohorts				
7-day unplanned readmission, %	0.010***	(0.003)	0.105***	(0.016)
30-day in-hospital mortality, %	0.001	(0.001)	−0.003	(0.006)
Length of stay, days	0.000	(0.001)	−0.075***	(0.005)
Transfers out, %	−0.001	(0.003)	0.002	(0.011)
Discharges to home, %	−0.009	(0.007)	0.020	(0.015)

Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (4); (2) All specifications include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (3) N = 2,765,866 for elective patients, and N = 1,513,937 for emergency patients; (4) Standard errors clustered at the hospital-level (177 clusters); (5) ***/**/* indicates statistical significance at the 1/5/10% level.

Table 6: OLS estimates of the effect of emergency admissions on readmission diagnoses of emergency discharge cohorts (ordered by frequency of diagnosis)

	ACE	SE	Mean	ACE % baseline
Mechanical complication of internal joint prosthetic	0.0017	(0.0016)	0.0946	2%
Infection following procedure (excl. relating to prosthetic)	0.0022	(0.0013)	0.0929	2%
Fracture of neck of femur	0.0030*	(0.0015)	0.0735	4%
Fracture of lower end of radius	0.0151***	(0.002)	0.0632	24%
Soft tissue disorders	0.0003	(0.0007)	0.0554	1%
Cellulitis	0.0001	(0.001)	0.0489	0%
Haemorrhage following procedure (excl. relating to prosthetics)	0.0013**	(0.0006)	0.0462	3%
Fracture of lower leg including ankle	0.0045***	(0.0014)	0.0454	10%
Urinary system disorders	0.0008	(0.0013)	0.0449	2%
Infection and inflammation due to internal orthopedic device	0.0010	(0.0009)	0.0415	2%
Unknown or unspecified causes of morbidity	0.0110	(0.0102)	0.0397	28%
Other complications of internal orthopedic prosthetic device	0.0001	(0.0007)	0.0356	0%
Pain in limb	0.0003	(0.0007)	0.0294	1%
Other complications of procedures	−0.0011**	(0.0004)	0.0285	−4%
Phlebitis and thrombophlebitis of deep vessels of lower extremities	0.0001	(0.0006)	0.0282	0%
Lumbar and other intervertebral disc disorders	0.0005	(0.001)	0.0274	2%
Pulmonary embolism without mention of acute cor pulmonale	0.0007	(0.0007)	0.0272	2%
Lobar pneumonia	−0.0002	(0.0008)	0.0260	−1%
Disruption of operation wound	−0.0011	(0.0007)	0.0254	−4%
Pain in joint	−0.0005	(0.0009)	0.0250	−2%

Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (4); (2) Dependent variables are indicators for whether a readmission event is associated with a particular ICD-10 diagnosis code, shown for the top 20 diagnoses; (3) All specifications include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (4) Standard errors clustered at the hospital-level (177 clusters); (5) ***/**/* indicates statistical significance at the 1/5/10% level.

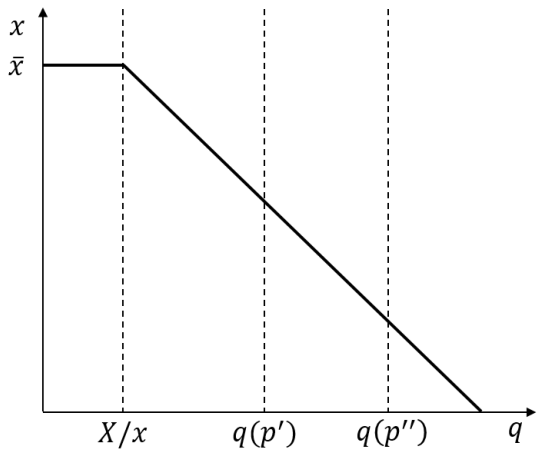
Table 7: OLS estimates of hospital-level ACE estimates on hospital characteristics

	(1)		(2)	
	Length of stay, ACE		Readmission, ACE	
Mean number of physicians	−0.108	(0.106)	0.097	(0.096)
Mean overnight patient census	−0.233**	(0.095)	−0.106	(0.088)
Mean emergency admissions	−0.027	(0.113)	−0.051	(0.102)
Predicted mortality, emergencies	0.345***	(0.082)	0.225***	(0.079)
Predicted mortality, electives	−0.031	(0.090)	0.016	(0.082)
RTT target performance	−0.110*	(0.066)	−0.061	(0.061)
Scheduling performance	−0.012	(0.084)	−0.074	(0.076)
Length of stay, ACE			0.583***	(0.071)

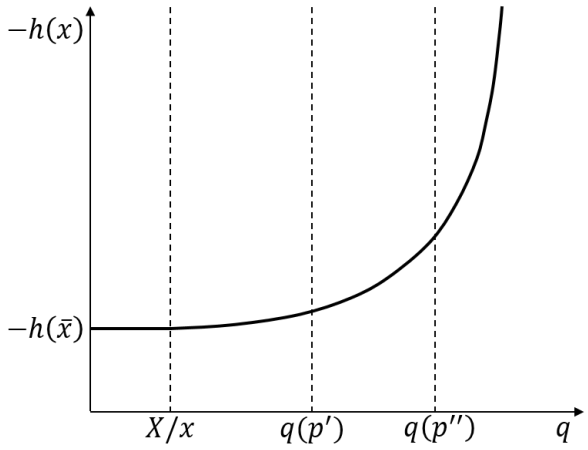
Notes: (1) All variables standardized as absolute z-scores; (2) 'RTT target performance' is the referral-to-treatment wait time target for elective patients; (3) ***/**/* indicates statistical significance at the 1/5/10% level.

Figure 1: Hospital incentives and crowding outcomes

(a) Production inputs

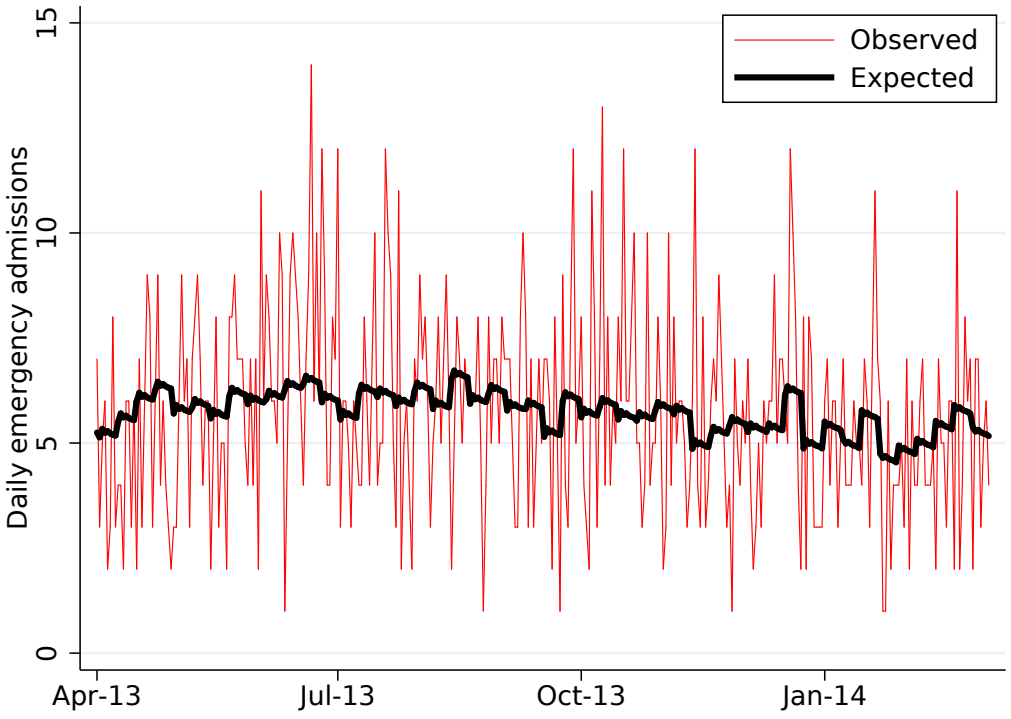


(b) Health outcomes (negative)



Notes: Panel (a) shows the impact of admissions q on production inputs x, where crowding effects occur once $q > X/x$. Panel (b) shows the impact of admissions q on (negative) health outcomes $-h(x)$, where crowding effects occur once $q > X/x$. An increase in admission incentives from p' to p'' increases admissions q and in turn reduces inputs per patient x and worsens health outcomes $h(x)$.

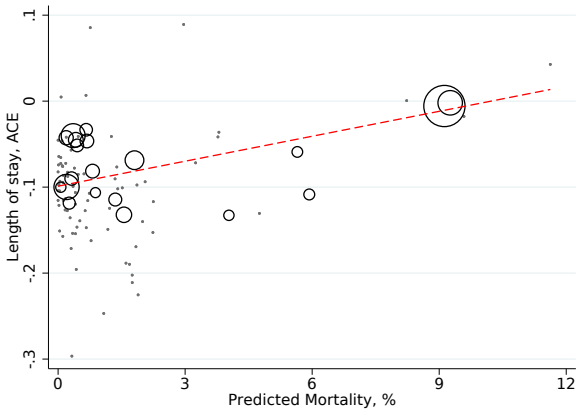
Figure 2: Example of the decomposition of daily emergency admissions



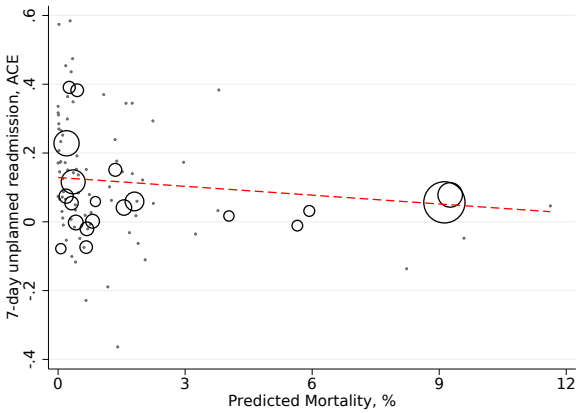
Notes: (1) Data shown for one hospital in one year; (2) Expected emergency admissions defined by a regression of emergency admissions on hospital-specific year, week-seasonal, and day-of-the-week fixed effects.

Figure 3: Average crowding effects for emergency discharge cohorts split by diagnosis group

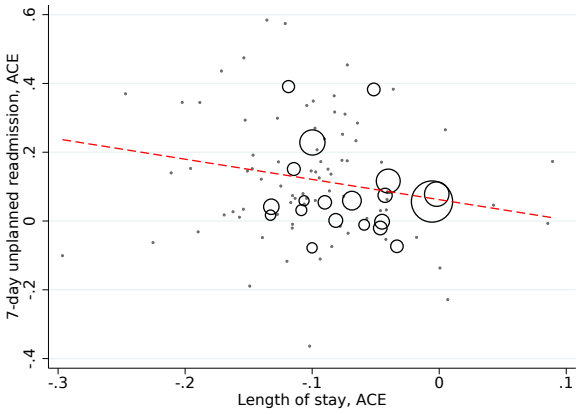
(a) Length of stay



(b) 7-day unplanned readmission



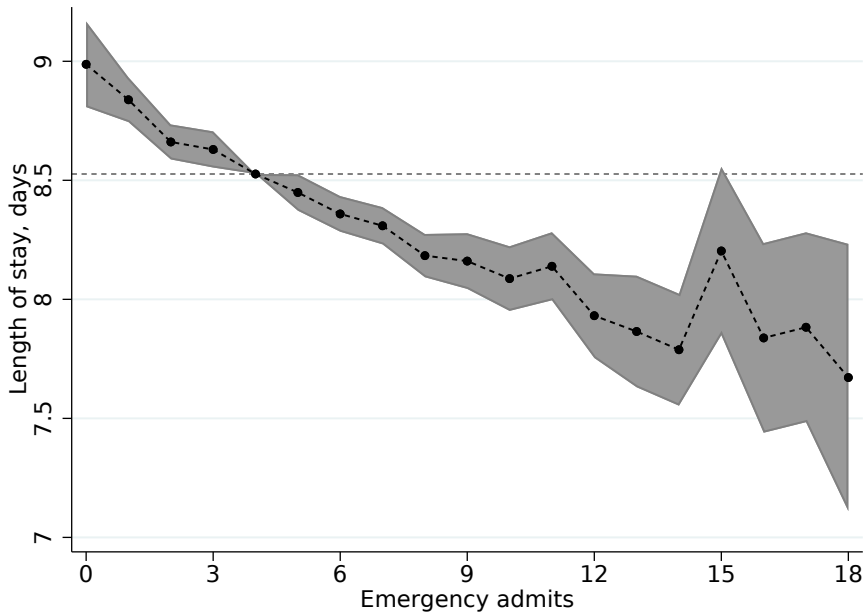
(c) Length of stay vs. 7-day unplanned readmission



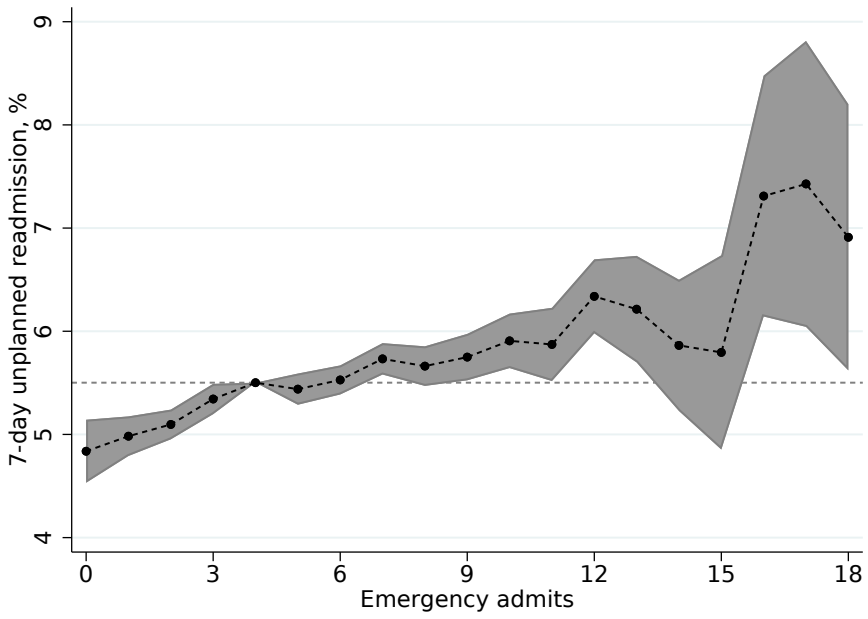
Notes: (1) Each dot represents a diagnosis group for emergency patients, where the largest 20 groups are shown as circles that are proportional to the number of patients; (2) ACE is the average crowding effect, an estimate of the β parameter in Equation (4); (3) Red dashed lines are weighted-regression estimates of the slope of each relationship, where each estimate is statistically significant at the 5% level.

Figure 4: Non-parametric estimates of the effect of emergency admissions on 7-day unplanned readmission and length of stay for emergency discharge cohorts

(a) Length of stay



(b) 7-day unplanned readmission



Notes: (1) Base category of 4 emergency admissions normalized to the unconditional mean of emergency admissions; (2) Estimates for values of emergency admissions above 15 omitted from the figure and are mostly statistically insignificant; (3) Standard errors clustered at the hospital-level (149 clusters) with 95% confidence intervals shown in the shaded region.

A Additional charts and tables

Table A1: OLS estimates of the effect of emergency admissions on health outcomes by different measurement windows

	All patients		Elective		Emergency	
	ACE	SE	ACE	SE	ACE	SE
Panel A: Admission cohorts						
7-day unplanned readmission	0.009**	(0.004)	0.003	(0.003)	0.018**	(0.008)
15-day unplanned readmission	0.011**	(0.004)	0.002	(0.004)	0.024**	(0.009)
30-day unplanned readmission	0.009*	(0.005)	0.000	(0.005)	0.024**	(0.012)
7-day in-hospital mortality	0.002	(0.002)	0.001	(0.001)	0.004	(0.005)
15-day in-hospital mortality	0.002	(0.002)	0.001	(0.001)	0.002	(0.006)
30-day in-hospital mortality	0.001	(0.002)	0.001	(0.001)	0.002	(0.006)
Panel B: Discharge cohorts						
7-day unplanned readmission	0.048***	(0.007)	0.010***	(0.003)	0.105***	(0.016)
15-day unplanned readmission	0.048***	(0.006)	0.009**	(0.004)	0.106***	(0.016)
30-day unplanned readmission	0.044***	(0.007)	0.008*	(0.005)	0.095***	(0.016)
7-day in-hospital mortality	−0.002	(0.002)	0.001	(0.001)	−0.005	(0.005)
15-day in-hospital mortality	0.000	(0.002)	0.001	(0.001)	−0.002	(0.005)
30-day in-hospital mortality	−0.001	(0.002)	0.001	(0.001)	−0.003	(0.006)

Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (4); (2) All specifications include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (3) N = 4,279,803 for all patients, N = 2,765,866 for elective patients, and N = 1,513,937 for emergency patients; (4) Standard errors clustered at the hospital-level (177 clusters); (5) ***/**/* indicates statistical significance at the 1/5/10% level.

Table A2: OLS estimates of the effect of emergency admissions on health outcomes and treatment decisions with different sets of control variables

	All patients					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Admission cohorts						
Daycase operation, %	0.022**	0.022**	0.022**	0.020**	0.020**	0.020**
Delayed operation, %	0.187***	0.186***	0.186***	0.19***	0.189***	0.19***
Number of procedures	−0.002***	0.000***	0.000***	0.000***	0.000***	0.000***
7-day unplanned readmission	0.009**	0.004**	0.004**	0.004***	0.004***	0.004***
30-day in-hospital mortality	0.001	0.002	0.002	0.002	0.002	0.002
Panel B: Discharge cohorts						
Length of stay, days	−0.030***	−0.030***	−0.030***	−0.025***	−0.025***	−0.025***
Transfers out, %	0.000	0.000	0.000	0.001	0.001	0.001
Discharges to home, %	0.003	0.003	0.003	−0.003	−0.003	−0.003
7-day unplanned readmission	0.048***	0.048***	0.048***	0.050***	0.050***	0.05***
30-day in-hospital mortality	−0.001	−0.001	−0.001	0.003	0.003	0.003
Control variables (both panels)						
Baseline fixed effects	✓	✓	✓	✓	✓	✓
Gender fixed effects		✓	✓	✓	✓	✓
Ethnicity fixed effects			✓	✓	✓	✓
Diagnosis count				✓	✓	✓
Co-morbidity index					✓	✓
Past ED admits						✓

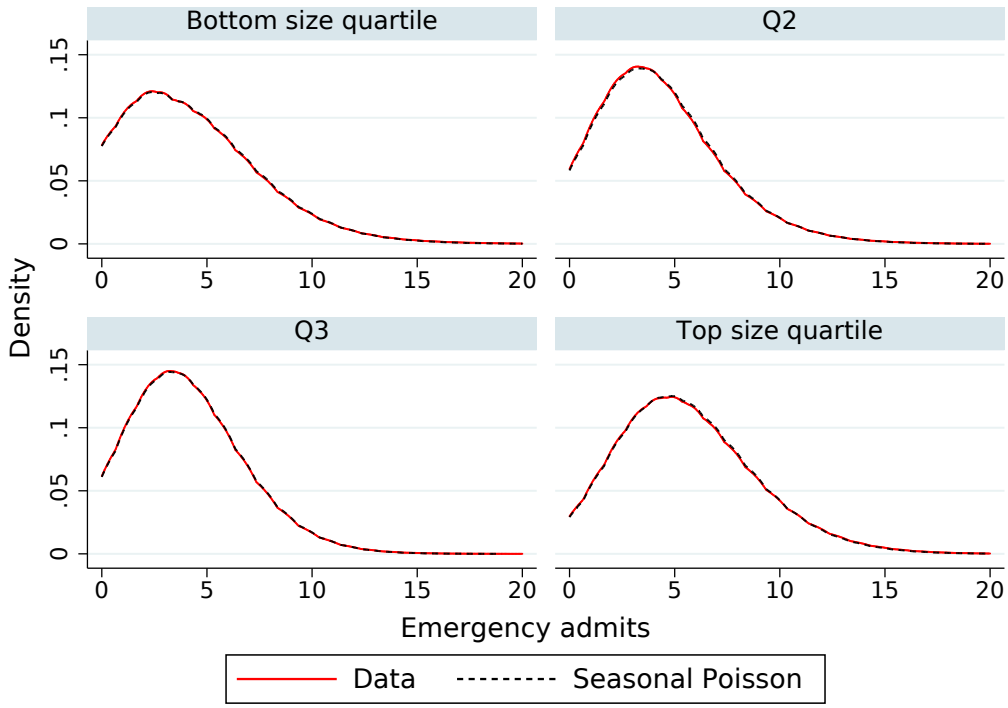
Notes: (1) Estimates are the average crowding effect (ACE), an estimate of the β parameter in Equation (4); (2) Baseline fixed effects include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (3) N = 4,279,803; (4) Standard errors clustered at the hospital-level (177 clusters); (5) $^{***}/^{**}/^{*}$ indicates statistical significance at the 1/5/10% level.

Table A3: OLS estimates of the effect of emergency admissions on readmission specialties (ordered by frequency of specialty)

	ACE	SE	N
Panel A: Admission cohorts			
Trauma and orthopedics	0.0095***	(0.0024)	
General medicine	0.0001	(0.0018)	
Geriatric medicine	0.0002	(0.0010)	
Accident and emergency	−0.0005	(0.0009)	
Plastic surgery	0.0000	(0.0010)	
General surgery	0.0009	(0.0006)	
Neurosurgery	−0.0002	(0.0006)	
Rehabiliation service	−0.0003	(0.0005)	
Cardiology	0.0004	(0.0004)	
Respiratory medicine	0.0001	(0.0004)	
Panel B: Discharge cohorts			
Trauma and orthopedics	0.0445***	(0.0059)	
General medicine	0.0029*	(0.0016)	
Geriatric medicine	−0.0010	(0.0010)	
Accident and emergency	0.0003	(0.0008)	
Plastic surgery	0.0010	(0.0012)	
General surgery	0.0003	(0.0006)	
Neurosurgery	0.0022***	(0.0006)	
Rehabiliation service	−0.0004	(0.0004)	
Cardiology	0.0000	(0.0004)	
Respiratory medicine	0.0001	(0.0004)	

Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (4); (2) Dependent variables are indicators for whether a readmission event was associated with a particular medical specialty, shown for the top 10 specialties; (3) All specifications include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (4) Standard errors clustered at the hospital-level (177 clusters); (5) ***/**/* indicates statistical significance at the 1/5/10% level.

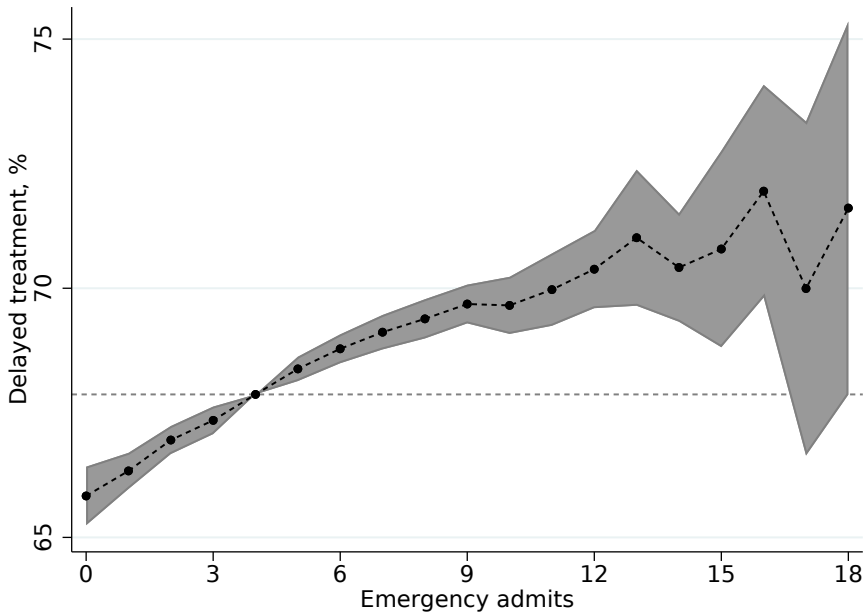
Figure A1: Poisson property of daily emergency admissions



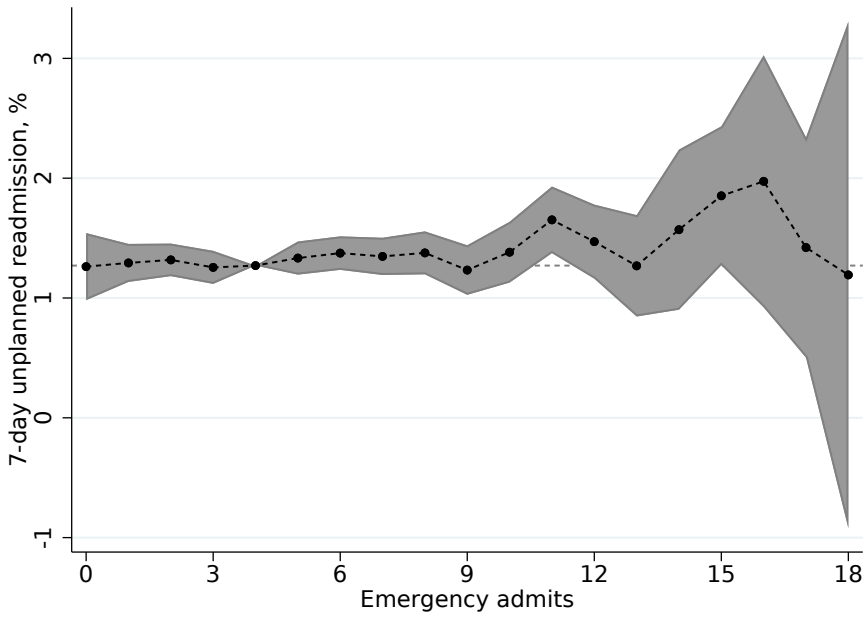
Notes: (1) Size quartiles are defined by the mean daily emergency admissions of each hospital; (2) Poisson data simulated for each hospital with a mean equal to expected emergency admissions, which is defined by a regression of emergency admissions on hospital-specific year, week-seasonal, and day-of-the-week fixed effects.

Figure A2: Non-parametric estimates of the effect of emergency admissions on 7-day unplanned readmission and delayed treatment for emergency admission cohorts

(a) Delayed treatment



(b) 7-day unplanned readmission



Notes: (1) Base category of 4 emergency admissions normalized to the unconditional mean of emergency admissions; (2) Estimates for values of emergency admissions above 15 omitted from the figure and are mostly statistically insignificant; (3) Standard errors clustered at the hospital-level (149 clusters) with 95% confidence intervals shown in the shaded region.

B Serial correlation in emergency admissions

To study the time-series properties of emergency admissions, I decompose the emergency admissions using the following specification

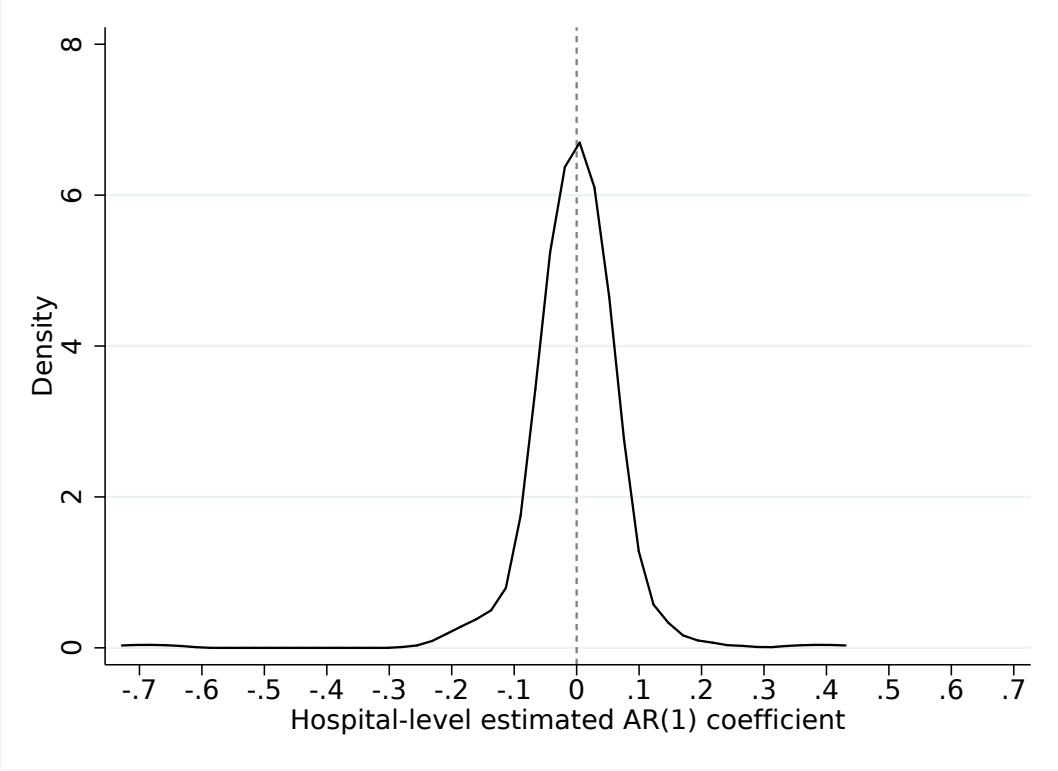
$$q_{hs} = \theta_{hy} + \eta_{hw} + \kappa_{hd} + z_{hs}, \tag{5}$$

where q_{hs} is the number of emergency admissions at hospital h on day s, θ_{hy}, η_{hw} and κ_{hd} are hospital-specific year, weekly-seasonal, and day-of-week fixed effects (consistent with the baseline specification) and z_{hs} is an error term. I refer to predicted values from this regression as 'expected admissions' and the residuals as 'emergency shocks'. The variation in z_{hs} is equivalent to the variation in q_{hs} in Equation (4) after partialing out the hospital-specific fixed effects. The adjusted R^2 statistic from estimating Equation (5) is 0.44 indicating that the hospital-specific seasonality explains just under half of the variation in emergency admissions. Across the full sample, emergency shocks have a mean of zero and a standard deviation of 1.99 (approximately 2 patients).

To evaluate the first-order serial correlation in the shocks, for each hospital I regress the emergency shocks on its lag, and then obtain the distribution of AR(1) coefficients across hospitals. Figure B1 presents this distribution. The distribution is centered at zero and is symmetric, consistent with the shocks being pseudo-random at each hospital and estimation error creating the variation around zero. A similar result is found if the emergency shocks are squared and the analysis repeated. The emergency shock today is therefore independent of the emergency shock yesterday or tomorrow.

There are two implications of the shocks being pseudo-random. First, it allows for the impacts of crowding to be studied in a static framework, because the daily variation in emergency admissions can be considered a random shock that is unrelated to its previous realizations. Second, conditioning on the hospital-specific fixed effects will mitigate endogeneity concerns that could arise through hospitals making scheduling scheduling decisions ahead of the shock - either because of changes to supply-side resources or patient selection - because they are unable to forecast the shocks.

Figure B1: Hospital-level tests of first-order serial correlation



Notes: (1) Figure shows the density of estimated AR(1) coefficients from regressions of emergency shocks on their lag for each hospital separately; (2) Emergency shocks are defined as residuals from a regression of daily emergency admissions on hospital-specific year, weekly seasonal, and day-of-the-week fixed effects.