Does Hospital Crowding Matter?

Evidence from Trauma and Orthopedics in England*

Thomas P. Hoe†

November 19, 2018

Abstract

Hospital crowding is often associated with poor health outcomes. This paper estimates the impact of hospital crowding on patients and tests whether reductions in crowding would improve consumer welfare. Exploiting pseudo-random variation in emergency admits, I find that a one standard deviation admission shock increases the unplanned readmission rate by 4.1%. Further analysis suggests this is driven by emergency patients being discharged quicker and sicker. I then derive a sufficient statistic to test whether reducing crowding with a policy that rations non-emergency admissions would benefit consumers. I find that such policies can lead to widespread improvements in consumer welfare.

*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, 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). Whether or not regulators improve matters, however, depends on their ability or incentive to design appropriate incentive schemes (Stigler, 1971). Health care itself presents an especially difficult case when it comes to incentive design: preferences are hard to elicit, and production is uncertain and highly complex.

This paper examines a feature of hospital production 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. In response, regulators have enacted several policies that indirectly regulate crowding either through health outcomes (e.g. Gupta, 2017) or wait times (e.g. Gruber, Hoe and Stoye, 2018).

I use data from the England National Health Service (NHS), a prominent public health care system with perceived crowding problems, to estimate the effects of hospital crowding and analyze the welfare implications of a rationing policy that would reduce crowding.[1] I estimate the effects of hospital crowding using a quasi-experiment based on the natural variation in emergency admissions. I argue that this causes the number of patients a hospital must treat each day to vary while holding other production inputs fixed, and this corresponds to my definition of crowding. After establishing that crowding does have negative effects on patients, I test whether reductions in crowding would improve consumer welfare. I focus on a policy that reduces crowding by rationing the number of non-emergency ('elective') admissions; this would free up capacity and mitigate the effects of crowding. I evaluate this policy using a sufficient statistic that I derive from a simple model of consumer demand and hospital production.

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

The application focuses on inpatient care, an area that has been associated with crowding pressures and accounts for the large majority of secondary care spending. Inpatient care is also characterized by a separation between crowding and waiting, since elective patients wait for their inpatient appointments at home, and this motivates the rationing policy I study. 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 these departments is that the data records relevant health outcomes and previous studies report estimates of patient preferences. 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 use data from the Hospital Episode Statistics (HES), a comprehensive administrative dataset that covers all publicly-funded secondary care in England. The data links hospital emergency and inpatient departments, tracks patients across hospitals and over time, and contains extensive information about each patient's hospital visit. This allows for a detailed study of crowding, wait times, and admissions.

I identify the crowding effects using variation in emergency admissions. This approach rests on the idea that, in part, the volume of emergency admissions are unpredictable. Random admission shocks therefore change how many patients a hospital must treat each day but gives little scope for a hospital to change its production inputs. I show specifically, conditional on hospital-specific seasonality, shocks to emergency admissions are serially uncorrelated and approximate a Poisson process, which is consistent with typical causes of trauma injuries (e.g. road crashes, sports injuries, falls).

The random variation do 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 responses across several margins and using 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 impacts are especially pronounced for

emergency patients, with the overall range of variation causing the readmission rate to vary by up to 49% of the baseline, but close to zero for elective patients.

Turning to treatment decisions, I find that the salient effect of crowding is to cause emergency patients to be discharged earlier. A one standard deviation shock causes the average length of stay to reduce by 0.060 days (1.4% of baseline), which corresponds to approximately 1 in 15 patients being discharged a day early. Using a non-parametric specification, I show the length of stay effects are linear across much of the emergency admit distribution. This result is consistent with hospitals facing bed constraints and increases (decreases) in admissions feeding directly into increases in earlier (later) discharges. The readmission impacts are also broadly linear, suggesting length of stay may causally influence readmission odds and mean patients are being discharged 'quicker and sicker' (Morrisey et al., 1988). I also find some evidence of delays prior to surgery but no substantive impacts on the number or type of treatments patients receive.

I then turn to the question of whether reductions in crowding would improve consumer welfare. To do so, I set up a static model of consumer welfare that incorporates key features of the hospital market in England, such as the distinction between emergency and elective patients and the latter having the outside option of receiving care at a private hospital. I derive a sufficient statistic that characterizes whether consumers would benefit from reductions in crowding. This statistic equates the marginal rate of technical substitution (MRTS) between wait times and readmission outcomes with the sum of the marginal rate of substitution (MRS) and a term I refer to as the marginal cost of technical substitution (MCTS). While the MRTS and MRS comparison is analogous to the optimality condition for a club good (Buchanan, 1965), the MCTS incorporates the fact that rationing public care, by changing the equilibrium wait times and thus demand from elective patients, will affect the cost of financing public hospitals (Besley and Coate, 1991).

I estimate the sufficient statistic by combining my crowding estimates with estimates of wait time responses and preferences. Obtaining precise estimates of the latter two parameters is challenging. Queuing theory predicts the direction of wait time responses but offers no guide as to the magnitude. Similarly, it is clear that consumers dislike waiting and readmissions but the relative scale is difficult to elicit. I proceed by estimating bounds on the wait time responses, and adopting two complementary approaches to estimating preferences. The first approach to preferences draws on revealed preference estimates from a study of hospital demand in the same setting as mine (Beckert and Kelly, 2017). The second draws on stated preference estimates of

the willingness-to-pay for wait time reductions from a survey in a similar setting (Propper, 1990), and combines this with estimates of the opportunity cost of a readmission. Both approaches produce similar results, implying that patients are willing to wait several months to avoid a readmission.

The test strongly indicates that reductions in crowding would improve consumer welfare. This effect is generated through three channels. First, emergency patients benefit from reduced crowding because they are less likely to experience a readmission. Second, the rationing of elective admissions means that elective patients wait longer for an appointment, and this means they are worse off. Third, the increased wait times make public care less attractive for elective patients, and their demand response thus lowers the tax burden on all consumers of financing public hospitals. The key factor that causes this to net out as a positive impact is the tax burden. I repeat this test across a range of patients subgroups, at the regional and demographic level, and find similar results indicating that the welfare gains would be widespread.

Overall I conclude that hospital crowding has negative effects and that consumers would benefit from reduced crowding. This involves trading off health benefits for emergency patients with increased wait times for elective patients, and a lower tax burden for all patients.

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 examine how crowding can affect consumer welfare. The rationing policy I study also brings together the crowding literature with the literature on hospital wait times (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, as identifying the specific mechanisms

and trade-offs is critical for effective policy design.

The remainder of the paper is organized as follows. Section 2 describes the institutional setting and describes inpatient departments and the English NHS. Section 3 describes the HES data. Section 4 sets out the empirical analysis of crowding. Section 5 sets out the model of consumer welfare and derives and implements the sufficient statistic test. Section 6 concludes.

2 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 incentives of public hospitals in this system, followed by a description of how hospital inpatient departments operate.

2.1 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 are operational targets that focus on wait times. These apply to many areas of health care delivery, including inpatient care (the topic of this paper) and emergency care (Gruber, Hoe and Stoye, 2018). For inpatient care, wait time targets specify that patients should wait no longer than 3 months between the time of their referral from primary care to their inpatient surgery. 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.

Public hospitals have an obligation to break even financially and are also subject to financial incentives. 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 in England often running a financial deficit but do not facing closure.

Parallel to the NHS, there is a private health care system that provides elective inpatient care (but not emergency care). Individuals can obtain care at a private hospital by either paying a list price for treatment if uninsured ('self pay'), or by purchasing private medical insurance in advance and then meeting the cost of any co-payments for hospital care. Approximately 12-13% of the population has private medical insurance at the start of the sample period (Olivella and Vera-Hernández, 2013), and 14% of hip replacements conducted over the period 2008-2013 were privately funded (Kelly and Stoye, 2015). Private hospitals are not subject to the same regulations as public hospitals.

2.2 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). Public hospital staff in England are salaried and gain no direct financial benefit from their treatment of patients. Hospitals will typically schedule a fixed number of staff each hour of the day. At certain times and depending on the hospital, there will be one senior physician 'on call' who may have be off-site but available remotely, and can come on-site if required.

3 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 the ED records for the period 2010 to 2013.[2]

3.1 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' (ORIF). An ORIF is a common surgical procedure used to reassemble and fix broken bones into place.

[2]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. Another advantage of inpatient care relative to emergency care is that there is a clean separation between crowding and waiting, since elective patients wait for their inpatient appointments at home rather than in the hospital. This separation motivates a simple policy response to crowding pressures that I discuss in Section 5.

There are also specific advantages to trauma and orthopedic departments: the data records relevant health outcomes and I am able to obtain benchmarks of patient's preferences from the existing literature. These properties allow me to establish the effects of crowding on health outcomes and make welfare statements regarding these 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).

3.2 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. These hospitals are not subject to the same regulations as public hospitals and are therefore excluded from the study.

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. And 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 from aggregated data, the average number of trauma and orthopedic departments per hospital organization in England is approximately 65 beds.[3]

[3]The data at the organizational level is available from 2010. This reports an average trauma and orthopedic bed capacity of 65, compared to 61 by the end of 2013.

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).

3.3 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.

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.

4 The impact of hospital crowding on patients

The goal of this section is to obtain estimates of how hospital crowding affects patient health outcomes. I define the degree of crowding as the volume of patients being treated simultaneously in a hospital given a fixed level of capacity (beds) and other inputs (labor, machinery). Identifying the effects of crowding therefore require exogenous variation in patient volumes.[4] I exploit variation in emergency admissions for this purpose. I show that these admissions are in part pseudo-random and, under certain conditions, also exogenous which makes them suitable for identifying the effects of crowding.

4.1 Empirical specification

I use the following baseline specification

$$y_{iht} = \alpha_d + \gamma_{ht} + \beta q_{hs} + u_{iht} \tag{1}$$

where y_{iht} is an outcome for patient i at hospital h in cohort t, α_d is a series of fully interacted diagnosis, age category, and emergency status fixed effects (over 45,000 patient types), γ_{ht} is series of hospital-specific year, weekly-seasonal and day-of-the-week fixed effects, q_{hs} is the number of emergency admissions at hospital h on day s, and u_{iht} is an error term. I examine two cohorts of patients: those arriving at hospital each day ('admission cohorts', where $t = s+1$ and s is the admission date) and those leaving the hospital each day ('discharge cohorts', where

[4] An alternative approach would exploit variation in capacity. However, I do not observe department-specific bed capacity and the data on aggregate bed capacity suggests there was very little change over the period I study.

$t = s$ and s is the discharge date).[5]

The parameter of interest is β. This reflects the average causal impact of emergency admissions on patient outcome y_{iht}, which I refer to as a crowding effect. I estimate the baseline specification using an algorithm designed for high-dimensional fixed effects by Correia (2016).[6] Standard errors are clustered at the hospital-level (177 clusters).

4.2 Identification

I make the following identification assumption: $\mathbb{E}[q_{hs}u_{iht}] = 0$. 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).

4.2.1 Threats to identification

Threes 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. An alternative response might be to accept only the most severe patients to minimize any

[5] In principle one could think about analyzing the impact of q_{hs} on decisions during a patient's hospital stay. The data does not allow for this possibility as it only records admission and discharge dates and not daily treatment activity.

[6] OLS produces near-identical estimates.

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 group of possibilities as 'natural endogeneity'.

I now turn to a series of tests that provide support for the identification assumption.

4.2.2 Pseudo-random variation in emergency admissions

To begin with, I show that, conditional on the hospital-specific fixed effects, emergency admissions are pseudo-random. This property partially mitigates against each threat to identification since it implies that hospitals will be unable to forecast deviations from the seasonal expectation of emergency admits ('emergency shocks'). The shocks are therefore independent of any prior scheduling decisions made by the hospital (e.g. over resources and patients admissions) and, by definition, will be orthogonal to seasonal variations in patient composition.[7]

To illustrate the pseudo-random property, I explicitly decompose the emergency admissions using the following specification

$$q_{hs} = \theta_{hy} + \eta_{hw} + \kappa_{hd} + z_{hs}, \tag{2}$$

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 (1) after partialing out the hospital-specific fixed effects. The adjusted R^2 statistic from estimating Equation (2) 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).

[7] A further benefit of the pseudo-random property is that it motivates the static baseline specification. Pseudo-random shocks imply that there are no complex dynamics in emergency admits, and so it is reasonable to study how shocks impact patients in a static framework. This approach thus captures the impact of a random shock relative to the hospital being in an equilibrium, which would be the conditions when an expected, mean-zero, shock occurs.

Figure 1 provides an example of the decomposition for one hospital in one year. Observed emergency admissions (thin red line) have a mean of around five and exhibit significant variation with low admission days and high admission days often in close succession. The expected admissions (thick black line) show that the seasonal pattern is slightly higher in summer and lower in winter, with minor variations across days of the week. This pattern is consistent with the causes of many trauma admissions, which involve outside activities such as road traffic accidents, slips and falls, and sports injuries. Emergency shocks are defined as the difference between the observed data and the expected admissions.

The time-series properties of the emergency shocks are now straightforward to analyze. To illustrate the pseudo-random property I focus on first-order serial correlation. For each hospital I regress the emergency shocks on its lag, and then obtain the distribution of AR(1) coefficients across hospitals. Figure 2 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.

The implication is that conditioning on the hospital-specific fixed effects mitigates several endogeneity concerns that arise through hospital scheduling - either because of changes to supply-side resources or patient selection - and seasonal variations in patient composition.

4.2.3 Tests for within-day responses

The pseudo-random property mitigates scheduling responses but does not guarantee identification since hospitals may still respond to shocks during the day that the shock occurs. These short-notice, within-day responses are likely to be more limited than the scheduling responses but still threaten identification. For example, if hospitals cancel elective appointments or diverting incoming ambulance patients this may influence the selection of patients that are treated when there are emergency shocks. To test for these types of response, I use the baseline specification and replace the dependent variable with alternative outcomes.

Table 2 presents the results for a range of responses. Panel A considers supply-side responses and shows the response of senior physicians presence to emergency shocks. The estimate indicates that there is around 1 additional senior physician present for every 6 unexpected emergency admissions. This effect is small, with a shock of that magnitude corresponding to a 3 standard

deviation event. 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. The results therefore indicate that there is a supply-side response to the shocks, at least in terms of senior physicians, but it is limited in magnitude and, if anything, this would bias my crowding estimates towards zero.[8]

Panel B presents results for mechanisms relating to emergency patients. I find that emergency admits have no impact on the likelihood of inpatient admission and no convincing evidence of effects on hospital choice (which includes ambulance diversion) and 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, and the sign of the estimate is counter-intuitive. 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 orthopaedic patients produces similar results.

Panel C presents the results for mechanisms relating to elective patients. If hospitals are canceling or rescheduling appointments in response to emergency shocks then I should detect these responses in the characteristics of the admitted patients. I therefore use the baseline specification with the mean characteristics of admitted patients as the dependent variable and I remove the patient type 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 (0.2% of baseline). This result is consistent with physicians prioritizing elective patients according to their wait time to meet wait time targets but having only a small amount of discretion when canceling or rescheduling patients. One explanation for this is that there may be a small number of pre-surgery patients to choose from at the point when the need for cancellations becomes apparent.

Finally, I consider the potential for within-day natural endogeneity. The concern here is that the patients on high-shock days differ from those on low-shock days. This is partially mitigated by the patient type fixed effects, which are extremely granular and cover 45,000 groups. But

[8]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.

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.[9] Figure A1 in Online Appendix A presents the results of this exercise, showing that the simulated Poisson and the observed data are near-identical. 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 summarize, these results generally 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 type, 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.[10]

4.3 Results

I start by presenting average crowding effect estimates for health outcomes and then turn to estimates for treatment decisions that explore the mechanisms behind the health impacts.

4.3.1 Health outcomes

Table 3 presents the average crowding effect estimates for health outcomes. Panel A presents the results for admission cohorts and Panel B for discharge cohorts. I present the estimates for all patients in the first column, followed by separate results for elective and emergency patients in subsequent columns.

Looking first at the results for all patients, I find no effects 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

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

[10]Table A2 in Online Appendix A shows that the main estimation results are robust to additional control variables, which also suggests that within-day bias is not substantial.

contrast, I find positive and statistically significant effects on unplanned readmission, indicating that more patients return to hospital in an emergency 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 for admission cohorts by 0.018 percentage points (0.6% relative to the baseline) and for discharge cohorts by 0.096 percentage points (3.4%). The sum of these effects is 0.114 (4.1%). I investigate the mechanisms behind these different effects below.

Turning now to the elective and emergency patient estimates, there are clear differences. The crowding effects on elective patients are very close to zero. The estimate is only statistically significant for discharge cohorts, although it remains economically insignificant. The estimates for emergency patients are much larger in magnitude and statistically significant for both admission and discharge cohorts. As with the aggregate results, the impacts for discharge cohorts are around fives times larger in magnitude than for admission cohorts.

In summary then, crowding does not impact patient mortality but it does lead to increases in readmissions. These effects primarily impact emergency patients that are being discharged on the day of the crowding impact.[11]

4.3.2 Mechanisms

To understand what may be driving the health impacts, Table 4 presents average crowding effect estimates for treatment decisions in the inpatient department. Panel A presents results for decisions relating to admission cohorts and Panel B for those relating to discharge cohorts. The three columns present results for all patients, elective patients, and emergency patients.

Consider first the results for all patients. For admission cohorts, crowding causes delays prior to surgery. 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 receive their primary procedure. There is also some weak evidence of effects on the number and type of procedures received, but these estimates are very small in magnitude. Turning to the discharge cohorts, I find that crowding causes reductions in length of stay (patients are being discharged sooner) but no impacts on the likelihood of transfer to another hospital or being sent home rather

[11] Table A1 in Online Appendix A shows that these results are robust to changes in the window (7, 15 or 30 days) during which readmission and mortality is measured.

than to another care provider. A one standard deviation shock causes average length of stay of discharge cohorts to reduce by 0.060 days (1.4% of baseline), which corresponds to around 1 in 15 patients being discharged a day early.

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). Alternatively, there may be other unobservable responses that are generating the additional readmissions. For example, it is plausible that the effects on admission cohorts are caused by medical outliers (where patients are placed on inappropriate hospital wards) and the effects on discharge cohorts effects may be due to mistakes in the discharge process (such as incorrectly or inadequately dispensed drugs).

To try and get a better sense of the relationship between readmissions and these mechanisms, I conduct two exercises: first, I compare the elective and emergency patients again; and second, I estimate non-parametric versions of the baseline specification. For the latter, I replace the emergency admission variable in Equation (1) with a series of indicator variables for each discrete value of emergency admissions.

The results for elective and emergency patients are shown in the right-hand columns of Table 4. Looking at delays and length of stay, the impacts on emergency patients are substantially larger than on elective patients. The length of stay of elective patients is unaffected by emergency shocks, while the likelihood of a delay is around eight times smaller than for emergency patients. This pattern mirrors the results for readmissions and suggests that delays and earlier discharges may indeed cause some of the readmissions.[12]

Figure 3 presents non-parametric estimates for length of stay (Figure 3a) and readmission (Figure 3b) for discharge cohorts. Over most of 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. These estimates are consistent with the quicker and sicker hypothesis and length of stay being a causal factor in readmission outcomes. They also suggest hospitals are facing binding bed constraints since the alternative situation,

[12]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.

where there is slack bed capacity, would lead to length of stay adjustments only being required when there are large positive emergency shocks. 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).

In contrast, the non-parametric estimates for admission cohorts do not support delays being a causal factor in readmission outcomes. These results are shown in Figure A2 in Online Appendix A. The readmission impacts (Figure A2b) are small, with effects only present at large positive emergency shocks. In contrast, the effects on delayed treatment (Figure A2a) are broadly linear.

These results show that the salient, observable crowding mechanism is that emergency patients are discharged more quickly when there are unexpected increases in admissions. Similarly, unexpected decreases in admissions lead to emergency patients having a longer stay. These changes are associated with the changes in readmissions, suggesting length of stay is a causal factor in readmission outcomes.

5 Crowding and consumer welfare

The preceding results demonstrate that crowding has a negative impact on patient health outcomes, in part caused by early discharges. I now consider a rationing policy that has the potential to mitigate crowding effects: by reducing the number of elective admissions each period, hospitals can free up capacity and reduce crowding pressures.

The expected effects of this policy are as follows. A cost of the policy is that admitting fewer elective patients each period will raise the expected wait time faced by elective patients. This follows directly from any basic queuing model.[13] A benefit of the policy is that crowding would be reduced, since greater capacity would be available for the same distribution of emergency admissions, and the empirical results indicate this would result in fewer readmissions of emergency patients. On top of these costs and benefits there is a third impact: by affecting the characteristics of public hospitals, elective patients may change their decision of whether to attend a public or private hospital in the first place. The substitution to the private outside

[13] A queuing model would typically characterize the issue in terms of the capacity dedicated to elective patients ('servers'). I adopt the convention of referring to the number of admissions for simplicity. The difference between the quantity of and capacity dedicated to elective patients is only in periods where few elective patients demand service, in which case the admissions could be lower than the dedicated capacity. I ignore this difference as it does not feature in the equilibrium outcomes.

option may therefore affect the cost of financing public hospitals (Besley and Coate, 1991). In the case where elective patients are only affected by their wait times, and not the likelihood of readmission (as the empirical results indicate), the policy change will make public hospitals less attractive to elective patients and thus lower costs of financing.

The objective of this section is to test whether pursuing this policy option would improve consumer welfare. I set up a simple model that incorporates the key features of hospital care in England, such as patient heterogeneity (the differences between elective and emergency care) and endogenous demand (the choice between public and private hospitals). I use the model to derive sufficient statistic to test whether reductions in crowding would improve consumer welfare. I conclude this section by implementing the test and describing the results.

5.1 A simple model of consumer welfare

There is a fixed population of N consumers. Let care of type j be denoted by subscript s for elective care and e for emergency care. The model is static and consumers demand care of type j with probability ρ_j. The utility in the state when neither elective nor emergency care is required is normalized to zero.

Denote the number of elective admissions as q.[14] Let the patient health outcomes at public hospitals be $h_j(q)$, where $h_j(.)$ is decreasing convex. The $h_j(q)$ function reflects the crowding mechanism that was characterized by the earlier empirical results. Let $w_j(q)$ be the equilibrium wait times at public hospitals, where $w_j(.)$ is decreasing concave. I suppress the distribution and volume of emergency admissions in the model since it is assumed that this is not a control variable for hospitals or regulators in this setting.

Patients requiring emergency care attend public hospitals as there is no private alternative. The utility of emergency patients is defined as

$$U_e(q) = \psi h_e(q) - \lambda w_e(q). \tag{3}$$

I use a linear utility function to simplify the presentation, and assume additive separability which is plausible for acute but non-deteriorating conditions such as trauma and orthopedics.

Patients requiring elective care have a choice between attending a public and private hospital.

[14] Alternatively, the model can be framed in terms of how much capacity is dedicated to elective patients or, if it is assumed that length of stay is the only cause of the readmissions, the length of stay of patients.

The utility of attending a public hospital is

$$U_s^{pub}(q) = \psi h_s(q) - \lambda w_s(q) - \varepsilon_i. \tag{4}$$

where $\varepsilon_i \sim N(0,1)$ is a preference shock for public relative to private care. In contrast, the utility of attending a private hospital is $U_s^{priv} = \tau$. The fixed private utility assumes that private hospitals do not adjust their characteristics, such as wait times or prices, to changes in public elective admissions. Assuming patients are well-informed and rational, the utility maximizing demand for public care will be

$$\begin{aligned} Pr(\text{public} \mid q) &= \Pr\left(U_s^{pub}(q) > U_s^{priv}(q)\right) && (5)\\ &= \Pr\left(\psi h_s(q) - \lambda w_s(q) - \tau > \varepsilon_i\right) && (6)\\ &= \Phi(\Delta), && (7) \end{aligned}$$

where $\Phi(.)$ is the standard normal *cdf* and $\Delta \equiv \psi h_s(q) - \lambda w_s(q) - \tau$ is the difference in utility due to the characteristics of public and private hospitals.

Public hospitals are assumed to be financed by a per capita tax, t, that is levied on the entire population and in all states of the world. The required tax burden depends on the number of elective patients that choose public care and can be written $t \equiv \rho_s \Phi(\Delta) c$, where c is the marginal cost of a public treatment. The taxation terms were omitted from the above utility functions since they do not impact the choice of public or private elective care. This approach implicitly assumes that consumers do not internalize the impact of their choice on the level of taxes.

Consumer welfare is the expected utility of a randomly selected consumer prior to requiring any hospital care and can be written

$$\begin{aligned} W(q) &= \rho_s\Big\{\Phi(\Delta)E[U_s^{pub}(q) \mid \varepsilon_i > \tau] + \big(1-\Phi(\Delta)\big)U_s^{priv}\Big\} + \rho_e U_e(q) - t, && (8)\\ &= \rho_s\Big\{\Phi(\Delta)\big(\underbrace{\psi h_s(q) - \lambda w_s(q) - I(\Delta)}_{\text{Elective public care}}\big) + \big(1-\Phi(\Delta)\big)\underbrace{\tau}_{\text{Elective private care}}\Big\} && (9)\\ &\quad + \rho_e\Big\{\underbrace{\psi h_e(q) - \lambda w_e(q)}_{\text{Emergency public care}}\Big\} - \underbrace{\rho_s\Phi(\Delta)c}_{\text{Tax burden}}, && (10) \end{aligned}$$

where $I(\Delta) = \phi(\Delta)/\Phi(\Delta)$ is the inverse Mill's ratio. A social planner that seeks to maximize

consumer by setting q will therefore solve

$$\max_{q} \quad W(q). \tag{11}$$

This problem can also be framed a regulator problem, where instead of setting q directly, the regulator sets a bundle of policies that influence hospitals' incentives over q; for example, through a combination of wait time targets and penalties for poor health outcomes. Taking the first order condition, and noting that terms involving the derivative of $\Phi(\Delta)$ cancel through an envelope condition, gives

$$\rho_s \Phi(\Delta)\big(\psi h_s'(q) - \lambda w_s'(q)\big) + \rho_e\big(\psi h_e'(q) - \lambda w_e'(q)\big) = \rho_s \phi(\Delta)\big(\psi h_s'(q) - \lambda w_s'(q)\big)c. \tag{12}$$

This condition shows that a social planner would increase q, thereby intensifying the crowding pressures, when the marginal benefits in terms of the expected utility from public hospitals (left-hand side) exceed the marginal costs that accrue via the tax burden (right-hand side). The former accrue because crowding can affect both wait times and health outcomes, while the latter because elective patients may change their demand for public care. Note that in this general case, the tax burden may increase or decrease with changes in q, depending on the elasticity of demand for public hospitals: $\phi(\Delta)\big(\psi h_s'(q) - \lambda w_s'(q)\big)$. The sign of this elasticity will depend on whether elective utility increases or decreases with crowding, and thus on the balance of the wait time and health outcome effects for these patients.

The above condition gives the general trade-off that will be present with crowding. This condition, however, can be simplified in the current setting by imposing two assumptions that are supported by the data. First, crowding affects emergency patient health outcomes but not elective patient health outcomes (Table 3). Second, the wait times of emergency patients are unaffected by crowding (Table 2, Panel B). Imposing these conditions and rearranging gives

$$\underbrace{\frac{\rho_e}{\rho_s \Phi(\Delta)} \frac{\psi}{\lambda}}_{\text{Weighted-MRS}} + \underbrace{\frac{w_s'(q)}{h_e'(q)} I(\Delta) c}_{\text{MCTS}} = \underbrace{\frac{w_s'(q)}{h_e'(q)}}_{\text{MRTS}}. \tag{13}$$

This revised condition helps to further understand the intuition in this setting. The term on the right-hand side is a MRTS between elective wait times and emergency health outcomes, and specifies a trade-off implied by the hospital production function. The left-hand side is

a (probability-weighted) MRS and a term that I refer to as the marginal cost of technical substitution (MCTS). The MRS specifies the preferences of consumers over elective wait times and emergency health outcomes and weights those preferences by the ex-ante probability of being in each state. In contrast, the MCTS specifies the cost that consumers must pay via the tax burden to trade off a change in elective wait times for emergency health outcomes.

Consider the case of a decreasing in crowding (a reduction in q) and, at first, ignoring the MCTS. This then corresponds to a condition which is similar to that of a club good (Buchanan, 1965). The MRTS specifies the scale of the increase in wait times relative to the reduction in readmission odds. This will improve consumer welfare when the MRTS is smaller than the MRS since then the gains from expected emergency health outcomes will outweigh the expected losses from elective wait times. Now incorporate the MCTS: this term essentially reflects the penalty consumers must pay, through the tax burden, to make a trade-off between waits and health outcomes. Given that a decrease in q on the margin only affects elective patients through increased wait times, without any improvement in health outcomes, this will make elective public care less attractive and reduce the tax burden. The MCTS, by increasing the left-hand side of Equation (13), effectively raises the threshold at which it will be beneficial to increase crowding.

In summary, reductions in crowding will benefit consumer welfare when the sum of the MRS and MCTS is greater than the MRTS. I test this sufficient statistic below.

5.2 Implementing the test

I now describe how I obtain appropriate estimates and benchmarks for each component of Equation (13).

5.2.1 Demand probabilities

Using national population statistics and the HES data, I estimate that $\rho_e = 0.31\%$ and $\rho_s\Phi(\Delta) = 0.53\%$. In relative terms this indicates that the average patient is around 70% more likely to demand public elective care than emergency care in a given year. Using data published by Kelly and Stoye (2015), and assuming that the public and private mix of all orthopedic procedures is similar to that for hip replacements, I estimate that $I(\Delta) = 0.26$.[15]

[15]Kelly and Stoye (2015) publish annual data on public and private hip replacement volumes for the period 2008-2012. I also use this data to calculate that $\Phi(\Delta) = 0.84$ and then infer that $\phi(\Delta) = 0.22$.

5.2.2 Preferences

I require patients' preferences for health outcomes, readmissions in this case, and wait times for elective care. Since these preferences are non-standard and in some cases difficult to measure, I take several approaches and compare a range of estimates. Together I interpret the estimates as providing an order-of-magnitude for the MRS rather than a precise number.

The first approach draws on revealed preference estimates from a hospital choice study by Beckert and Kelly (2017). The study examines hip replacement patients in the England NHS during 2012, and therefore closely approximates my setting, population, and time period. Beckert and Kelly (2017) estimate mixed logit demand models that include the readmission rate and wait time as hospital characteristics. Combining their estimates of preferences, I obtain an estimate of MRS $= -102$.[16]

The second approach uses separate approaches for each preference. I use a stated preference estimate for the wait time preference and benchmarks of opportunity cost for the readmission preferences. 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 gives an estimate of MRS $= -240$.

Despite a number of conceptual differences, these two approaches produce broadly similar results.[17] Applying the probability weightings to an average of the above MRS figures gives a weighted-MRS of -101; this can be interpreted as indicating that the average consumer would be willing to wait an additional 14 weeks, or around 3 months, for elective care in order to avoid a readmission.

[16] To obtain the estimate I divide the mean of their estimated preferences for readmission and wait times. I use results from Table 5 in Beckert and Kelly (2017). The table reports estimates based on two different definitions of the hospital choice sets, and I average the results from both definitions to approximate the MRS.

[17] The first approach will understate preferences for wait times if this variable suffers from endogeneity bias (Beckert, 2018). The second approach may over or understate both the wait time and readmission preferences. Wait times preferences may be overstated because during the time of the survey wait times were substantially higher than they are during my period of study and, conversely, they may be understated if individuals responded strategically or out of protest to the survey (see Propper (1990) for more details). For readmissions, opportunity cost may understate the full cost of a readmission since it neglects any impacts on health and well-being.

5.2.3 Crowding derivatives

I use the crowding estimates from Section 4 as estimates for $h'_e(q)$. In doing so I make two implicit assumptions. First, I assume that unplanned readmissions and mortality capture all of the health impacts of crowding on patients. While these are standard health outcomes when working with administrative data, they may not capture all of the impacts of crowding. For example, crowding may impact health outcomes in more subtle ways that readmissions and mortality do not capture or, conditional on health outcomes, it may impact patient well-being during a hospital stay. I discuss the implications of this potential bias when presenting the results.

Second, I assume that my crowding estimates, which are identified from variation in *emergency* admissions, are comparable to the crowding effects that would occur when a hospital changes its *elective* admissions. The empirical results suggest this assumption is plausible: if the primary crowding mechanism is through length of stay and bed shortages, then whether it is an emergency or elective patient that arrives, the need to discharge a recovering patient and consequence of doing so will be similar.

5.2.4 Wait time derivatives

For the elective wait time derivative, ideally one would evaluate this derivative by exploiting exogenous variation in elective admissions and tracing out the impacts on the equilibrium wait time. Finding such variation, however, is difficult because many of the factors that shift wait times are also likely to induce responses from admissions at public hospitals, especially in the presence of fixed wait time targets.

In the absence of a convincing instrument, I evaluate the wait time derivative with an empirical analysis that produces a lower bound on the wait time response. I aggregate the HES data to the region-quarter level and estimate the following equation

$$w_{rt} = \kappa_r + \delta q_{rt} + v_{rt} \tag{14}$$

where w_{rt} is the log of average wait time in region r in quarter t, κ_r are regional fixed effects, q_{rt} is the number of elective admissions (000s) at public hospitals in the region-quarter, and v_{rt} is an error term. The parameter of interest is δ which is the average wait time derivative.

OLS estimates of δ are likely to be subject to two sources of bias. The first is standard

endogeneity bias. This will be present if hospital admissions respond to unobservable shocks to the wait time. An example of a shock could be new primary care practices (a demand shock that may increase referrals to public hospitals). The fact that public hospitals face fixed wait time targets make it likely that these shocks would be met by changes in admissions, generating a correlation between q_{rt} and u_{rt}. This correlation is expected to be positive: to continue meeting wait time targets, hospitals will have an incentive to increase (decrease) admits in response to positive (negative) wait time shocks. The implication is that endogeneity bias is expected to generate upward bias in the OLS estimates. Since $\beta < 0$ the OLS estimates are therefore expected to understate the magnitude of the wait time derivative.

The second potential bias is that the wait time specification is log linear in admissions. This functional form restriction, combined with estimation over a relatively long time period, may mean I estimate the average slope of a relationship that is not stable over time. To address this issue I estimate a non-parametric version of Equation (14) and consider the extremes of the wait time derivative over the admissions distribution. This again helps to bound the true parameter value.

Figure 4 presents the OLS and non-parametric estimates of Equation (14). The plot shows the region-quarter data after conditioning out the regional fixed effects, and the OLS estimates (black line) and non-parametric estimates (blue dashed line). The OLS estimate of β is -0.108 (0.013), indicating that increasing elective admits in a region by 1,000 each quarter would reduce the elective wait time by approximately 11.2% (a 9.4 day reduction from a baseline mean of 84 days). The non-parametric estimates, in comparison, have a similar slope to the OLS estimate for the left-hand parts of the admit distribution, but flatten out in the right of the distribution. The derivative on the right of the distribution, where admissions are above 1 on the plot, implies a wait time response that is around half of the OLS estimate at 5.7 days. This estimate provides a very conservative benchmark for the wait time derivative since it is the minimum derivative from the non-parametric estimates, and any endogeneity bias is expected to bias this estimate towards zero.

These wait time derivatives require rescaling to incorporate into Equation (13). This is necessary because the crowding estimates are at the hospital-day level and the waiting time estimates are at the region-quarter level.[18] I rescale to a scenario where hospitals on average

[18]The different levels of aggregation also imply differences in the regression weighting and it is possible to re-weight both regressions to be on a comparable basis (e.g. by regional population). In practice I found that this makes little difference.

accommodate one additional patient per day throughout the year. The crowding estimates already satisfy this, so it only involves rescaling the waiting time estimates. After rescaling, a one patient increase in hospital occupancy is estimated to decrease mean waiting times for elective patients by between 7.7 days (non-parametric) and 12.6 days (OLS).[19]

5.2.5 Costs

Using the national government tariffs for hospital payments, I estimate that $c = £3,100$.

5.3 Results

Combining the above estimates gives the following for each of the terms in Equation (13). The probability-weighted MRS is -101. I estimate that the MCTS is between 5,067,251 (non-parametric) and 8,263,721 (OLS), indicating that increasing elective admits would substantially increase the tax burden. Finally, I estimate that the MRTS is between 3,718 (non-parametric) and 6,063 (OLS), indicating that increasing elective admits would lower equilibrium wait times substantially (between 7.7 and 12.6 days) while only increasing readmissions marginally (by 1 in 814 patients). Putting these estimates together indicates that the sum of the MRS and MCTS substantially exceeds the MRTS. This result, which is statistically significant, implies that consumer welfare would be improved by reducing crowding.[20]

The intuition for this result can be understood by examining two features of the estimates. First, if the MCTS is not incorporated, then the MRS is actually substantially below the MRTS. On this basis, which corresponds to the pure club good case, it indicates that increasing crowding benefits consumer welfare. Yet when the MCTS is incorporated then this reverses that conclusion and implies that reductions in crowding are in fact beneficial. Second, the MCTS is particularly influential due to the sharp differences between patient types, and specifically the fact that crowding effects on elective patients are limited to the wait time. This implies that an increase in crowding, while causing health outcomes to deteriorate for emergency patients, only makes public hospitals more attractive to the elective patients because of the lower wait times,

[19]The waiting time estimates give the impact of 1,000 admissions in a region-quarter. I scale this by the number of elective patients required to occupy one bed at a single hospital throughout the year (i.e. 365 divided by the average elective length of stay), and then by the number of hospitals per region (around 7).

[20]Computing the first-order condition directly, by using Equation (12) and subtracting the right-hand side from both sides, gives an estimate of -57 (non-parametric) or -116 (OLS), with 95% confidence intervals of $[-90, -143]$ and $[-25, -89]$ respectively. To compute these confidence intervals, I take the demand probabilities, costs and preferences are given, and assume a zero covariance between the crowding and wait time estimates. A negative first-order condition indicates that reducing crowding would improve consumer welfare.

and so causes the tax burden to increase for all consumers. The MCTS therefore provides a strong disincentive for increases in crowding and generates significant monetary benefits from reductions in crowding.

There are two main sources of bias that could affect my results. First, each of the preference benchmarks, crowding estimates and wait time responses may individually be subject to certain biases in magnitude. In practice, however, an order-of-magnitude change up or down in these estimates would not change my conclusion. A second area of potential bias relates to the elasticity of demand for public hospitals. While the estimates establish that elective utility from public hospitals decreases as crowding reduces, the demand response is implied by the functional form assumption on the distribution of preferences. In the absence of estimating this directly, I note that previous work has shown patients are responsive to NHS wait time changes (Besley et al., 1999) and my conclusion will hold even for very small demand responses.[21] The key result that creates this situation is the fact that elective patients only lose out from reductions in hospital crowding, through increased wait times, and do not experience improvements in health outcomes. This ensures that the demand response to reductions in crowding is necessarily negative and, given the cost of public treatment, only a very small elasticity is required to generate the result.

The results so far provide an view of whether changes in crowding at a national level would benefit consumers. It is also interesting to ask whether this result masks significant heterogeneity across regions or patients, such that the first-order condition may be different for certain subgroups. While I do not have measures of the MRS across subgroups, it is possible to estimate the remaining components at more granular levels to explore the heterogeneity.

I conduct this exercise for regions and age-gender groups. I focus on Equation (12), which is the general first-order condition, since this incorporates the potential for non-zero effects on the health outcomes of elective patients. I compute this condition for each subgroup, assuming that preferences are constant, and by imposing some assumptions on the wait time effects to avoid introducing highly imprecise estimates. For the age-gender analysis, I assume that the wait time response is equivalent across demographic groups, and for the regional analysis, I maintain the aggregate derivative but apply this to the region-specific baseline wait time. The former assumption will hold if patients are treated on a first-come-first-served basis in the queue, which

[21] For example, even if the demand response is an order (or two order) of magnitude smaller than implied by the model, the same result will hold.

is approximately true, while the latter introduces a degree of heterogeneity in a parsimonious way. I compute the demand probabilities and crowding impacts for each demographic group and region separately.

Figure 5 presents the results of the heterogeneity analysis. The chart shows the distribution of first-order condition for demographic groups and regions and how this compares to the zero (red dashed vertical line). In all cases, the group-specific first-order condition is negative and in many cases by a significant margin. These findings suggest that reducing crowding would benefit welfare for consumers across regions and different demographic groups.

To sum up, the implication of these results is that reducing crowding would improve consumer welfare. This will improve health outcomes, particularly for emergency patients, and lower the costs of financing public hospitals. Together these two effects generate benefits that exceed the cost of increased wait times for the elective patients. This finding is consistent across regions and demographic groups, suggesting that the benefits will be widespread.

6 Conclusion

In spite of efforts by regulators, hospital crowding remains a major policy issue and has recently been described as an international crisis (Hoot and Aronsky, 2008). Numerous public health care systems are routinely cited as facing crowding pressures despite existing efforts to ration demand. I analyze crowding in one of the most well-known public health care systems, the English NHS. This paper presents estimates of the causal effects of crowding on patients and then tests whether reductions in crowding would improve consumer welfare.

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 (2 emergency admits, relative to an average day at a 50-day hospital department), 7-day unplanned readmissions increase by 0.114 percentage points (4.1% relative to baseline). The impacts are especially large for emergency patients but close to zero for elective patients. I find strong evidence that reducing crowding by rationing elective care would improve consumer welfare. This follows not only because of the improvement in health outcomes for emergency patients, but also because of the tax savings created by elective patients going outside of the public system for treatment. I show that this finding holds universally across regions and demographic groups.

The implication of these results is that policies which ration elective care, thereby reducing

crowding but increasing elective wait times, can benefit consumers. One way to implement such a policy in England would be to relax the existing waiting time targets for elective care.

There are two main caveats to the analysis. First, it is unclear how far to proceed in the direction of rationing care, with the results here only indicating which direction would improve matters. Policy change should therefore proceed with caution. Second, these conclusions do not generalize to other departments, providers or countries. Determining factors in these other settings include the levels of capacity, the health conditions involved, and patients' preferences. Gaining an empirical sense of these factors and how they interact will be essential for designing effective health care policy.

A further question raised by this study is if consumer welfare can be improved through reduced crowding, what has prevented this policy option from being implemented? One potential explanation is that there may be political influence, with the patients exposed to elective wait times holding more sway when it comes to setting wait time targets. An alternative is that regulators may simply fail to maximize their intended criterion, perhaps because of incomplete information over the production process or preferences. While there is recent work exploring the former explanation, (Bloom et al., 2015; Chan and Dickstein, 2018), the fact that my results hold similarly across regions and demographic groups, suggests that the latter explanation is perhaps more plausible in this case. With this in mind, further research on the efficacy of health care regulation is essential.

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.

Beckert, Walter. (2018). 'Choice in the presence of experts: The role of general practitioners in patients' hospital choice', *Journal of Health Economics* 60, 98–117.

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* .

Besley, Timothy and Coate, Stephen. (1991). 'Public Provision of Private Goods and the Redistribution of Income', *American Economic Review* 81(4), 979–984.

Besley, Timothy, Hall, John and Preston, Ian. (1999). 'The demand for private health insurance: do waiting lists matter?', *Journal of Public Economics* 72(2), 155–181.

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'.

Buchanan, James M. (1965). 'An Economic Theory of Clubs', *Economica* 32(125), 1–14.

Chan, David and Dickstein, Michael. (2018). 'Industry Input in Policymaking: Evidence from Medicare', *NBER Working Paper No. 24354* .

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.

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.

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* .

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* .

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.

Olivella, Pau and Vera-Hernández, Marcos. (2013). 'Testing for Asymmetric Information in Private Health Insurance', *Economic Journal* 123(567), 96–130.

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).

Stigler, George J. (1971). 'A theory of economic regulation', *Bell Journal of Economics and Management Science* 2(1), 3–21.

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: Patient selection–emergencies				
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
Panel C: Patient selection–electives				
Age	0.002	(0.006)	0.00%	2,765,848
Male, %	0.012	(0.015)	0.02%	2,765,848
White, %	0.021^{*}	(0.012)	0.02%	2,765,848
Diagnosis count	-0.001^{*}	(0.001)	−0.03%	2,765,848
Co-morbidity index	0.000^{*}	(0.000)	−0.03%	2,765,848
Past ED visits	0.000	(0.001)	0.03%	2,765,848
Elective wait time, days	0.075^{***}	(0.027)	0.09%	2,765,848

Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (1); (2) Estimates in Panel A use the hospital-day panel, those in Panel B use the ED dataset, and those in Panel C use the inpatient 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 in Panel B 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; (6) Specifications in Panel C include 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

	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)
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)
30-day in-hospital mortality	−0.001	(0.002)	0.001	(0.001)	−0.003	(0.006)

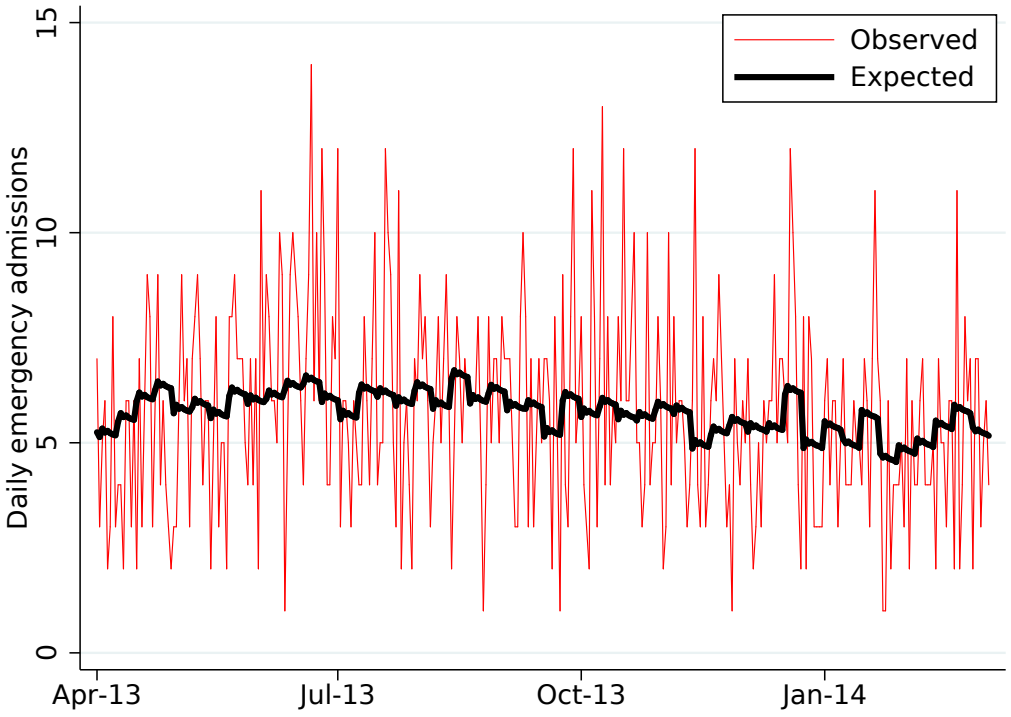
Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (1); (2) All specifications include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (3) N = 4,279,803 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 4: OLS estimates of the effect of emergency admissions on inpatient treatment decisions

	All patients		Elective		Emergency	
	ACE	SE	ACE	SE	ACE	SE
Panel A: Admission cohorts						
Daycase operation, %	0.022**	(0.01)	−0.002	(0.011)	0.005	(0.006)
Delayed operation, %	0.187***	(0.015)	0.053***	(0.016)	0.398***	(0.022)
Number of procedures	−0.002***	(0.000)	−0.001***	(0.000)	−0.003***	(0.000)
Panel B: Discharge cohorts						
Length of stay, days	−0.030***	(0.002)	0.000	(0.001)	−0.075***	(0.005)
Transfers out, %	0.000	(0.005)	−0.001	(0.003)	0.002	(0.011)
Discharges to home, %	0.003	(0.009)	−0.009	(0.007)	0.020	(0.015)

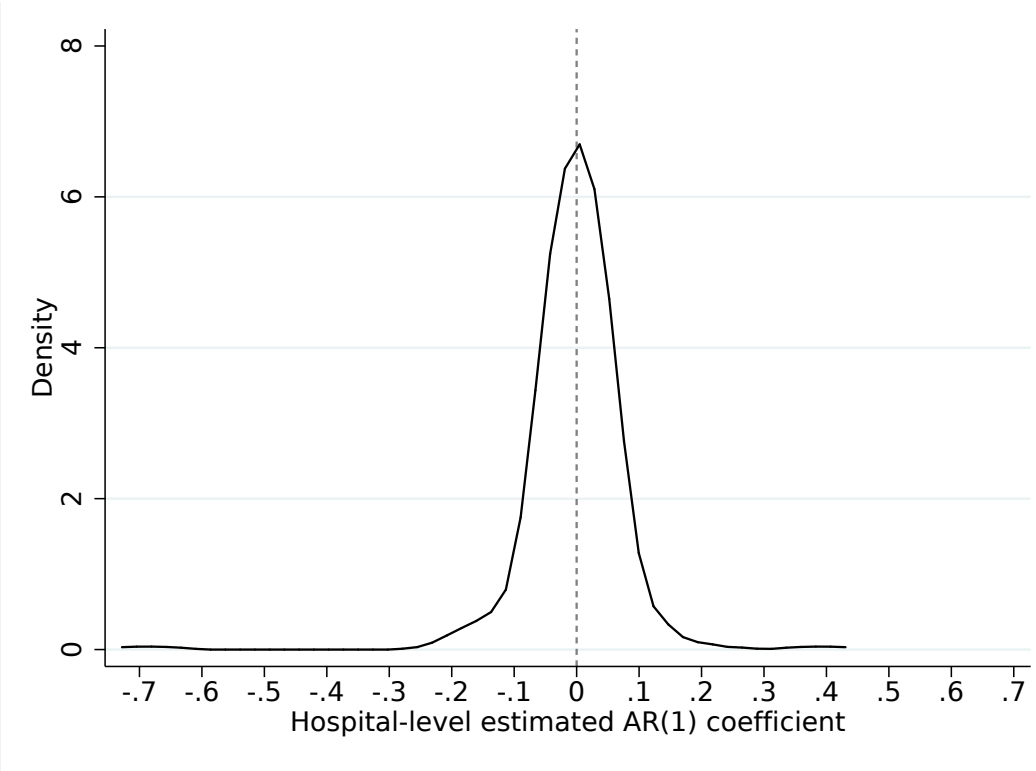
Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (1); (2) All specifications include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (3) N = 4,279,803 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.

Figure 1: 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.

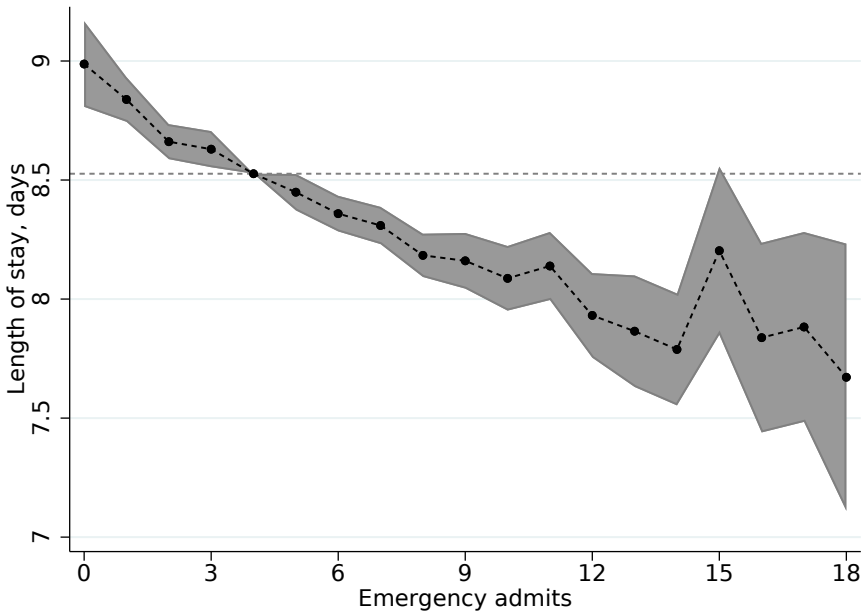
Figure 2: Hospital-level tests of first-order serial correlation



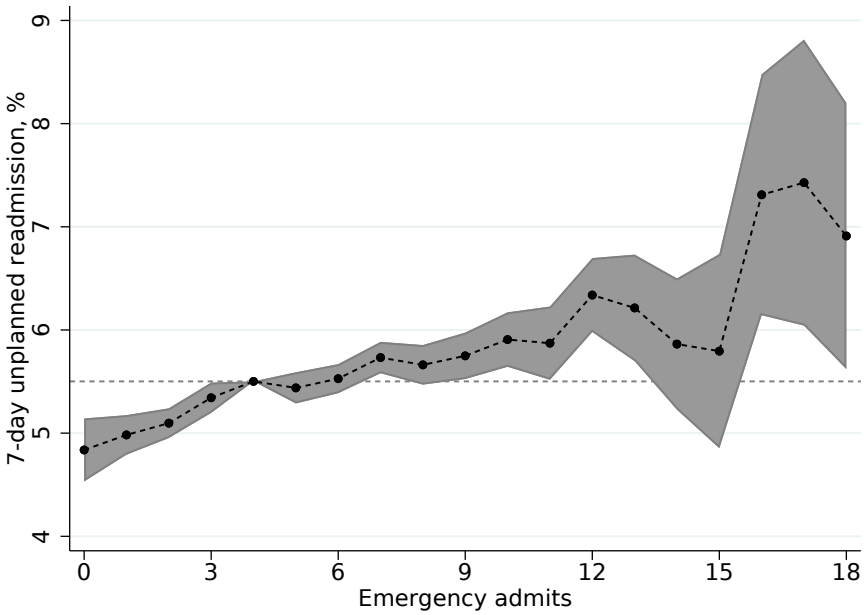
Notes: (1) Figure shows the density of estimated AR(1) coefficients from regressions of emergency shocks on their lag for each hospital separately; (2) Emergency shocks are defined as residuals from a regression of daily emergency admissions on hospital-specific year, weekly seasonal, and day-of-the-week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

Figure 3: 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

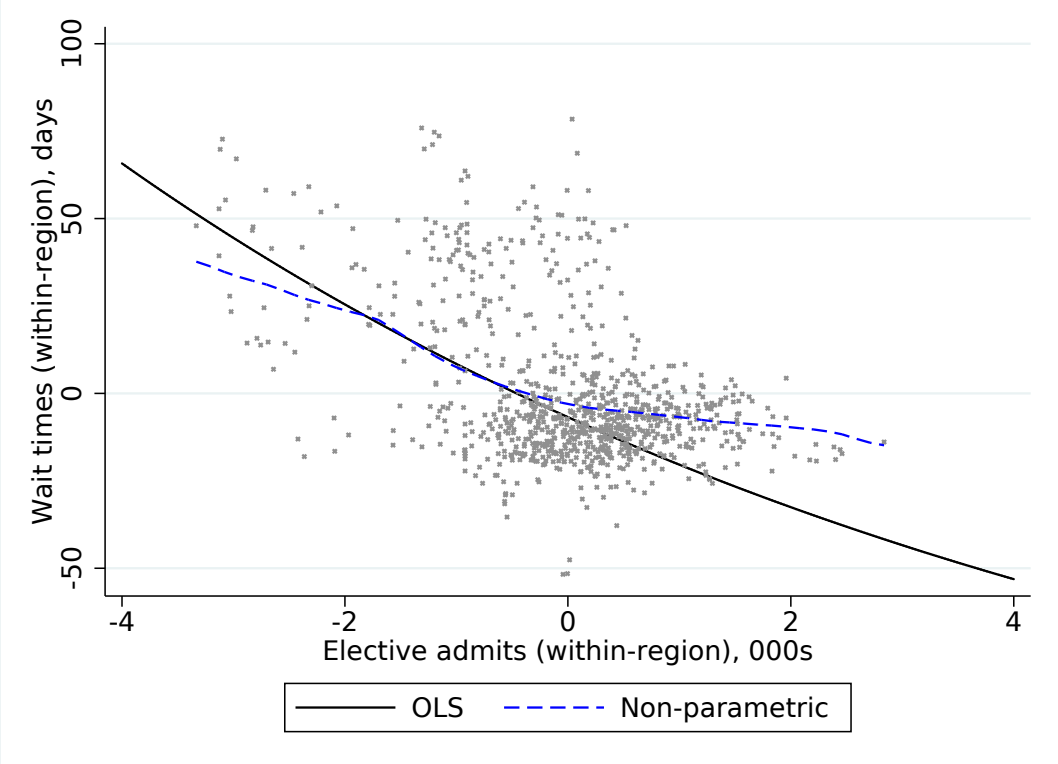


(b) 7-day unplanned readmission



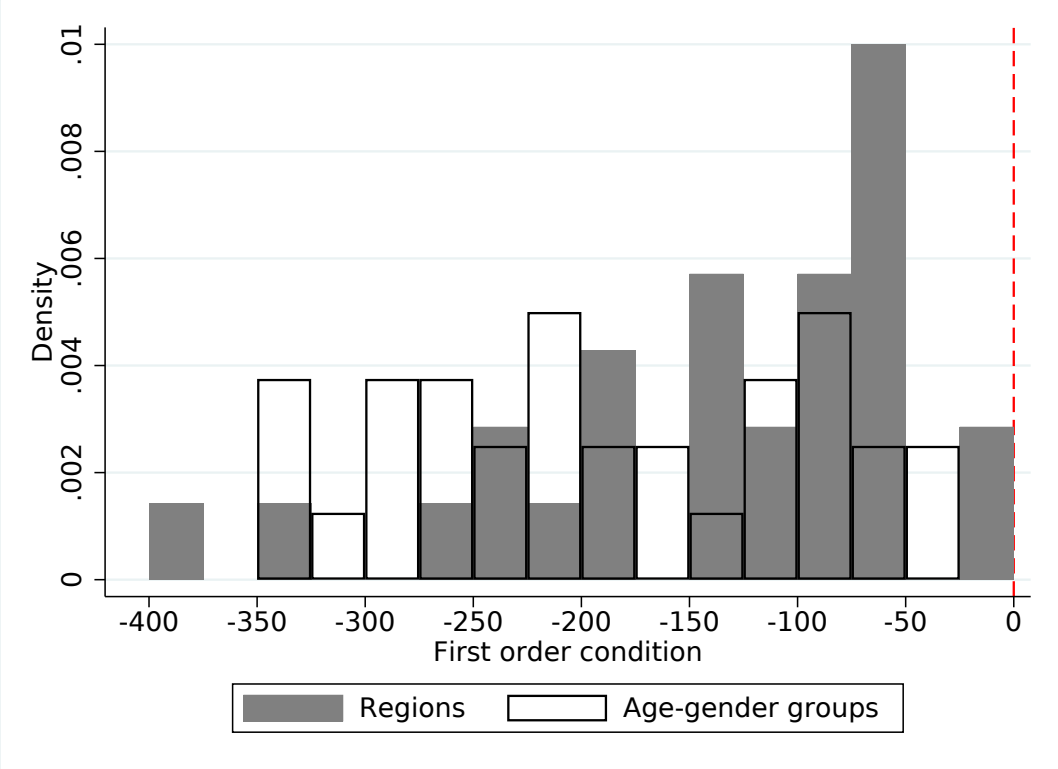
Notes: (1) Base category of 4 emergency admissions normalized to the unconditional mean of emergency admissions; (2) Estimates for values of emergency admissions above 15 omitted from the figure and are mostly statistically insignificant; (3) Standard errors clustered at the hospital-level (149 clusters)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

Figure 4: OLS and non-parametric estimates of the effect of elective admissions on wait times



Notes: (1) OLS estimates obtained from a regression of log wait times on elective admissions and regional fixed effects, with the estimates shown here adjusted via the constant term to be on the same visual scale; (2) The OLS parameter estimate on elective admissions is $-0.112(0.014)$ where standard errors are clustered at the regional level; (3) Non-parametric estimates a local polynomial regression of wait times on elective admissions, where I partial out the regional fixed effects from first variables in a first-stage regression.

Figure 5: Sufficient statistic tests by region and demographic group



Notes: (1) If the first-order condition is negative it indicates that reductions in hospital crowding would improve consumer welfare; (2) Regions and age-gender groups are unweighted; (3) The aggregate first-order condition estimate is -116.

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)

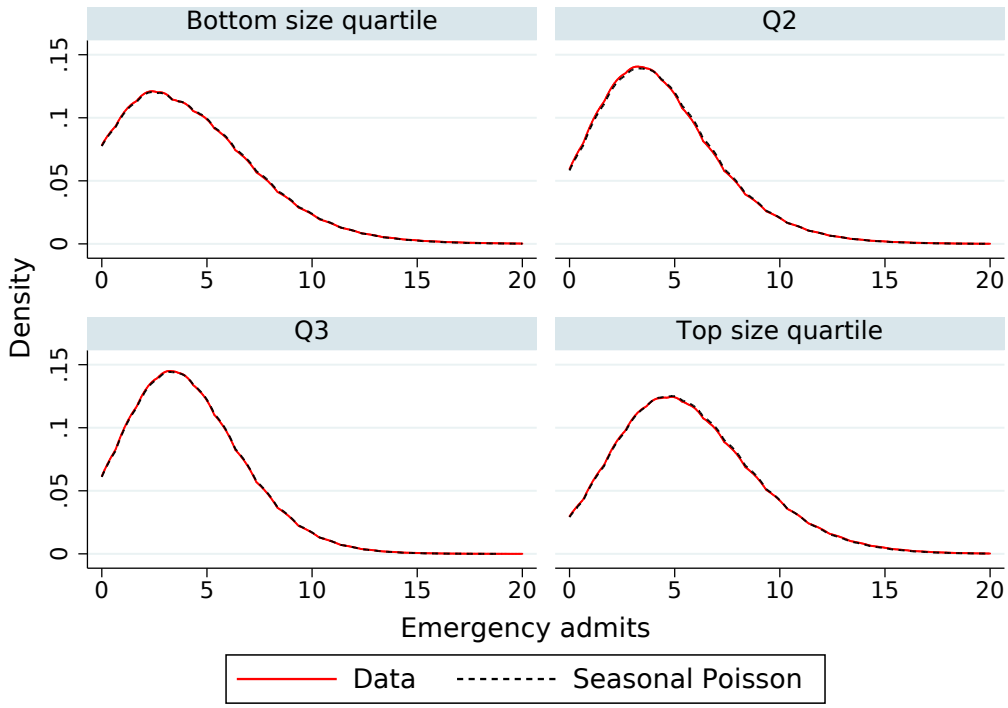
Notes: (1) ACE is the average crowding effect, an estimate of the β parameter in Equation (1); (2) All specifications include a fully interacted set of diagnosis, age category, and emergency status fixed effects, and hospital-specific year, weekly-seasonal, and day-of-week fixed effects; (3) N = 4,279,803 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 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 (1); (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.

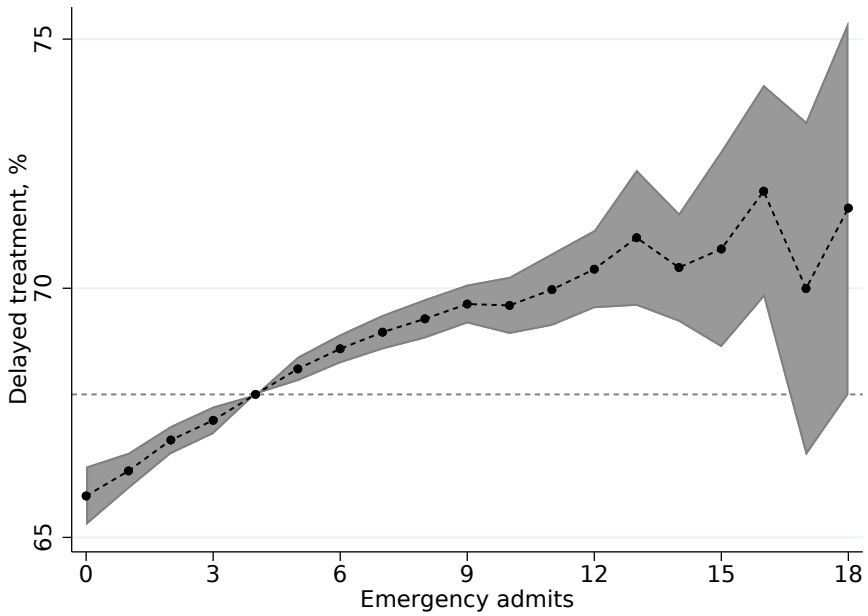
Figure A1: Poisson property of daily emergency admissions



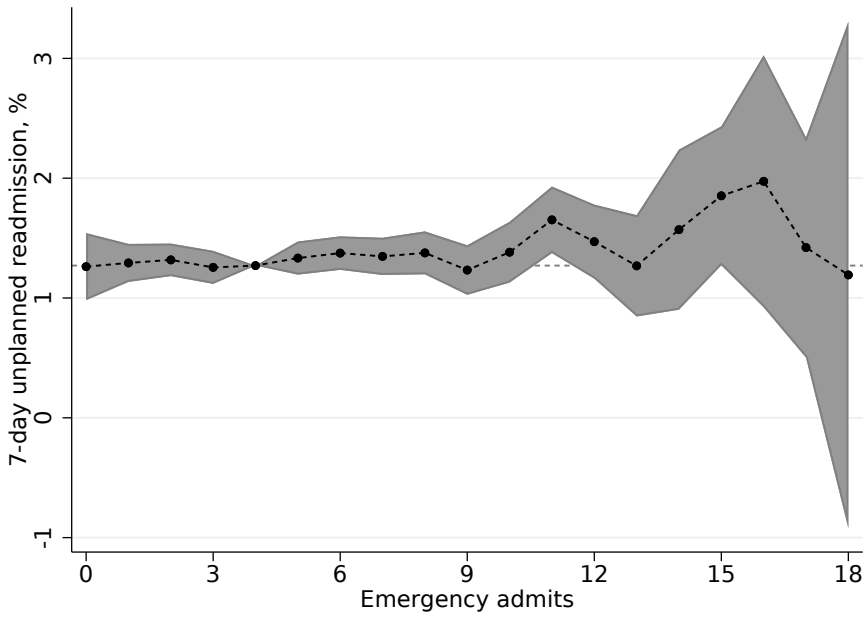
Notes: (1) Size quartiles are defined by the mean daily emergency admissions of each hospital; (2) Poisson data simulated for each hospital with a mean equal to expected emergency admissions, which is defined by a regression of emergency admissions on hospital-specific year, week-seasonal, and day-of-the-week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