Are Public Hospitals Overcrowded?

Evidence from Trauma and Orthopaedics in England

Thomas P. Hoe
University College London and Institute for Fiscal Studies*

September 15, 2017

Abstract

Hospitals face a trade-off between how many non-emergency patients to admit each period (‘crowding’) and how long these patients must wait for an appointment (‘waiting’). This paper evaluates whether there is too much crowding at public hospitals in England. I first exploit pseudo-random variation in emergency admissions to estimate the short-run effect of admissions on health outcomes. I find that crowding has a large adverse effect on patients, causing the rate of unplanned readmission to vary by up to 22%. The most plausible mechanism for this effect is that physicians routinely vary how soon they discharge patients. I then conduct a marginal welfare analysis that compares the crowding effect with a waiting time effect. I estimate the waiting time effect by exploiting within-region variation in non-emergency admissions caused by technological change. I compare the two effects using a model of consumer welfare and conservative benchmarks of preferences. I find that policies which reduce crowding and increase waiting times may lead to substantial net welfare gains.

*Contact email: thomas.hoe.12@ucl.ac.uk. I am extremely grateful to Richard Blundell for advice and suggestions, and to Orazio Attanasio, David Chan, David Cutler, Alejandro Estefan, Jonathan Gruber, Amanda Kowalski, Magne Mogstad, Lars Nesheim, David Oliver, Aureo de Paula, Imran Rasul, John Van Reenen, Marcos Vera-Hernández, and Heidi Williams for helpful discussions. I thank seminar participants at the Centre for Health Economics (Manchester), The Department of Health, EMCON 2016, NHS Improvement, RES 2016, TADC 2017 and University College London. I also thank several 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.
1 Introduction

Hospital capacity has fallen dramatically in recent years leading to widespread concerns about hospital crowding. Beds per capita in OECD countries fell by 13% between 2000 to 2013, with pronounced drops of over 30% in a number of countries (OECD, 2015). Physician groups have argued this has created crowding pressures and caused quality of care to deteriorate (British Medical Association 2017; Royal College of Physicians 2017). Yet despite these concerns, previous studies have not found a consistent association between measures of hospital crowding and health outcomes (Eriksson et al., 2017).

This raises challenging questions about whether and how to moderate crowding pressures. One option is simply for hospitals to admit fewer patients, which for non-emergency (‘elective’) care is often feasible. By admitting fewer elective patients crowding is reduced but these patients must then wait longer for a hospital appointment. This creates a trade-off between the quality of care (crowding) and access to care (waiting). In many countries this trade-off is already regulated in public hospitals through explicit waiting time targets (Viberg et al., 2013) and other policies that impose financial payments or penalties (e.g. prospective payment tariffs, readmission penalties). A question facing policymakers in these settings is how to make the appropriate trade-off between crowding pressures and waiting times.1

This paper examines this trade-off in the context of the one of the largest public healthcare systems, the English National Health Service. I first address the question of whether hospital crowding affects quality of care. This presents an endogeneity problem because the number of patients in the hospital (the level of ‘crowding’) is correlated with patient composition and hospital scheduling decisions. I deal with this by exploiting emergency admissions which I show contain pseudo-random variation that is uncorrelated with several key confounding factors. I find that crowding causes large adverse effects on patient health outcomes.

I then turn to the question of whether reductions in crowding would be desirable. I evaluate this by comparing the crowding effect with the effect of elective admissions on waiting times in a marginal welfare analysis. I estimate the waiting time effect by exploiting variation in admissions induced by technology change. I then compare this estimate with the crowding effects in the context of a simple model of consumer welfare. Relative to conservative benchmarks, I find that healthcare policies in England overvalue preferences for waiting times substantially. This

---

1Another policy question is whether to change hospital capacity. Yet conditional on capacity, one may still want to assess the trade-off between crowding and waiting times.
suggests that hospitals are overcrowded and that policies which reduce crowding and increase waiting times would lead to net welfare gains.

Throughout the paper I use a unique administrative dataset that records the universe of medical records for all publicly funded hospitals in England. The data tracks patients throughout the secondary healthcare system and individuals can be linked across hospitals and over time. For each hospital visit, I observe extensive information about a patient’s health conditions and treatments received. This permits a detailed study of admissions, waiting times, and many other dimensions of hospital care. I focus on trauma and orthopaedic departments over the period 2006 to 2013. These are the third largest inpatient department in England and treat diseases and injuries of the musculoskeletal system such as arthritis and broken bones. The setting is well suited to the study, with emergency trauma admissions occurring frequently and unplanned readmission events, a relevant health outcome for these patients, observed in the data.

I begin by studying daily emergency admissions at each hospital. I decompose admissions into a systematic component, based on the seasonal and within-week pattern, and a shock component. This is intended to separate out the non-random variation in emergency admissions that may be correlated with patient composition and hospital scheduling decisions. I present evidence consistent with this notion, showing that on days with more predicted admissions, hospitals schedule more physicians and, conditional on physician presence, fewer elective appointments. I then examine the statistical properties of the ‘emergency shocks’. I show that the shocks are uncorrelated across time and with key confounding factors. These pseudo-random shocks therefore provide a rich source of variation to assess the effects of crowding.

I find that emergency shocks lead to large variations in health outcomes. A one standard deviation emergency shock (approximately 2 admissions) increases the likelihood of unplanned readmission by 4.2% per cent among discharged patients. I show that these effects are approximately linear over the shock distribution, and shocks cause the rate of readmission to vary by up to 22%.

The rich heterogeneity in the data allows me to explore several mechanisms that could drive these health impacts. I find that bed constraints and physicians varying how soon they discharge patients is the most plausible cause of readmission effects. This is evident in correlations between the effect of shocks on length of stay and readmission: bigger reductions in length of stay are associated with bigger increases in readmissions. This correlation holds across subgroups of
patients and the cross-section of hospitals. This mechanism suggests that hospitals do not reserve sufficient bed capacity to absorb even small variations in emergency admissions.

I can also rule out changes in patient composition as a potential explanation. The estimates are robust to the inclusion of a high-dimensional set of fixed effects and I show that hospital admission decisions do not create selection among the inflows of elective or emergency patients. While hospitals can insure against the shocks by cancelling elective appointments, this occurs only in response to extreme positive shocks and I show it does not induce selection along a range of observable characteristics. Hospitals can also divert ambulances, deny incoming emergency admissions, or transfer existing patients to other hospitals. But I find no evidence that any of these responses are used to manage emergency shocks. I do find that shocks cause delays - both in the emergency department and prior to inpatient surgery - but these effects are not associated with the impacts on health outcomes.

These results show that hospital crowding has adverse effects on health outcomes. The question for policymakers is whether reducing crowding will improve consumer welfare. I focus on policy options that maintain hospital capacity but influence the number of elective appointments, so that the same capacity is simply utilised differently. A salient example is a policy that imposes an upper bound on how long patients wait for elective appointments. By relaxing such policies, policymakers can moderate hospital crowding at the expense of increasing waiting times. I evaluate this trade-off by estimating the impact of elective admissions on waiting times and then comparing this with the crowding effects in a marginal welfare analysis.

To estimate the impact of elective admissions on waiting times I exploit within-region variation in technological change. During the sample period hospitals gradually adopted ‘fast track surgery’ - an innovation in post-operative recovery procedures for elective patients - that led to major reductions in length of stay without impairing health outcomes (Kehlet, 2013). The shorter hospital stays led to large increases in elective admissions. I therefore use length of stay as an instrument for elective admissions in a regression on elective waiting times. This instrument is plausibly exogenous to other factors that may affect waiting times such as use of the fringe private market. I find that an annual increase of 1,000 admissions in a region reduces average waiting times by approximately one week. Comparing this to the crowding estimates, after appropriate rescaling, indicates that a marginal change in elective admissions will reduce readmissions by 1.2% at the cost of increasing waiting times by 6.2%. This naturally raises the question of whether this trade-off leads to net welfare gains.
I use a simple model of consumer preferences and hospital technology to derive an optimality condition that characterises the welfare maximising level of elective admissions. This states that the ratio of the marginal effect of admissions on waiting times and crowding should be proportional to consumer preferences over these outcomes. In conjunction with my reduced-form estimates I use this condition to estimate the revealed preferences are implicit in healthcare policies in England. This provides a basis to assess whether hospitals are overcrowded: if the revealed preferences overvalue waiting times then reductions in elective admissions will be welfare improving.

I compare the revealed preference estimates and to conservative benchmarks of true preferences. I find that policies in England overvalue preferences for waiting times by at least four times. This creates incentives for hospitals to admit high volumes of elective patients, which keeps waiting times low but hospital crowding high. The overall implication is that there may be substantial welfare gains from policies that reduce elective admissions. This will increase waiting times and lower unplanned readmissions but the net effect on consumer welfare is predicted to be positive. Furthermore, by repeating the analysis at the regional level I show there are large disparities across the country and this suggests that reducing hospital crowding may be best pursued locally.

This paper contributes to two literatures that together shed light on how demand variation affects hospitals and patients. The first literature focused primarily on hospital costs. Friedman and Pauly (1981) highlighted that demand volatility imposed additional costs on hospitals because of the need to insure against quality deteriorations. They argued that reserving bed capacity creates this insurance. A series of papers have estimated the cost of this insurance, often referred to as the 'cost of an empty bed'. Examples include Friedman and Pauly (1983), Gaynor and Anderson (1995), Keeler and Ying (1996), and Hughes and McGuire (2003).

The second literature examines how demand variation affects patients. This has a long history in medical research where studies typically focus on the association between hospital occupancy or similar measures and health outcomes. This has produced mixed findings (see Eriksson et al. (2017) for a recent review). In contrast the economics literature has emphasised causality but has tended to focus on the impact of demand variation on the likelihood of hospital
In early work, Joskow (1980) examined the likelihood of being denied hospital admission, while Windmeijer et al. (2005) found evidence of substitution between elective and emergency admissions. Two recent studies are closest to my work: Fiedler (2016) and Freedman (2016). These papers study intensive care units in different settings in the U.S. and examine how the likelihood of admission to these units is affected by shocks to daily occupancy. Both papers find that occupancy has a negative impact on the likelihood of admissions and that the impact is greater for healthier patients. Freedman (2016) argues that capacity expansions may therefore lead to additional and more marginal admissions.

I make the following contributions to these literatures. First, I show how hospital crowding affects patient health outcomes using a new source of empirical variation: emergency admissions. These admissions offer pseudo-random variation that is plausibly exogenous. This has clear advantages over measures of demand variation that include elective admissions or predictable variations in emergency admissions. Second, using the rich administrative data in my setting I explore a range of mechanisms related to the crowding effects. I find that bed constraints and physicians varying their discharge decisions is the most plausible cause of the health impacts. This is consistent with hospitals in England adopting little or no insurance against demand volatility. Third, I develop a framework for evaluating whether there would be consumer welfare gains from reducing hospital crowding. I apply this in my setting and find that policies which reducing crowding may lead to substantial welfare gains.

The paper proceeds as follows. Section 2 provides information about hospital inpatient departments and the institutional setting. Section 3 describes the data. Section 4 sets out the empirical analysis of crowding. Section 5 sets out the marginal welfare analysis including the analysis of waiting times. Section 7 concludes.

2 Background

2.1 Hospital inpatient departments

Inpatient departments are where the majority of 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
Inpatient departments account for a large part of physical hospital capacity as many patients require accommodation for overnight stays.

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 waiting 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 additional treatment, but those that do require treatment are admitted to an appropriate inpatient department.

Upon admission, both elective and emergency patients experience a similar overall pathway: a surgical or medical procedure is provided on or shortly after admission, after which they are monitored and nursed through the recovery process until they are considered fit for discharge. The specifics of a pathway will vary according to the diagnosis. In the case of high volume elective surgeries, such as a total hip replacement procedure, the pathway can be very standardised. These patients will often have set goals for each day of their stay and will be discharged as soon as they can navigate a flight of stairs unaided. In contrast, emergency patients with more complex and varied health conditions require a more flexible pathway. Examples include patients with multiple or very severe injuries, who will be assessed on a day-by-day basis according to their needs.

Hospitals have a degree of control over 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 cancelled or rescheduled at short notice. There is far less control over the inflow of emergency patients. For urgent and severe cases there is often no option but to accept patients. For less urgent or severe cases, there is potentially more control, as hospitals can divert ambulances to alternative hospitals or adjust the threshold for inpatient admissions from the ED, although these responses have potential for adverse effects on patients.
The outflow of elective and emergency patients is controlled by discharge decisions. Patients are evaluated daily and discharged once they are medically fit and able to leave the hospital. Upon discharge they may either return to their home residence or be transferred to another hospital or an alternative care facility.

Decisions over patient flow are made by a combination of physicians and managers. Physicians are responsible for all decisions about individual patients. This includes whether to admit (following an outpatient consultation or an ED visit), all treatment decisions, and when a patient is fit for discharge. Managers are responsible for operational decisions such as when to cancel elective appointments or divert ambulances. Patient flow is monitored closely throughout each day and managers will communicate information to physicians through meetings and via electronic means.

2.2 Institutional setting

The empirical application focuses on public hospitals in the English National Health Service. This is a single-payer healthcare system funded through the proceeds of general taxation. All approved hospital treatments are provided to residents for free. Public hospitals provide the large majority of elective inpatient care and all emergency care in England. These hospitals are centrally managed and regulated by a number of government departments. Policies are set that specify targets and incentives for hospitals to operate by and apply to financial, operational, and clinical performance. The majority of policies are set at the national level and apply equally to all public hospitals.

The sample covers the period 2006 to 2013. During this period two policies had a major influence on the incentives of hospitals to admit elective admissions. The first policy is the ‘Referral to Treatment’ waiting time target which specified the maximum time between a referral and admission for all elective patients. The target was introduced in 2006, setting a maximum of six months, and then tightened to three months in 2008. The target was strongly enforced through senior management incentives (Propper et al., 2008) and financial penalties ($390 per patient that waits above the threshold). It proved very effective: average waiting times fell by over 50% between 2000 and 2010 for trauma and orthopaedic elective patients.

The second policy is the ‘Payment by Results’ tariff that determines hospital payments for treating elective payments. Hospitals were paid on the basis of a prospective payment system.

4The exception to this is prescription drugs which are subject to a small co-payment.
(PPS) which, similar to the DRG system in the US, specified fixed payments per admission of each diagnosis type (Department of Health, 2012). The tariff, which was implemented for most hospitals in 2006, created a financial incentive for hospitals to treat higher volumes of elective patients.\(^5\) In many cases this incentive mattered because public hospitals have a financial target to break even and many hospitals were running a financial deficit. For example, by the end of 2005, hospitals were on average running a financial deficit of 2.5% of total revenue (with 10-90th percentile range of -8.5% to 0.5%). This improved marginally over the sample period yet by 2013 hospitals were still running an average financial deficit of 1.6% of total revenue (with 10-90th percentile range of -7.0% to 1.0%).\(^6\)

There were also a number of other policy changes that were implemented around the beginning of 2006. This includes: the removal of restrictions on hospital choice (Cooper et al. 2011, Gaynor et al. 2013); increased monitoring of clinical performance, especially for elective patients (NHS Digital, 2017); and, capacity expansions achieved by enabling private hospitals to conduct publicly-funded elective care (Kelly and Stoye, 2015).

3 Data

I use data on medical records for inpatient and ED visits from the Hospital Episodes 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.\(^7\)

The empirical application focuses on trauma and orthopaedic departments at general acute hospitals with an active ED in England. These departments treat musculoskeletal conditions such as broken bones and arthritis and are the third largest department measured by admissions

---

\(^5\)The tariff was implemented on a limited scale between 2003 and 2005, covering a small number of hospitals and a limited range of activities. It was rolled out to all hospitals and all activity over the period 2006 to 2008.

\(^6\)Financial data obtained from the annual accounts of NHS Trusts. Data for 2005 and 2013 was available for 127 and 115 of the 138 NHS Trusts present in the datasets described in Section 3. Financial figures exclude any financial support.

\(^7\)The dates refer to financial years beginning in April and ending in March the following year. This convention is used throughout the paper.
(6.6% in 2013). The trauma and orthopaedic setting is well suited to the analysis: they are strongly influenced by the policy pressures discussed above, relevant outcome measures can be constructed from the data, and emergency admissions, which I use as a source of identification for the analysis of crowding, are common in these departments.

3.1 Sample construction

I construct three data samples for the analysis. In each sample I identify general acute hospitals using the Estates Return Information Collection data and define hospitals by their postcode, which references a specific geographic location. I define an ED as active in a year if the trauma and orthopaedic department received on average five emergency admissions per week in each quarter of the year. Trauma and orthopaedic patients are identified by the medical specialty that they are treated under.

The first sample, referred to as the panel dataset, contains hospital-day level information on the number of elective and emergency admissions to trauma and orthopaedic departments. I exclude hospital-years with incomplete information, which can occur if the specific hospital site is not recorded accurately. After making these exclusions I further exclude departments with fewer than three years of data. These exclusions, which together account for 24% of hospital-days, ensure that each department has a reliable and reasonably long time-series of data which is important when I decompose emergency admissions. The qualitative results are robust to changes in these exclusion rules.

The second sample, referred to as the inpatient dataset, contains medical records for patients admitted to trauma and orthopaedic departments. I limit this dataset to patients admitted and discharged on days contained in the panel dataset. I construct the following variables: an indicator for whether the primary operation received involves no overnight stay for the median patient (‘daycase operation’); an indicator for whether surgery occurred after the day of admission (‘delayed operation’); a count of the number of medical procedures received; length of stay; an indicator for discharge to another hospital (‘transfers out’); an indicator for discharge

---

8 I define a hospital-year as incomplete using two rules: (i) the data contains fewer than 51 weeks in a year; (ii) the data contains a week where emergency and elective admissions are both at least 80% below the annual average. After these exclusions I also remove four weeks either side of any data break to ensure that the data is not missing information from adjacent but excluded periods.

9 The majority of results are also robust to removing all of the exclusion rules. One exception to this is the analysis of elective admissions. Periods of incomplete data, which cause elective and emergency admissions to move together simultaneously, unduly influences this analysis by forcing a positive correlation between the two types of admission.
to the patient’s home residence (‘home discharge’); a count of the number of diagnoses recorded; the Charlson co-morbidity index (a proxy for the severity of underlying health conditions); the count of ED admissions in the past year (another proxy for underlying health conditions); 7-day unplanned readmission; and, 30-day in-hospital mortality. Unplanned readmissions are defined as any emergency inpatient admission to any hospital within a specified time horizon from the previous discharge. I use a 7-day horizon in the baseline analysis and conduct robustness tests using other horizons.

The third sample, referred to as the ED dataset, contains medical records for all visits to emergency departments. I match this data to the inpatient dataset which later allows me to evaluate how inpatient crowding affects ED outcomes. This matching process is incomplete because the ED data does not always contain information on the specific hospital site. I am able to match 64% of hospitals in the ED dataset to the inpatient dataset. I also exclude patients that visit the same ED multiple times on the same day as these patients cannot be matched uniquely to the inpatient data (2.5% of visits) and limit the data for matched hospitals to the days present in the panel dataset. I compute three variables using the ED data: an indicator for whether a patient attended their nearest hospital; the (straight-line) distance travelled to hospital; and, an indicator for whether admission was to the trauma and orthopaedic department (rather than another inpatient department).

Together these three samples provide information on 158 trauma and orthopaedic departments, 4.0 million inpatient visits (2006-2013), 101 emergency departments, and 22.2 million ED visits (2010-2013). Appendix A presents basic summary statistics for each sample.

3.2 Descriptive statistics

3.2.1 Characteristics of trauma and orthopaedic patients

Table 1 presents the mean characteristics for trauma and orthopaedic patients and compares them to the patients in other specialties at general acute hospitals (excluding maternity and paediatric care). Trauma and orthopaedic patients are similar along several dimensions to patients admitted to other specialties. The demographic mix is comparable in terms of age, gender and ethnicity, but trauma and orthopaedic patients wait significantly longer for elective care, and are healthier in terms of pre-existing conditions (diagnoses, co-morbidities, past ED admissions) and health outcomes (likelihood of unplanned readmission, in-hospital mortality). Trauma and orthopaedic patients are on average 53 years old, with an even gender balance, and
patients are predominantly white.

### 3.2.2 Differences between elective and emergency patients

There is substantial heterogeneity between different trauma and orthopaedic patients. The differences are along two principle dimensions: elective and emergency cases, and the specific diagnosis. Patient demographics and hospital treatment vary significantly across these two dimensions.

Figure 1 illustrates the heterogeneity between patients in the data for 2010. It shows the most common diagnosis groups for elective and emergency patients, plotted by average length of stay and age with the size of the marker indicating the number of patients. There are two notable differences between elective and emergency patients. First, they are located in different regions of the length of stay and age space. Elective patients typically have shorter stays and are on average between 45 and 70 years old. The most common elective diagnosis is arthrosis (commonly known as osteoarthritis) and the majority of these patients will require a hip or knee replacement. In contrast, emergency patients stay longer and the age distribution is bimodal: there is a group with an average age of around 40 presenting with broken arms and legs, and a group with an average age of around 80 presenting with broken hips. These emergency patients will often receive an ‘open reduction and internal fixation’ which involves open surgery and the use of metal plates or screws to realign and secure a broken bone. The second difference between elective and emergency patients is the degree of heterogeneity among diagnoses: there is significantly more heterogeneity for emergency patients. The figure contains equal volumes of elective and emergency patients, and the elective patients are concentrated within 8 diagnosis groups while the emergency patients are spread over more than 30 groups.

In some cases elective and emergency patients have a similar or the same diagnosis. Even here there are strong differences between the two patient types. In Appendix Table A4 I show that hospital stays for emergency patients are on average 94% longer than elective patients and after controlling for the observable characteristics of patients, including the specific diagnosis, this difference is still as large as 47%.

### 3.2.3 Health outcomes

I use unplanned readmission as my primary measure of health outcomes. This is widely used in academic studies and by healthcare regulators including NHS Improvement in England and
the CMS in the US. Unplanned readmission is also used specifically in relation to trauma and orthopaedic patients; for example, in quality monitoring surveys (NHS Digital, 2017), by the CMS to set readmission penalties, and in medical research to evaluate trauma and orthopaedic surgery (Kehlet, 2013).

Common diagnoses among readmitted trauma and orthopaedic patients include complications with internal devices (e.g. mechanical components of a hip replacement), infections, inflammation, and bleeding (see Appendix Table A5). The average length of stay for a readmission is between 2 and 20 days depending on the diagnosis. The average length of stay across all readmissions is 7.3 days, which is approximately equal to the length of stay in the index admission of readmitted patients.

Alongside readmission, I report results using 30-day in-hospital mortality. Relative to readmission this outcome has two drawbacks: first, it is a very extreme outcome that does not occur very often in the sample; and second, I only observe mortality that occurs within the hospital, which makes this outcome conditional on other events such as admission and length of stay. Across trauma and orthopaedic patients, the 7-day unplanned readmission rate is 2.8% and the 30-day in-hospital mortality rate is 1.1%.

4 The impact of hospital crowding on patients

I now turn to the question of how crowding affects patients, and define crowding as the number of patients admitted to hospital. The empirical challenge when trying to estimate the effects of crowding is that admissions are endogenous. This is because admissions will typically be correlated with factors such as patient composition and inputs to hospital production. These correlations can be caused by seasonal fluctuations, such that the type and volume of patients presenting varies naturally, and hospital scheduling, where resources and workload are organised to match peaks and troughs in admissions. The analysis seeks to separate the effects of crowding from changes in these other factors. I do this by exploiting variation in emergency admissions. I show that these admissions can be decomposed to produce ‘emergency shocks’ which are a pseudo-random source of variation in admissions that is plausibly exogenous.
4.1 Pseudo-random variation in emergency admissions

I decompose the time-series of emergency admissions for each hospital into a seasonal component and a random shock component. The idea is that while the seasonal component of admissions may be correlated with patient composition and hospital scheduling decisions, the remaining shock component may be exogenous to these factors. I validate the decomposition by examining whether the properties of each component are in line with this notion. This shows that hospitals do schedule their resources and workload around the seasonal component, consistent with this being part of their information set, but that the shock component is pseudo-random and uncorrelated with several key factors.

I decompose emergency admissions by using the panel dataset and estimating the following regression

\[ q_{hs} = \lambda_{hy} + \phi_{hw} + \pi_{hd} + z_{hs}, \]  

where \( q_{hs} \) is the number of emergency admissions at hospital \( h \) on day \( s \), \( \lambda_{hy} \), \( \phi_{hw} \) and \( \pi_{hd} \) are hospital-specific year, weekly-seasonal, and day-of-week fixed effects, and \( z_{hs} \) is the ‘emergency shock’. I use estimates of this regression to predict emergency admissions and the shock to emergency admissions for each hospital-day in the sample.

Figure 2 provides an example of the decomposition. Panel (a) shows the time-series of daily emergency admissions for one hospital in one year. Average daily admissions are around five but the time-series exhibits significant variation with low admission days and high admission days often in close succession. Panel (b) superimposes the predicted emergency admissions over the observed admissions. The seasonal pattern is slightly higher in summer and lower in winter, and the minor variations are due to differences 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. The difference between the predicted admissions and observed admissions is the shock component.

4.1.1 Predicted emergency admissions

I first examine the predicted admissions. If these are known to the hospital then it should be apparent that hospitals respond to these predictions. Figure 3 shows a simple comparison of the seasonality in emergency and elective admissions, which provides suggestive evidence of these responses. Elective admissions higher in periods when emergency admissions are lower
(early and late in the year) with exceptions around holiday periods (April near Easter, August in the vacation season, Christmas and New Year) when there are sharp reductions in elective admissions but little change in emergency admissions.

To probe this relationship further, I conduct two tests. The first test compares how the number of physicians at a hospital varies with predicted emergency admissions. The second test compares how the number of elective admissions, conditional on the number of physicians, varies with predicted emergency admissions. If hospitals are aware of the pattern in predicted emergency admissions, then this should show up as a positive correlation with the number of physicians working (more physicians are scheduled when it is expected to be busier) and a negative correlation with the number of elective appointments (fewer elective patients are admitted to moderate overall admissions). Note that it is important to control for the number of physicians in the second test, since hospitals may schedule fewer elective appointments in periods when there is less staff availability and this may correlate with emergency admissions (e.g. during holiday periods).

Table 2 presents the results of these tests. Column (1) contains the first test. Column (2) and (3) contain the second test with and without the control for the number of physicians. Each specification contains hospital fixed effects. The results in column (1) are as expected: more physicians are present on days with higher predicted levels of emergency admissions. Each additional predicted admission is associated with 0.64 additional senior physicians. This specification also shows that more physicians are also present on days when the shock component of emergency admissions is higher although the effect is smaller: one additional emergency patient shock is associated with 0.14 additional physicians. Column (2) shows that elective admissions are positively associated with predicted emergency admissions. Column (3) shows, however, that once the number of physicians has been controlled for, elective admissions are negatively associated with predicted emergency admissions: each additional predicted admission is associated with 0.26 fewer elective admissions. These tests indicate that hospitals are at least partially aware of the seasonal pattern in emergency admissions and plan the scheduling of physicians and elective admissions around expected admissions.

4.1.2 Emergency shocks

I now examine the emergency shocks. Figure 4 shows the distribution of emergency shocks across hospital-days. The distribution is centred around zero and approximately normally dis-
tributed, with a standard deviation of 2.2. To study the time-series properties of these shocks I estimate the serial correlation in the emergency shocks for each hospital by regressing the shocks on their lag. Figure 5 shows a density plot of the estimated AR(1) coefficients obtained from this process. The coefficients are centred around zero indicating that there is no serial correlation for the majority of hospitals. The estimated coefficients are statistically insignificant in 128 of 158 cases and the average estimate for the statistically significant cases is -0.03. Repeating the same analysis using squared values of the shocks produces similar results. Together these results suggest that the shocks are pseudo-random: the shock today is uncorrelated with the shock yesterday and the shock tomorrow.

Two implications follow from the pseudo-random property. First, hospitals are unable to forecast emergency shocks even with short-term information about recent shock realisations. The best forecast is the seasonal expectation based on previous years of data. The inability to forecast shocks prevents hospitals from scheduling different levels of resourcing on high and low shock days. While this does not rule out hospitals making adjustments at short notice, the evidence in Table 2 suggests that these effects are likely to be small for labour inputs. For other resources, such as bed capacity, it is likely that no short term adjustments can be made in response to shocks. The implication of this is that hospital resourcing will be similar on high or low shock days.

The second property is that the composition of existing patients at the hospital when the shock occurs will be uncorrelated with the magnitude of the shock. This follows because the lack of serial correlation means that, even if shocks are correlated with patient composition, the current shock will be uncorrelated with previous arrivals. I illustrate this in Appendix Figure B1, which shows that the predicted mortality of patients at the hospital each day is largely uncorrelated with the magnitude of the emergency shock.10

It is also important to assess how shocks in trauma and orthopaedics relate to admissions in other departments. This could be relevant if shocks across departments are correlated and there are spillovers from joint production inputs. I assess this by regressing emergency shocks at trauma and orthopaedic departments on admissions at each of the eight largest other inpatient departments. The coefficient estimates in this regression are very small in magnitude and in almost all cases are statistically insignificant at the 1% level (see Appendix Table B1). These results indicate that even if there are spillovers between departments they are uncorrelated with

10There is a statistically significant negative correlation although the magnitude of the correlation is negligible.
the emergency shocks and trauma and orthopaedic can be suitably analysed in isolation.

Overall these results indicate that emergency shocks are highly suitable for the empirical analysis: they occur pseudo-randomly, cause changes in crowding, and are uncorrelated with several key factors.

4.2 Empirical specifications

I use the following baseline specification

\[ q_{hs} = \lambda_{hy} + \phi_{hw} + \pi_{hd} + z_{hs} \] (2)

\[ y_{iht} = \alpha_{ade} + \beta \tilde{z}_{hs} + u_{iht}, \] (3)

where \( y_{iht} \) is an outcome for patient \( i \) in cohort \( t \) (defined below) at hospital \( h \), \( \alpha_{ade} \) represent fully interacted age category, diagnosis, and emergency status fixed effects (over 55,000 categories), and \( \tilde{z}_{hs} = z_{hs}/\sigma_h \) is the standardised emergency shock. I standardised the shocks to allow for a consistent interpretation across hospitals.

Equation (2) is the same specification used in Section 4.1 and defines emergency shocks as the residual from the annual, weekly-seasonal and within-week average of emergency admissions. Equation (3) relates the shocks to the outcomes of interest \( y_{iht} \). In this specification the relationship is approximated with a linear functional form and the parameter of interest is \( \beta \). An alternative interpretation of \( \beta \) is as a weighted average of treatment effects (Angrist and Krueger, 1999).

I also use a step-function specification that replaces Equation (3) with the following

\[ y_{iht} = \alpha_{ade} + \sum_{p=-3, p \neq 0}^{3} \gamma_p \mathbb{1}\{\tau_{p-1} < \tilde{z}_{hs} \leq \tau_p\} + u_{iht}, \] (4)

where the \( \tau_p \) parameters define a step-function. I set \( \tau_{-3} = -\infty \), \( \tau_3 = \infty \) and \( \tau_{p \neq \{-3,3\}} = p + 0.5 \) such that the steps group together shocks located around standard deviation multiples. This specification relaxes the linear functional form in the baseline specification and allows me to examine the impact of shocks across the shock distribution. The parameters of interest are \( \gamma_p \), which give the impact of shocks in a particular magnitude range on outcomes relative to the omitted category (near-zero shocks).

I use these specifications to examine four groups of outcomes: health outcomes; inflows of
emergency patients; inflows of elective patients; and treatment and discharge decisions in the inpatient department. I am primarily interested in the impact on health outcomes and the remaining outcomes help to understand the mechanisms behind the health impacts. Depending on the particular outcome I specify s and t accordingly, and examine either admission cohorts \( (t = s + 1, \text{where } s \text{ is the admission date}) \) or discharge cohorts \( (t = s, \text{where } s \text{ is the discharge date}) \).

### 4.2.1 Identification

The identifying assumption in the outcome equation is that \( \tilde{z}_{hs} \) is exogenous such that \( E[\tilde{z}_{hs}u_{ihlt}] = 0 \). This states that the emergency shock is uncorrelated with other period-specific shocks. Potential threats to this assumption include changes in patient composition, labour and capital inputs, and emergency shocks to other inpatient departments. These concerns are largely ruled out by the pseudo-random property of the shocks (see Section 4.1.2).

Despite the shocks being pseudo-random, however, patients arriving on high shock days may still differ to those arriving on low shock days. This has potential to cause bias when analysing discharge cohorts because some patients that arrive (affecting \( \tilde{z}_{ht} \)) are discharged on the same day (affecting \( y_{ihlt} \)). If the magnitude of the shock is correlated with the type of arrivals, then changes in patient composition may directly affect patient outcomes through these ‘daycase’ patients.\(^{11}\) The role of the age-diagnosis-emergency fixed effects \( (\alpha_{ade}) \) is to partial out any correlation introduced by these patients. These variables allow for over 55,000 categories of patient and include five digit ICD-10 diagnosis codes, which record the exact location of the injury (e.g. fracture of lower end of tibia) and severity (e.g. whether the wound is open or closed). The identifying assumption for these cohorts is that any variation in patient composition within an age-diagnosis-emergency category is uncorrelated with the magnitude of the emergency shock.\(^{12}\)

### 4.2.2 Estimation

I estimate Equation (2) using OLS and Equations (3) and (4) using an algorithm developed and implemented by Correia (2016). The algorithm delivers the same point estimates as OLS but

\(^{11}\)This issue does not arise with admission cohorts because there is a clear separation in the timing between the shock (patients arriving at \( t - 1 \)) and the outcomes (patients arriving at \( t \)) and the shocks are serially uncorrelated.

\(^{12}\)An alternative approach is to exclude the daycase patients from the analysis. However, this introduces selection concerns because hospitals may respond to emergency shocks by discharging patients as a daycase when they would otherwise have stayed overnight. Despite these concerns, the qualitative features of the results are robust to taking this approach.
avoids the computationally intensive task of inverting a high-dimensional matrix. To account for the generated regressor in the outcome equation, I compute standard errors using a non-parametric bootstrap procedure with clustering at the hospital-level (500 replications).

4.3 Results

I now present the results for each group of outcomes and then the results of a subgroup analysis which further explores the mechanisms behind the impacts on health outcomes.

4.3.1 Health outcomes

Table 3 presents estimates of the baseline specification for readmission and mortality using admission cohorts (panel A) and discharge cohorts (panel B). For admission cohorts I find that the effect of emergency shocks on readmission and mortality are small in magnitude and statistically insignificant at the 5% level. The mortality result for discharge cohorts is similar. The striking result is for readmissions of discharge cohorts. Here the effect of emergency shocks on readmission is positive, statistically significant at the 1% level, and large in magnitude. A one standard deviation shock is estimated to increase readmissions among discharge cohorts by 0.115 percentage points (4.2% relative to the baseline). These results are robust to various robustness checks.\textsuperscript{13}

I further examine the readmissions of discharge cohorts using the step-function specification. Figure 6 presents the estimates. Both positive and negative shocks have an impact on readmissions and the response of readmission is approximately linear, although shows some weak signs of convexity for extreme positive shocks. Across the entire shock distribution the readmission rate varies between 2.66 and 3.28 (with a mean of 2.80). The implication is that the likelihood of readmission varies by 22% depending on the day of discharge.

Three features of these results stand out. First, the impact of emergency shocks on health outcomes affects patients being discharged. There is no evidence of effects for admission cohorts. Second, mortality is unaffected by emergency shocks and the health effects are evident through readmissions. Third, even small variations in the number of daily admissions have ma-

\textsuperscript{13}Appendix Table C1 shows that the estimates are slightly larger in magnitude if the fixed effects are excluded from the specification. This is consistent with there being some correlation between emergency shocks and patient composition, but the differences are small and the estimates are not statistically different from the baseline results. Appendix Table C2 also shows that the results are robust to the inclusion of a wide range of additional control variables, including further demographics and measures of patient severity. Appendix Table C3 shows that the results are robust to changing the time horizon for readmission and mortality.
terial effects on readmission, suggesting that hospitals are daily adjustments that affect health outcomes.

4.3.2 Inflows of emergency patients

Table 4 presents estimates of the baseline specification using the ED dataset, where I focus on visits that occur the day after an emergency shock and examine three outcomes: the likelihood a patient attends their nearest hospital; time spent in the ED; and the likelihood of inpatient admission. For hospital attendance, I use the emergency shock at the nearest hospital which allows me to test whether hospital choice is affected by the how busy the local trauma and orthopaedic department is. The estimates show that emergency shocks have no statistically significant impact on hospital attendance or inpatient admission (at the 5% level) but do have a small and statistically significant impact on time spent in the ED. A one standard deviation shock is estimated to increase average time spent in the ED by around 20 seconds. I explore these results further in Appendix Table C4 and show that the estimates are similar for various subgroups of patients (e.g. those more likely to be admitted to a trauma and orthopaedic department, ambulance arrivals).

These results show that emergency shocks have only very limited impacts on patients in the ED. Patients may experience delays but will otherwise attend the same hospital and have the same likelihood of receiving inpatient treatment. The implication is that the inflow of emergency patients to inpatient departments is not moderated in the ED, and emergency patients arriving at the inpatient department are not subject to any pre-selection by the hospital.14

4.3.3 Inflows of elective patients

Table 5 presents estimates using daily elective admissions as the dependent variable. As this variable is measured at the hospital-day level I use the panel dataset and omit the fixed effects from the baseline specification. In column (1) I report estimates of a specification containing only the contemporaneous emergency shock. The estimates show that emergency shocks have a negative and statistically significant effect on the number of elective admissions. In column (2), I use the same specification but include additional lags of the emergency shock. These estimates show that, as well as the contemporaneous effect, emergency shocks lead to fewer

---

14One surprising finding here is that ambulance diversion is not used to insure hospitals against the volatility in emergency admissions. This may be possible in some cases (e.g. patients without time sensitive injuries and where the spatial correlation between shocks is low).
elective admissions for up to two days after the shock. In subsequent days the effects become smaller in magnitude and statistically insignificant.

Figure 7 presents estimates of the step-function specification for elective admissions where I include separate step-functions for each lag of the emergency shock. The estimated relationships are distinctly non-linear. Only extreme positive and negative shocks have an impact on elective admissions. The timing of these effects differs. For example, an extreme positive shock today has no impact on elective admissions today but decreases elective admissions tomorrow and the day after. This compares to an extreme negative shock today which increases elective admissions today and tomorrow but not thereafter. These results are consistent with future appointments being cancelled on days when there are extreme positive shocks and these appointments being rescheduled to days when there are extreme negative shocks.

The magnitude of the effects is relatively modest. A three standard deviation positive shock (equivalent to 6.6 emergency admissions) cumulatively leads to 1.6 fewer elective admissions and a two standard deviation negative shock (4.4 emergency admissions) leads to 1.2 more admissions. The fact that the cancellation and reschedule estimates do not net out suggests some reschedules are made to days irrespective of the shock size.

In Appendix Table C5 I explore whether cancellations creates selection among admitted elective patients. There is only very weak evidence of selection across a wide range of observable characteristics. One potential explanation for this is that the timing of the pseudo-random shock (throughout the day) and the elective admissions (typically in the morning) may constrain the choice of cancellations.

These results show that hospitals partially insure themselves against emergency shocks by cancelling elective admissions. This happens only in response to extreme positive shocks, and relatively few cancellations are made (approximately one elective cancellation for every four emergency admissions). This insurance is therefore modest and operates by shuffling patients between the extreme tails of the shock distribution. This reshuffling does not appear to introduce any selection among the pool of admitted elective patients.

4.3.4 Inpatient care

Table 6 presents estimates of the baseline specification for several aspects of inpatient care. For admission cohorts I examine: the likelihood of receiving a daycase operation; the likelihood of having a delayed operation; and the number of procedures. For discharge cohorts I examine:
length of stay (logged); the likelihood of being transferred to another hospital; and the likelihood of being discharged to home. The results show that emergency shocks have three statistically significant effects on inpatient care. First, shocks lead to delays, causing a higher proportion of patients to wait at least a day before receiving their primary operation. Second, shocks cause patients to receive fewer procedures. Third, shocks cause patients to have shorter hospital stays implying that patients are discharged earlier. I find no statistically significant evidence for other outcomes. Appendix Table C6 shows that these results are robust to the inclusion of additional control variables.

Figure 8 presents estimates of the step-function specification for delays, procedures and length of stay. The magnitude of the effects on admission cohorts are modest. Across the shock distribution, delays and procedures vary by 5.7% and 1.8%, and the effects on procedures are statistically insignificant for negative shocks. In contrast the effects on length of stay, which relate to discharge cohorts, are larger in magnitude. Across the shock distribution length of stay varies by 10.2%. This difference is equivalent to around 3 in 7 patients being discharged a day early (4.2 is the average length of stay).

These results show that shocks have a greater impact on inpatient care than in other areas of the hospital. The effects here differ between patients that are arriving and those already in the hospital when the shock occurs. The impacts on admission cohorts are modest and mostly create delays prior to surgery. There are also effects on procedures, but these effects are small and may also be overstated if physicians simply record fewer procedure codes at busy times. These modest effects are consistent with the earlier results that showed health outcomes of admission cohorts are unaffected by emergency shocks.

The most pronounced effects are on discharge cohorts, where shocks cause patients to be discharged earlier. There are two potential explanations for this effect. First, as physicians become busier and deal with more patients, this may result in fewer checks or tests and, in turn, more mistakes when discharging patients (a ‘staff constraints’ effect). The second possibility is that as the hospital becomes busier physicians are required to lower the threshold at which they discharge patients to free up space for newly arriving patients (a ‘bed constraints’ effect). The staff constraints explanation suggests that the effects are likely to more prominent for extreme positive shocks and negligible for negative shocks, while the bed constraints explanation suggests that shocks would feed through linearly to length of stay as new arrivals imply increases

15To incorporate patients with a length of stay of zero, I use log(length of stay+1).
or decreases in bed availability. The step-function estimates, which are approximately linear, indicate that bed constraints are the most likely explanation.

The bed constraints explanation implies that new arrivals effectively ‘push out’ recovering patients. This would be self-reinforcing if it leads to emergency readmissions that then crowd the hospital in future periods. To assess whether this is the case I construct a variable that measures the total hospital bed-days of a patient from initial admission and over the subsequent 90 days. I then estimate the effect of emergency shocks on this variable (logged). The coefficient estimate on emergency shocks from this regression is -0.018 (0.001), indicating that a one standard deviation shock leads to a 1.8% decrease in total hospital bed-days (0.11 days relative to the baseline mean). Comparing this to the the effect on current length of stay, 2.0% (0.08 days), indicates that discharging patients early is not self-reinforcing: discharging patients early may lead to some readmissions but the net effect is to decrease total bed-days.

Together these results indicate that the primary impact of shocks is to cause patients to be discharged earlier. There is no evidence that patients are discharged to alternative locations, they are simply discharged to the same location but sooner. This suggests that discharging patients early may be driving the increases in readmission.

4.3.5 Subgroup analysis

To further investigate the link between the length of stay and readmission effects, I segment the patient population into different groups and assess how the two effects correlate. I examine the effect of shocks across two dimensions: elective and emergency status, and expected mortality risk. To define expected mortality risk I use predicted values from a regression of 30-day in-hospital mortality on a fully interacted set of diagnosis, age category, and emergency status indicators.

Table 7 presents the results split by elective and emergency patients. There is a sharp distinction between the impacts on these two patient types: the estimated effects for emergency patients are consistently an order of magnitude higher than for elective patients. For example, a one standard deviation shock has an impact on readmission of 0.255 percentage points for emergency patients compared to just 0.020 percentage points for elective patients. Similarly the impact of a shock on length of stay is near-zero for elective patients but a large 5.2% decrease for emergency patients. The estimates across other outcomes show similar features, with the effects for elective patients being close to zero or statistically insignificant.
Table 8 presents the results split by mortality risk. I present these for emergency patients only since there is essentially no variation in mortality risk for elective patients, where risk is always near zero. The estimates show that the impact of shocks is correlated with mortality risk. The magnitude of the length of stay effects is monotonically decreasing with risk and it is similar with the readmission effect, where the effect is larger for low and medium risk patients and smaller and less precisely estimated for high risk patients.

Both sets of results support the notion that bed constraints cause the readmission effects. Along both dimensions considered, the magnitude of the length of stay effect and the readmission effect are positively correlated. This correlation is even more apparent if the analysis is further segmented by hospital. Figure 9 presents hospital-level estimates for emergency patients, which are grouped together by the magnitude of readmission effects. The correlation is clear: hospitals that make greater reductions in length of stay in response to shocks are the same hospitals that exhibit greater increases in readmissions.

These subgroup results indicate that the effects on length of stay are a plausible mechanism behind the readmission effects. It also suggests hospitals are running at high levels of utilisation, such that small variations in admissions cause physicians to discharge patients early. With this mechanism in mind, it is also informative to revisit the step-function estimates. Figure 8(c) shows that the effects on length of stay are approximately linear, while Figure 6 shows some signs of convexity. Together these results indicate that the returns to length of stay, in terms of reducing readmission odds, are diminishing. This is a relatively intuitive finding that also supports length of stay as a mechanism.

These subgroup results also raise a separate question about how physicians prioritise patients. There are clear distinctions between elective and emergency patients and within emergency patients yet it is not clear what incentives create these distinctions. It may be driven by clinical incentives - prioritising patients according to expected health benefits - or non-clinical incentives such as for financial gain.\footnote{Examples of non-clinical incentives include financial motivations created by competition (Cooper et al. 2011; Gaynor et al. 2013), malpractice (Kessler and McClellan, 1996), and concerns over their reputation (Kolstad, 2013). Physicians may also simply make systematic mistakes (Abaluck and Agha, 2016).} I explore some of these non-clinical incentives through further subgroups analysis but find that the distinction between elective and emergency patients is remarkably consistent across hospitals and throughout other cuts of the data (see Appendix Tables C7-C11 and Figures C1-C4). Without precise knowledge of the clinical benefits from length of stay, or relevant heterogeneity between hospitals, it is difficult to conclude on whether
the elective and emergency distinction is driven by clinical or non-clinical incentives. I leave this matter for future research.

5 Marginal welfare analysis

The results to now show that hospital crowding has adverse effects on patients, mostly notably through increases in readmissions. I now turn to the question of whether it would be desirable to reduce hospital crowding. I take the perspective of consumer welfare and consider policy options that affect the level of elective admissions. Examples of such policies include waiting time targets and financial targets. Using these policies to reduce elective admissions will necessarily reduce hospital crowding. This creates benefits for patients because lower levels of crowding will mean fewer readmissions and other adverse events (delays, cancellations). But it also creates costs for patients. By decreasing the rate at which patients are served, the waiting time for elective appointments will increase. Policymakers must therefore trade off quality of care (a crowding effect) with access to care (a waiting time effect).

To analyse this trade-off I compare my estimates of the crowding effects with estimates of the waiting time effect. To simplify matters I focus solely on crowding effects relating to readmission. I compare this the effect of elective admissions on waiting times, which I estimate by exploiting variation in elective admissions caused by technological change. I first compare the crowding and waiting times effects purely on the basis of magnitude, which gives an intuitive sense of how much each outcome would adjust in response to a marginal change in elective admissions. I then address consumer welfare. I outline a simple model of welfare that allows me to estimate the relative preferences for readmission and waiting times that are implicit in English healthcare policies. I compare this estimate to benchmarks of true preferences to assess whether there would be net welfare gains from policies that decrease elective admissions.

5.1 The impact of elective admissions on waiting times

Over the sample period there have been major changes in both elective admissions and waiting times for elective surgery. Average waiting times fell by 50% from 2000 to 2010 and elective admissions almost tripled over the same period. Several factors are behind these trends. The first is technological change: hospitals implemented ‘fast track surgery’, an innovation in post-surgical care for elective procedures which led to substantial reductions in length of stay.
without impairing health outcomes (Kehlet, 2013). Shorter stays allowed hospitals to treat
greater volumes of patients in the same amount of time and this reduced waiting times. The
second factor is that private hospitals entered the market for publicly-funded elective surgery.
As these hospitals treated more patients, waiting times at public hospitals fell (Kelly and Stoye,
2015). Third, use of the private healthcare market contracted over this period in part due to the
global financial crisis (Competition and Markets Authority, 2014). I do not observe information
about the private market but there is a natural concern that its decline is correlated with the
increases in elective admissions. Conditional on privately-provided but publicly-funded activity,
I use the variation in admissions created by the introduction of fast track surgery to identify
the effect of elective admissions on waiting times. This variation is plausibly uncorrelated with
the changes in the private healthcare market.

I aggregate the HES data to the regional-year level and include all hospitals that treat trauma
and orthopaedic patients. This includes some hospitals that were previously excluded from the
analysis and I use regional definitions that correspond to the 28 local healthcare authorities at
the start of the sample period. I estimate the following equation

\[ w_{rt} = \kappa_r + \tau q_{grt} + \psi q_{prt} + v_{rt} \]

where \( w_{rt} \) is the mean waiting time in days for elective surgery in region \( r \) during year \( t \), \( \kappa_r \) are
regional fixed effects, \( q_{grt} \) is the number of elective admissions at general acute hospitals, and
\( q_{prt} \) is the number of elective admissions at other hospitals (e.g. private hospitals conducting
publicly-funded work, specialist hospitals with no ED). The parameter of interest is \( \tau \), which
is the impact of elective admissions at general acute hospitals on the waiting times for elective
surgery. I estimate this equation by instrumenting for \( q_{grt} \) with the average length of stay
of elective patients at general acute hospitals, \( l_{grt} \). The identification assumption is that \( l_{grt} \)
is uncorrelated with other factors that affect waiting times contained in \( v_{rt} \). Given that I
control for the activity of other hospitals conducting publicly-funded surgery, the key threat to
this assumption is the changes in the private healthcare market. These changes are plausibly
unrelated to the changes in length of stay caused by the roll-out of fast-track surgery.

Table 9 presents estimates of Equation (5). In the first column I present OLS estimates.
These show that elective admissions have a negative and statistically significant impact on
elective waiting times: 1,000 additional admissions in a region over a year is estimated to reduce
average waiting times by around 2 days. In the second and third columns I present the first-stage and reduced-form regressions. These show that the instrument is relatively strong, with an F-statistic of 13.6 in the first-stage regression, and it has a statistically significant positive impact on waiting times in the reduced-form regression. In the final column I present the IV estimates. Similar to the OLS estimates, these estimates indicate a negative and statistically significant impact of admissions on waiting times but the effect is larger in magnitude: 1,000 additional admissions is estimated to reduce elective waiting times by around 7 days. The difference between the IV and OLS estimates suggests that the OLS estimates are upward biased, which is consistent with usage of the private market being an omitted variable.

In Appendix Figure D1 I show estimates of non-parametric versions of Equation (5). These show some evidence of non-linearities but the relationship is approximately linear over much of the sample variation. As an alternative specification I estimate Equation (5) using a log dependent variable and this produces similar results. Finally, I cross-check the estimates by comparing them to Kelly and Stoye (2015). They exploit the increase in privately provided but publicly-funded activity to assess the impact on waiting times at public hospitals and find results that are similar in magnitude.\(^{17}\)

I now compare the waiting times estimates to the earlier estimates of the crowding effect on readmission. To do this requires carefully rescaling the estimates, since both use different sources of variation and the estimation is conducted at different levels of aggregation. I rescale to a scenario where hospitals on average accommodate one additional patient per day throughout the year. The waiting time effect is scaled to an effect per admission (rather than 1,000 admissions) and then again by the average number of hospitals in a region and the number of elective patients required to increase occupancy by one over the year. The readmission effect is scaled down by the average number of patients in an average emergency shock which then gives the impact of an additional admission every day on the likelihood of readmission.\(^{18}\)

After rescaling, I estimate that reducing hospital occupancy by one elective admission would

\(^{17}\)Kelly and Stoye (2015) use a difference-in-distance identification strategy to estimate the impact of the elective admissions at private hospitals on public hospital waiting times. Over the period 2002 to 2010 they find that an additional 12,000 admissions at private hospitals reduced public waiting times by between 7.6 (using OLS) and 33 days (IV). Rescaling these estimates implies that the effect of 1,000 admissions would reduce waiting times by between 0.6 and 2.8 days.

\(^{18}\)Strictly speaking there are two additional scaling factors to consider. The first is the weighting in each regression. The different levels of aggregation imply differences in the weighting and it is possible to re-weight both regressions on a comparable basis (e.g. by regional population). In practice I find this makes little difference. The second factor is the potential impact on occupancy of the readmission events themselves. Since the majority of patients are not readmitted I do not incorporate this second-order effect.
decrease readmissions by 0.03 percentage points (1.2% relative to the baseline mean) and increase waiting times by 5.3 days (6.2%). These changes imply a combination of welfare benefits and costs for patients. This motivates the question of whether the net effect of these changes leads to welfare gains.

5.2 A model of consumer welfare

To evaluate the net welfare implications I use a static model of consumer welfare. The purpose of the model is to derive a condition that can be used to test whether the net benefits are positive or negative. I consider consumers that have exogenous demand for elective and emergency care. The utility they receive from attending hospital includes the healthcare services received (a benefit) and the waiting time before receiving care (a cost). Hospital technology means the quality of care and the waiting time depend on how many elective admissions the hospital makes and these two hospital characteristics are in direct conflict. I assume that hospitals do not fully internalise the welfare cost of patients waiting for care which provides an economic rationale for policymakers to regulate elective admissions. A regulator is assumed to set policies that determine elective admissions with an objective of maximising social welfare. It is the solution to the regulator’s problem that gives me an optimality condition that I later take to the data.

5.2.1 Consumer preferences

There is a population of $N$ consumers and consumer $i$ demands inpatient care with probability $\rho_{ei}$, where $e = 0$ for elective care and $e = 1$ for emergency care. The probabilities are independent across patients. Consumer utility from receiving inpatient care is a function of the hospital characteristics which include health benefits $h^e$ and a waiting time of $w$ days. I assume that $w = 0$ for all emergency patients. Utility in each state of inpatient care can be written as

$$u_{0i} = h_0 - \gamma_i w,$$

$$u_{1i} = h_1,$$  \hspace{1cm} (6)

(7)

\[19\] In a dynamic model these assumptions would imply Poisson demand as $N \to \infty$. 

28
where $\gamma_i$ reflects the welfare cost of waiting a day for elective care (e.g. impaired mobility, pain, any absence from work) normalised in units of health benefits. Utility when no care is demanded is normalised to zero.

### 5.2.2 Hospital technology

I model hospital technology using reduced-form functions for the health benefits and waiting times. These functions can be thought of as representing a single hospital or a group of hospitals. An advantage of taking a reduced form approach here is that it avoids the need to specify complex mechanisms or extraneous details such as how the health benefits differ between patients.

Health benefits are affected by elective admission through a crowding mechanism. I assume health benefits are weakly decreasing and concave in admissions, which is consistent with the earlier estimates (see Figure 6). Let $q_0$ be the number of elective admissions and denote the health benefits for elective and emergency patients as $h_e(q_0)$. The assumptions on crowding effects imply $h_e'(q_0) \leq 0$ and $h_e''(q_0) \leq 0$. These functions could also incorporate the effect of emergency admissions, which will also have crowding effects, but I omit this from the notation as it is not central to the model.

Waiting times are affected by elective admissions through a queuing mechanism. I assume that waiting times are weakly decreasing and convex in the number of elective admissions. This is intuitive: increasing admissions implies patients are seen more quickly which will cause the queue to decrease, but eventually these benefits will diminish as the queue tends to zero. A similar property follows from standard models of queuing.\(^{20}\) Denoting waiting times as $w(q_0)$, this assumption implies $w'(q_0) \leq 0$ and $w''(q_0) \geq 0$.

\(^{20}\) For example, in a dynamic model this could be set up as a $M/d/c$ queue. In this model the queue would have Poisson demand (‘M’), a deterministic service time (‘d’, length of stay in this setting), and multiple servers (‘c’, number of beds in this setting). Simulating this type of queuing model shows that expected waiting times will be weakly decreasing and convex in the number of servers. One difference between these types of model and my reduced form approach is that I focus on admissions (the actual flow) and not servers (the capacity to accommodate the flow). I effectively ignore the possibility that in some periods the hospital may wish to admit $q_0$ elective patients but there may be insufficient demand to do so. This does not occur frequently in my setting and does not fundamentally change the analysis so I avoid introducing additional notation. The reason queues do not reach zero in practice is because the queue has more features than basic queuing models (e.g. there are multiple queues for different types of elective patients, there is seasonal variations in emergency and elective demand).
5.2.3 Regulator behaviour

The regulator is assumed to maximise social welfare by setting a lower bound on elective admissions for hospitals. Since the hospital does not fully internalise the effect of waiting times, this lower bound will always bind and the regulator effectively sets elective admissions. The regulator problem can be derived as follows.

After including the hospital technology functions the expected utility of consumer \( i \) can be written

\[
E[u_i] = E \left[ \rho_0 (h_0(q_0) - \gamma w(q_0)) + \rho_1 h_1(q_0) \right].
\] (8)

Assuming that the demand probabilities are independent from the welfare costs of waiting, social welfare can be written

\[
U = \frac{1}{N} \sum_{i=1}^{N} E[u_i] = \rho_0 (h_0(q_0) - \gamma w(q_0)) + \rho_1 h_1(q_0),
\] (9)

where \( \rho_e \) and \( \gamma \) are population averages.\(^{21}\) The regulator problem is therefore

\[
\max_{q_0} \rho_0 (h_0(q_0) - \gamma w(q_0)) + \rho_1 h_1(q_0),
\] (10)

which has the following first order condition

\[
\rho_0 (\gamma w'(q_0) - h_0'(q_0)) = \rho_1 h_1'(q_0).
\] (11)

This condition gives the basis on which the regulator will set the level of admissions. It is composed of four components: the relative demand for elective and emergency care; the welfare cost of waiting; the impact of elective admissions on waiting times; and the impact of elective admissions on health outcomes for each type of patient. The marginal benefit of an elective admission is the reduction in waiting times for elective patients, and the marginal costs are the reductions in health benefits to elective and emergency patients. The regulator set elective admissions such that these marginal benefits and costs, weighted by the demand probabilities, are equal.

\(^{21}\)If the demand probabilities and the welfare costs of waiting are not independent then the previous equation will include an additional covariance term. If the covariance is negative - patients more likely to demand care also have a lower costs of waiting, which is plausible for older, retired workers - the revealed preference estimates I present later will be upwards biased. This would make the conclusions conservative.
5.3 Revealed preference estimates

I now explain how the reduced form estimates from earlier in the paper can be combined with Equation (11) to estimate the regulator’s revealed preferences for the cost of waiting. I denote revealed preferences by a tilde, \( \tilde{\gamma} \), to distinguish them from the true consumer preferences, \( \gamma \). The revealed preference estimates provide a basis for assessing hospital overcrowding. If \( \tilde{\gamma} \neq \gamma \) then it implies that Equation (11), which characterises the welfare maximising level of elective admissions, does not hold. By implication, changes in elective admissions will then lead to net welfare gains. In particular, if \( \tilde{\gamma} > \gamma \) then waiting times are valued too highly, hospitals will be overcrowded as a result of the regulator’s policies, and there will be welfare gains from reducing elective admissions.

The key step for estimating Equation (11) is dealing with the terms that reflects the crowding effects of admissions on health benefits. I discuss this first and then describe benchmarks for the true cost of waiting based on opportunity cost, and finally compare my estimates to these benchmarks.

5.3.1 Incorporating the estimated effects of crowding in the model

The revealed preference condition contains the impact of elective admissions on health benefits. Estimating this term directly is not possible in the absence of variation in elective admissions that is exogenous to health outcomes. I therefore substitute this term for the earlier estimates of the impact of emergency admissions on readmission. This substitution has two implications. The first is a scaling issue and presented itself earlier when comparing the magnitude of the two marginal effects (see Section 5.1). I deal with this in the same way here and rescale the effects to a scenario where hospitals accommodate one additional patient per day throughout the year.

The second implication of the substitution is that readmission only partially capture the effects on health outcomes. This implies that my estimates are conservative from perspective of crowding because they neglect any impacts on delays, unobservable aspects of health, or other aspects such as patient satisfaction. But more importantly, the use of readmission affects how the revealed preference estimates are interpreted. To account for this I revisit the regulator problem and replace the health benefit terms, \( h_e(q_0) \), with a readmission term, \( -r_e(q_1)c \), where \( r_e(q_1) \) is the likelihood of readmission as a function of emergency admissions and \( c \) is the per diem welfare cost of a readmission (e.g. a second hospital visit, a second operation, days spent in...
hospital after the readmission). The minus sign reflects readmission being a negative indicator of health benefits and expressing the readmission cost in days aligns its units with waiting times. I reflect this per diem adjustment in the rescaling described above.

Working through the regulator problem with these modifications gives a revised revealed preference condition. This can be simplified by noting that the empirical results show that crowding has no impact on elective patients. The revised condition is therefore

\[
-\frac{\rho_0}{\rho_1} \frac{w'(q_0)}{\theta r_1'(q_1)} = \frac{c}{\gamma},
\]

where \( \theta \) is the scaling term (including the normalisation by readmission days). One interpretation of this modified equation is to think of the regulator having imperfect information about health benefits. If the regulator can only observe readmissions, for example, then it will set policies on the basis of Equation (12) and the analysis aligns directly with regulator behaviour.

It is now straightforward to estimate Equation (12) to obtain the revealed preferences \( \tilde{c}/\tilde{\gamma} \).

The left-hand side contains the ratio of two reduced form estimates, the ratio of the demand probabilities, and the scaling factor. The reduced form estimates have already been obtained and the latter two components can be computed using the HES data and population data. I estimate that \( p_0/p_1 = 2.6 \) and \( \theta = 14.5 \) (see Appendix Table D1 for details).

5.3.2 Opportunity cost benchmark

The right-hand side of Equation (12) is the welfare cost of a day spent in hospital following readmission relative to the welfare cost of a day spent waiting for elective surgery. I compare these welfare costs from the perspective of opportunity cost. This necessarily neglects other important aspects of welfare, which are much harder to assess from an economic standpoint, but the difference in these factors is unlikely to invalidate the qualitative conclusions drawn from this analysis.\(^{22}\)

To begin with I note that in the case of readmission a patient spends a number of days in hospital recovering. This period is 7.3 days on average. Each of these days has a clear opportunity cost since the patient cannot attend work or otherwise spend the day as normal. This suggests wages or other estimates of the value of time may serve as proxies for \( c \). These

\(^{22}\)Other welfare costs would include the clinical implications for the patients future health, and the pain and discomfort for the patient on the specific day. Informal discussions with physicians from trauma and orthopaedics have suggested that these costs may be higher in the case of a readmission, at least in the context of the relatively low waiting times for elective care present in the sample period.
benchmarks suggest that \( c \) is in the region of $80-120\textsuperscript{23}.

In contrast, a day spent waiting for elective surgery does not involve physically waiting at the hospital. The patient can therefore attend work or be at home, albeit potentially with impaired mobility. It is more difficult to gauge the pure opportunity costs in this scenario, although previous studies provide some indication of the order of magnitude. For example, Propper (1990) uses a stated preference survey to evaluate willingness to pay for reducing the days spent waiting for a (hypothetical) low acuity diagnosis in England. More recently, Beckert et al. (2012) use revealed preference data to estimate a demand model for hip replacement patients in England. The demand estimates, combined with additional assumptions, provide another way to estimate of the value of a day spent waiting. These studies suggest \( \gamma \) is relatively low and in the region of $2-4\textsuperscript{24}.

These figures suggest that the true value of \( c/\gamma \) could be at least as high as 20, which is approximately the ratio of the lower range of \( c \) estimates and the upper range of \( \gamma \) estimates. This implies that the average individual would be willing to wait around three weeks to avoid spending an additional day in hospital. I use this as a conservative benchmark to assess the empirical results against.

5.3.3 Results

I estimate that \( \bar{c}/\bar{\gamma} = 4.9 \) (2.7). This indicates that healthcare policies in England assume patients would wait five additional days to avoid an additional day in hospital. This is an entire order of magnitude below the opportunity cost benchmark and I can statistically reject the null hypothesis that \( \bar{c}/\bar{\gamma} \geq 20 \) at the 1% level\textsuperscript{25}. This comparison may also understate the true difference because both the benchmark and the estimates were constructed in a conservative fashion to underestimate the crowding effects.

These results suggest that hospitals in England are overcrowded in the sense that there may be substantial welfare gains from reducing elective admissions. Policies that lower the incentive to admit elective patients are predicted to reduce unplanned readmissions but increase waiting

\textsuperscript{23}The national median net weekly wage is approximately $713 (Office of National Statistics) to which I apply a tax rate of 21%, and recent estimates of the value of travel time in England are approximately $15 per hour (Department of Transport, 2015) which I multiply by 8 to represent a ‘full day of travelling’.

\textsuperscript{24}Propper (1990) finds that the average willingness to pay to reduce waiting times by one month is approximately $85 (uprated to 2017 prices). Beckert et al. (2012) finds that patients are willing to travel an additional 10.7km to obtain waiting times that are 1 day lower, which can be translated into prices by assuming an average driving speeds and an average value of time. I use an average driving speed of 60kph and an average value of time of $15. Both of these studies effectively incorporate opportunity cost and other welfare effects.

\textsuperscript{25}Standard errors were computed using a non-parametric bootstrap approach with 500 replications.
times, but the net impact of this is expected to create consumer welfare gains.

I also compute estimates of $\frac{\hat{c}}{\hat{\gamma}}$ at the regional level. I use the same methodology applied separately for each each region. The estimates vary between -4.7 and 64.3, with a median of 4.7. Only 15% of regions have an estimate that is above 20. In Appendix Figure D2 and D3 I compare these estimates with region wages, as a proxy for opportunity cost. This is a simple way to assess whether the variation in $\frac{\hat{c}}{\hat{\gamma}}$ aligns with consumer preferences. I find only a weak positive correlation between the two measures. The comparison with wages provides some indication that the regional differences are not aligned with consumer preferences.

Overall these results indicate that there would be welfare gains from policies that reduce elective admissions. Given the wide variation between regions, these reductions may be best pursued at a local level.

6 Conclusion

In this paper I show that variations in the level of hospital crowding affect patient health outcomes. To deal with measures of crowding being endogenous, I exploit pseudo-random variation in emergency admissions. This variation is plausible exogenous and offers a rich source of variation for the analysis. I find that the shocks cause the likelihood of unplanned readmission to vary by up to 22% and this is likely to be driven by bed constraints and physicians varying the discharge threshold.

This raises the question of whether policymakers should reduce hospital crowding. By moderating elective admissions, the adverse effects of hospital crowding can be reduced at the cost of increasing waiting times. I estimate the effect of elective admissions on waiting times using within-region variation in technological change, and compare this estimate to the crowding effects. Using a model of consumer welfare and conservative benchmarks of preferences, I find that reductions in hospital crowding may lead to substantial welfare gains.

There are two main caveats to the results. The first is that the analysis is limited to trauma and orthopaedic departments in England. While similar mechanisms may operate within other healthcare settings, the health conditions and preferences of patients may differ substantially. This paper provides a framework for how to assessing hospital crowding in these other settings. The second caveat is that the analysis relies primarily on unplanned readmission as a measure of health outcomes. Despite this being widely used it should not be taken as the only rele-
vant outcome and policies would ideally be set using more holistic measures. The availability of such measures is growing (NHS Digital, 2017) and the present work underlines the importance of obtaining these measures across the spectrum of patients including those admitted as emergencies.

I conclude by noting two policy issues that merit further research. First, the marginal welfare analysis revealed that national healthcare policies may have very different implications across regions. This may suggest a greater role for policies that rebalance resources across regions. Second, ambulance diversion and hospital transfers offer ways to insure hospitals against the volatility in emergency admissions. Implementing these forms of insurance in a centrally managed healthcare system may be feasible and policies that do so may offer low cost increases in available hospital capacity.

References


35
Department of Health. (2012), ‘A simple guide to Payment by Results’.


Tables and figures

Table 1: Mean characteristics of patients in trauma and orthopaedics and other specialties

<table>
<thead>
<tr>
<th></th>
<th>Trauma and orthopaedics</th>
<th>Other specialties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>52.9</td>
<td>56.3</td>
</tr>
<tr>
<td>Male, %</td>
<td>48.5</td>
<td>45.4</td>
</tr>
<tr>
<td>White, %</td>
<td>85.4</td>
<td>89.4</td>
</tr>
<tr>
<td>Emergency, %</td>
<td>39.4</td>
<td>36.9</td>
</tr>
<tr>
<td>Elective waiting time, days</td>
<td>84.4</td>
<td>58.9</td>
</tr>
<tr>
<td>Diagnosis count</td>
<td>3.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Charleston comorbidity index</td>
<td>1.7</td>
<td>2.8</td>
</tr>
<tr>
<td>ED admissions within past 12 months</td>
<td>0.8</td>
<td>1.1</td>
</tr>
<tr>
<td>7-day unplanned readmission, %</td>
<td>2.8</td>
<td>4.1</td>
</tr>
<tr>
<td>30-day in-hospital death, %</td>
<td>1.1</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Notes: (1) ‘Other specialties’ excludes paediatrics and maternity care and is based on a 1% sample of the full inpatient HES data.

Table 2: Estimated effects of predicted admissions and emergency shocks on hospital scheduling

<table>
<thead>
<tr>
<th></th>
<th>Physician count (1)</th>
<th>Elective admissions (2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted emergency admits</td>
<td>0.64*** (0.072)</td>
<td>1.262*** (0.242)</td>
<td>−0.262*** (0.079)</td>
</tr>
<tr>
<td>Shock to emergency admits</td>
<td>0.143*** (0.005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of physicians</td>
<td></td>
<td>2.381*** (0.074)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>340,816</td>
<td>340,816</td>
<td>340,816</td>
</tr>
</tbody>
</table>

Notes: (1) All specifications include hospital fixed effects; (2) Standard errors clustered at the hospital-level; (3) ***/**/ indicates statistical significance at the 1/5/10% level.
Table 3: Estimated effects of standardised emergency shocks on health outcomes

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Coeff</th>
<th>Std error</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Admission cohorts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-day unplanned readmission</td>
<td>0.02152*</td>
<td>(0.012)</td>
<td>4,019,288</td>
</tr>
<tr>
<td>30-day in-hospital mortality</td>
<td>0.00671</td>
<td>(0.007)</td>
<td>4,019,288</td>
</tr>
<tr>
<td><strong>Panel B: Discharge cohorts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-day unplanned readmission</td>
<td>0.11517***</td>
<td>(0.022)</td>
<td>4,019,288</td>
</tr>
<tr>
<td>30-day in-hospital mortality</td>
<td>−0.00214</td>
<td>(0.007)</td>
<td>4,019,288</td>
</tr>
</tbody>
</table>

Notes: (1) Reported coefficients are parameter estimates on the standardised emergency shock variable; (2) Non-parametric bootstrapped standard errors clustered at the hospital-level (500 replications); (3) ***/***/* indicates statistical significance at the 1/5/10% level.

Table 4: Estimated effects of standardised emergency shocks on inflows of emergency patients

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Coeff</th>
<th>Std error</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attended nearest hospital with T&amp;O department</td>
<td>0.019*</td>
<td>(0.010)</td>
<td>22,231,970</td>
</tr>
<tr>
<td>Time spent in the ED, mins</td>
<td>0.346***</td>
<td>(0.071)</td>
<td>22,231,970</td>
</tr>
<tr>
<td>Inpatient admission</td>
<td>0.013</td>
<td>(0.013)</td>
<td>22,231,970</td>
</tr>
</tbody>
</table>

Notes: (1) Reported coefficients are parameter estimates on the standardised emergency shock variable (time spent in the ED, inpatient admission) or the standardised emergency shock variable at the nearest hospital (attended nearest hospital); (2) The nearest hospital is defined according to straight-line distances from the patient’s home to the set of general acute hospitals in the panel dataset; (3) Non-parametric bootstrapped standard errors clustered at the hospital-level (500 replications); (4) ***/***/* indicates statistical significance at the 1/5/10% level.

Table 5: Estimated effects of standardised emergency shock on inflows of elective patients

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th></th>
<th>(2)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>Std error</td>
<td>Coeff</td>
<td>Std error</td>
</tr>
<tr>
<td>Emergency shock, t</td>
<td>−0.039**</td>
<td>(0.0181)</td>
<td>−0.039**</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Emergency shock, t-1</td>
<td>−0.058***</td>
<td>(0.018)</td>
<td>−0.058***</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Emergency shock, t-2</td>
<td>−0.038**</td>
<td>(0.019)</td>
<td>−0.038**</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Emergency shock, t-3</td>
<td>−0.019</td>
<td>(0.018)</td>
<td>−0.019</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Emergency shock, t-4</td>
<td>−0.013</td>
<td>(0.019)</td>
<td>−0.013</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Emergency shock, t-5</td>
<td>−0.006</td>
<td>(0.018)</td>
<td>−0.006</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Emergency shock, t-6</td>
<td>−0.006</td>
<td>(0.017)</td>
<td>−0.006</td>
<td>(0.017)</td>
</tr>
</tbody>
</table>

| N               | 326,668   | 326,668   |

Notes: (1) Samples restricted to periods when all lagged values are available; (2) Residual bootstrapped standard errors clustered at the hospital-level (500 replications); (3) ***/***/* indicates statistical significance at the 1/5/10% level.
Table 6: Estimated effects of standardised emergency shocks on inpatient care

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Coeff</th>
<th>Std error</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Admission cohorts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daycase operation</td>
<td>0.051</td>
<td>(0.038)</td>
<td>4,019,288</td>
</tr>
<tr>
<td>Delayed operation</td>
<td>0.443***</td>
<td>(0.029)</td>
<td>4,019,288</td>
</tr>
<tr>
<td>Number of procedures</td>
<td>−0.004***</td>
<td>(0.000)</td>
<td>4,019,288</td>
</tr>
<tr>
<td><strong>Panel B: Discharge cohorts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of stay (log)</td>
<td>−0.020***</td>
<td>(0.001)</td>
<td>4,019,288</td>
</tr>
<tr>
<td>Transfers to other hospitals</td>
<td>0.004</td>
<td>(0.010)</td>
<td>4,019,288</td>
</tr>
<tr>
<td>Discharges to home</td>
<td>0.005</td>
<td>(0.020)</td>
<td>4,019,288</td>
</tr>
</tbody>
</table>

Notes: (1) Reported coefficients are parameter estimates on the standardised emergency shock variable; (2) Length of stay defined as log(length of stay + 1); (3) Non-parametric bootstrapped standard errors clustered at the hospital-level (500 replications); (4) ***/***/* indicates statistical significance at the 1/5/10% level.

Table 7: Estimated effects of standardised emergency shocks on inpatient care and health outcomes by patient type

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Electives</th>
<th></th>
<th>Emergencies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff</td>
<td>Std error</td>
<td>Coeff</td>
<td>Std error</td>
</tr>
<tr>
<td><strong>Panel A: Admission cohorts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daycase operation</td>
<td>0.099*</td>
<td>(0.06)</td>
<td>−0.021</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Delayed operation</td>
<td>0.141***</td>
<td>(0.032)</td>
<td>0.895***</td>
<td>(0.045)</td>
</tr>
<tr>
<td>Number of procedures</td>
<td>−0.002***</td>
<td>(0.001)</td>
<td>−0.005***</td>
<td>(0.001)</td>
</tr>
<tr>
<td><strong>Panel B: Discharge cohorts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of stay (log)</td>
<td>0.001</td>
<td>(0.001)</td>
<td>−0.052***</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Transfers to other hospitals</td>
<td>−0.006</td>
<td>(0.005)</td>
<td>0.018</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Discharges to home</td>
<td>−0.011</td>
<td>(0.015)</td>
<td>0.028</td>
<td>(0.033)</td>
</tr>
<tr>
<td>7-day unplanned readmission</td>
<td>0.020***</td>
<td>(0.007)</td>
<td>0.255***</td>
<td>(0.033)</td>
</tr>
<tr>
<td>30-day in-hospital mortality</td>
<td>0.002</td>
<td>(0.002)</td>
<td>−0.017</td>
<td>(0.013)</td>
</tr>
</tbody>
</table>

Notes: (1) Reported coefficients are parameter estimates on the standardised emergency shock variable; (2) N = 3,848,785 (2,756,091) in all specifications for elective (emergency) patients; (3) Length of stay defined as log(length of stay + 1); (4) Non-parametric bootstrapped standard errors clustered at the hospital-level (500 replications); (5) ***/***/* indicates statistical significance at the 1/5/10% level.
Table 8: Estimated effects of standardised emergency shocks on inpatient care and health outcomes for emergency patients by predicted mortality risk

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Admission cohorts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daycase operation</td>
<td>0.034</td>
<td>-0.065</td>
<td>-0.016</td>
</tr>
<tr>
<td>Delayed operation</td>
<td>0.927***</td>
<td>0.890***</td>
<td>0.969***</td>
</tr>
<tr>
<td>Number of procedures</td>
<td>-0.005**</td>
<td>-0.001</td>
<td>-0.005</td>
</tr>
<tr>
<td><strong>Panel B: Discharge cohorts</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of stay (log)</td>
<td>-0.065***</td>
<td>-0.058***</td>
<td>-0.027***</td>
</tr>
<tr>
<td>Transfers to other hospitals</td>
<td>0.005</td>
<td>-0.027</td>
<td>0.101</td>
</tr>
<tr>
<td>Discharges to home</td>
<td>0.039</td>
<td>-0.028</td>
<td>0.018</td>
</tr>
<tr>
<td>7-day unplanned readmission</td>
<td>0.262***</td>
<td>0.275***</td>
<td>0.172*</td>
</tr>
<tr>
<td>30-day in-hospital mortality</td>
<td>-0.006</td>
<td>0.005</td>
<td>-0.005</td>
</tr>
</tbody>
</table>

Notes: (1) Reported coefficients are parameter estimates on the standardised emergency shock variable; (2) Predicted mortality computed from a regression of 30-day in-hospital mortality on a fully interacted set of diagnosis, age category, and emergency status indicators; (3) Length of stay defined as log(length of stay + 1); (4) N = 751,116 / 290,831 / 522,834 in all specifications for low/medium/high risk groups; (5) Non-parametric bootstrapped standard errors clustered at the hospital-level (500 replications); (6) ***/**/* indicates statistical significance at the 1/5/10% level.

Table 9: OLS and IV estimates of the effect of elective admissions on elective waiting times

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>First-stage</th>
<th>Reduced-form</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elective admits at GAHs, 000s</td>
<td>-2.058***</td>
<td></td>
<td></td>
<td>-6.966***</td>
</tr>
<tr>
<td></td>
<td>(0.668)</td>
<td></td>
<td></td>
<td>(2.02)</td>
</tr>
<tr>
<td>Length of stay at GAHs, days</td>
<td>-3.332***</td>
<td>23.209***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.904)</td>
<td>(4.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>220</td>
<td>220</td>
<td>220</td>
<td>220</td>
</tr>
</tbody>
</table>

Notes: (1) GAH is an abbreviation for general acute hospital (i.e. those included in the panel dataset); (2) All specifications include regional fixed effects and a control for the number of elective admissions at non-GAHs; (3) The first-stage regression uses elective admits at GAHs as the dependent variable; (5) Standard errors clustered at the regional level (28 clusters); (6) ***/**/* indicates statistical significance at the 1/5/10% level.
Figure 1: Heterogeneity between trauma and orthopaedic patients

Notes: (1) Each marker is a three-digit ICD-10 category; (2) Market size indicates the relative number of admissions; (3) Labels shown for the three largest diagnosis groups for elective and emergency patients; (4) Data extracted from the inpatient dataset in 2010 for the top tertile of elective and emergency patients when sorted by diagnosis frequency.

Figure 2: Decomposing emergency admissions

Notes: (1) Data shown for one hospital in one year; (2) Panel (b) predictions based on a regression of daily emergency admissions on hospital-specific year, weekly-seasonal and day-of-the-week fixed effects using all years of data in the panel dataset.
Figure 3: Seasonal components of elective and emergency admissions

Notes: (1) Seasonal estimates based on local polynomial estimates of demeaned average admissions on day-of-the-year.

Figure 4: Distribution of emergency shocks

Notes: (1) Emergency shocks are defined as residuals from a regression of hospital-day emergency admissions on hospital-specific year, weekly-seasonal and day-of-week fixed effects.
Figure 5: Hospital-level estimates of first-order serial correlation in emergency shocks

![Figure 5](image)

Notes: (1) Estimated AR(1) coefficients from a regression of emergency shocks on their lag, where 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 6: Estimated effects of standardised emergency shocks on 7-day unplanned readmission

![Figure 6](image)

Notes: (1) Point estimates are relative to the base category (near-zero shocks) which is normalised to the sample mean; (2) Standardised emergency shocks are the mean value within the relevant segment of the step-function defined by Equation (4); (3) N = 4,019,288; (3) Non-parametric bootstrapped standard errors clustered at the hospital-level (500 replications) with 95% confidence intervals shown.
Figure 7: Estimated effects of standardised emergency shocks on inflows of elective patients

Notes: (1) Point estimates are relative to the base category (near-zero shocks) which is normalised to the sample mean; (2) Standardised emergency shocks are the mean value within the relevant segment of the step-function defined by Equation (4); (3) N = 326,668; (3) Residual bootstrapped standard errors clustered at the hospital-level (500 replications) with 95% confidence intervals shown.
Figure 8: Estimated effects of standardised emergency shocks on inpatient care

(a) Delayed operation

(b) Number of procedures

(c) Length of stay

Notes: (1) Point estimates are relative to the base category (near-zero shocks) which is normalised to the sample mean; (2) Standardised emergency shocks are the mean value within the relevant segment of the step-function defined by Equation (4); (3) N = 4,019,288 in all specifications; (4) Non-parametric bootstrapped standard errors clustered at the hospital-level (500 replications) with 95% confidence intervals shown.
Figure 9: Estimated hospital-level effects of standardised emergency shocks on length of stay and 7-day unplanned readmission for emergency patients

Notes: (1) Estimates of the baseline specification for length of stay and 7-day unplanned readmission conducted separately for each hospital in the sample; (2) Each point on the plot represents approximately 10 hospitals, where hospitals are grouped into 16 quantiles according to the magnitude of the hospital-level readmission effects.