Essays in the economics of healthcare

George Stoye

Thesis submitted to the Department of Economics
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

University College London

London
7^{th} September 2020

Declaration

I, George Stoye, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. Chapters 1, 2 and 5 are sole-authored. Chapter 3 is joint work with Jonathan Gruber and Thomas Hoe. Chapter 4 is joint work with Elaine Kelly.

Date: September 7, 2020

Abstract

How to provide and fund healthcare is becoming an increasingly important debate in many countries due to widespread demographic changes. This has led to a growing focus on inequalities in the amount and quality of care provided to different groups. This thesis contains three papers that examine the roles played by medical staff and institutional frameworks in explaining this variation.

Throughout, I use the English National Health Service as a testbed to examine these roles, exploiting the institutional features of this universal public health system and its rich administrative data.

In the first paper, I examine the extent to which individual doctors explain variation in patient outcomes. Studying consultants treating heart attack patients, I exploit within-hospital random assignment of patients to doctors, and the movement of staff between hospitals, to estimate the effect of individual doctors on patient survival. I show considerable variation in the quality of individual doctors, and examine potential improvements in patient survival from reassigning doctors across patients.

In the second paper, I study the impacts of external regulation on the performance of doctors in English emergency departments. I extend a 'bunching' methodology commonly used in the tax literature to examine the impacts of the four-hour waiting time target that applies to all English hospitals. I show the regulation was successful in reducing waiting times and drastically reduced mortality. This shows that changes to the incentives of doctors can be successful in improving care quality.

In the final paper, I examine the impact of reforms that allowed pre-existing private hospitals to enter public healthcare markets. I exploit historical locations of hospitals to instrument for potentially endogenous hospital entry. I show private hospital entry sizeably expanded the market, but led to little competition between new and existing hospitals, and therefore did not impact care quality.

Impact Statement

The research contained in this thesis has the potential to have an impact both on public policy and within the academic community. My research helps to improve the understanding of the drivers of variation in the quality of care that is provided to different patients. This is important for policy in identifying and reducing inequalities in care, an explicit goal of the National Health Service in England since the 2012 Health and Social Care Act. It also provides an important contribution to the academic understanding of what drives variation in the productivity of health care providers, quantifying differences in the productivity of individual medical staff and building knowledge on how incentives for healthcare providers can be used to improve patient outcomes.

I will seek to disseminate the findings of this work to policymakers and practitioners by engaging with them directly and publicising the work in the popular press, and to academics through publication in leading peer-reviewed economics journals.

In Chapter 2, I show that senior doctors in England vary considerably in their effectiveness in treating patients, and that aggregate patient outcomes could be improved by reorganising emergency cardiac treatment to better match patients with the appropriate doctors. This is important for hospital managers when scheduling doctors' shifts. It also demonstrates that routinely collected administrative data could be used by policymakers to improve patient outcomes.

Chapter 3 explores how the incentives faced by doctors working in emergency departments can directly impact mortality rates by studying a high profile waiting time target in England. The four-hour target has been of huge media and policy focus in recent years, with hospitals regularly failing to meet the target since 2017. This has led to policy moves to replace the target with alternatives that are focused on specific conditions. I have already disseminated this work directly to NHS England and NHS Improvement, helping to inform the ongoing changes to the target, and through the national media broadcast and press media, and will continue to do so going forward.

Chapter 4 explores the impact of private provider entry to the public market for elective healthcare in England. Private provision within the NHS has been a huge political issue for years. This debate is only likely to intensify following the coronavirus crisis, with the private sector providing crucial capacity to address build-ups in public waiting times. This research crucially improves our understanding about the consequences of private sector provision for health outcomes and health inequalities.

Throughout the course of my PhD, I have regularly communicated with analysts at the Department of Health and Social Care, NHS England and NHS Improvement to disseminate and discuss my research. I hope to use these already established channels to continue to disseminate existing work, and to build links to focus future work on key policy issues.

Acknowledgements

The completion of this thesis has been aided by financial support from the Economic and Social Research Council, through the Centre for the Microeconomic Analysis of Public Policy (CPP) at the Institute for Fiscal Studies (grant numbers RES-544-28-5001 and ES-M010147-1) and 'The impact of medical labour on variation in patient outcomes: evidence from English public hospitals' research grant (ES-S003118-1).

I am grateful to my supervisors, Eric French and Richard Blundell, for their input and guidance throughout my PhD, and comments on specific work. Thank you for all that you have done. I also wish to thank other members of the UCL faculty for help at different points in university life: specific thanks to Imran Rasul and Marcos Vera-Hernández for helpful comments and friendly advice at various stages, and to Wendy Carlin for encouraging me to initially apply for a job at the IFS and to pursue a career in research.

Research is much more fruitful and fun when surrounded by excellent colleagues and co-authors. Thank you to Jon Gruber, Tom Hoe and Elaine Kelly for your help and guidance with this work. I have learnt a lot from you all. I am also enormously grateful to Carol Propper, who has been extremely generous with her time in helping me with much of this work.

Throughout my PhD I have had the pleasure of working with a great set of colleagues at the IFS. In particular, I wish to thank: Gemma Tetlow, Rowena Crawford and Carl Emmerson for being supportive managers, and enabling me to develop my own research agenda along the lines that I wished; Cormac O'Dea for sage advice (with a smile) at various points on the way; Rory McGee for much-needed help with programming and weekend office company; and Jonathan Cribb, Peter Levell, Barra Roantree and Kate Smith for your friendship and encouragement from day one of the PhD.

Last but certainly not least, I am eternally grateful for the support of my friends and family in this and all other endeavours. Natalie - I'll try to work fewer weekends in future! Dave and George - sorry for the never-ending questions about the inner workings of hospitals (I make no promises to stop). And to Mum and Dad - thank you for all you do; none of this would be possible without your love and support over the last 31 plus years.

Data Provision

This thesis makes use of a number of data sources. Chapters 2, 3 and 4 all use the Hospital Episode Statistics. Chapters 2 and 3 use Office for National Statistics (ONS) mortality data. Chapter 4 uses data from the National Joint Registry.

The Hospital Episode Statistics contain pseduo-anonymised administrative patient records for the National Health Service (NHS) in England. These data were made available by NHS Digital under data sharing agreement CON-205762-B8S7B. Chapters 2, 3 and 4 use data from the Admitted Patient Care Hospital Episode Statistics dataset. Chapters 2 and 3 use data from the Accident and Emergency Hospital Episode Statistics dataset.

ONS mortality data contain information on the date of death for all deceased UK citizens. Chapters 2 and 3 use ONS mortality data linked to the Hospital Episode Statistics by NHS Digital, and made available under data sharing agreement CON-205762-B8S7B. Chapter 2 uses these data for all patients treated in NHS hospitals following a heart attack between April 2005 and March 2018. Chapter 3 uses these data for all patients treated in an NHS Accident and Emergency department between April 2011 and March 2013.

The National Joint Registry contains information on all hip replacements conducted by the NHS and the private sector in England. Data are made available through Northgate Public Services.

Neither the owner nor the distributor of any of these data bear any responsibility for their further analysis and interpretation. Research datasets may not reproduce National Statistics aggregates, or data published by NHS Digital. Any errors and omissions are the responsibility of the author.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis consists of three self-contained papers that explore the determinants of the widespread variation in patient outcomes that has become increasingly documented across most developed countries. In particular, I study the role of medical staff, and the influence of the institutional frameworks in which they work, in explaining variation in patient outcomes in England. In this introduction, I briefly summarise the approach and findings of each of these papers, before discussing the cross-cutting themes arising from across the entire thesis.

Chapter 2 ("The Distribution of Doctor Quality in England") examines the role played by senior doctors in determining patient outcomes. Attempts to quantify the effectiveness of specific doctors have typically struggled to account for potential selection between patients and doctors: for example, if sicker patients choose to be treated by more able doctors, then estimates of doctor quality will be biased (Glance et al., 2008). I address this concern by using the institutional features of heart attack treatment in the English public hospital system, where patients are randomly assigned to senior doctors after conditioning on the hospital that they attend for treatment.

Exploiting this within-hospital quasi-random assignment of patients to senior doctors following a heart attack, I estimate the persistent differences in patient outcomes associated with each individual physician, and the returns to experience from treating these patients. More than a quarter of doctors worked in multiple hospitals in England over a 13-year period, enabling me to separately identify the impacts of receiving treatment from a specific doctor from broader hospital effects. Using this approach reveals that there is considerable variation in the quality of doctors: a one standard deviation improvement in doctor quality leads to a reduction in mortality of 4.2 percentage points, or 29% of mean mortality between 2005 and 2018. I also find that there are some returns to specific experience, but these are small compared to the persistent differences in productivity between these doctors.

These findings suggest that patient outcomes could be improved if badly performing doctors were replaced with better performing peers. However, this is hard to achieve in practice, with long training periods for such highly qualified staff. Instead, I consider an alternative question: could patient outcomes be improved by reassigning patients to different doctors, while keeping the available pool of doctors unchanged?

In order to answer this question, I extend my model to estimate the quality of each doctor when treating 'low' and 'high' severity patients. I find that doctors do vary in their quality when treating alternative patient types. I then analyse the potential mortality reductions from reallocating patients and doctors, so that doctors who are comparatively better at treating high severity patients are matched to these patient types. The results of this exercise suggest that mortality could be substantially reduced by reassigning doctors to patients: with no constraints, mortality could be reduced by 19%, while mortality could be reduced by 9% if doctors were reassigned across patients within the hospitals that they currently work.

These findings indicate that doctors play a key role in explaining variation in patient outcomes. The estimates quantify the substantial variation in the performance of doctors, even after undergoing large amounts of training. This also suggests that there are considerable improvements in patient survival that could be achieved by a more efficient allocation of existing senior staff resources.

This chapter contributes to two linked literatures that examine variation in the quality of care. It extend works on variation in care quality that has typically focused on the performance of hospitals (Gowrisankaran and Town, 1999; Geweke et al., 2003; Doyle et al., 2015; Hull, 2020) to estimate the quality of individual doctors. It also contributes to a growing literature that has studied individual elements of doctor behaviour, including their beliefs (Cutler et al., 2019), practice styles (Currie et al., 2016; Molitor, 2018; Chan et al., 2019; Currie and Macleod, 2020), and spending (Doyle et al., 2010; Van Parys, 2016). In this case, I provide a summary measure of doctor productivity as opposed to examining individual behaviours or characteristics.

Chapter 3 ("Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers") explores how the incentives faced by doctors providing emergency care affect their behaviour, and ultimately what this means for patient outcomes. Since 2010, all emergency departments (EDs) in England have been subject to a waiting time target that required 95% of patients to be admitted, discharged or transferred to another hospital within 4 hours of arrival at the hospital. This arbitrary target places significant constraints on doctors treating patients in the ED. I use this target as a natural experiment to analyse

the impact of these constraints on the treatment provided to patients, and their health outcomes.

A significant challenge in analysing the impacts of the target is that the policy was rolled out nationally, without obvious treatment and control groups. In addition, there are no detailed pre-policy data available to examine changes in outcomes. The chapter therefore adopts a different approach to estimating the impacts of the target on treatments, costs and patient outcomes. It applies the bunching techniques that have been used widely in other contexts (see Kleven, 2016) to analyse wait times and outcomes using data from the period when the policy was already in place. This allows me to model the counterfactual outcomes under a set of identifying assumptions, estimating the short-term impact of changing wait times on patient outcomes.

In a first step, I show that the target reduced waiting times significantly, with a mean reduction of 21 minutes (8%) for all patients affected by the policy. For patients whose waiting times were reduced from over 4 hours in the counterfactual scenario, these wait time reductions are even larger at 59 minutes.

The chapter then examines the wider impacts on doctor behaviour by examining changes in the treatment provided to patients, and the outcomes they experience. Plotting these treatments and outcomes conditional on wait times reveals a substantial spike just prior to the four-hour wait time. I decompose this spike into two separate channels: first, a 'composition effect' generated by moving patients with different characteristics forward in the waiting time distribution, and second, a 'distortion effect' that captures changes to the treatment or outcome generated by the target. I show that this distortion effect can be estimated from the data under a 'no selection' assumption, using the observed outcomes for patients whose waiting times were not affected by the policy to adjust for the composition of patients treated just prior to four hours.

The results show that the target distorts doctor behaviour to reduce wait times and to increase the probability of admission to hospital. It also results in very large reductions in mortality, with 30-day mortality falling by 14%. I also show that these mortality reductions are associated strongly with wait time reductions, and not increases in treatment intensity, by exploiting heterogeneity in the target impacts on different treatments and outcomes across different conditions and periods of time.

The work contributes to two academic literatures. There has been a growing focus on explaining the production function of emergency care in different settings (Chan, 2016, 2017; Silver, 2016). Chapter 3 contributes to this literature by showing how ED production is affected when doctors are pressured to make decisions quickly. The results show that incentives do have substantial impacts on doctor

behaviour and on their patient outcomes. While the target did lead to increases in costs for hospitals, it did also deliver substantially better care for patients. This is a key result for policymakers to understand when designing regulations for hospitals.

The chapter also makes a significant methodological contribution to the literature using bunching estimators (Saez, 2010; Chetty et al., 2013; Kleven and Waseem, 2013; Diamond and Persson, 2016; Einav et al., 2018; Best and Kleven, 2018), applying these estimators to the setting of emergency healthcare provision. I adapt these estimators to study outcomes indirectly affected by a running variable, rather than impacts on the running variable itself as has previously been standard in the literature.

Chapter 4 ("The Impacts of Private Hospital Entry on the Public Market for Elective Care in England") explores how the entry of private hospitals impacted the size and shape of the elective healthcare market in England, and the role played by these providers in driving increasing socio-economic inequality in elective surgeries. Traditionally, all publicly funded care in England was provided by large, publicly owned hospitals. Following promises to reduce waiting times and increase quality, a series of reforms in the mid-2000s allowed pre-existing private hospitals to enter the public market in order to compete with existing public hospitals by treating public patients alongside their own privately funded patients. I explore the impacts this had on existing providers and their patients.

I study the impact of this entry on the market for hip replacements by comparing changes in outcomes across areas differentially exposed to private hospital entry. Provider entry is instrumented with the pre-reform location of private hospitals. The results indicate that private hospital entry increased the number of publicly funded hip replacements by 12%, but did not reduce volumes at incumbent public hospitals, and had no impact on readmission rates. This suggests new entrants exerted little competitive presssue on incumbents. Instead, the market expanded with more marginal patients receiving treatment at an earlier point in time, resulting in a fall in average patient severity.

I also explore whether these impacts varied across areas with different levels of local deprivation. This shows that the impacts did not vary substantially across areas with different levels of deprivation, and suggests that private provider entry played only a very small part in the increasing socio-economic inequality in joint replacements that emerged in England during the late 2000s and early 2010s.

These findings contribute to two literatures. First, it builds on a small number of papers that examine private provider entry in England (Cooper et al., 2012, 2018) and the US (Courtemanche and Plotzke, 2010; Munnich and Parente, 2018), extending this work to examine the impacts of pre-existing private hospitals rather

than the introduction of purpose-built facilities. Second, it builds on a wider literature that examines the impact of hospital competition and choices on patient outcomes (Kessler and McClellan, 2000; Propper et al., 2004; Cooper et al., 2011; Gaynor et al., 2013, 2016). Private provider entry was a key component of large-scale reforms to hospital competition in England in the 2000s, and until now the effects of this entry have been poorly understood.

While each of these chapters has sought to answer a specific question, they are linked by an overarching research agenda that seeks to understand how and why the healthcare provided to different patients varies. As policymakers around the world seek to understand and address health inequalities within their countries, exploring the drivers of this variation and potential solutions are of key importance.

This thesis as a whole clearly demonstrates the important role played by providers in determining patient outcomes, and the role that regulation can have to reduce inequalities and improve patient outcomes. Chapter 2 shows that there is significant variation in the performance of individual doctors, but outcomes for patients could be improved by simply changing the allocation of existing staff. Chapter 3 shows that targets can be used very successfully to influence the behaviour of doctors and hospitals in order to deliver better outcomes for patients. Chapter 4 shows that rapid increases in the supply side can be achieved to reduce waiting times for patients without worsening health inequalities.

Each chapter is also linked by a common setting: the English National Health Service. Much of the existing work on the behaviour and performance of health care providers comes from the US. However, the US healthcare system varies dramatically from much of the developed world. The complicated network of providers and insurers means that studying some concepts - such as the productivity of individual providers - is complicated by selection issues and partial data coverage. The NHS provides an ideal institutional setting in which to study many of the big economic issues in healthcare, with both rich administrative data and a variety of institutional features to exploit to answer these important research questions. My work therefore provides new evidence from a large, universal healthcare system with similarities to many other systems around the world.

The remainder of this thesis is structured as follows. Chapters 2 to 4 each contain one self-contained paper. Chapter 5 then concludes by setting out future research questions that build upon the research included here.

Chapter 2

The Distribution of Doctor Quality in England[1]

2.1 Introduction

Widespread variation in patient outcomes across place and providers has been well documented in many developed countries (Skinner, 2011). An important driver of this variation is likely to be the behaviour and performance of doctors, who play a key role within all healthcare systems - diagnosing patients, and deciding on and administering treatment. A growing literature has documented wide variation in the beliefs of physicians (Cutler et al., 2019), their preferred styles of treatment (Currie et al., 2016; Molitor, 2018; Chan et al., 2019; Currie and Macleod, 2020) and differences in the resource utilisation of different doctors (Doyle et al., 2010; Van Parys, 2016). Given the essential roles that doctors play, these variations in behaviour can literally mean the difference between life and death for patients. Better understanding of the variation in the performance of doctors is therefore key in providing higher quality and more efficient care.

This paper provides new empirical evidence on the role played by senior doctors in explaining variation in patient outcomes. I exploit two key institutional features of the English National Health Service (NHS) to estimate the distribution of individual doctors fixed effects on patient survival following a heart attack and the returns to experience in treating such patients. I show that the performance of doctors varies when treating patients of different severity, and use these type-specific estimates of doctor quality to show that substantial improvements in patient outcomes could be attained by reallocating doctors to specific patients.

[1]I am grateful to Richard Blundell, Eric French, Carol Propper, Imran Rasul, Marcos Vera-Hernández, and seminar participants at IFS, UCL and King's College London for helpful discussions and comments on this paper.

I focus on patients being treated following a heart attack in English public hospitals. This provides an ideal setting to minimise concerns about selection of patients to specific doctors, or vice versa, that could potentially bias estimates of doctor quality (Glance et al., 2008). Heart attack patients require rapid care, with outcomes worsening significantly following delays. Patients in England can seek care in any of the public hospitals in the country, with no available private market for emergency medicine. Upon admission, patients are assigned to an available cardiologist. This therefore provides a setting where within-hospital patient assignment to doctors is essentially random, while also being free of the potential complications (such as partial data coverage) of provider networks that exist in insurance-based systems such as the US.

A second advantage of the English hospital setting is that doctors are public, salaried employees. Doctors are not paid on the basis of the procedures that they carry out and therefore do not face obvious financial incentives to deviate from what they believe to be medical best practice. Senior doctors in England are routinely contracted to multiple hospitals, while moves across hospitals over time are also common. As a result, more than a quarter of senior doctors are observed treating heart attack patients in multiple hospitals over the 13-year period studied here. This provides a setting in which the underlying individual productivity of doctors can be measured separately from the hospitals in which they work, and free of the influence of financial incentives that could mask differences in innate productivity.

My analysis uses administrative hospital data from the Hospital Episode Statistics (HES). These data contain the census of publicly funded inpatient hospital care in England, and provide rich information on patient diagnoses and treatments. Patients are tracked over time to create detailed health histories, and all admissions are tagged to an identifier of the senior doctor with overall responsibility for the patient. These data are linked to official mortality statistics that record mortality for patients up to a year after hospital admission occurs. This enables me to study the treatments provided to, and the long-run mortality outcomes of, all patients in England who received treatment from a cardiologist following a heart attack between April 2005 and March 2018.

My analysis proceeds in two parts. In the first part of my analysis, I set out a simple model for patient survival following a heart attack. I estimate this model by regressing an indicator of patient survival on a rich set of patient characteristics, in addition to a full set of physician and hospital dummy variables, and measures of doctor experience. This is analogous to the two-way fixed effects models commonly used to study worker and firm effects in wage settings, as first proposed by Abowd et al. (1999), with an additional term that captures doctor experience. Under a

set of identifying assumptions, this approach recovers estimates of the persistent differences in patient outcomes across doctors, and the returns to specific and general experience in treating cardiology patients.

Identifying these parameters depends on two key conditions. First, that after conditioning on hospital, time of treatment, and observed characteristics, patients are randomly assigned to doctors. The institutional features outlined above suggest that this is highly likely for heart attack patients within the English system. I provide further evidence in support of this assumption by showing that there is no significant relationship between a wider set of patient characteristics (that are excluded from the initial control variables) and the estimated quality of their assigned doctor. Second, doctor and hospital fixed effects can only be separately identified within a connected set of doctors and hospitals, formed by doctors treating patients in multiple hospitals. I show that working in multiple hospitals - either contemporaneously or over the course of a career - is common, and identifies a common set that captures the majority of heart attack patients in England.

The results suggest that doctors play an important role in explaining the variation in patient outcomes. Decomposing variation across doctors, hospital, patient attributes and unobserved characteristics indicates that doctors explain 5.5% of the overall variation in one-year mortality rates, substantially greater than the share explained by hospitals (3.5%). Furthermore, the estimates suggest there is considerable variation in the quality of individual doctors: a one standard deviation increase in quality reduces one-year mortality rates by 4.2 percentage points, equivalent to 29% of mean mortality, or 0.1 standard deviations in mortality. This suggests that considerable improvements in patient outcomes could be obtained by improving average doctor quality.

The estimates also suggest that there are some gains to specialisation among cardiologists, but the improvements in performance with experience are small relative to permanent differences in the performance of these doctors. I find that an one standard deviation increase in the number of heart attack patients treated over a three-year period increases doctor quality by the equivalent of 6% of a standard deviation in permanent quality.

In the second part of my analysis, I extend this model to allow doctors to vary in their effectiveness in treating different patient types. I split patients into low and high severity groups based on their predicted mortality, and estimate fixed effects for each doctor when treating each patient type. Using these estimates, I rank doctors according to their comparative advantage in treating high severity patients and reallocate high severity patients to these doctors under two different scenarios. This includes an 'unconstrained' scenario where doctors can be assigned to patients in any hospital, and a 'constrained' scenario where doctors can only be

reassigned to patients within the hospitals in which they work. In line with similar past exercises that examine the potential gains from replacing poorly performing teachers (Hanushek, 2009), I also estimate the gains from replacing the worst performing 10% of doctors when treating each patient type as a comparison case.

These estimates show that doctors do vary in their ability to treat different patient types, and that there are potentially substantial gains to reallocating patients to more appropriate doctors. The estimates suggest that, in 2017, deaths in the year following heart attack treatment could have been reduced by 19% if doctors were reallocated in the unconstrained scenario and 9% when restricting reallocations to be only within hospital. These effects are actually larger than the 5% reduction from simply replacing the worst performing 10% of doctors. This indicates that even among highly trained doctors, there is variation in their ability to carry out different tasks. Changing shifts patterns to better match patients and doctors is therefore likely to improve patient outcomes.

My work contributes to two literatures. First, I contribute to a literature on variation in care quality across providers. Much of the past work has focused on differences in care quality across hospitals, without explicitly considering the roles played by individual doctors (Gowrisankaran and Town, 1999; Geweke et al., 2003; Doyle et al., 2015; Hull, 2020). I contribute to this work by estimating quality measures for individual doctors.

Second, I contribute to a literature that examines the individual behaviour of physicians (Chandra and Staiger, 2007; Van Parys, 2016; Molitor, 2018; Currie and Macleod, 2020). This work typically focuses on the impact of individual decisions or behaviours on costs and patient outcomes. I extend this work by providing a summary measure of physician productivity in order to demonstrate the wide variation in the persistent performance of these doctors.

My empirical approach also relates to past papers that use two-way fixed effects models, which are usually identified by changes in firm of employment or place of residence by individuals, in order to identify individual effects separately from broader factors. In particular, this relates to the large body of work that examines the individual productivity of workers and firms (e.g. Abowd and Kramarz, 1999; Card et al., 2013; Bonhomme et al., 2019). More recently, a large literature has also developed to examine the impact of teachers on their students' test scores and later life outcomes (e.g. Aaronson et al., 2007; Mansfield, 2015; Chetty et al., 2014a,b). My approach extends a similar approach to examining the behaviour of a new set of highly skilled workers.

The rest of this paper is organised as follows. Section 2.2 describes the institutional features of the English public hospital system and the rich administrative data used in the analysis. Section 2.3 sets out a model of the determinants of

patient outcomes following a heart attack. Section 2.4 describes my empirical approach and identifying assumptions. Section 2.5 presents the baseline results, before discussing a series of robustness checks and alternative specifications. Section 2.6 sets out a more flexible model to estimate doctor quality when treating different patient types, and estimates the potential improvements in patient mortality outcomes when reallocating patients across doctors. Section 2.7 concludes.

2.2 Background and data

2.2.1 Emergency health care in England

Emergency health care in England is publicly funded through general taxation, and is available free at the point of use for all residents. Acute care is provided by large, publicly owned hospitals, who are reimbursed by the government for the care they provide, and staffed by public employees.[2] There are no restrictions on where patients can access emergency care, but patients typically attend their nearest hospital. There is no private market for emergency care.[3] As a result, patients who require acute emergency care are all treated by a public hospital.

This paper focuses on patients receiving treatment for an acute myocardial infarction (AMI) or heart attack. Heart attack patients require immediate care, with marked improvements in survival rates among patients receiving treatment within hours of onset (Maxwell, 1999). Emergency treatment for these patients typically follows one of two pathways, with different types of treatment given at different stages. The majority of patients will first be treated in the emergency department (ED) before being admitted to a specialist cardiology department.[4] Patients either arrive independently at the ED, or by ambulance following an emergency call. In 2017/18, 71% of heart attack patients who received ED treatment arrived by ambulance. Upon arrival, patients undergo an initial assessment to establish the severity of their condition, followed by a series of investigations and (if necessary) treatments to stablise their conditions. When required, patients are then admitted as an inpatient, and sent to the cardiology department where they will be assessed and treated by a trained cardiologist.

Alternatively, some ambulance patients may be admitted directly to the cardiology department, without first attending an ED. These patients will undergo

[2]Payments to hospitals vary according to the care provided. Treatments are assigned to a Healthcare Resource Group (HRG), similar to Diagnosis-related Groups (DRGs) in the US. Each treatment is assigned to a nationally set tariff, with small adjustments made across regions with varying fixed costs and for very long length of stays.

[3]There is a small private market for elective health care in England, with treatment funded out-of-pocket or by private medical insurance.

[4]EDs are known as Accident and Emergency (A&E) departments in England.

initial tests and treatments in the ambulance, who communicates these results directly to the hospital. In 2017/18, a third of heart attack patients treated by a cardiologist were admitted without attending an ED first.

Patients can receive a number of treatments as a cardiology inpatient depending on the nature of the heart attack and the preferences of the staff treating them. Two types of treatments are commonly used, and sometimes combined. First, intravenously administered thrombolytic drugs can be used to dissolve blood clots that have blocked blood flow through major arteries and veins. This treatment is most effective when provided to patients within 12 hours of the onset of the heart attack (Windecker, 2014). Second, more invasive treatments can also be used to restore blood flow to the heart, with 77% of AMI patients undergoing a procedure when being treated by a cardiologist in 2017/18. The most commonly used procedure for this purpose is angioplasty, where a balloon is used to open blocked arteries, and which can be combined with the insertion of a stent to maintain blood flow (a process known as a percutaenous coronary intervention, or PCI). Where angioplasty is not appropriate, patients may instead undergo more invasive surgery, such as a coronary artery bypass graft (CABG) which diverts blood flow around the blockage in the artery.

2.2.2 The role of senior doctors

All inpatient care in public hospitals is overseen by a senior doctor, known as a consultant.[5] These doctors are legally responsible for patients, and are in charge of the treatment given to patients. They will manage the overall treatment of patients, either directly treating patients or overseeing decisions made by more junior staff. This will include assessing patient severity, deciding on the course of treatment, and performing individual surgeries. They will also set out plans for secondary prevention of future heart attacks and related conditions, including providing lifestyle advice and prescribing post-discharge medication. This paper focuses on estimating differences in the performance of senior doctors treating heart attack patients, and will combine all of these elements.[6]

All inpatients are assigned to a responsible consultant upon admission. In the case of a heart attack, patients will be assigned to the cardiologist consultant who is working at the time of admission.[7] Hospitals schedule, for every shift,

[5]Consultants are equivalent to 'attending physicians' in the US.

[6]Junior staff are not included in the data, and no indication is provided of who actually performs surgery. While organisation of staff varies across hospitals, consultants will typically share registrars rather than always working with the same doctor on every shift. The estimates of doctor performance will therefore combine managerial ability, underlying medical skill and effort.

[7]Consultants may also be 'on call' for emergencies, with specific consultants available to attend the hospital in the case of a heart attack patient arriving and requiring treatment.

a consultant to be available to carry out emergency treatment if a heart attack patient arrives. The identity of the consultant is scheduled in advance but is not publicly available, and so would be unknown to patients ahead of time. Patients do not receive a choice about which consultant they are assigned to (conditional on attending a particular hospital at a particular time), while consultants cannot refuse to treat specific patients. As a result, conditional on hospital and the time attended, patients are essentially randomly assigned to consultants.

Consultants are all experienced doctors. At a minimum, cardiologist consultants will have undergone at least five years of medical school, two years (or more) of basic training, and a five-year cardiology residency (which can be extended if consultants also undertake research during this period).

An important feature of the English National Health System (NHS) is that doctors regularly work in multiple hospitals. Doctors are contracted by each (publicly owned) hospital organisation separately.[8] However, many consultants hold positions at multiple hospitals simultaneously, or move between hospitals over time. As a result, doctors are observed treating patients at multiple hospitals. This enables me to separately study the impact of doctors and hospitals on patient outcomes, as set out in detail below.

2.2.3 Data

The primary source of data for the analysis comes from the inpatient Hospital Episode Statistics (HES). These data cover all public hospital admissions between April 1997 and March 2018. Observations are recorded at the episode level, with an episode recording the period of care under the responsibility of a single consultant. The data record detailed information about the patient and the care they receive, including their age, sex, local area of residence, admission and discharge dates, a primary diagnosis and up to 19 secondary diagnoses, an extensive list of procedure codes, and a hospital identifier. All patients are tracked by a pseudonymised patient identifier, which can be used to create detailed histories of past hospital treatment and diagnoses.

From March 2003, all patients are also linked to a consultant identifier. This enables the assignment of patients to consultants, and also allows me to derive histories of consultant activity. Doctor experience is observed imperfectly: patient records between March 2003 and April 2018 can be used to track activity of all consultants during this period, but does not provide information on experience in other roles or prior to this date.

[8]Hospitals in the same area are grouped into 'Trusts', with shared management. In most cases, trusts have only one or two major acute hospitals, with smaller hospitals providing specialist non-acute care. Throughout the text, I refer to trusts as 'hospitals'.

Mortality outcomes are recorded by the UK Office for National Statistics (ONS), and linked to individual patients through anonymised identifiers based on patient National Insurance (Social Security) numbers. These data include the date of death for all individuals who died in the UK, or UK citizens who die abroad, between April 2005 and March 2019. I use these data to create indicators of whether a heart attack patient died within 30 days and a year of initial treatment.

Table 2.1 reports summary statistics for all patients included in the analysis, who were treated in public hospitals between April 2005 and March 2018.[9] This shows the mean and standard deviation for a range of patient characteristics and outcomes. Mean patient age was 67.8 years. 67% of patients were male, and 81% were white. In the previous year, patients had received £1,915 of inpatient treatment. 3.6% of patients had received treatment for a previous heart attack since 1997. The average Charlson index score was 1.94.

Table 2.1: *Descriptive statistics of AMI patients*

	Cardiology	
	Mean	S.D.
Age	67.8	14.0
Male	0.67	0.47
White	0.81	0.39
Prev treatment cost	1,915.23	8,730.55
Past AMI	0.036	0.185
Charlson index	1.94	1.30
30-day in-hospital death rate	0.063	0.244
30-day death rate	0.066	0.248
1-year death rate	0.145	0.352
Number of patients	566,148	

Notes: (1) Treatment costs include inpatient treatment costs in the past year, and are reported in 2018 GBP; (2) Past AMI records whether the patient has been admitted to an English hospital since April 1997 for heart attack treatment.

The central focus of this paper is to understand how individual senior doctors affect the mortality outcomes of their patients, and how these outcomes could be achieved if existing staff resources were reallocated. Table 2.1 shows that mortality is a common outcome following a heart attack. 6.6% of patients died within 30 days of admission, with the vast majority of these patients dying in hospital (the in-hospital mortality rate during the same period is 6.3%). Over the course of a year, the mortality rate more than doubled, to 14.5%.

[9] I discuss in detail the construction of this sample in Section 2.4.2.

2.3 Health production function

After patient i suffers a heart attack in period t, they attend hospital k and are treated by doctor j. Survival up to one year after treatment (Y_{ijkt}) is determined as follows:

$$Y_{ijkt} = \beta X_{it} + \mu_{jt} + \psi_{kt} + \epsilon_{ijkt} \tag{2.1}$$

X_{it} captures observable individual determinants of patient survival, including patient demographic characteristics and health history. μ_{jt} is the impact of being treated by doctor j in period t. ψ_{kt} is the impact of being treated in hospital k in period t. ϵ_{ijkt} captures any remaining factors that are unobserved and affect patient survival.

The impact of being treated by a specific doctor in each period (μ_{jt}) can be decomposed into three terms:

$$\mu_{jt} = \mu_j + \delta Exp_{jt} + \nu_{jt} \tag{2.2}$$

where μ_j is a doctor fixed effect. It is interpreted as the permanent quality of doctor j, as measured by their impact on the survival probability of their patients. The paramter captures the effect of assigning overall responsibility for a patient's care to a particular senior doctor. As noted above, senior doctors play several roles in planning and providing patient care. I do not attempt to separate the extent to which this parameter captures underlying medical skill, managerial ability or physician effort.

Exp_{jt} measures the experience of the doctor in each year. δ therefore represents an experience profile in treating heart attack patients that is common across all doctors. ν_{jt} captures any year-specific shocks in doctor performance. I assume that this shock has a mean of zero. This means that doctor performance only systematically varies across years on the basis of experience.

This model has two further important restrictions. First, the effects of hospital and doctors are additively separable. This is a strong assumption: it rules out that doctors perform better or worse in particular hospitals. This would be violated, for example, if hospitals differ in the technology available for treating heart attack patients, and doctors vary in their ability to use different technologies. I discuss this assumption further in Section 2.5.3 and present empirical evidence that suggests that any match effects between doctors and hospitals are small.

Second, the model also restricts doctor quality to be fixed across all patients. A more flexible model would allow heterogeneity in the ability of each doctor to treat patients with different conditions or characteristics. In Section 2.6, I explicitly

relax this assumption to examine variation in doctor quality across different patient types (based on their predicted mortality). I then use this to examine potential gains in patient survival from reallocating doctors to treat patients of specific types.

2.4 Empirical strategy

2.4.1 Baseline estimation

The main focus of this paper is to produce empirical estimates of permanent quality for each doctor (μ_j) and the common experience profile across doctors (δ). To do this, I estimate the following specification:

$$Y_{ijkt} = \beta X_{it} + \mu_j + \delta Exp_{jt} + \psi_k + \xi_{ijkt} \tag{2.3}$$

Y_{ijkt} is a binary variable that takes the value of one if patient i survives for one year after receiving treatment.[10] X_{it} captures a rich set of patient demographic and health characteristics. This include a quadratic in age, sex (also interacted with age), ethnicity, charlson comorbidity index, the cost of any hospital treatment in the previous year, indicators of whether the patient has experienced a previous heart attack or stroke, a full set of fixed effects capturing primary and secondary diagnoses recorded at the time of admission, indicators of the day of the week, month and year when the patient is admitted, and an interaction between month and year of admission.[11]

μ_j and ψ_k are doctor and hospital fixed effects respectively. The inclusion of a hospital fixed effect will control for permanent differences in the casemix of patients treated in each hospital, the supply side of hospitals (e.g. management and equipment), and other elements of care quality. Patients are free to choose which hospital they attend, and although they often attend their nearest hospital, there may be selection into hospitals based on the characteristics of patients who live in the local area. I therefore view the inclusion of hospital fixed effects as an important control, but do not make any causal claims about the impact of receiving treatment at a particular hospital on patient mortality outcomes.[12]

[10] I also repeat this using 30-day survival.

[11] For the period after 2009, when complete ED data are available, I also include indicators of whether the patient arrived by ambulance, the exact time of arrival at the ED and discharge to the cardiology department, and the number and type of investigations and treatments. The inclusion of these variables has a negligible impact on the estimates of μ_j. Results available upon request.

[12] Note that in Equation (2.3) I do not allow the effect of hospitals to vary over time (as specified in Equation (2.1)). As I discuss below, this allows me to identify fixed effects for a greater number of doctors.

Experience (as measured by the number of years worked as a cardiologist) is imperfectly measured in the data. The senior doctors included in the data will also all be relatively experienced already. As a result, I focus on two measures of activity undertaken by these doctors in the baseline specification as measures of specific and general experience. First, I include the number of heart attack patients treated by doctor j in the previous three years. This measures specific experience in treating heart attack patients in a recent period. Second, I include the number of all other patients treated in the previous three years. This measures recent general experience.

The error term (ξ_{ijkt}) captures any further factors that influence patient mortality outcomes. As discussed below, the validity of the estimates of doctor quality will rely on a lack of correlation between this error term and the identity of the doctor treating the patient. As a result, it is helpful to consider a further decomposition of this error term into three parts:

$$\xi_{ijkt} = \nu_{jt} + \lambda_{kt} + \upsilon_{ijkt} \tag{2.4}$$

As in equation (2.2), ν_{jt} represents year-specific shocks in doctor performance that could arise due to a range of factors, including their own health and personal life, or the influence of past performance that is not fully captured by the experience measures included in equation (2.3). λ_{kt} represents year-specific shocks in hospital performance. These include changes in wider management practices or in the equipment available for staff (e.g. if a hospital installs a catheterisation lab). Finally, υ_{ijkt} includes any remaining individual survival factors, including luck, and any measurement error.

2.4.2 Identifying assumptions

Estimation of permanent doctor quality relies on two key assumptions. The first concerns the within-hospital assignment of patients to doctors:

Assumption 1 (Quasi-random conditional assignment of patients to doctors)**.** *Conditional on hospital and time of treatment, and patient observables, patients are as good as randomly assigned to senior doctors*

$$Cov(\mu_j, \xi_{ijkt}) = 0 \tag{2.5}$$

This assumption rules out both that patients choose to be treated by a particular doctor, or that doctors choose or are systematically allocated to patients with particular (unobserved) characteristics. Satisfying this assumption has often been problematic in the past when attempting to measure doctor performance, and has

been a common criticism of attempts by policymakers to introduce report cards for hospitals or individual doctors (Marshall et al., 2000; Dranove et al., 2003).

Importantly, this assumption is about within-hospital assignment. Patients may choose to attend specific hospitals. Indeed, patient mix is likely to vary across hospitals, with some hospitals treating sicker or more deprived populations than others. This assumption does not rule out patients systematically sorting into particular hospitals, but only that doctors working at a particular hospital are not assigned to particular patients within the pool of patients attending the hospital.

Two institutional factors make this assumption likely to hold in practice. First, heart attack patients require rapid treatment, with outcomes deteriorating as time passes (Maxwell, 1999). Treatment is therefore unlikely to be routinely delayed in order to find alternative physicians (either at the request of the patient or the doctor). Second, as noted in Section 2.2.2, patients are assigned to senior doctors who are either physically present in the hospital at the time of arrival, or who are specifically 'on call' for emergency cases. Scheduling of these doctors are not publicly available, and patients would be unlikely to choose a particular time and hospital to seek heart attack treatment from even if the identity of doctors was known in advance.

In Section 2.5.3 I provide empirical evidence in support of this assumption by examining the relationship between the estimated quality of doctors and the characteristics of patients that are not included in the baseline control (e.g. not included in X_{it} in Equation (2.3)). I show that there is no significant relationship between any of these observables and the estimated quality of the assigned doctor, suggesting that the assumption holds in practice.

In addition to the assumption on patient assignment, the doctor and hospital effects set out in equation (2.3) are only identified within a set of hospitals that are connected by physicians who worked at multiple hospitals (Abowd and Kramarz, 1999). This gives rise to the second identifying assumption:

Assumption 2 (Connected set of doctors and hospitals)**.** *There is a connected set of doctors and hospitals, where doctors treat patients in multiple hospitals, such that all doctors in this set can be identified separately from the hospitals in which they work.*

This assumption allows me to compare doctors both within and across different hospitals, and is equivalent to having doctors who treat more than a minimum number of patients in at least two hospitals over the 13 years considered. Intuitively, this allows me to compare the performance of doctors within each hospital, and then use doctors who work in two hospitals as a benchmark for comparing

doctors in the two connected hospitals. Expanding this across the whole national system allows me to compare the performance of doctors working in any hospital which is linked to the others through a doctor working in more than one hospital.

Hospitals are linked both by doctors who transfer across jobs over time, and by doctors who work for multiple hospitals at the same time. The institutional features of the English public hospital system are important in providing these links between hospitals, with many doctors working in multiple hospitals at the same time. The size of the connected set, and the number of patients and doctors observed within it, depend on the restrictions placed upon the number of patients that 'linking' doctors must treat in each hospital to be included within this set. Doctors who treat very few patients in a specific hospital would have noisy outcomes, and would not provide good comparators with other doctors in this hospital.

In the baseline analysis, I create this connected set by including all hospitals where at least one cardiologist treated more than 10 heart attack patients in at least one other hospital between April 2005 and March 2018.[13] I also exclude any doctors who treated fewer than 10 heart attack patients in a single hospital. This provides a final analysis sample of 566,148 patients treated by 1,764 consultants across 145 hospitals. This includes all cardiologists treating more than 10 patients in these hospitals: the 468 'linking' consultants who treated at least ten patients in two hospitals during this period of time (who accounted for the treatment of 174,543 patients, or 29% of all patients) and the 1,296 who are observed only working in a single hospital. It excludes around 4,000 heart attack patients who are treated by cardiologists outside of this connected set.

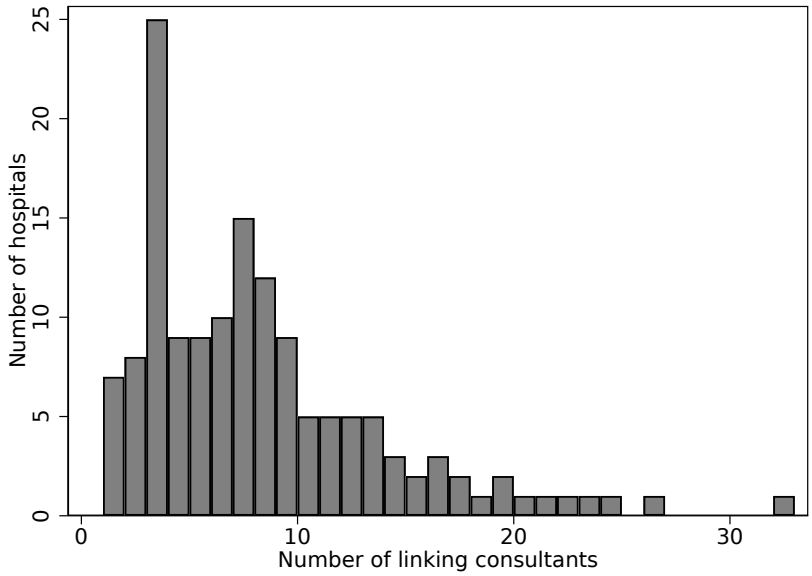
Figure 2.1 shows the distribution of 'linking' consultants across hospitals. This shows that there is variation across hospitals in the number of consultants that link the hospital to the rest of the set: some hospitals are linked by only 1 consultant, while others are linked by more than 20 consultants. Similarly, Figure 2.2 shows that there is wide variation in the number of patients treated by doctors in their 'secondary' hospital (e.g. where they treat the fewest patients). The modal caseload in the secondary hospital is between 10 and 20 patients, but some consultants treat hundreds of patients in multiple hospitals over 13 years.

Table 2.2 shows summary statistics for the 1,764 cardiologists included in the analysis. On average, these consultants treated 329 heart attack patients over a 13 year period. Consultants treated heart attack patients for an average of 7.2 years, and had an average caseload of 48.6 heart attack patients in each years. The mean mortality rate was slightly higher (at 15.0% over a year after treatment) than

[13]Using an alternative minimum caseload to define this set does not alter the results substantially. Results are available upon request.

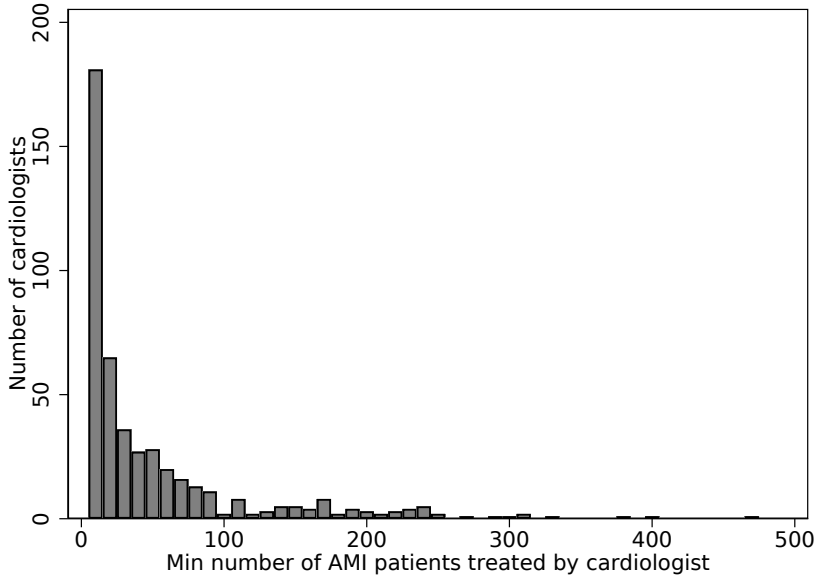
across all patients, suggesting that patients with higher caseloads have slightly lower mortality rates than doctors with low caseloads.

Figure 2.1: *The distribution of linking consultants across hospitals*



Notes: (1) Observations are at the hospital level; (2) Number of linking cardiologists indicate the number of consultants working at the hospital who treated at least 10 patients in another hospital within the connected set between April 2005 and March 2018.

Figure 2.2: *Patient distribution of linking consultants in 'secondary' hospitals*



Notes: (1) Observations at the consultant level; (2) The sample includes all linking cardologists who treat patients in multiple hospitals; (3) 'Secondary Hospitals' are defined as the hospital in which the consultant treats the fewest patients over the period.

Table 2.2 also shows that many doctors are observed treating patients in multiple hospitals. Over the 13 year period, 28.7% of consultants are observed treating at least one patient in multiple hospitals. On average, consultants treated patients at 1.4 hospitals over the entire period. These working patterns provide the important variation that enable the separate identification of doctors and hospitals within the connected set.

Table 2.2: *Descriptive statistics of cardiologists*

	Cardiology	
	Mean	S.D.
Volume	329.34	329.88
Annual Volume	48.57	34.87
30-day in-hospital death rate	0.064	0.037
30-day death rate	0.068	0.039
1-year death rate	0.150	0.064
Multiple hospital	0.287	0.44
Number of hospital	1.38	0.97
Number of years in the data	7.21	4.16
Number of consultants	1,764	

Notes: (1) Includes all consultants in the connected set who treated a minimum of 10 AMI patients between April 2005 and March 2018; (2) Multiple hospital takes the value of one if a consultant is observed treating at least ten patients in more than one hospital over the period

2.4.3 Empirical implementation issues

A common concern in literatures that conduct similar exercises for other workers (most notably teachers) is that the variation in worker fixed effects may be overestimated when such estimates are based on small samples for individual workers (Card et al., 2013). Intuitively, doctors who are observed treating very few patients may be lucky or unlucky in their outcomes. In this case, these doctors will either have very large positive or negative estimated fixed effects but these estimates will simply reflect statistical noise as opposed to being a good signal of the true ability of these doctors.

To address the impact of statistical noise on the estimates of each doctor fixed effects, I implement an Empirical Bayes shrinkage estimator. This has been commonly used in the teacher value-added literature to adjust estimates for low class sizes (Kane and Staiger, 2008; Kane et al., 2008; Chetty et al., 2014a). I shrink estimates according to a shrinkage factor, λ_j, that measures the proportion of the variation in the average doctor residual that is due to signal variance:

$$\lambda_j = \frac{\sigma_u^2}{\sigma_u^2 + (\sigma_e^2/n_j)} \tag{2.6}$$

σ_u^2 is the between-doctor variance in mortality outcomes, σ_e^2 is the within-doctor variance, and n_j is the total number of heart attack patients treated by doctor j over the entire period. Estimates are shrunk towards zero if doctors treat few patients, or if the overall share of the variation is attributed to within-teacher variation.

In my baseline estimates, I estimate these parameters as sample analogues from my fixed effects regression. Alternative methods have been used in the teacher value-added regression, including a two-step method that uses the covariance within-teachers across different class years in order to estimate the reliability of the estimates (Kane and Staiger, 2008; Chetty et al., 2014a; Bitler et al., 2019). I show in Section 2.5.3 that while the exact magnitude of the variation in doctor fixed effects does vary across methods, the qualitative conclusions are unchanged.

2.5 Results

I first set out a variance decomposition of one-year survival rates to examine how much of this variance is explained by doctors as compared to other factors, before reporting and discussing the estimates of permanent doctor quality and returns to experience from equation (2.3). I then present evidence that the key identifying assumption of within-hospital random assignment of patients to doctors holds, and discuss a series of robustness checks and alternative specifications that explore other threats to identification.

2.5.1 Variance Decomposition

Table 2.3 shows the results of a decomposition in the variation of one-year survival rates between doctors, hospitals, and patient observed and unobserved characteristics. This indicates that senior doctors account for 5.5% of the total variance. This is around two-thirds larger than the share of the variance accounted for by hospitals (3.5%). Patient observables account for just under a quarter of the variance. This means that almost three-quarters of the variation in patient outcomes is explained by unobserved factors.

It is important to note that well established sampling errors in the doctor and hospital terms are likely to overstate the estimates of the variation associated with these terms (Card et al., 2013). Correlations in the sampling errors of the doctor and hospital fixed effects are also likely to result in negatively biased estimates of the covariance between these effects (Andrews et al., 2008). I do not attempt

to correct for potential biases in the variance or covariance terms from sampling errors. These results should therefore be taken as upper bounds for the share of the variation explained by these factors.

Table 2.3: *Variance decomposition of patient one-year survival rates*

	Variance (1)	Standard Deviation (2)	Share of Total Variance (3)
Total	**0.124**	**0.352**	100.0%
Doctors (μ_{jt})	0.007	0.082	5.5%
Hospitals (ψ_k)	0.004	0.066	3.5%
Patient observables ($X\beta$)	0.028	0.166	22.4%
Residual (ϵ_{it})	0.094	0.306	75.9%
2 x Cov(Doc, Hosp)	-0.008		-6.4%
2 x Cov(Doc, Obs)	0.000		-0.2%
2 x Cov(Hosp, Obs)	-0.001		-0.6%

Notes: (1) Decomposition does not adjust for any biases from low sample sizes in the variance or covariance terms; (2) Patient observables include a quadratic in age, sex, an interaction between age and sex, race, charlson index, an ambulance dummy, treatment costs in the previous financial year, whether the patient suffered an AMI or stroke in the previous year, a set of dummy variables for primary and secondary diagnoses, day of the week, and an interaction between month and year of admission.

2.5.2 Estimates of doctor quality and returns to experience

Table 2.4 shows the estimates of doctor quality (μ_j) obtained by estimating equation (2.3). The dependent variable in the first two columns is 30-day survival following a heart attack, with raw estimates reported in the first column and the shrunken estimates in the second column. Columns three and four repeat this for 365-day survival as the dependent variable. In both cases, I report the standard deviation, the variance, and the 10th, 25th, 50th, 75th and 90th percentiles. Bootstrapped standard errors are reported for each point of the distribution.

The unadjusted results indicate that a one standard deviation improvement in doctor quality is related to a 4 percentage point reduction in 30-day mortality. This is equivalent to 61% of the mean 30-day mortality rate of 6.6%, or 0.16% of a standard deviation. The adjusted estimates are smaller, with a one standard deviation improvement in doctor quality equal to a 2.7 percentage point reduction in 30-day mortality. This is equivalent to 41% of the mean mortality rate, or 0.11 standard deviations. In both cases, the estimates are statistically significant at the 1% level.

For 365-day survival, the unadjusted results indicate that a one standard deviation improvement in doctor quality is related to a 4 percentage point reduction

in 30-day mortality, or 41% of the mean one-year mortality rate of 14.5%. The adjusted estimates indicate that a one standard deviation improvement in doctor quality is equal to a 4.2 percentage point reduction in one-year mortality. This is equivalent to 29% of the mean mortality rate, or 0.1 standard deviations.

Table 2.4: *Estimated doctor fixed effects, 30-day and 365-day survival rates*

	Doctor fixed effects			
	30-day survival		365-day survival	
	Unadjusted (1)	Adjusted (2)	Unadjusted (3)	Adjusted (4)
Std Deviation	0.040***	0.027***	0.060***	0.042***
	(0.0034)	(0.0039)	(0.0065)	(0.0067)
Variance	0.002***	0.001***	0.004***	0.002***
	(0.0004)	(0.0003)	(0.0010)	(0.0008)
10th percentile	-0.039***	-0.029***	-0.057***	-0.044***
	(0.0027)	(0.0025)	(0.0040)	(0.0038)
25th percentile	-0.020***	-0.015***	-0.024***	-0.020***
	(0.0016)	(0.0014)	(0.0023)	(0.0022)
50th percentile	0.001	0.001	0.004**	0.003**
	(0.0016)	(0.0014)	(0.0018)	(0.0016)
75th percentile	0.024***	0.017***	0.034***	0.026***
	(0.0021)	(0.0019)	(0.0029)	(0.0028)
90th percentile	0.051***	0.034***	0.072***	0.053***
	(0.049)	(0.0044)	(0.0070)	(0.0069)
Number of patients	566,146	566,146	566,146	566,146
Number of doctors	1,764	1,764	1,764	1,764
Number of hospitals	145	145	145	145

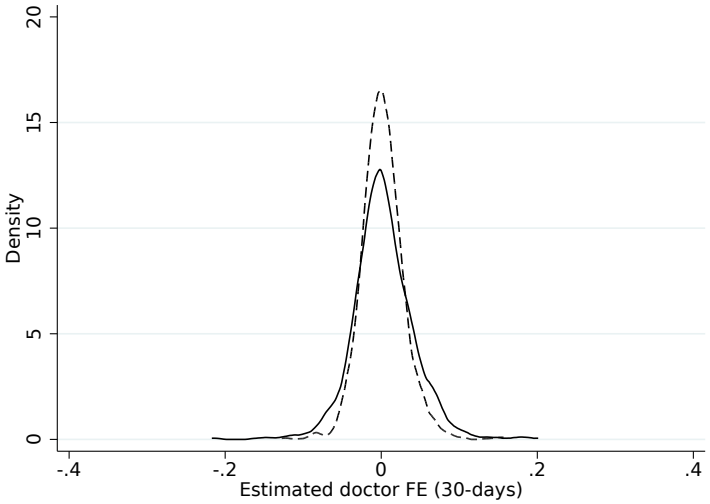
Notes: (1) Controls include a quadratic in age, sex (also interacted with age), ethnicity, charlson comorbidity index, the cost of any hospital treatment in the previous year, indicators of whether the patient has experienced a previous heart attack or stroke, a full set of fixed effects capturing primary and secondary diagnoses recorded at the time of admission, indicators of the day of the week, month and year when the patient is admitted, and an interaction between month and year of admission; (2) Bootstrapped standard errors clustered at the doctor level (199 repetitions).

Figure 2.3 shows the distribution of unadjusted and adjusted doctor fixed effects for 30-day (Panel A) and one-year (Panel B) patient survival. This demonstrates the impact on the distribution when the shrinkage procedure is applied, with both the left- and right-tail of the distribution becoming less pronounced when adjusting the estimates for statistical noise. It also displays the large differences in the performance of doctors at either end of the distribution. For example, using the adjusted estimates, moving between a doctor at the 10th percentile to

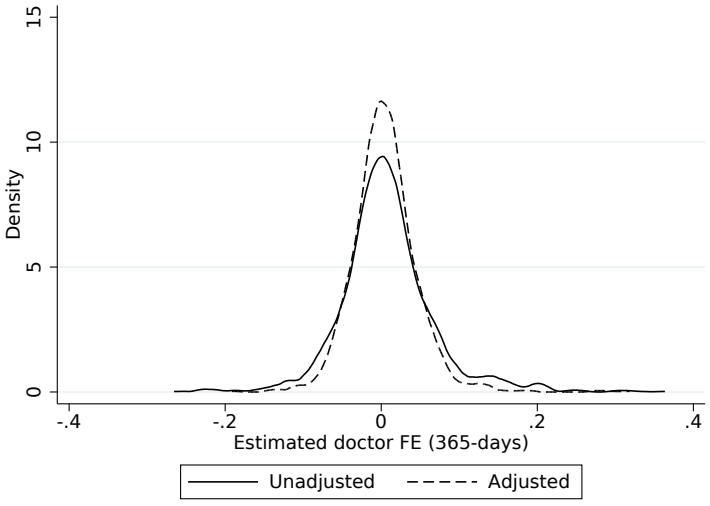
one at the 90th percentile reduces mortality by 9.7 percentage point, or two thirds of the mean mortality rate.

Figure 2.3: *The distribution of the raw and adjusted estimated of doctor fixed effects, by survival length*

(a) *30-day survival*



(b) *365-day survival*



Notes: (1) Doctor fixed effects from same regression as described in Table 2.4; (2) Adjusted fixed effects are adjusted using the procedure set out in Equation (2.6).

Comparing the estimated impacts over 30 days and a year suggest that the impact of being treated by a more effective doctor is persistent over time. In absolute terms, a one standard deviation increase in doctor quality reduces mortality by a larger amount over the course of a year than it does over a month. As a share of mortality, this effect is reduced as mean mortality rises over a longer period. Figure A.1.1 in Appendix A.1 plots the estimated 30-day impact on survival against the estimated 365-day impact for each doctor in the sample. There is a strong positive correlation (0.71). This suggests that doctors who perform well in the short-term are also likely to be those with good long-term outcomes.

Table 2.5 shows the estimates of the returns to specific and general experience from the same fixed estimates regressions as above. The first column shows the results when using 30-day survival as the dependent variable and the second columns shows results for 365-day survival. For the count of AMI and other patients, the variables are standardised so that they have mean zero and standard deviation one.

The results suggest that there is a small return to specialisation. Treating more AMI patients over the previous three years is associated with a small decrease in mortality over both a 30-day and a year long period. For example, a one standard deviation increase in the number of AMI patients treated over the previous three years (or 280 patients) reduces mortality by 0.24 percentage points over a year. This is equivalent to an improvement of just under 6% of a standard deviation in (adjusted) doctor quality.

The coefficient on the volume of other patients treated over the same period is negative for both 30-day and 365-day survival, but is only statistically significantly different from zero for the shorter period of time. The estimates suggest that a one standard deviation increase in the number of non-AMI patients leads to an increase in 30-day mortality treating of 0.6 percentage points, or 15% of a standard deviation reduction in doctor quality. This estimate reduces in magnitude over a longer period of time, and is no longer statistically significantly different from zero. Again, this provides weak evidence that focusing on treating a particular condition leads to better patient outcomes.

Taken together, these results suggest that there are some returns to specialised experience when treating heart attack patients, but these are relatively small when compared to the permanent differences in the performance across doctors.

Table 2.5: *Estimated impact of recent experience on one-year survival rates*

	Patient survival	
	30-day	365-day
	(1)	(2)
Experience (last 3 years)		
AMI patients	0.0019***	0.0024**
	(0.0007)	(0.0010)
Other patients	-0.0068***	-0.0043
	(0.0020)	(0.0028)
Observations	566,146	566,146
Number of doctors	1,764	1,764
Number of hospitals	145	145
R-Squared	0.148	0.214

Notes: (1) Estimates taken from same regression as described in Table 2.4; (2) Variables record the number of AMI and other patients treated over the past three years, and are standardised to have mean zero and a standard deviation of one; (3) Standard errors are clustered at the doctor level.

2.5.3 Robustness checks

The results reported above rely on the identifying assumptions outlined in Section 2.4.2. Below, I discuss these assumptions further and present evidence that such assumptions are likely to hold in practice. I first examine the empirical evidence around within-hospital quasi-random assignment of patients to doctors. I then explore other threats to identification, including the presence of hospital-specific trends in performance and potential match effects between hospitals and doctors. Finally, I set out alternative approaches to address sampling error arising from small samples.

Evidence of quasi-random assignment

As discussed in Section 2.4.2, identification of the doctor fixed effects relies on an assumption of within-hospital random assignment of patients to doctors. This seems highly plausible given the nature of heart attack treatment and the set-up of English hospitals. In addition, while this assumption cannot be tested directly, I can provide strongly suggestive evidence in support of it by examining the relationship between the estimated quality of doctors and characteristics of patients that are not included in the baseline control variables (e.g. in X_{it} in Equation (2.3)). If conditional assignment is random, then there should be no relationship between these observables and the estimated quality of the assigned doctor.

I demonstrate this using two sets of patient characteristics. First, I examine non-medical characteristics based on the local area in which the patient lives.[14] This includes local area deprivation, population density and the annual number of house sales. Second, I use two variables that capture medical attributes that were not included in the initial regression: total days spent in hospital and the number of emergency hospital admissions in the previous year. These variables should both capture information about the underlying health of patients, with sicker patients using more hospital services in the recent past.

I first show the evidence visually. Importantly, the assumption is that assignment is random conditional on the hospital and time of admission. I therefore first regress each characteristic on the full set of hospital dummy variables, admission day-of-the-week, and an interaction between admission month and year.[15] I then plot the residuals from these regressions against the estimated doctor quality for each patient estimated in Equation (2.3) in Figure 2.4. In all cases there are no obvious patterns that suggest that certain types of patients are systematically assigned to doctors with better or worse estimated quality.

To test this more formally, I estimate the following specification:

$$C_{ijkt} = \beta\hat{\mu}_j + \psi_k + \tau_t + u_{ijkt} \tag{2.7}$$

where C_{ijkt} is one of the patient characteristics described above. $\hat{\mu}_j$ is the estimated quality of doctor j, which has been standardised to have a mean of zero and a standard deviation of one. ψ_k is a hospital fixed effect, and τ_t includes indicators for admission day-of-the-week, and an interaction between admission month and year. Standard errors are clustered at the doctor level.

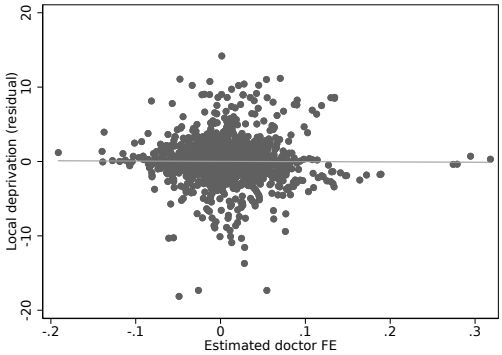
Table 2.6 shows the results. In all columns, the outcome is the estimated doctor fixed effect. In the first three columns, I regress this on each of the non-medical characteristics separately. Column 4 includes all three variables together. I then repeat this for the past health variables in columns 5-8, before including all variables together in column 8. In all cases, the magnitude of the coefficients are close to zero, and none are statistically significant. Taken together with the institutional factors discussed above, the evidence suggests that there is little meaningful selection between doctors and patients.

[14]Residence is located at the Lower Layer Super Output area (LSOA). This provides a small area around where the patient lives. In 2011, there were 32,000 LSOAs in England, with an average population of 1,500 people.

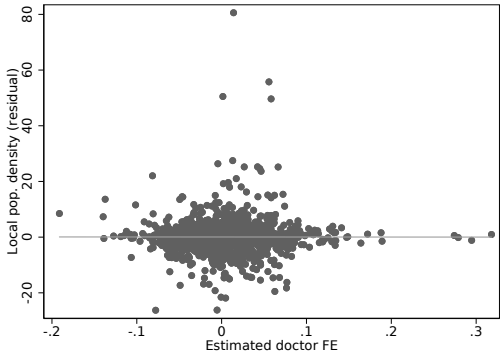
[15]The controls for hospital and time of admission are the same as in Equation (2.3). The assumption is that assignment is random conditional on these factors: one would expect a correlation between patient characteristics and estimated doctor quality when these controls are removed if populations with particular characteristics have access to better or worse quality doctors.

Figure 2.4: *Correlation between estimated doctor effects and selected mean patient characteristics*

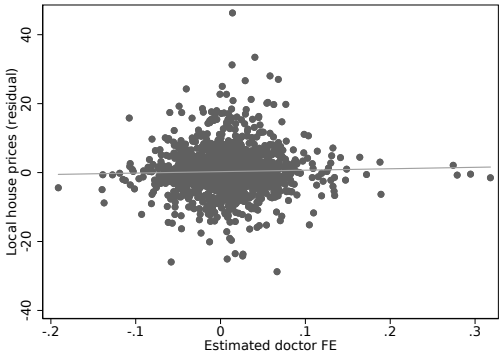
(a) *Local area deprivation*



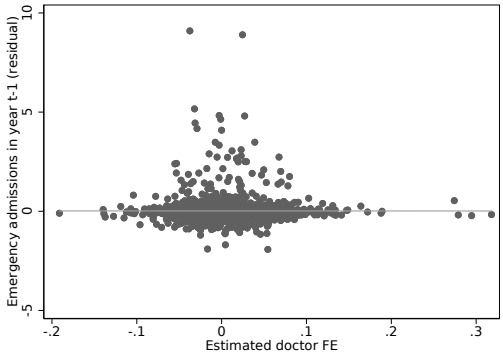
(b) *Population density*



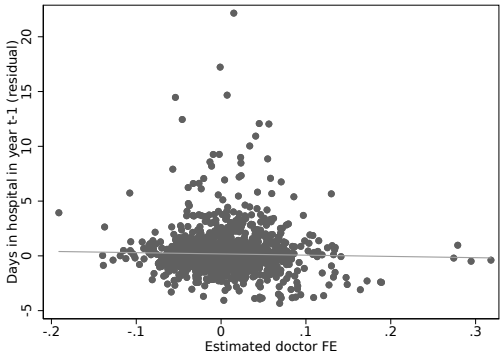
(c) *House sales*



(d) *Previous emergency admissions*



(e) *Previous days in hospital*



Notes: (1) All outcomes are residuals from an initial regression of the outcome of interest on hospital dummies, day of the week, and an interaction between month and financial year; (2) Panel A uses the Index of Multiple Deprivation (2004) from the local Lower Super Output Area (LSOA), Panel B uses LSOA population density from 2004, Panel C, captures the number of house sales in the Middle Super Output Area in 2004, Panel D uses the number of emergency admissions to hospital in the previous year, and Panel E uses the total number of days spent in hospital in the previous year.

Exploring further threats to identifcation

The evidence presented above suggests that patients and doctors are randomly matched within hospitals. However, two main threats to identification remain.

Table 2.6: *The relationship between estimated doctor effects and mean patient characteristics*

	Estimated doctor FE (1-year survival)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Non-medical characteristics								
Local area deprivation	-0.0008			-0.0017				-0.0017
	(0.0022)			(0.0023)				(0.0023)
Population density		0.0023		0.0033				0.0033
		(0.0028)		(0.0029)				(0.0029)
House sales			0.0019	0.0018				0.0018
			(0.0017)	(0.0018)				(0.0018)
Past health variables								
Days in hospital					-0.0002		-0.0003	-0.0003
					(0.0010)		(0.0010)	(0.0010)
Number of emergency admits						0.0011	0.0011	0.0011
						(0.0007)	(0.0007)	(0.0007)
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	566,148	566,148	566,148	566,148	566,148	566,148	566,148	566,148

Notes: (1) All variables are standardised to have mean zero and standard deviation equal to 1 (2) Local deprivation is measured by the 2004 Index of Multiple Deprivation score for the Lower Super Output Area of the patient (Local Deprivation); (3) Population density refers to the LSOA level in 2004; (4) House sales measure the number of house sales in the MSOA in 2004; (5) Past health variables include the total days in hospital and number of emergency admissions from the year prior to receiving AMI treatment; (6) All specifications control for hospital fixed effects, day-of-the-week and an interaction between month and year; (7) All specifications clustered at the doctor level, *** p< 0.01, ** p<0.05, * p<0.1.

First, the estimates may be biased if doctors systematically move away from hospitals that are becoming worse in quality. The estimates of μ are identified by comparing the performance of doctors within the same hospital over the entire 13-year period. If doctors leave a hospital that is becoming worse over time, or is taking on an increasingly severe mix of patients, their observed performance will improve relative to colleagues at the hospital at a later period of time. In the baseline estimates, hospital performance and patient mix is assumed to be constant over time, with deviations captured by λ_{kt} in the error term. Staff movements like this could therefore lead to a correlation between observed doctor performance and the unobserved deviations in hospital performance over time.

One way to examine how problematic hospital-specific performance is for the estimates to directly include hospital-year dummy variables in the estimation. Identification for doctor performance then arises from comparing the performance of doctors working in the same hospital in the same year, and by a connected set of doctors who work in multiple hospitals simultaneously. This makes the conditions for the connected set more restrictive, and limits the sample to a smaller number of doctors and hospitals. For this reason, I do not include hospital-year effects in the baseline estimates but instead present this as a robustness check. When

implementing these restrictions, the connected set is reduced to include 1,657 doctors treating 327,604 patients across 140 hospitals.

Table A.1.1 in Appendix A.1 shows the results of this exercise. In column 1, I repeat the baseline analysis for the smaller connected set, where I include time-invariant hospital fixed effects. This shows that the distribution of doctor quality is more widespread for this sample than in the baseline: a one standard deviation improvement in doctor quality when using the adjusted estimates leads to a reduction in mortality of 5.8 percentage points. Column 2 repeats this estimate with time-varying hospital-year effects included. The estimated standard deviation in quality is slightly larger, at 6.1 percentage points. However, the difference between the estimates is not particularly large. This suggests that hospital-specific deviations in outcomes over time do not have a meaningful impact on the estimates of permanent differences in doctor quality.

The second remaining threat to identification relates to potential match effects between doctors and hospitals. In the baseline model, doctors and hospitals are assumed to have additively separable impacts on patient survival. This rules out a scenario where doctors are more effective when working in particular hospitals, either because they are better suited to the available technology in each hospital or because they work better with other staff at specific locations. This assumption has been routinely criticised within the larger literature that examines worker and firm effects in wages and productivity (Eeckhout and Kircher, 2011; Woodcock, 2015). If doctors systematically choose to work in hospitals where they are better matched then this would bias the estimates of their persistent quality in any setting.

To explore whether an assumption of additive separability is plausible, I consider two pieces of evidence. Following Card et al. (2013), I first re-estimate the model with a fully saturated model that includes separate dummy variables for each hospital-doctor pair instead of separate doctor and hospital fixed effects. If the match effects are quantitatively important, then this model should provide a much better statistical fit of the data, as measured by an increase in the estimated R-squared. Carrying out this exercise suggests that there is essentially no difference between the two models, with the R-squared rising from 0.214 in the baseline estimated to 0.215 in the model with hospital-doctor pairs included.

An alternative way of testing whether there is a match effect between doctors and hospitals is to examine the difference in patient outcomes across hospitals for doctors who are observed treating patients in multiple locations. In the absence of specific physician-hospital effects, the difference in mortality outcomes for patients treated by doctor j in two separate hospitals should be explained by differences in the characteristics of patients treated in the two hospitals, and the estimated

hospital effect. I therefore test for evidence of match effects by examining variation in the differences between predicted and actual survival for each doctor across hospitals they work in.

I do this in the following way. First, I estimate a fixed effect for each hospital by regressing one-year survival on patient characteristics and a full set of hospital fixed effects. Second, I predict survival for each patient (absent hospital or doctor fixed effects) and take the mean predicted survival rate for each doctor in every hospital that they work in. This is then adjusted by the estimated hospital fixed effect to calculate the hospital-specific predicted survival rate of patients treated by every doctor in each hospital they work in. Taking the difference between the observed and predicted survival rates for each doctor-hospital pair provides an indication of doctor performance in each hospital that they work in.

While performance is unlikely to perfectly correlate due to differences in sample sizes (and associated measurement error) across hospitals, doctors who perform well in one hospital would be expected to do similarly well in another if match effects are small. I therefore sort doctors on the basis of the differences between predicted and actual survival in the lowest and highest ranked hospital that they work in, where hospitals are ranked on the basis of their fixed effect.[16] I then compare these doctor rankings across hospitals. In the absence of match effects, doctors should perform similarly across the different hospitals that they work in.

Table 2.7 shows that this indeed is the case. The table plots the quartile of doctors working in their 'worst' performing hospital against the quartile of the same doctor working in their 'best' performing hospital. Quartile 1 contains the 25% of doctors with the worst outcomes (after accounting for the characteristics of their patient and the hospital) while quartile 4 contains the 25% of doctors with the best outcomes. Each cell reports the fraction of doctors in a particular quartile in their worst performing hospital who appears in each quintile of performance in their best performing hospital. If performance at the two hospitals were unrelated, one would expect 25% of each quartile in the lower ranked hospital to appear in each quartile of the highest ranked hospital. This is clearly not the case here: 37.6% of doctors who perform best in their lower ranked hospital also appeared in the top quartile of their best performing hospital, while only 16.2% appeared in the lowest performing quintile. Similarly, 39.3% of those in the bottom quartile in their lower ranked hospital were also found in the bottom quartile of their top ranked hospital, while only 12.8% appeared in the top quartile.

Taken together, this evidence suggests that match effects between physicians and doctors are unlikely to be large in practice.

[16] This hospital 'ranking' will in part reflect differences in patient severity rather than quality of care.

Table 2.7: *Within-consultant correlation of differences between actual and predicted survival across hospitals*

	Quartiles in best performing hospital			
Quartiles in worst performing hospital	1	2	3	4
1	39.3	29.9	18.0	12.8
2	22.2	30.8	29.1	18.0
3	22.2	16.2	29.9	31.6
4	16.2	23.1	23.1	37.6

Notes: (1) Hospitals are ranked by fixed effect (from the most negative to the most positive) and reflects differences in quality and permanent patient mix; (2) Doctors are ranked within their lowest and ranked hospitals, and are assigned to quartiles on the basis of the difference between actual survival rates and predicted survival rates (on the basis of patient characteristics and hospital fixed effects); (3) Patient survival is predicted by regressing one-year survival on a quadratic in age, sex, an interaction between age and sex, race, charlson index, an ambulance dummy, treatment costs in the previous financial year, whether the patient suffered an AMI or stroke in the previous year, a set of dummy variables for primary and secondary diagnoses, day of the week, and an interaction between month and year of admission; (4) Hospital fixed effects are estimated by regressing patient survival on the set of controls listed in (3) and a full set of hospital fixed effects.

Addressing small samples

As noted in Section 2.4.3, variation in the estimated doctor fixed effects may be overestimated if these estimates are based on small samples for individual doctors. If sample sizes for each doctor are too small, then individual estimates of quality will be very unreliable due to statistical imprecision. This has been a common criticism of the practical use of estimates of teacher value-added to evaluate the performance of teachers (McCaffrey et al., 2009; Schochet and Chiang, 2013). I therefore now examine in detail whether my estimates of doctor quality are likely to be driven by statistical noise.

I examine this in two ways. First, I examine how the (adjusted) distribution of results change when I progressively restrict the sample to include only doctors who treat a large number of patients. Columns 1-3 of Table 2.8 show the results of this exercise. The first column shows the baseline results, when all doctors treating a minimum of 10 patients within the connected set are included. In column 2, I increase this minimum number of patients to 50. This has no meaningful impact on the estimate of a standard deviation of doctor quality, or on other parts of the distribution. In column 3, I increase this further to a minimum number of 100 patients. Again, there is little impact on the estimated distribution: under the more restrictive set, a one standard deviation increase in doctor quality is associated with a 3.9 percentage point reduction in one-year mortality, compared to 4.2 percentage points in the baseline case. The 90-10 ratio also becomes slightly

smaller. However, these differences are relatively small and do not change the qualitative conclusions of the results.

Table 2.8: *Estimated distribution of doctor quality under alternative empirical implementations*

	365-day survival			
	Minimum number of patients per consultant			
	10 1-step (1)	50 1-step (2)	100 1-step (3)	10 2-step (4)
Std Deviation	0.042	0.044	0.039	0.035
Variance	0.002	0.002	0.002	0.001
10th percentile	-0.044	-0.043	-0.04	-0.035
25th percentile	-0.020	-0.02	-0.021	-0.013
50th percentile	0.003	0.002	0.001	0.001
75th percentile	0.026	0.024	0.023	0.017
90th percentile	0.053	0.5	0.047	0.038
Number of patients	566,146	556,804	540,125	566,146
Number of doctors	1,764	1,399	1,159	1,764
Number of hospitals	145	145	143	145

Notes: (1) All specifications include the same controls as the results in Table 2.4; (2) Column 4 uses the 2-step procedure outlined in Section 2.5.3.

Second, I implement an alternative shrinkage procedure. Bitler et al. (2019) show that standard shrinkage techniques can fail to fully account for statistical noise, as demonstrated by showing non-zero estimated effects of teachers on outcomes that they cannot feasibly impact (i.e. height). They find that a zero result is only recovered when using the covariance in effects from multiple years of classroom data to estimate the signal-to-noise ratio. This follows an approach by Kane and Staiger (2008) that can be implemented only for teachers who are observed working in multiple years or classrooms. In this case, the covariance between teacher residuals across years or classrooms can be used to estimate the signal component.

Such an approach is possible to implement in my setting, with doctors observed treating patients over multiple years. I therefore compute estimates using the method of Kane and Staiger (2008). This uses a two-step approach: first, I regress patient survival on my full set of controls and hospital fixed effects. This yields a residual for each doctor in each year, which is then regressed on a full set of doctor fixed effects. I take from this estimates of σ_u and σ_e, which I use to estimate the

shrinkage estimator in Equation (2.6), and apply this to the estimates of doctor quality.

Column 4 in Table 2.8 shows the results. Using this method does decrease the variance in doctor fixed effects somewhat: a one standard deviation increase in doctor quality reduces mortality by 3.5 percentage points, or 24% of mean mortality. This is 17% smaller than the baseline estimates. The tails of the distribution are also less extreme: moving from the 10th to the 90th percentile of the estimated distribution would reduce mortality by 7.3 percentage points as compared to 9.7 percentage points in the initial estimates. However, once again this does little to alter my original conclusions: there remains substantial variation in the permanent differences in patient outcomes achieved by individual doctors.

2.6 Reallocating doctors to reduce mortality

As noted in Section 2.3, the baseline model restricts doctors to be equally effective at treating all patients, regardless of the underlying characteristics of patients. However, doctors may vary in their ability to treat different patients based on the severity of their condition, or their wider characteristics. In particular, there are multiple ways to treat heart attack patients, with some treatment types more suitable for patients with particular symptoms. Previous work has shown that doctors often prefer to treat patients for heart attacks in a particular way, regardless of the suitability of such treatment in the case of that specific patient (Chandra and Staiger, 2007; Currie et al., 2016). More generally, some doctors may be more effective when treating different patient types. As a result, patient outcomes could potentially be improved if patients were reassigned to doctors who perform best when treating patients of their particular type.

To examine this further, I modify the model to allow doctors to vary in the quality of care that they provide to patients on the basis of observed patient severity. I then examine the potential reductions in patient mortality that could be achieved by reallocating the existing set of doctors across patients of different types under different scenarios.

2.6.1 Varying doctor quality across patient types

To examine whether doctors vary in their effectiveness when treating patients with different needs, I estimate the quality of each doctor when treating patients of two different types. To do this, I first predict mortality for each patient using all the patient characteristics included in the baseline specification (but omitting any information about the doctor or hospital of treatment). I then split the sample in

half on the basis of this predicted mortality to define two types of patients: 'low severity' (L) patients are those with below-median predicted mortality, while all other patients are classified as 'high severity' (H).[17]

Using this classification, I modify Equation (2.3) to estimate the following specification:

$$Y^s_{ijkt} = \beta X_{it} + \mu^s_j + \delta.Exp_{jt} + \psi_k + \epsilon_{ijkt}, \quad \forall s = [L, H] \tag{2.8}$$

Y^s_{ijkt} is a binary variable that measures survival over 365 days for a patient of type s. μ^s_j now measures the permanent difference in patient outcomes for doctor j when treating a patient of type s. All other elements of the specification are the same as in Equation (2.3).

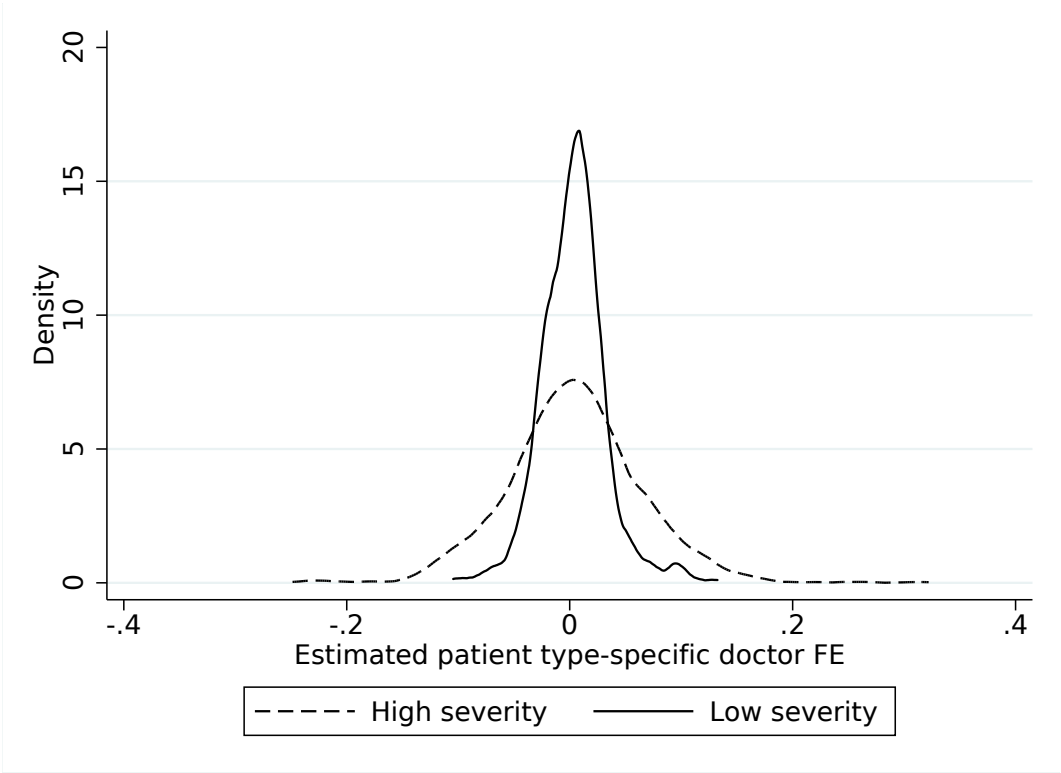
This yields two estimated fixed effects for each doctor (μ^L_j, μ^H_j).[18] Figure 2.5 shows the distribution of the estimated effects for each patient type. Given the higher mortality rates among high severity patients, the figure shows that there is greater (absolute) variation in the performance of doctors when treating high severity patients: improving quality by one standard deviation when treating high-severity patients is equivalent to a reduction in mortality of 6.0 percentage points compared to a reduction in mortality of 2.9 percentage points when increasing doctor quality by one standard deviation when treating low-severity patients.

Table 2.9 shows how the ranking of doctors compares across low and high severity patients, correlating the quartiles of the two fixed effects for each doctor. Quintile 1 shows the 20% of doctors with the lowest (i.e. most negative) impact on patient survival, while quintile 5 shows the 20% of doctors with the highest impact on patient survival. Each cell reports the fraction of doctors in a particular quartile when using the low severity fixed effects who appears in each quintile of the high severity fixed effect ranking. It shows that doctors ranked in the top quintile in treating low severity patients are also more likely to be ranked in the top quintile: 31.3% of doctors ranked in the top quintile of the low severity distribution appear in the highest quintile of the high severity distribution (compared to 20% if this was drawn at random). The worst performers when treating low severity patients are the least likely to be among the top performers when treating high severity patients, with only 11.3% of those in the lowest ranked 20% of doctors for low severity patients appearing in the top quintile when treating high severity patients. The worst performing doctors when treating low severity patients are also more likely to be in the bottom quintile for doctors treating high severity patients.

[17] The one-year mortality rate among low severity patients is 3.7% compared to 25.2% among high severity patients.

[18] 11 doctors only treat high severity patients. To simplify the reallocation exercise below I assign a zero effect to those with missing values for $\hat{\mu^L}$.

Figure 2.5: *Distribution of estimated doctor fixed effects, by patient type*



Notes: (1) Estimates from Equation (2.8) are adjusted using the one-step method used for the baseline results shown in Table 2.4.

These correlations indicate that some doctors are better than others when treating all patients types. However, for some doctors their performance is significantly better when treating a particular type of patient: for example almost a fifth of doctors who were ranked in the top quintile when treating low severity patients are ranked in the bottom fifth of doctors treating high severity patients. This suggests that patient mortality could potentially be reduced by allocating doctors to treating particular patient types where they have a comparative advantage.

2.6.2 Reallocating doctors across patient types

I now consider different scenarios where doctors can be matched to specific patient types on the basis of their comparative advantage in treating each patient type. Intuitively, matching doctors who are comparatively better at treating high severity patients to those with a high predicted mortality could substantially reduce mortality among heart attack patients.[19]

[19]Throughout this exercise I assume that the experience effects of treating different patient types are the same: if doctors improve more rapidly when treating high severity patients then there would be a trade off between matching consultants in the earlier stage of their career with high severity types in order to reduce future mortality regardless of their innate comparative advantage in treating these patients.

Table 2.9: *Correlation of quintile rankings of doctors when treating low and high severity patients*

	Quantiles of $\hat{\mu}^H$				
Quantiles of $\hat{\mu}^L$	1	2	3	4	5
1	24.7	19.3	25.2	19.6	11.3
2	18.5	22.4	21.0	21.9	16.2
3	21.3	19.9	19.6	20.7	18.5
4	16.2	22.5	20.2	18.0	23.1
5	19.3	15.8	13.7	19.9	31.3

Notes: (1) Doctors are ranked according to their estimated fixed effect for each patient type, with the worst performing 20% of doctors in quantile 1 and the best performing 20% of doctors in quantile 5; (2) Fixed effects estimates are the same as those shown in Figure 2.5.

The estimated comparative advantage of doctor j in treating a high severity patient rather than a low severity patient is simply the difference between the estimated fixed effects (e.g. $\hat{\mu}_j^H - \hat{\mu}_j^L$). I use the estimates from Equation (2.8) to estimate the comparative advantage for each doctor. I then rank doctors according to this difference, with those with the greatest comparative advantage in treating high severity patients ranked higher. These rankings enable me to match doctors with particular patients under different constraints.

I consider three scenarios. First, I consider the potential reductions in mortality if the worst performing 10% of doctors for low and high severity patients were replaced with a doctor that had zero impact. This is similar to classic experiments in the teacher value-added literature (Hanushek, 2009; Chetty et al., 2014b). This provides a baseline case of mortality reductions that could be reduced by simply replacing the worst performing doctors with better performing ones.

However, simply replacing doctors would be hard in practice, with long training periods required. I therefore examine two further scenarios where I reallocate existing doctors across patients. This includes an 'unconstrained' and 'constrained' exercise. In the 'unconstrained' reallocation exercise, I hold the caseload of each doctor constant and reallocate doctors towards patients of particular types regardless of the hospital in which these doctors work. Specifically, I allocate high severity patients to doctors with the highest marginal benefit, allocating patients first to the doctor with the highest comparative advantage in treating high severity patients, and then to the next highest ranked doctor until all high severity patients are allocated. All remaining patients are assigned to unallocated doctors. This scenario is unconstrained in the sense that I do not place limits on the location where doctors can treated patients: high severity patients can be located in many hospitals in which the doctor does not work in practice. This therefore represents

an upper bound for reductions in mortality that could be achieved by reallocating existing staff, but would unlikely to be fully workable in reality.

As a result, I also examine a third scenario. In this 'constrained' reallocation exercise, doctors can now only be reassigned to patients within a hospital that they currently work in.[20] This is a more feasible scenario in the sense that doctors would be better able to access these patients.[21] In this case, doctors are ranked on their comparative advantage within-hospital. Patients are then reassigned to doctors in the same way within hospital.

Under the first scenario where I simply replace the worst performing dcotors, the number of deaths within a year for patients treated between April 2017 and March 2018 would be reduced by 376, a 4.8% reduction from the 7,762 deaths observed in the data. This would reduce the mortality rate from 12.8% to 12.1%. Such a change is equivalent to improving the average doctor quality by 0.17 standard deviations as reported in the baseline results (Table 2.4).

In the 'unconstrained' scenario, matching doctors to patients in this way reduces the number of deaths by 1,475 in 2017-18, or a reduction of 19% on the observed number of deaths. This reduces the mortality rate from 12.8% to 11.7%. This is equivalent to improving average doctor quality by a half a standard deviation, and represents a substantial improvement in patient outcomes.

The 'constrained' scenario also produces substantial improvements in patient outcomes compared to observed mortality rates, albeit smaller than in the scenario when doctors can be reallocated to patients anywhere. The estimates indicate that allocating doctors to patients within-hospital could reduce deaths in 2017-18 by 680 patients (9%). This is equivalent to a reduction in the one-year mortality rate from 12.8% to 11.7%. The same improvement could be made by raising the average doctor quality by 0.26 standard deviations.

Taken together, the results suggest that substantial improvements in patient survival rates could be achieved by reallocating doctors towards patients to which they are better matched with. In practice, these gains may be hard to fully achieve given the time constraints inherent in treating heart attack patients. However, these exercises do suggest that some gains could be made by rescheduling the shifts of senior doctors or moving doctors across hospitals to better match doctors and patients.

[20]For example, if a doctor treats 10 patients in hospital A and 15 patients in hospital B in the observed data in 2017, they will have the same caseloads in the simulation.

[21]I do not constrain the number of days, or the time between patients. In practice doctors would not be available at all times. A further restriction would therefore be to reassign doctors to particular days or shifts where a greater mix of patients are of a particular type.

2.7 Conclusion

Variation in patient outcomes across places and providers is increasingly becoming a concern for policymakers around the world. These differences in the quantity and quality of care provided to different patients risk exacerbating already existing health inequalities. Medical staff - and doctors in particular - are likely to play a key role in driving this variation. Understanding this role in greater detail is therefore an important step towards addressing these inequalities.

In this paper, I exploit features of the English National Health Service to estimate the quality of individual doctors. I show that doctors account for 5.5% of variation in patient survival rates in the year after suffering a heart attack. This is almost twice the share of variation explained by the wider features of the hospital where treatment is delivered. I also find that there is substantial variation in the quality of individual doctors: a one standard deviation improvement in doctor quality reduces mortality by 4.2 percentage points, or more than a quarter of mean mortality rates over a year.

I then extend this model to show that doctors vary in their quality when treating patients of different severity. Dividing patients into low and high severity groups, I estimate the comparative advantage of each doctor in treating high severity patients. I use these estimates to study several scenarios where doctors are replaced or reassigned across patients. This exercise reveals that there are large potential gains for patients in reallocating doctors: reallocating doctors across patients within the hospitals that they work in to better match with the most appropriate patients is estimated to reduce mortality by 9%.

These findings have a number of implications. First, the results suggest that patient outcomes could be improved by reallocating existing staff resources to treat certain patients. Even within a narrow field of medicine, doctors require different skills in treating different patients. Matching doctors to the appropriate patients therefore offers significant efficiency gains. However, reallocating these doctors may face significant practical challenges. In this work, I have abstracted from the other tasks that senior doctors are doing: reallocating doctors across hospitals or shifts may take them away from other tasks that they perform well. Similarly, even the exercise that considers reallocating doctors within hospitals does not take into account the difficulties of scheduling doctors to be available at the times when particular patients enter the hospital. Future work could consider whether patient types vary systemically across different times or days of the week, with doctors reassigned to shifts where the greatest number of appropriate patient types are likely to attend.

Second, the results show that doctors vary considerably in the results that they achieve for patients, even when they are already very experienced and highly qualified. This variation is likely to be even greater among more junior staff in earlier parts of their career. Future work should concentrate on studying why these differences emerge. Access to information on the earlier careers of these doctors, including histories of where and who they have worked with, is essential in understanding the formation of skills over the course of an entire career. Previous work suggests that peer effects are important in the formation of doctor habits (Molitor, 2018), and any future extension of this work should consider the scope for peer effects in this setting. Such work would help to answer the question of what makes a 'good' physician, and could ultimately lead to improvements in the way that physicians are recruited and trained.

Chapter 3

Saving Lives by Tying Hands: The Unexpected Effects of Constraining Health Care Providers[1]

3.1 Introduction

Perhaps the most complicated node of health delivery in any modern health care system is the emergency department (ED). Patients arrive at the ED with a wide array of different problems. ED nurses and physicians must quickly assess where patients should slot in what can be a very large queue, deciding almost instantly who needs to be treated right away and who can wait. And ultimately these providers need to decide whether those going to the ED are to be admitted to the hospital or sent back to their homes – a decision that can, in many instances, have life or death consequences.

Despite its critical role, EDs often face budgetary pressures and a shortfall in resources. These pressures have been especially acute in recent years, with ED performance having been described as an international crisis in several developed economies (Hoot and Aronsky, 2008). Practising doctors are especially vocal, referring to 'battlefield medicine' and 'third world conditions' caused by ED overcrowding in England.[2] Alongside these tensions, EDs are increasingly facing public pressure to advertise and reduce their wait times. U.S. cities are replete

[1]This paper is joint work with Jonathan Gruber and Thomas Hoe. Thanks to Richard Blundell, Auro de Paula, Eric French, Peter Hull and Henrik Kleven for helpful comments on earlier drafts, as well as seminar participants at the IFS, MIT, UCL, NBER Summer Institute 2018, the Kellogg Healthcare Market Conference 2019 and NASMES 2019.

[2]https://www.nytimes.com/2018/01/03/world/europe/uk-national-health-service.html?smid=tw-share&_r=0

with digital billboards highlighting wait times at local EDs. And other nations use regulatory and financial tools to reward reductions, or penalize increases, in wait times.

Many are concerned that external pressures on wait times could reduce the ability of EDs to maximize the quality of the care that they provide. At the same time, however, it is not clear that ED personnel would maximize patient quality in the absence of such pressures. Emergency rooms are not directly compensated for shortening wait times. Moreover, while health-maximizing ED personnel may internalize the costs of waiting to the extent that they impact patient outcomes, this may only be partial if physicians have incomplete knowledge or are imperfect agents for their patients. Theoretical ambiguities such as this have motivated a growing number of empirical studies of hospital production in the ED setting (Chan, 2016, 2017; Silver, 2016).

In this paper, we provide new evidence on the impacts of regulating doctors in the ED - and in particular, putting pressure on them to make decisions more quickly - on treatment decisions and patient outcomes. To do this, we use the 'four-hour wait' policy in England. This policy was first announced in 2000 as part of a wide ranging set of government pledges to decrease wait times for different types of care, and came into force in all English public hospitals in 2004.[3].[4] The ability of hospitals to meet this target became an important part of overall hospital evaluation in England, with managers in some cases losing their jobs because of poor wait time performance. In addition, there were strong financial penalties associated with breaching the target – hospitals were penalized by an amount that was more than twice the average revenue of an ED patient, and total fines for missing ED and elective wait time targets were equivalent to a third of hospital deficits.[5]

Despite this focus on the target, there is little consistent evidence from either the UK or other nations that have introduced wait time targets on the impact of those targets on patient costs and health outcomes. This is because the policies are generally introduced nation-wide, with no 'hold-out' or control populations, making it impossible to apply quasi-experimental methods such as difference-in-difference estimation. An additional challenge in the case of the English wait time

[3]Other targets included maximum limits on wait times for elective surgery. The policy sets arbitrary targets for wait times, with 95% of all patients required to be treated within four hours of arrival

[4]The initial target stated that 98% of patients should be treated within 4 hours but this level was reduced to its current level of 95% in 2010.

[5]English hospitals have no other financial incentive to shorten wait times, or monitor the impacts on patient outcomes. Hospitals receive payments that vary by ED diagnosis group but not by wait time or health outcome.

policy is that no systematic data on wait times are available before the policy was introduced in 2004.

We therefore take a different approach in order to estimate the effects of the policy on treatments, costs and patient outcomes. We apply the bunching techniques that have been used widely in other contexts (see Kleven, 2016) to analyze wait times and outcomes using administrative hospital data from 2011 to 2013, a period when the policy was already in place. This approach allows us to model how the four-hour target impacts wait times, costs and outcomes, conditional on the underlying hospital technology in place to monitor patient wait times without using pre-policy data. That is, we estimate here the short term impact of changing wait times, but hold constant the underlying technological changes that might be associated with the introduction or removal of a wait time target and the prioritization of patients treated in the ED. This counterfactual focuses attention on the impact of incentives rather than technology adoption or on broader changes to the way that hospitals treat ED patients.

We initially examine the distribution of wait times around the four-hour target, where we define 'wait times' as total time spent in the ED (including the time being examined and treated) consistent with the definition of the policy. We find a very large spike right at four hours. We then turn to estimating counterfactual distributions of wait times in order to measure the effect of the four-hour policy. We estimate that, relative to the counterfactual, the four-hour target led wait times to be 21 minutes (8%) lower for patients affected by the policy, and for those patients that move from after-to-before the four hour position, the wait time reductions are large and average 59 minutes.

The regulations may also change the treatments provided by doctors and the outcomes of their patients. For example, doctors may order fewer tests or treat patients less intensely as a result of the policy. This could have negative effects on health outcomes, while reductions in wait times may be beneficial for patients. We therefore use the data to also study the impact of the policy on patient treatment and outcomes. Without pre-period data and exogenous variation in policy effects across hospitals, we cannot directly use data on treatments and outcomes to identify policy effects. But we argue that under a set of testable assumptions we can directly identify policy effects from bunching at the four-hour target.

Plotting these treatments and health outcomes conditional on the wait time reveals spikes just before the four-hour wait time. We can then decompose these spikes into two separate channels. First, there is a 'composition effect'. If the target causes patients to be moved from later to earlier in the distribution of wait times, and the characteristics of patients also vary across this wait time distribution, then the observed change in outcomes prior to the four hour target

will in part reflect this movement of patients. For example, admission probability is increasing with wait time (as more severe patients undergo more extensive testing and treatment in the ED before they are admitted). Moving patients from just after four hours to just before will cause the average admission probability of patients seen prior to the target to increase, even if the target has no impact on the probability of admission for each individual patient. Second, there may be an additional 'distortion effect' if the target itself leads to direct changes in the treatment received by patients, or in their outcomes.

In order to separately identify the distortion effect, we estimate a 'composition-adjusted counterfactual outcome' by imposing a 'no-selection' assumption on the distribution of patients that obtain shorter wait times because of the policy. This assumes that patients who are moved forward as a result of the wait time target are representative of those who are not. Under this assumption, we can use the observed outcomes of patients treated just after four hours to adjust the observed outcomes of patients treated just before four hours for these compositional changes. Comparing these 'composition-adjusted counterfactual outcomes' with observed outcomes therefore provides an estimate of the 'distortion effects' of the policy.

We can test this 'no selection' assumption directly using patient characteristics such as age, sex and past health status. These variables cannot be changed by the hospital at the time of the ED visit, and so by definition, any observed spikes in these outcomes should be due solely to a composition effect (i.e. the distortion effect is zero). Consistent with our assumption, we show that along multiple dimensions there is little meaningful difference between patients who are moved forwards and not. In the rare cases where there does appear to be non-random movement of patients, the evidence suggests that patients who experience wait time reductions are slightly more severe than those who do not. Importantly, if this does reflect differences in the unobserved severity among these patients then this would suggest that any estimated health benefits of the policy are underestimates of the true effects of the target.

Our analysis also relies on a 'local effects' assumption. This assumes that the wait time and treatments of patients outside of an 'exclusion window' around the four-hour mark are unaffected by the target. This would be violated if doctors substitute resources away from patients in the early part of the wait time distribution in order reduce waits for patients in danger of breaching the target. We argue that institutional factors make such behaviour unlikely, and present a range of empirical tests to support this. We also show that while the exact magnitude of our estimates are sensitive to some choices of parameters used in the estimation, our overarching conclusions are extremely robust.

We estimate that there is a significant distortion effect of the English policy. We find that there is more intensive testing of patients in the ED, leading to a modest rise in ED costs. We also find that there is a significant increase in hospital admissions as a means of meeting the target, with corresponding reductions in those discharged to home. Among those marginal admits, inpatient resource use is insignificant, suggesting that such admissions were just placeholders to meet the four-hour target. These admissions were not costless, however, and we estimate that inpatient payments from the government to hospitals rose by roughly 5% due to the target.

Most interestingly, we find significant improvements in patient outcomes associated with the four hour policy. We estimate that 30-day patient mortality falls by 14% among patients who are impacted by the wait time change, a very sizeable positive effect. This effect falls slightly over time while baseline mortality rises, so that by one year after ED admission this amounts to a 3% mortality reduction, which is still quite large.

We then turn to understanding the mechanism behind the outcome improvement that we observe. To do so we exploit heterogeneity across patient groups that are affected along different margins. The first is patients of different severity: across severity groups, the four-hour policy is associated with differential impacts on wait times, but not admission probabilities. The second is patients facing different levels of crowding of the inpatient department when they arrive at the ED: across different levels of crowding, the four-hour policy is associated with differential impacts on admission probabilities but little variation in the wait times impacts. We then show that the mortality effect we estimate varies strongly across patient severity, but not across inpatient crowding. Taken together, this evidence suggests that it is the wait time mechanism, and not the admissions mechanism, that is driving our mortality effect. As a final check, we examine whether the reductions in mortality occur among patients with potentially time-sensitive conditions, and find that the majority of these reductions are found among conditions which are known to benefit from rapid treatment.

We contribute to two literatures. First, there is a growing literature that has begun documenting features of hospital production relevant for incentive setting (Chan, 2016, 2017; Silver, 2016). Chan (2016) and Chan (2017), for example, study how ED physicians respond to team environments and work schedules, while Silver (2016) studies peer effects in the ED. Adjacent to these studies, a medical literature has documented robust correlations between mortality rates and measures of ED crowding and wait times (Hoot and Aronsky, 2008). Our contribution is to show how ED production is affected when doctors are put under pressure to make decisions quicker. We find that the wait time policy generated cost-effective mor-

tality improvements through reduced wait times but at the expense of distorting medical decisions. These findings are consistent with the medical literature and highlight that ED wait times are an important input to the health production process. The findings also illustrate how constraining healthcare providers through regulatory interventions can improve health outcomes even in the presence of significant distortions.

The second contribution we make is to the literature using bunching estimators. From its origins in the tax setting (Saez, 2010; Chetty et al., 2013; Kleven and Waseem, 2013), these estimators have now been deployed in other settings such as health insurance (Einav et al., 2015, 2017, 2018), mortgage markets (Best et al., 2017; Best and Kleven, 2018) and education (Diamond and Persson, 2016). We apply these estimators in a healthcare provision setting, adapting them to study outcomes indirectly affected by a discontinuity in the incentives associated with the running variable, and devise new empirical tests to evaluate the credibility of the bunching assumptions required in our context.

Our paper proceeds as follows. Section 3.2 provides background information on emergency care in England and on the four-hour target policy. Section 3.3 describes the data. Section 3.4 sets out our methodology. Section 3.6 describes our results for wait times, treatment decisions and health outcomes. Section 3.7 explores heterogeneity and mechanisms. Section 3.8 concludes.

3.2 Background

3.2.1 Emergency care in England

Emergency care in England is publicly funded and is available free at the point of use for all residents. There is no private market for emergency care. The majority of care is provided at emergency departments (EDs) attached to large, publicly owned hospitals. These major emergency departments are physician-led providers of 24-hour services, based in specifically built facilities to treat emergency patients that contain full resuscitation facilities. In 2011/12, 9.2 million patients made 13.6 million visits to 174 emergency departments. In addition, 2.1 million patients made an additional 2.7 million visits to specialist emergency clinics and 'walk in' or minor injury centres where simple treatment is provided for less serious diagnoses; as discussed below, we exclude patients from these centres due to the minor nature of their injuries and our results are unaffected if they are included.

EDs provide immediate care to patients. Hospitals are reimbursed by the government for the care they provide, receiving a nationally fixed payment for

providing certain types of treatment.[6] In 2015/16, there were 11 separate tariffs for ED treatment depending on the severity of the patient and the type of treatments administered.[7] These tariffs ranged from $77 to $272 (£57 to £200) per visit.[8] Revenue from the ED accounted for 5.3% of total hospital income in 2015/16.[9]

Treatment in the ED follows one of two pathways depending upon the method of arrival. Non-ambulance patients register at reception upon arrival, where they must identify themselves and provide basic details of their condition. Patients then undergo an initial assessment to establish the seriousness of their condition. This triage process is carried our either by a specialist triage nurse or doctor, and includes taking a medical history, and, where appropriate, conducting a basic physical examination of the patient. Patients are then prioritized according to severity.

Alternatively, patients can arrive at the ED by ambulance following an emergency call out. In 2011/12, 29.4% of ED patients arrived by ambulance. For these patients, ambulance staff collect medical details en route, and report these details to hospital staff upon arrival.[10] This information feeds into a separate triage process, where patients will be categorized by their severity.

These triage processes sort patients into 'minor' and 'major' cases. Minor cases require relatively simple treatment, and can often be treated in a short space of time. Major cases are often those who arrive by ambulance, although there are some exceptions to this (for example, a patient with chest pain may arrive independently at the hospital). Major cases will receive treatment more quickly, as they often present with more severe symptoms, but will usually require more treatment and investigations within the ED, and are therefore likely to spend longer in the ED. Treatment of the two types often requires the use of different resources (including staff and machines), and in most large hospitals, treatment for minor conditions will take place in a separate part of the emergency department (for example, in the hospital's 'urgent care centre'). In particular, senior staff time is typically concerned on treating major cases, with these staff having little

[6]Treatments are assigned to a Healthcare Resource Group (HRG), similar to DRGs in the US, with a set of national tariffs for each HRG announced each year by the Department of Health.

[7]https://www.gov.uk/government/publications/confirmation-of-payment-by-results-pbr-arrangements-for-2012-13

[8]All cost figures in 2017/18 US Dollars. Figures are deflated using the UK GDP deflator, and then converted from sterling to dollars using an exchange rate of 1GBP:1.35USD (US Treasury, 31st Dec 2017, https://www.fiscal.treasury.gov/fsreports/rpt/treasRptRateExch/currentRates.htm).

[9]Figures calculated from the 2015/16 UK Department of Health Reference Costs. See: https://www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016

[10]Ambulance staff also provide emergency treatment in the ambulance to patients where required.

interaction with minor cases (except to sign off admission or discharge decisions made by junior staff).

Following triage, patients are placed into a queue on the basis of their severity and time of arrival. Patients are not aware of their position in the queue. Patients are assigned to individual doctors as they become available. These doctors will carry out a series of further examinations and tests. The nature of these investigations depend on the symptoms presented by the patients, and range from physical examinations to tests such as x-rays or MRI scans. Patients can also receive treatment in the ED, ranging from sutures to resuscitation, before being admitted for further treatment in an inpatient ward, or discharged from the hospital.[11]

3.2.2 The 4-hour target

All public hospitals with EDs in England are subject to a wait time target. This target specifies that 95% of ED patients must be admitted for further inpatient treatment, discharged or transferred to another hospital within four hours of their arrival. While the target is officially a 'wait time' target, the definition employed – which includes the time being examined and treated in the ED – corresponds more precisely to the total time a patient spends physically in the ED. We use the terminology and definition of 'wait times' consistent with the policy throughout this paper. The target level was initially set at 98% when it was first introduced in December 2004, before being relaxed to its current level in November 2010.[12]

This target is important to hospitals in two ways. First, the target is widely used by policy makers and the media as a measure for the wider performance of the public health service in England.[13] Hospital managers who consistently fail to meet this target are likely to be fired, and therefore have a strong incentive to organise emergency care in a way that minimises the number of patients who take more than four hours to treat.

Second, hospitals face significant financial incentives to meet the target. As the target came into force between March 2004 and March 2005, hospitals were offered payments (to be used only for hospital investment) if they met the target level early (National Audit Office, 2004). In recent years, significant financial penalties have been imposed for missing the target. In 2011/12, hospitals were fined $300

[11]See Appendix B.3 for further details on the range of treatments and investigations received by ED patients.

[12]Interviews with hospital managers, doctors and regulators suggest that it is the 'four-hour' component of the target that matter to hospitals rather than the absolute level of the target. Hospitals attempt to meet the target on a daily basis, and aim to achieve the highest proportion possible. This suggests that certain behaviours, such as relaxing or improving performance in later parts of the reporting period, are unlikely.

[13]For example, see http://www.mirror.co.uk/news/uk-news/ae-crisis-exposed-only-three-9801509.

(£220) for every patient who failed to be treated within 4 hours if the hospital missed the overall 95% target during that week.[14] This compares to an average payment of just over $140 (£100) per patient in the same year. In 2015, a report commissioned by a number of hospitals indicated that public hospitals paid $325 million (£250 million) in fines due to missed performance targets (including the 4 hour target), with total penalties equal to around a third of the average deficit of public hospitals in that year.[15]

Hospital staff therefore face pressure from hospital management to meet the target. As a result, the organisation of EDs has changed significantly since the target was introduced.[16] Changes include the use of new IT systems, which track patient wait times in real time. The exact systems vary by hospital, but will indicate when patients reach particular waiting thresholds (e.g. 3 hours) and alert physicians (for example through changing the colour of the computer screen).

3.2.3 How do EDs respond to the target in the short-run?

Setting aside these longer-run changes to ED organisations, the target incentivizes doctors to speed up treatment for at least some patients. This could be achieved in two different ways, which will have important implications for the validity of our identifcation strategy.

First, doctors may proceed with treatment as they would do in the absence of the target and only change their behaviour if patients wait long enough so as to begin to approach the target. In this case, as wait times exceed a certain point, EDs may speed up treatment by reducing the number of investigations or treatment conducted in the ED (either pushing these into inpatient treatment after admission, or discharging patients with less information), reducing the waiting time between receiving results and implementing treatment decisions, or by reallocating senior doctors to make clinical decisions quickly about long-wait patients. This approach would alter only the wait times of those approaching the four-hour mark, and leave the treatment of other patients unchanged.

Alternatively, EDs may more fundamentally change the way that patients are treated and prioritized. For example, doctors may substitute resources (e.g. doctor time) away from more minor patients in order to concentrate on major patients that are more likely to breach the target. This would increase the wait times of

[14] This penalty was decreased to $170 (£125) in 2015.

[15] https://www.theguardian.com/society/2016/mar/29/nhs-bosses-slam-600m-hospital-fines-over-patient-targets

[16] Interviews with senior member of the Emergency Care Improvement Programme (ECIP), a clinically led programme intended to improve the performance of EDs, clearly describe significant changes to the technology used in EDs since the target was introduced. One manager in the programme claimed that "This [the target] is the most monitored part of the entire healthcare system with software specifically designed for it."

those in the earlier part of the distribution in order to reduce wait time for others. Alternatively, doctors anticipating the target, may speed up the treatment of minor patients in order to free up capacity for other patients. In both cases, the target would have implications for the wait times of patients across the wait time distribution.

We assume the first of these scenarios in our empirical methodology and discuss the implications of this in detail in Section 3.4. In Section 3.5.1, we provide three pieces of empirical evidence in support of this assumption that suggest there was little wholesale changes to the way in which all patients are treated as a result of the target.

In addition, there are four institutional details that suggest that such behaviour by doctors and managers in the ED is unlikely. The first is simply that in practice, there is little scope for substitution between patients since there is no incentive to substitute between patients that are not within the same four hour window. Our analysis aggregates patients across many hospitals, days, and time periods. However, there are on average only 33 patients who arrive at a hospital within a given four-hour window. This limits the potential extent of dynamic responses.

Second, it is important to note that the definition of wait times in our setting, which corresponds to total time in the ED, allows for the possibility that physicians can shorten a patient's wait time by simply discharging them or admitting them to an inpatient department. Shortening wait times therefore does not necessarily involve a physician spending more time with a patient than they would do otherwise (or even substituting time from one patient to another). Physicians therefore have little incentive to shorten or lengthen waits for patients who are unlikely to approach the target.

Third, ED staff are organized in a way that further limits scope for substitution. Physicians and nurses in English EDs are generally separately assigned to minor or major units within the ED, and this physical separation limits the prospect of substitution between early and late exit patients. It is of course possible that as a major unit becomes busy, staff could be diverted from the minors unit to assist. In this case the presence of the target may incentivize more staff to be moved to treat majors than would otherwise occur. We test for this directly in Section X, and find no obvious evidence of substitution in these cases.

Finally, hospitals are likely to be maximizing an objective function that, at least in part, contains patient mortality. This will naturally place limits on their willingness to substitute between different types of patients. For example, patients with clear and life-threatening injuries (e.g. knife wounds) will always be treated immediately, and for a similar length of time, irrespective of the target. Similarly, patients with very minor injuries will always be sent home shortly after initial

assessment. These unambiguously high and low severity patients are likely to account for a significant proportion of exits from the ED in the early part of the wait time distribution. This suggests that hospitals are unlikely to change the way they treat these patients in light of the target, with any potential substitution of resources between patients occurring further up the wait time distribution.

3.3 Data

3.3.1 Hospital Episodes Statistics

Our primary source of data are the Hospital Episode Statistics (HES). These contain the administrative records of all visits to public hospitals between April 2011 and March 2013, and include information on both ED visits and inpatient admissions.[17]

The ED data record treatment at the visit level, and include information on the precise time of arrival, initial treatment and the admission decision. We define ED 'wait times' as total time spent in the ED, consistent with the definition of the policy. This includes time being examined and treated. We calculate ED wait times as the time elapsed between arrival and the admission decision, where the arrival time is recorded as patients enter the ED.[18,19]

The data also include a hospital identifier, whether the patient is admitted or discharged, details of basic diagnoses, the number and types of ED investigations and treatments, whether the patient arrived by ambulance, and some basic patient characteristics such as age, sex and local area of residence.

Patients are identified by a psuedo-anonymized identifier that allows patients to be followed over time and across hospitals, and enable linkage between ED and inpatient records. Inpatient records contain detailed information on treatment undergone in the hospital. The data contain the dates of admission and discharge, and information on up to twenty diagnoses and procedures undertaken. Treatment is recorded at the episode level, defined as a period of treatment under the care

[17]Data on EDs is available prior to 2011, covering 2008 and 2010, although data from the earlier period is less complete than in the years we study.

[18]For non-ambulance patients, this time is recorded when they first speak with the receptionist.

[19]Hospitals may attempt to manipulate wait times to meet the target. We evaluated one possibility in this regard, namely that hospitals simply miscode the timing of the admission decision, such that the total wait time is 4 hours or less. Following Locker and Mason (2006), we analyzed the distribution of 'final digits' in wait times (e.g. the digits 0 to 9 at the end of each wait time value) which in the absence of manipulation should be uniformly distributed. Relative to this benchmark, we found that less than 1% of records were likely to be miscoded and that this would have a negligible impact on our analysis.

of a single senior doctor.[20] We combine information across all episodes within the same admission to create visit-level variables for total length of stay (in days) and number of inpatient procedures. Each episode also contains a Healthcare Resource Group (HRG) code, similar to Diagnosis Related Groups (DRGs) in the US. English hospitals are compensated by the government through a system of national tariffs for each HRG.[21] We calculate 'costs' for each episode by matching tariffs to the appropriate HRG, which gives us a measure of the cost to the government, and revenue received by the hospital, associated with each visit. We then sum all treatment costs over a 30 day period to estimate the cost associated with each ED visit and any follow-up treatment.

Mortality outcomes are recorded in administrative records made available by the UK Office for National Statistics (ONS). These records are linked to HES through anonymized identifiers based on patient National Insurance (Social Security) numbers. The data include the date of death for all individuals who died in the UK, or UK citizens who died abroad, between April 2010 and March 2014. We create indicators of whether a patient dies within 30, 90 and 365 days of an ED visit.

Sample construction

Our analysis focuses on a sample of emergency patients treated in 'major' emergency departments.[22]

We keep all patients with full information relating to the timing of treatment and their exit route from the ED, in addition to their age, gender and whether they arrived by ambulance. Dropping patients with some missing information reduces the number of visits in the sample by 14.5%.[23] This yields an analysis sample of 14.7 million patients, who made 24.7 million visits to 184 EDs between April 2011 and March 2013.

[20]Senior doctors in England are known as 'consultants', and are equivalent to attending physicians in the US.

[21]National tariffs are calculated for each HRG on the basis of annual cost reports submitted by hospitals to the UK Department of Health. These tariffs are meant to reflect the average cost of providing the procedure. Payments are then adjusted for unavoidable regional differences in providing care, and unusually long hospital stays.

[22]Major emergency departments are defined as consultant-led providers of 24-hour services, based in specifically built facilities to treat emergency patients that contain full resuscitation facilities. We exclude patients treated at specialist clinics that treat only particular diagnoses (e.g. dental) and minor injury ('walk in') centres. Patients treated by these units typically have simple diagnoses and short wait times, and are therefore unlikely to be affected by the target. This excludes 18% of emergency visits.

[23]Results are unaffected by the inclusion of patients with full information relating to treatment times and decisions, but who are missing demographic information.

Summary statistics

Table 3.1 reports summary statistics. The first two columns present the mean and standard deviation for a range of patient characteristics, treatments and outcomes for all ED patients in the sample. Mean ED patient age was 39 years, and 51% of patients were male. 29% of patients arrived by ambulance. 5.8 million visits, or 24% of all ED episodes, resulted in an inpatient admission at the same hospital. 58% of visits did not require further hospital treatment and led to a patient being discharged. The remaining visits resulted in a transfer to an outpatient clinic or another hospital for further treatment. Mean 30-day treatment costs were $1,676 (£1,240), of which 89% was accounted for by subsequent inpatient treatment. In the short term, mortality among ED patients is relatively rare. 2% of patients died within 30 days of visiting the ED. This increases to 3% over a 90 day period, and 5% during the following year.

Table 3.1 also shows summary statistics separately for visits that result in an inpatient admission. As expected, these case are typically more severe, with an older average age (55 years) and twice the likelihood of arriving in an ambulance (60%). Mortality rates (5% over 30 days, 16% over a year) are substantially higher than in the main sample. ED treatment is more intense for this sample, with a higher mean number of treatments and investigations than in the main sample. Their treatment is also more expensive, with an average total cost over a 30-day period of $4,762 (£3,530).

Inpatients also experienced longer mean wait times in the ED than those who are not admitted. Mean wait times were 223 minutes for patients who were eventually admitted as inpatients, compared to a mean of 155 minutes for all ED patients. This demonstrates that the level of patient complexity, and the intensity of treatment for these patients, is likely to vary by wait time. This variation is important to account for when analysing the impact of the target.

Figure 3.1 shows the distribution of ED wait times. There is a noticeable discontinuity in the proportion of patients who exit the ED in the period immeditely prior to 4 hours. This spike is unlikely to naturally occur, and is instead induced by the target. We cannot illustrate the absence of this spike prior to the wait times target, since we do not have systematic data available from that period. But it is worth noting, as we do in Appendix Figure A1, that such a spike is not present in data on ED wait times from a major U.S. hospital.[24]

One possibility is that this spike in wait times simply reflects recoding and is not a real change in patient wait times. Two features suggest that this is not the

[24]Of course, different ED objectives and technologies across countries means that the U.S. data does not provide a natural comparison group, but the lack of any spike confirms our conclusion that the large spike here is particular to the wait time policy.

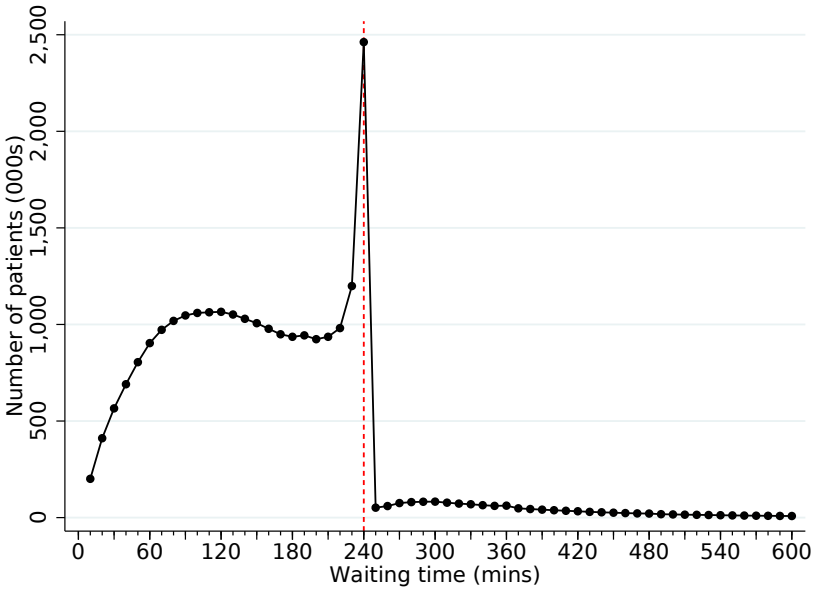
Table 3.1: *Summary statistics*

	All patients		Admitted inpatients	
	Mean	Std. dev.	Mean	Std. dev.
Patient characteristics				
Age	38.99	26.22	54.64	27.84
Male	0.51	0.50	0.48	0.50
Ambulance arrival	0.29	0.45	0.60	0.49
Past-CCI	0.20	0.78	0.47	1.20
Treatment decisions				
Inpatient admission	0.24	0.42	1.00	0.00
ED discharge	0.58	0.49	0.00	0.00
ED referral	0.19	0.39	0.00	0.00
Wait time (mins)	154.56	100.20	222.50	120.46
ED treatment count	1.81	1.38	2.22	1.68
ED investigation count	1.54	2.03	3.18	2.50
Inpatient length of stay (days)	1.28	5.63	5.41	10.58
Inpatient procedure count	0.16	0.64	0.69	1.18
Costs				
30-day ED cost	172.35	117.21	203.98	114.98
30-day inpatient cost	1,503.58	5,321.99	4,558.00	8,524.53
30-day total cost	1,675.93	5,358.37	4,761.98	8,559.73
Mortality outcomes				
30-day mortality	0.02	0.13	0.05	0.23
60-day mortality	0.03	0.16	0.09	0.29
365-day mortality	0.05	0.22	0.16	0.37

Notes: (1) Costs reported in 2018 USD and refer to payments from the government to hospitals based on the prospective payment system; (2) All inpatient variables (e.g. length of stay, costs) take on the value zero for patients that are not admitted.

case. First of all, a sizeable share of hospitals pay large penalties and are publicly criticized as a result. Indeed, a substantial number of hospitals only just miss the target, with 23% of hospitals missing the target by less than two percentage points in 2011/12. If recoding explained the spike then those hospitals should do more recoding to avoid the penalty altogether. Second, we show below that there are comparable spikes in a number of real outcomes, such as hospital admissions, costs, and mortality, which are inconsistent with this simply being a coding response.

Figure 3.1: *Distribution of wait times*



Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy.

3.4 Empirical methodology

A key challenge when analysing the four-hour target is that without pre-policy data or a control sample, quasi-experimental methods cannot be used to construct the counterfactual outcome. To address this issue we use and extend bunching estimators that were developed in the tax literature (Saez 2010, Chetty et al. 2013). We argue that these methods can be used in our setting to estimate the counterfactual outcomes that would occur if the target were removed but other aspects of hospital production were held constant. This allows us to quantify the short-run impact of the policy.

We now set out our empirical methodology. We begin by setting out a bunching estimator for waiting times before giving an overview of our analysis of treatment decisions and health outcomes. More details on this methodology are set out in Appendix B.4.

3.4.1 Wait times

We first apply a bunching estimator to the distribution of wait time outcomes. Let w be the wait time in minutes, where $w^* = 240$ (the target threshold). Denote the density function of w in the targeted regime as $f_t(w)$ where $t = \{0, 1\}$ signifies

whether the function relates to the targeted or non-targeted regime. We observe data on $f_1(w)$ and use a bunching estimator to obtain $f_0(w)$.

To implement the bunching estimator we aggregate the data to 10-minute wait time bins and then interpolate parts of the distribution using a polynomial regression. Following Kleven (2016) we define $\hat{f}_0(w) \equiv \sum_{i=0}^{p} \hat{\beta}_i w^i$ and obtain the estimates $\hat{\beta}_i$ from the following regression

$$c_j = \sum_{i=0}^{p} \beta_i (w_j)^i + \sum_{i=w^-}^{w^+} \gamma_i \mathbf{1}[w_j = i] + u_j, \tag{3.1}$$

where c_j is the number of individuals in wait time bin j, w_j is the maximum wait time in bin j (e.g. $w_j = 10$ for the 1-10 minute wait time bin, $w_j = 20$ for the 11-20 minute wait time bin, etc), p is the order of the polynomial, and $[w^-, w^+]$ is an 'exclusion window' that contains w^* and is the period during which we assume that the target may have had local effects on the wait time. This regression fits a polynomial to the wait time distribution in periods outside of the exclusion window, where the window is captured by the indicator variables which then do not feature in $\hat{f}_0(w)$.

Equation (3.1) makes the following assumption in relation to the exclusion window.

Assumption 3 (Local wait time effects)**.** *Wait times of patients outside of an 'exclusion window', defined locally around the threshold w^*, are unaffected by the target:*

$$f_0(w) = f_1(w) \qquad \forall w \notin [w^-, w^+]. \tag{3.2}$$

This assumption will hold if hospitals do not respond to the target by substituting resources between patients that are inside and outside of the exclusion window.[25] We discuss this assumption at length in the next section.

To establish the bounds of the exclusion window, we follow Kleven and Waseem (2013) and set w^- visually by examining when the distribution changes sharply and determine w^+ using an iterative procedure that equates the excess mass in the period $[w^-, w^*]$ with the missing mass in the period $(w^*, w^+]$.[26] An advantage of this iterative approach is that we make no assumption about w^+ and let the

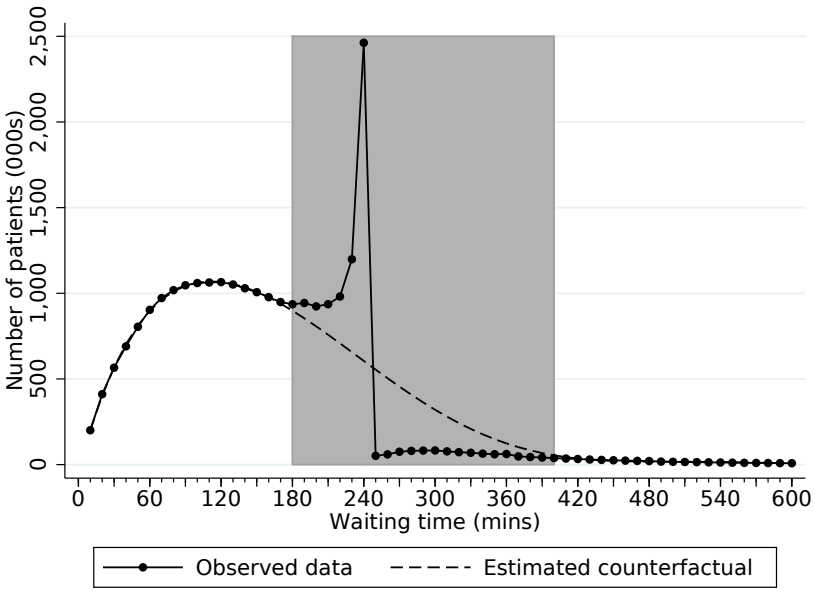
[25] A comparable assumption is required when using bunching techniques to study taxable income responses. In that setting the local effects assumption is often innocuous because the income distribution is the result of optimization decisions of many unrelated individuals, with those situated far from the tax scheme discontinuity having no incentive to adjust their behaviour. In our setting, the distribution of patient wait times is not determined by patients' decisions but by the decisions of doctors and nurses, and this raises the concern that there may be an incentive to substitute wait times between patients across different parts of the wait time distribution.

[26] This implicitly assumes that the target does not affect patient demand for ED care in the short-term.

data determine where the effects on the wait time distribution end. In the baseline analysis we use a polynomial of order 10 and set $w^- = 180$. After applying the iterative procedure this produces an upper cut-off of $w^+ = 400$. We show below that while the exact magnitude of the results are somewhat sensitive to the choice of parameters, our conclusions are qualitatively robust to variations in the choice of polynomial and w^- (see Appendix Tables A4 and A5).

The observed data and our estimated counterfactual distribution are shown in Figure 3.2, which indicates that the target moves a number of patients from the post-threshold period to the pre-threshold period ('post-threshold movers'). We later use these distributions to estimate the impact of the target on wait times.

Figure 3.2: *Estimated counterfactual wait time distribution*



Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy; (4) The estimated counterfactual is obtained from a polynomial regression that omits the exclusion window shown in grey.

Interpreting the counterfactual

The counterfactual that the bunching estimator delivers in our context is the short-run outcome that would occur if the four-hour discontinuity in incentives were removed. The counterfactual holds constant other aspects of hospital production, such as patient prioritization, capital and labour inputs, and government funding. As a benchmark, the counterfactual focuses attention on the role of incentives in determining outcomes rather than the specifics of the production function in our setting. We see it as a logical benchmark for understanding how wait time incentives affect outcomes.

Our counterfactual differs from the pre-policy or long-run outcomes. To give an example of the difference, we know from anecdotal evidence that the pre-policy outcome had different production inputs (particularly the volume of staff) and different production technology (e.g. IT systems). It also rules out wholesale changes in the treatment and prioritisation of patients in the earlier part of the wait time distribution. The full policy impact relative to the pre-policy situation would potentially include the impact of these changes as well as the discontinuity in incentives introduced by the target.

We refer to our results as the 'impact of the target' for brevity but with the above understanding in mind. This interpretation applies to the results for wait times and other outcomes.

3.4.2 Treatment decisions and mortality outcomes

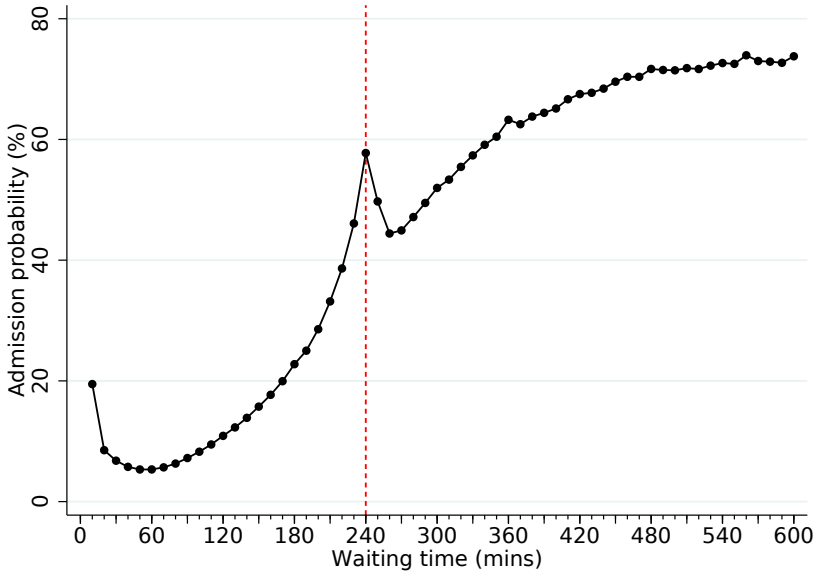
We now extend the analysis to consider outcomes other than the wait time, such as treatment decisions (e.g. inpatient admission) and mortality outcomes. Plotting these outcomes conditional on the wait time shows that they also exhibit 'bunching' at the four-hour discontinuity point. Figure 3.3 gives an example for the likelihood of inpatient admission. The plot shows that admission probability is generally increasing with wait times, and there is a clear spike in admission probability at 240 minutes. Our analysis decomposes this spike into two channels.

The first channel is the 'composition effect'. As Figure 3.1 suggests, the target causes a substantial number of patients to be moved from later to earlier in the distribution of wait times (a group we refer to as 'post-threshold movers'). Since admission probabilities are increasing with wait time, this movement of patients would increase the observed pre-threshold admission probability even if the target led to no additional admissions. This effect arises purely because the target changes the composition of patients observed at each wait time.

There is also potential for a 'distortion effect' if the target has a direct effect on treatment decisions and health outcomes. The distortion effect implies identical patients receive different treatment depending on whether or not the target is in place. In the case of admissions, for example, it would imply that part of the spike in observed outcomes is because the target causes additional admissions, in addition to the composition effect shifting some admissions from after to before the target.

To decompose these two effects we construct a 'composition-adjusted counterfactual' (CAC). This is the outcome that would occur in the presence of composition effects but the absence of distortion effects. Since the observed data contains both effects, the difference between the observed data and the CAC identifies the

Figure 3.3: *Inpatient admission probability conditional on wait time*



Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy.

distortion effect. Estimates of the distortion effects and tests of whether these are significantly different from zero are the central results of this paper.

We construct estimates of the CAC as a weighted average of counterfactual outcomes for patients who are observed in the pre-threshold part of the wait time distribution (i.e. between w^- and w^*). This includes two separate groups: patients who are shifted by the target from the post-threshold to the pre-threshold period ('post-threshold movers') and patients who would have been treated prior to the threshold even in the absence of the target ('pre-threshold non-movers'). From the wait time analysis, we know how many patients are moved from the post-threshold part of the wait time distribution to the pre-threshold part of the distribution as a result of the target. The weights are therefore defined by the observed and counterfactual wait time distributions.

We then construct the required counterfactual outcomes by applying bunching techniques to the *expected outcomes conditional on the wait time*.[27] This relies on two key assumptions.

[27]This is in contrast to a typical bunching application that would work with the distribution of a variable that is subject to a discontinuity in incentives. Here we work with outcomes conditional on a variable that is subject to a discontinuity in incentives. Our approach is similar in spirit to Diamond and Persson (2016) and Gerard et al. (2018).

Assumption 4 (Local outcome effects)**.** *Outcomes outside of an 'exclusion window', defined locally around the threshold w^*, are unaffected by the target:*

$$E[y_1 \mid w_t] = E[y_0 \mid w_t] \qquad \forall w \notin [w^-, w^* + \varepsilon]. \tag{3.3}$$

where $E[y_t \mid w_t]$ is the expected outcome conditional on wait time in regime t.[28]

Assumption 4 rules out distortion effects outside of the pre-threshold period. It is the parallel of Assumption 3 for the conditional expectation function. In this case the exclusion window ends at $w^* + \varepsilon$, where ε is a small 'overhang period' that extends past the four-hour threshold.

The overhang period allows for the empirical fact that the bunching in outcomes extends slightly past the threshold (see Figure A2 in Appendix B.1). We interpret the overhang as being a case of distortion effects for patients that are narrowly discharged or admitted after the threshold. For example, it may be that doctors admit additional patients in attempts to meet the target but not all of the excess admits occur prior to the threshold as some patients may be delayed for unexpected reasons. We determine the size of the overhang period visually, setting $\varepsilon = 20$ in the baseline analysis, and note that our findings are robust to more conservative (larger) overhang periods.[29]

Assumption 5 (No-selection)**.** *Non-targeted regime outcomes conditional on the non-targeted wait time are comparable for post-threshold movers and post-threshold non-movers:*

$$E[y_0 \mid \underline{w}_1^-, w_0] = E[y_0 \mid \underline{w}_1^+, w_0] = E[y_0 \mid w_0] \qquad \forall w_0 \in \underline{w}_0^+. \tag{3.4}$$

where $E[y_0 \mid \underline{w}_1^-, w_0]$ is the expected outcome for post-threshold movers under the non-targeted regime, $E[y_0 \mid \underline{w}_1^+, w_0]$ is the expected outcome for post-threshold non-movers under the non-targeted regime, and $\underline{w}_0^+ = w* < w_0 < w^*$.

Assumption 5 rules out composition effects in the post-threshold period. It states that after conditioning on the non-targeted wait time, there is no selection when the post-threshold movers are assigned. The assumption is consistent with doctors randomly selecting which patients get a shorter wait time in response to the target, and in that sense it is equivalent to an unconfoundedness assumption

[28] $t = 0$ denotes the non-targeted regime, while $t = 1$ denotes that the target is in place. This allows us to express average outcomes (either in the targeted or non-targeted regime) for groups of patients located in different parts of the wait time distribution (either in the targeted or non-targeted regime). For example, the observed data can be written as $E[y_1 \mid w_1]$.

[29] Our estimates of the distortion effect, which relate to the pre-threshold period, do not capture distortions in the overhang period. These omitted effects are small: the number of patients in the overhang period is 1.3% of the number of patients in the pre-threshold period.

in traditional IV terminology.[30] While this is strong assumption we believe it is plausible and, most importantly, we are also able to evaluate the assumption empirically using placebo tests. We discuss this assumption and the results of these tests in detail shortly.

Together Assumptions 4 and 5 imply that there are no composition or distortion effects outside of the exclusion window $[w^-, w^* + \varepsilon]$. We can therefore apply the bunching estimator in the same way as for the wait times but to the conditional expectation function $E[y_1 \mid w_1]$. The estimated counterfactual delivered by the bunching estimator is then $E[y_0 \mid w_0]$. This is shown in Figure A2 in Appendix B.1. This directly gives us counterfactual outcomes for the pre-threshold non-movers and, given Assumption 5, also provides us with the counterfactual outcomes for the post-threshold movers. Taking the weighted average of these outcomes therefore yields an estimate of the composition-adjusted counterfactual.

Under these assumptions, we can test whether there are distortion effects by taking the differences in the observed outcomes and the estimated composition-adjusted counterfactuals (ΔD). Tests for distortion effects are then simply hypotheses tests that these differences are equal to zero. We compute statistical significance for the test using non-parametric bootstrapped standard errors clustered at the hospital organisation level.[31]

Figure 3.4 provides a visual example of how we construct the CAC and the test of distortion effects for the probability of inpatient admission. The pre- and post-threshold periods are shown in different shades of grey. In each of these periods the horizontal thin dashed line gives the counterfactual conditional expectation. The CAC, which is a weighted average of these two conditional expectations, is shown in the horizontal thick dashed line in the pre-threshold period.[32] In comparison, the horizontal thick solid line in the pre-threshold period is the mean observed outcome in the pre-threshold period. Finally, the difference between the thick solid and dashed line is the distortion effect, Δ_D, which shows that the observed admission probability in the pre-threshold period is too high to be explained by the composition effect alone. In this case we can reject the null hypothesis that $\Delta_D = 0$.

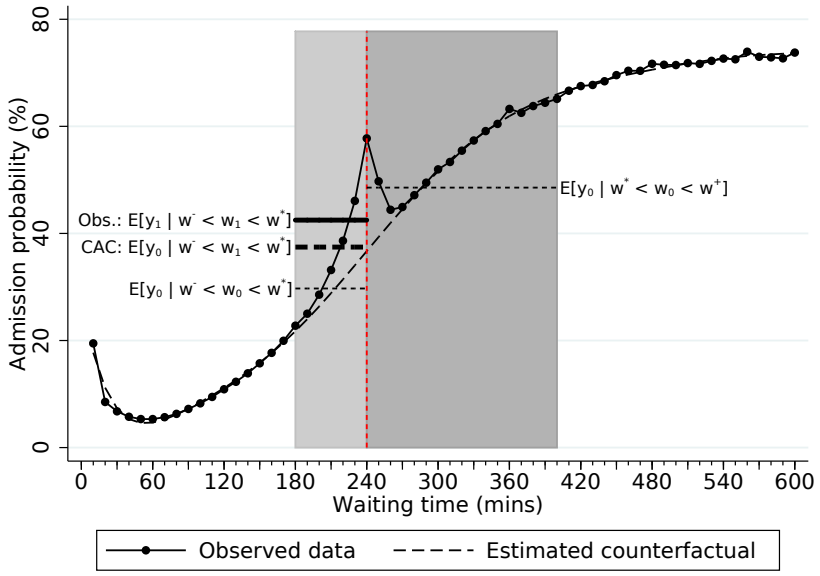
The next section explores the validity of these assumptions in detail. To evaluate the validity of the no-selection assumption, we devise a test based on observable

[30]Similarly, in IV terminology, the post-threshold movers would be compliers, the post-threshold non-movers would be never-takers, and the pre-threshold non-movers would be always-takers. We implicitly make the assumption that there are no defiers.

[31]Throughout the analysis we cluster results at the trust (organisation) level. NHS trusts include groups of one or more hospitals in close geographical proximitiy that share common management. We do not use hospital site codes due to some organisations entering data only at the trust level. All results are robust to clustering at the site level.

[32]The weights are obtained from the wait time distributions shown in Figure 3.2.

Figure 3.4: *Constructing the composition-adjusted counterfactual for admission probability*



Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy; (4) The horizontal thin dashed lines in the light grey (dark grey) region give the counterfactual outcome in the pre-threshold (post-threshold) period, $E[y_0 \mid w_0]$; (5) The horizontal thick dashed line in the pre-threshold period is the composition-adjusted counterfactual, $E[y_0 \mid \underline{w}_1^-]$; (6) The horizontal thick solid line in the pre-threshold period is the observed admission probability, $E[y_1 \mid \underline{w}_1^-]$; (7) The distortion effect is the gap between the thick solid and dashed line, $\Delta_D = E[y_1 \mid \underline{w}_1^-] - E[y_0 \mid \underline{w}_1^-]$.

patient characteristics that cannot be altered by the hospital. This includes age, sex, whether the patient arrived in an ambulance, three health measures based on hospital use in the prior year (Charlson comorbidity index (CCI), number of emergency admissions, days spent in hospital), and predicted mortality and admission. These variables, conditional on the wait time, also exhibit bunching at the four-hour point but in these cases the spike can only be explained by a composition effect since there is no distortion effect by definition. If the no-selection assumption is valid then for these variables the observed data and the CAC should be equal (i.e. the estimated distortion effect is equal to zero). We therefore estimate the distortion effect for each of these variables. This acts as a placebo test, where an estimated distortion effect significantly different from zero would suggest that the no-selection assumption has been violated.

Figure A3 in Appendix B.1 provides a visual example of the demographic test using age, which follows the same format as Figure 3.4. There is again bunching at the four-hour threshold but in this case it cannot be explained by any distortion effects because patient age is unaffected by hospital treatment decisions. Comparing

the observed data and the CAC shows that these now lie very close to one another and indeed a hypothesis test cannot reject the null hypothesis that $\Delta_D = 0$. This is consistent with the no-selection assumption: the mean age of post-threshold movers is comparable to the mean age of all post-threshold patients.

We repeat this analysis for each of the observed patient characteristics outlined above. We pass these placebo tests for the majority of the tested variables, and where they fail (sex, past-CCI) the magnitudes are very small. We discuss these results, and further tests of this assumption, in more detail in Section 3.5.2.

3.5 Validity of key assumptions

Our methodology rests on our assumptions about local effects (Assumptions 3 and 4) and selection (Assumption 5). We discuss these assumptions and supporting evidence below.

3.5.1 Local effects assumption

The local effects assumption is that wait times and treatment decisions prior to w^- are unaffected by the target, and we set w^- at 180 minutes in the baseline analysis. As noted earlier this will not hold if hospitals substitute time or resources between patients that exit before w^- ('early exit patients') and after w^- ('late exit patients'). This assumption rules out certain dynamic responses that may impact the wait time distribution. We suggest a set of factors that mitigate the importance of this issue, and then carry out two empirical tests that support the credibility of this assumption in our setting.

As noted in Section 3.2.3, there are a number of institutional factors that mitigate concerns about a violation of local effects. First, The fact that patients are treated across many hospitals, days and time periods limits the scope for dynamic responses. Second, physicians can potentially shorten ED treatment by admitting patients, and so shorten waits without taking resources away from other patients. Third, staff are organized to separately treat 'minor' and 'major' patients, limiting the ability to substitute resources to major patients at the expense of the treatment of minor patients. Finally, hospitals are already likely to be attempting to maximise an objective function that prioritises patient outcomes: changes to the ordering of very severe or very easy cases is therefore unlikely. These unambiguously high and low severity patients are likely to account for a significant proportion of exits prior to the exclusion window, and suggest that if there are substitution responses, then these are more likely to occur near to w^- rather than at the very start of the distribution.

Despite these mitigating factors, concerns about the local effects assumption may remain, so we conduct three empirical tests to further evaluate this assumption.

First, since we expect any dynamic responses to be concentrated near to w^-, a natural robustness test is to check whether our results are sensitive to the choice of w^-. We therefore vary our choice of w^- and assess how sensitive our results are. We show the results of this exercise in Appendix B.1 (Table A1). The results suggest that the results are qualitatively robust to changes in w^-, with the same sign and significance across all specifications for most variables. However, some point estimates do vary in size: for example, the estimated impacts on admission double in magnitude when moving from the earliest to latest starting point.[33] In contrast, the estimated mortality effects are not statistically significantly different from one another. This suggests there is some sensitivity in the magnitude of our estimates with respect to the starting point of the exclusion window but this does not change the overall conclusions.

In our second test of this assumption, we address the potential concern that there may be a more fundamental impact of the wait time target on the early part of the distribution. In particular, hospitals may approach the prioritization of patients entirely differently when facing the target compared to an unconstrained scenario. This would undermine the previous test by implying that there is no part of the distribution that is unaffected by the target (i.e. theoretically w^- should be set at zero, which in practice leaves no data for estimation).

If this concern is valid, however, it implies that hospitals should change the priority order assigned to patients based on how tightly the target is anticipated to bind - with a non-binding target, hospitals are unconstrained by the four hour rule. Our second test is therefore to exploit variation in the expected volumes of ED arrivals – with the target binding more tightly as volumes increase – to see if it impacts patient prioritization, especially at earlier wait times.[34]

More specifically, we plot the proportion of patients exiting in the early part of the distribution (prior to 180 minutes) against a proxy for whether patients are expected to exit in the early or late part of the distribution (predicted mortality)

[33]Importantly, reducing w^- does not result in statistically significant changes to the estimates. In contrast, increasing w^- is associated with larger changes in the estimates. This is unsurprising given that visual inspection of the waiting times distribution shows clear distortions as wait times approach 240 minutes, and these are already apparent at 200 minutes. We would therefore expect to capture some of these dynamic responses in our estimates when using a higher value of w^- in our estimates.

[34]While the volume of arrivals may be correlated with other factors, such as the number of doctors scheduled to be on shift, this would not necessarily impact the patient prioritization that we compare in this test. One potential concern could be that any increase in scheduled doctors may offset any increase in expected arrivals. If we repeat the same test but use shocks to ED arrivals then we find similar results.

for periods when the ED is more or less busy (measured by expected ED admits). First consider how this plot would look in the absence of any change in patient prioritization. Higher severity patients typically spend longer in the ED (requiring more tests and treatment) and therefore fewer exit in the early part of the wait time distribution. Moreover, due to congestion effects, in busier periods wait times will be longer in general and, assuming congestion affects all patients similarly, will mean fewer of all patient types exit prior to 180 minutes. Together, this will mean the proportion of exits prior to 180 minutes is decreasing in predicted mortality and this relationship experiences a parallel downward shift in busier ED periods.

Now consider how this relationship would be affected by hospitals responding to the target through changes in patient prioritization. The concern is that, as the target binds more, the ED may choose to delay low severity, early exit patients in order to benefit high severity, late exit patients. In this situation, the proportion of patients exiting prior to 180 minutes would decrease among the early exit patients (who experience wait time increases) by more than it does for the late exit patients (who experience wait time reductions). This would cause a change in the slope of the relationship between the proportion of exits prior to 180 minutes and predicted mortality, and mean a non-parallel downward shift in busier ED periods. The intuition that underlies this non-parallel shift is what underlies our test for changes in patient prioritization.

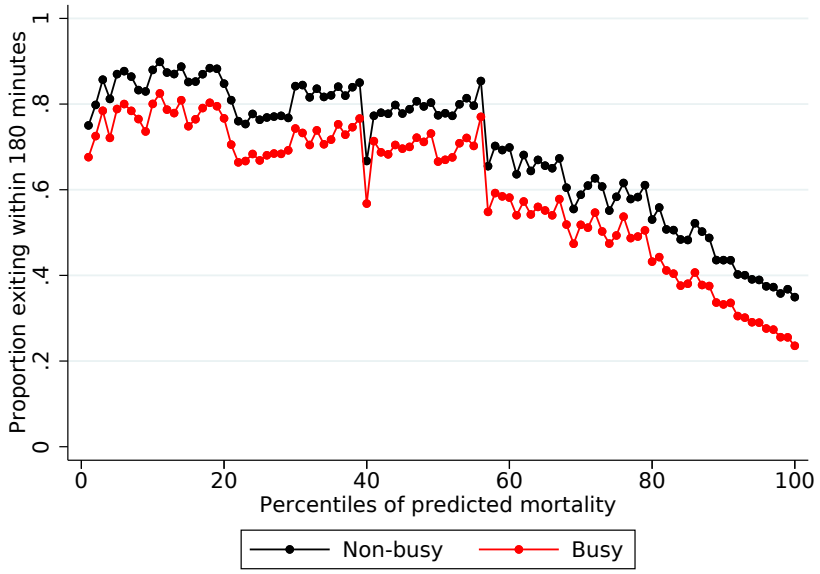
Figure 3.5 plots the proportion of patients who exit within 180 minutes for each percentile of predicted mortality for patients that arrive during 'busy' and 'non-busy' periods. We define busy periods by first predicting the number of patients present in the ED during each hour in our data, using a regression with hospital-specific week-of-year, day-of-week, and hour-of-day fixed effects. We then divide periods into the top-third of predicted volumes (busy) and bottom-third of predicted volumes (non-busy).

The plot shows that a smaller proportion of high severity patients leave the ED within 180 minutes. It also shows that busier periods have longer wait times for patients of all severity, with a smaller proportion of patients across the entire severity distribution leaving the ED within 180 minutes as the department becomes busier. Most importantly, the relative probabilities of exits within 180 minutes for high and low severity patients are very similar in both types of period – i.e. there is a parallel downwards shift in the relationship. This is precisely what should happen if hospitals do not change the patient prioritization in response to the target binding more or less tightly, as implied by the local effects assumption. The same is also true if you repeat this exercise for other waiting periods.[35] These

[35]We also examined the proportion of patients who exit the ED after 60 minutes, 120 minutes, 240 minutes and 400 minutes. This produces similar results in all cases, with a broadly parallel

results suggest that as the target binds more or less tightly hospitals do not change the prioritization of patients, consistent with our assumption.

Figure 3.5: *The probability of ED exit within 180 minutes by patient severity and expected volumes of ED arrivals*



Notes: (1) Wait times defined as the time from arrival in the ED to leaving the ED; (2) Predicted mortality defined using a regression of 30-day in-hospital mortality on a fully interacted set of age, gender, ambulance arrival fixed effects and diagnosis fixed effects; (3) Busy and non-busy periods defined by predicting the volume of ED arrivals during each hour in our data, using a regression with hospital-specific week-of-year, day-of-week, and hour-of-day fixed effects, and then dividing periods into the top-third of predicted volumes (busy) and bottom-third of predicted volumes (non-busy).

Our final test considers whether hospitals temporarily substitute resources between patients if they experience a demand shock – such as the ED being momentarily overrun with patients – and this causes them to make short-term deviations from planned priorities to meet the four-hour target. A specific example is that the hospital has a build up of patients that are close to breaching the target and they temporarily substitute resources away from newly-arriving patients to clear the backlog.

To test for evidence of such behaviour, we examine whether there is any evidence that hospitals substitute resources away from patients that we would expect to exit in the early part of the distribution in order to ensure that patients approaching the target do not wait over 4 hours. Intuitively, we compare wait times of newly arrived patients on the basis of how many existing patients have waited

shift downwards in the proportion of exits when moving from a non-busy to busy period. This shift is very small when examining the proportion of patients exiting within 240 and 400 minutes as the vast majority of patients have exited the ED by this point.

almost four hours. If there are temporary substitution effects between these individuals, we would anticipate large effects of the presence of existing patients near the four-hour threshold on the wait times of new patients.

We examine four groups of newly arriving patients on the basis of predicted waiting times: those predicted to have wait times below 150 minutes; 150-180 minutes; 180-210 minutes; and 210-240 minutes. For each group, we regress wait times on the volume of existing patients ahead of them at each 10-minute interval of the queue. We compare results between the early exit patients (those in the first two groups with predicted wait times below 180 minutes, such that they exit prior to the exclusion window) and late exit patients (predicted wait times above 180 minutes). The late exit groups act as control groups in the sense that Assumption 1 allows for temporary substitution effects to occur for these groups (inside the exclusion window) but not for the early exit groups (outside the exclusion window). We predict early or late exit using a regression of wait times on age, gender, diagnosis fixed effects and an ambulance indicator.

To implement the test we aggregate the data to the hospital-period level, where periods are defined at 10-minute intervals, and estimate the following equation:

$$w_{ht}^{g} = \sum_{k} \beta_k q_{h,t-k} + \mu_{hw} + \delta_{hd} + \gamma_{hp} + e_{ht} \tag{3.5}$$

where w_{ht}^{g} is the mean wait time for newly arriving patients of group g (as per the four categories described above) at hospital h in period t (e.g. between 12:01 and 12:10), $q_{h,t-k}$ is the number of existing patients waiting ahead in the queue at horizon $t-k$ (e.g. the number of patients that have been waiting 1-10 minutes, 11-20 minutes, and so on), and μ_{hw}, δ_{hd} and γ_{hp} are hospital-specific week-of-year, day-of-week, and period-of-day fixed effects.

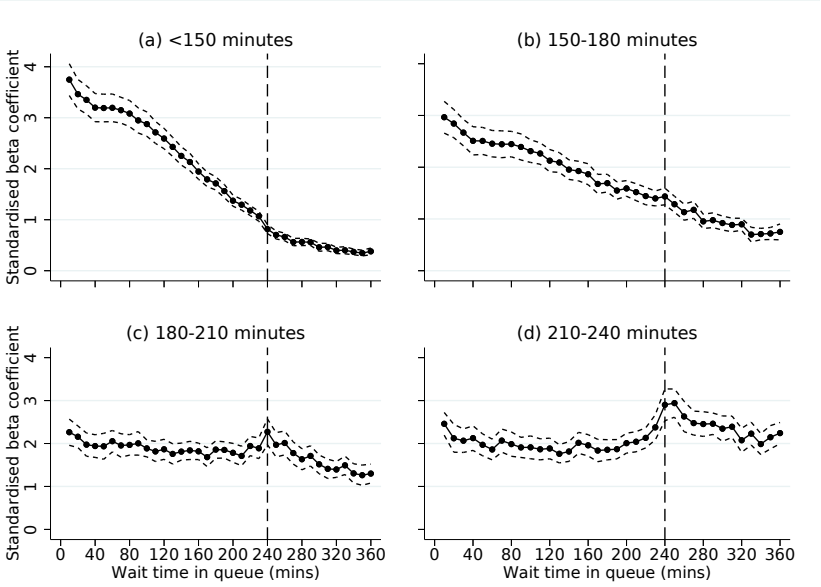
Figure 3.6 presents the estimated β_k coefficients from Equation (3.5). We normalise coefficients so they can be interpreted as the impact of a one standard deviation increase in the queue length at each horizon on newly arriving patients' wait times. Panels (a) - (d) show results for each group separately. Looking first at the early exit groups (panels (a) and (b)), the plot shows longer queues increase wait times and impacts decline with the time horizon. There is no evidence of disproportionate impacts around four hours in either group.[36] For the late exit groups (panels (c) and (d)), there is again evidence of longer queues increasing wait times but for these groups there is clear evidence of a discontinuity at the four-hour threshold. This indicates that, for the late exit groups, doctors actively

[36]We would expect to see a spike around the target in panel (b) if the exclusion window should start prior to 180 minutes. The similarities between panel (a) and panel(b) therefore suggest that 180 minutes is a reasonable choice for the lower bound of the exclusion window.

substitute resources away from newly arriving patients towards those patients that are at risk of breaching the target. The spike is largest for the group who are predicted to wait between 210 and 240 minutes (panel (d)), indicating that the greatest substitution occurs between the most similar patients. These results suggest that there are temporary substitution responses for patients predicted to be within the exclusion window (late exits) but not for those predicted to be in the earlier part of the distribution (early exits).

Taken together, we interpret these tests as providing strong empirical support for the plausibility of Assumptions 3 and 4 in our setting and proceed on that basis.

Figure 3.6: *Impact of queues on wait times for arriving patients by predicted waiting times*



Notes: (1) Wait times defined as the time from arrival in the ED to leaving the ED; (2) We normalise coefficients so they can be interpreted as the impact of a one standard deviation increase in the queue length at each horizon on newly arriving patients' wait times; (3) Predicted wait times are estimated using a regression of wait times on age, gender, diagnosis fixed effects and an ambulance indicator. Panel (a) contains all individuals with predicted wait times below 150 minutes. Panel (b) includes individuals with predicted wait times between 150 and 180 minutes. Panel (c) includes individuals with predicted wait times between 180 and 210 minutes. Panel (d) includes individuals with predicted wait times between 210 and 240 minutes.

3.5.2 No selection assumption

We set out our methodology for a test of Assumption 5, based on observable demographic and prior health variables, in Section 3.4.2. Table 3.2 presents the results of the relevant tests. Column (1) presents estimates of the distortion effect

and column (2) presents estimates of the distortion effect as a proportion of the counterfactual mean. Panel A presents results using individual variables, where we test using age, a male indicator, an indicator for whether the patient arrived via ambulance, and the number of emergency admissions, total number of days spent in hospital and the Charlson co-morbidity index ('past-CCI') based on the 12 months of hospital admissions prior to the beginning of our ED data. Each of these variables should be unaffected by decisions made in the ED, and thus allow us to test our selection assumption.[37]

For age, ambulance-arrival, past number of emergency admissions, and past days spent in hospital, we cannot reject the no-selection hypothesis. In contrast, we reject the no-selection hypothesis for gender and past-CCI. The gender result suggests that post-threshold movers are more likely to be female than the post-threshold non-movers. However, the extent of this selection effect is small: the difference between the observed and composition-adjusted counterfactual proportion of females in the pre-target period is 0.5 percentage points (1.1% of the baseline).[38] With regard to the past-CCI results, the positive estimate suggests post-threshold movers are on average less healthy than post-threshold non-movers, with a past-CCI score that is 4% higher. While this estimate is small in magnitude, this is consistent with physicians responding to the target by prioritizing patients with a worse health record.

Panel B in Table 3.2 presents results for variables that are linear combinations of the individual demographic variables. We use predicted admission and predicted mortality, where the predictions are obtained from linear regressions of the outcome on a flexible specification of the demographic variables (past-CCI score, previous hospital days and emergency admissions, and a fully interacted set of age, gender and ambulance-arrival fixed effects). The R^2 statistic from these predicted regressions is 0.22 and 0.06. An advantage of using these predicted variables is that they weight individual demographic variables according to their relative importance for clinical outcomes. Weighting factors on this basis is useful because selection on factors which do not impact these outcomes is unlikely to bias our estimates. Looking at the estimates, the demographic tests for these predicted variables cannot reject the hypothesis of no-selection. So even though the gender

[37]We considered several other variables for the demographic test. Unfortunately a number of variables have missing data around the threshold (e.g patient ethnicity, ED diagnosis). As a result we restrict our attention to examining health variables that were recorded prior to the ED visit, and whose recording will therefore not be affected by the target itself.

[38]Exploring the male indicator more carefully shows that, unlike the other variables we study, it is poorly correlated with ED wait times. This causes the polynomial regression, which we use to determine the counterfactual outcomes, to fit the data less robustly (i.e. it is sensitive to the choice of polynomial). The demographic test is therefore not reliable for this variable. The same is also true for other variables that we considered for the test, including whether the patient lives in an urban area and the deprivation level of the local area.

Table 3.2: *Demographic tests of the no-selection assumption*

	Distortion effect (Δ_D)		CAC mean
	Level	%	Level
	(1)	(2)	(3)
Panel A: Individual characteristics			
Age	0.417	0.009	46.47
	(0.284)	(0.006)	
Male	-0.005^{***}	-0.011^{***}	0.487
	(0.001)	(0.003)	
Ambulance	-0.002	-0.005	0.440
	(0.004)	(0.010)	
Past-CCI	0.013^{***}	0.043^{***}	0.300
	(0.005)	(0.016)	
Hospital days (2010)	0.066	0.018	3.67
	(0.058)	(0.016)	
Emergency admissions (2010)	0.020	0.052	0.385
	(0.014)	(0.037)	
Panel B: Predicted characteristics			
Predicted admission	0.003	0.008	0.308
	(0.002)	(0.007)	
Predicted mortality	0.000	0.015	0.019
	(0.000)	(0.015)	

Notes: (1) CAC mean is measured over the pre-threshold period, $E[y_0 \mid \underline{w}_1^-]$; (2) Predicted variables defined using a regression of the variable on past-CCI score, number of emergency admissions and days spent in hospital in 2010, and a fully interacted set of age, gender and ambulance-arrival fixed effects; (3) Bootstrapped standard errors clustered at the hospital trust level (199 repetitions).

and past-CCI tests reject the hypothesis, the contribution of these variables to salient medical outcomes and thus the likelihood of bias is low.

As a direct test of whether gender and past-CCI introduce meaningful bias to our estimates, we computed estimates conditional on these observables and compared them to our baseline estimates that we present below. The two sets of results were very similar, suggesting that any selection does not introduce substantive bias to the estimates.[39] We also note that any bias from selection on (unobservable) severity, if it mirrors the past-CCI result, would attenuate our estimates towards zero and thus make our mortality estimates conservative.

As a final probe of the no-selection assumption, we simulated how selection of different degrees would manifest itself in the observed data. To do this we built

[39]To compute the conditional estimates, we apply our methodology to subgroups of patients defined by gender and past-CCI, and then aggregate these results up to be comparable to the baseline estimates using the sample weights associated with each subgroup.

a simulated dataset using the counterfactual wait time and age distributions, and then artificially assigned post-threshold movers using different selection rules. We describe this process in more detail in Appendix B.2. The simulation highlights three facts about selection in our setting. First, the observed data on age looks very similar to the simulated data with a random selection rule. Second, even very modest selection is predicted to have a clear impact on the data, by creating a spike in outcomes in the pre-threshold period and a very pronounced 'dip' in outcomes in the post-threshold period, and neither of these features of selection are present in our data. Third, an advantage of our test is that it has potential to detect selection-on-unobservables even though it relies purely on observables. This follows since a test based on age, for example, would reveal selection on another unobservable variable as long as it is sufficiently correlated with age.

Together these results indicate that the no-selection assumption is plausible in this setting. On its face, this is perhaps a surprising finding. While patients themselves do not make the selection decision, hospitals do make these choices and selecting certain patients may be in their interest. But on a day-to-day basis, hospitals are treating patients at different times and we find this limits the scope for selection. If the data is segregated into hospital-hour periods, for example, then the number of patients approaching the target at any given point in time is actually small, at around 3 to 4. This compares to an average of 3.5 physicians that are on shift in a typical ED, suggesting physicians rarely have a choice between multiple 'potential breach' patients.[40] Rather than being a result of selection on patient characteristics, we view breaches of the target as more likely to occur due to idiosyncratic events and delays (e.g. staff shortages). While we cannot rule out that such events could be correlated with patient characteristics, our demographic tests suggest that this is not the case.

In practice, we therefore treat those patients observed with wait times in excess of 240 minutes (post-threshold non-movers) as comparable to those patients that would have had wait times in excess of 240 minutes in the absence of the target (post-threshold movers), and we can therefore use these post-threshold non-movers as the counterfactual for the post-threshold movers.

[40]We do not have detailed staffing data but obtained this number from a data request sent to all hospital trusts. We received responses from roughly 40% of trusts, and average physician figure is an average of physician numbers across all hospital-hours.

3.6 Results

We begin this section by first presenting the wait time results. We then present the results concerning treatment decisions and mortality outcomes. We explore the mechanisms behind the mortality outcomes in Section 3.7.

3.6.1 Wait times

Figure 3.2 shows the observed wait time distribution and our estimated counterfactual distribution. The shaded panel is the exclusion window where we estimate the effects of the policy, covering the period between 180 and 400 minutes. The solid line is the observed distribution of patients that exit at each interval and the dashed line is the estimated counterfactual distribution. The effect of the target on exit times is clear: a large proportion of patients from the post-threshold period (240 to 400 minutes) are moved to the pre-target period (180 to 240 minutes); these are the patients we refer to as post-thresholder movers. By comparing the observed wait time distribution with our counterfacutal we can compute the impact of the target on average wait times.

The results indicate that the target is successful in achieving its primary aim of reducing wait times. We estimate that the target reduces mean wait times by 7 minutes. This is equivalent to 4% of the estimated counterfactual mean. For patients affected by the target (i.e. in the exclusion window), we estimate that the target reduces wait times by 21 minutes, or 8% of their estimated counterfactual mean. Moreover, if we restrict our attention to those patients moved to the pre-threshold period from the post-threshold period (the post-threshold movers), then the average wait time reduction is 59 minutes.[41]

3.6.2 Treatments and mortality outcomes

Table 3.3 presents results of the distortion test for a range of treatment decisions, costs and mortality outcomes. Each row shows results for a separate outcome. Column (1) presents estimates of the distortion effect and column (2) presents estimates of the distortion effect as a proportion of the counterfactual mean.

Panel A presents estimates for treatment decisions in the ED. We find that, controlling for compositional changes, there is an increase in the probability of

[41] To obtain estimates of the distribution of wait time reductions would require further assumptions on the ordering of patients. We do not impose these assumptions, but note that the maximum wait time reduction could be as large as 200 minutes (i.e. a patient moved from 390 minutes to 190 minutes).

admission of 4.6%. This is 12.2% of the baseline composition-adjusted counterfactual value, which is sizeable. The results for discharges and referrals out of the ED to specialist clinics or hospitals offset these admission effects, with roughly three-quarters of the effect coming from decreased discharges, and one-quarter from decreased referrals, although as a percentage of the baseline these responses are of comparable magnitude.

We also show target effects on the number of investigations performed in the ED, such as x-rays, blood-test and CT scans. We find that investigations rose by 0.1 per patient, or 4.6% of the baseline. We do not, however, find any effect on the number of treatments performed in the ED. This suggests that doctors perform more tests in order to speed up the admission decision for individuals (i.e. they perform an extra test instead of monitoring the patient for a longer period of time) but has little effect on the treatments that they provide in the ED.

Panel B examines inpatient treatment decisions. For inpatient treatments, in order to avoid selection, we include all ED patients, even those who did not end up being admitted. As a result, the coefficient represents the incremental amount of treatment due to the four-hour target. We find no evidence of any statistically significant increases in length of stay or the number of procedures. This suggests that the extra admissions do not receive substantial amounts of care in the hospital. That is, these admissions appear to be largely placeholders in order to avoid the four-hour target.

Nevertheless, the additional admits are costly. Panel C of Table 3.3 examines the impact of the four-hour target on 30-day patient costs. There is a small rise in ED costs of $3 (£2), or two percent of ED costs. But there is a significant increase in inpatient costs of $126 (£93), which is 5% of inpatient costs. That is, even though most patients appear to be only housed in inpatient departments as a way of avoiding the four-hour target, these admissions generate transfers from the government to hospitals. Total costs rise by roughly 5% relative to the baseline.

Panel D then extends our analysis to look at patient mortality outcomes. We consider mortality at a variety of time frames, ranging from 30 days after entering the ED to 1 year later. We find significant short term declines in mortality. Mortality over 30 days declines by 0.41 percent, or 14% of baseline. The CAC for 30-day mortality is shown in Appendix Figure A4; here, after adjusting for the composition effect, we find that the observed data is lower than the CAC and this is what produces the negative estimate. This effect fades slightly over time and falls as a share of the baseline, so that at one year it is only 3.1% of baseline. This pattern suggests that the health benefits of the four-hour policy are seen in the short term.

Table 3.3: *Estimated distortion effects of the target on treatment decisions, costs and mortality*

	Distortion effect (Δ_D)		CAC mean
	Level	%	Level
	(1)	(2)	(3)
Panel A: ED treatment decisions			
Pr(admission)	0.046^{***}	0.122^{***}	0.379
	(0.008)	(0.022)	
Pr(discharge)	-0.033^{***}	-0.070^{***}	0.472
	(0.007)	(0.014)	
Pr(referral)	-0.013^{***}	-0.089^{***}	0.150
	(0.003)	(0.020)	
ED investigation count	0.108^{**}	0.046^{**}	2.369
	(0.048)	(0.021)	
ED treatment count	−0.033	−0.016	2.070
	(0.028)	(0.014)	
Panel B: Inpatient treatment decisions			
Length of stay (days)	0.035	0.015	2.302
	(0.048)	(0.021)	
Inpatient procedure count	0.000	0.001	0.290
	(0.006)	(0.020)	
Panel C: Hospital costs			
30-day ED cost	3.040^{***}	0.016^{***}	192.950
	(0.911)	(0.005)	
30-day inpatient cost	125.793^{***}	0.052^{***}	2,414.087
	(33.992)	(0.015)	
30-day total cost	128.833^{***}	0.049^{***}	2,607.037
	(34.389)	(0.014)	
Panel D: mortality			
30-day mortality	-0.0041^{***}	-0.138^{***}	0.029
	(0.0006)	(0.019)	
90-day mortality	-0.0040^{***}	-0.079^{***}	0.048
	(0.0010)	(0.019)	
1-year mortality	-0.0029^{*}	-0.031^{*}	0.090
	(0.0016)	(0.017)	

Notes: (1) CAC mean is measured over the pre-threshold period, $E[y_0 \mid \underline{w}_1^-]$; (2) All inpatient variables (e.g. length of stay, costs) take on the value zero for patients that are not admitted; (3) Bootstrapped standard errors clustered at the hospital trust level (199 repetitions).

This is a sizeable mortality decline given the modest increase in costs documented in Table 2. We find that total costs over 30 days from admission to the ER rise by 5%, while mortality falls by 3.1% over a year. Calculating the cost per year of life saved by the policy requires assumptions on how long-lasting is the impact on mortality and on any subsequent costs past 30 days. Assuming no subsequent costs, but also assuming that the mortality impact only lasts one year, this implies a cost per year of life saved of $43,000 (£31,850).[42] This is low relative to standard valuations of a life-year in the U.S., where typical benchmarks are around $100,000 (£74,000) (Cutler, 2003), and at the upper end of valuations in the U.K., where the national benchmarks are set at $28,000 to $42,000 (£20,000 to £30,000)(McCabe et al., 2008).

In summary, then, our analysis of the four-hour target shows that it led to shorter wait times, more admission, only marginal additional costs (due to little use of inpatient care for those admitted), and significant reductions in mortality. That is, it appears that constraining hospitals did save lives.

3.7 Mechanisms

3.7.1 Using Patient Heterogeneity to Identify Mechanisms

Our results so far show a number of effects of the wait time target on patient treatment – on wait times, admission probabilities, and treatment costs more generally. We also show a significant effect on patient mortality. Ideally we would like to uncover the mechanism through which the four-hour target impacts patient mortality. This is difficult since we essentially have one instrument (the target) and multiple changes in patient treatment.

To address this issue we turn to considering heterogeneous impacts across types of patients. That is, we examine whether there are groups of patients where there are differential effects of the four-hour target. If those groups have effects that are focused along one channel (e.g. wait times) but not another (e.g. admits), then we can use this to separate the effect of the two channels on outcomes.

We consider two natural sources of heterogeneity. The first is differences across diagnosis. In particular, we divide patients into 36 diagnosis groups.[43] It seems

[42]This reflects the cost to the government of the policy due to the increase in HRG transfers to hospitals. The actual cost in terms of resource-use will be even lower if the marginal admissions due to the policy use fewer resources than the average HRG cost. The calculation also omits the fines levied on hospitals for breaching the target. Incorporating these fines into the cost calculation reduces the cost per year of life saved (as the government effectively recoups some of its additional expenditure through collecting the fines) but this effect is very small because performance was close to the 95% target in the period we analyze.

[43]The data assign patients to 40 diagnosis categories, including a 'missing' category. We exclude four diagnoses (nerve injuries, electric shock, near drowning and visceral injury) as

likely that the largest wait time impacts of the target will show up for those who have the most severe diagnoses, since they are the most likely to hit the wait time target. This is indeed what we see in the data: Figure E1 in Appendix B.5 shows that the probability of hitting 240 minutes is in fact much higher for the most severe diagnoses. We therefore separately compute the wait time reduction effects, and distortion effects for admissions and 30-day mortality for each diagnosis group.[44] We then assess how the heterogeneity across diagnosis groups translates to each of these outcomes.

We first examine these results graphically. Panel A in Figure E2 in Appendix B.5 shows that higher severity diagnoses have larger wait time effects. This is sensible since they are most likely to wait the longest without the four-hour policy. But Panel B shows that the effects of the target on hospital admissions is no higher for more severe diagnoses. That is, the more severe diagnoses are getting treated sooner, but are no more likely than others to have that treatment resolve in an extra hospital admission. Panel C shows the absolute value of mortality reduction for each diagnosis group, and clearly shows that the mortality effect of the four-hour target is strongest for the most severe diagnosis. To ensure that selection is not driving our result, the graph also repeats this exercise for predicted mortality and finds no systematic relationship between the effects of the target on predicted mortality and the severity of the diagnosis.

The results of this exercise are formalized in columns 1-3 in Table 3.4. For each diagnosis group, we regress the distortion effect on mortality in absolute value for each group, on the estimated wait time reduction and the distortion for admission proability. Essentially, these regressions report associations between the estimated impact on mortality and the estimated impact on wait times and admissions, using a grouping estimator with groups defined by severity or inpatient crowding. A positive coefficient in these regressions can be interpreted as that margin being associated with a larger policy effect on 30-day mortality.

Column 1 shows that across the 36 diagnosis groups, those groups with larger wait time effects have larger mortality effects. The estimated coefficient suggests that each additional minute of wait time reduction increases the mortality reduction by 0.001 percentage points. Earlier, we estimated that wait times fell by 19 minutes on average. This suggests a mortality reduction of 2.2 percentage points.

small samples do not allow us to separately estimate the impact of the target for these groups. We also tested whether the missing data had an impact on the results, by conducting the same analysis for hospitals that had fewer missing data points and we found similar results to those presented here.

[44]We adjust the polynomial choice for each diagnosis to ensure that it both fits the data well and meets the condition that the excess and missing mass are equal. To do this we use an approach that maximizes the adjusted-R^2 of Equation (3.1) for each outcome. We use the same approach for the crowding analysis that follows.

Table 3.4: *OLS regressions of the estimated 30-day mortality reductions on other effects of the target*

	Diagnosis groups			Crowding-severity groups		
	(1)	(2)	(3)	(4)	(5)	(6)
Wait time	0.118***		0.115***	0.083***		0.066***
	(0.034)		(0.034)	(0.018)		(0.022)
Admission probability		−0.059	−0.029		−0.088***	−0.037
		(0.065)	(0.058)		(0.024)	(0.028)
N	36	36	36	95	95	95

Notes: (1) Dependent variable is the absolute value of the target impact on 30-day mortality measured as % of the CAC mean over the pre-threshold period; (2) Independent variables are the absolute value of the target impact on the respective variable, measured as a % of the CAC mean over the pre-threshold period.

This is of a similar magnitude to our reduced form estimate in Table 3.3 of 3-4 percentage points. Column 2, however, shows that there is no impact of the increase in admissions on mortality. And column 3 shows it is still the case that groups with larger wait time effects, but not larger admit effects, have larger mortality effects when we consider both variables together.

Given that there is an effect on wait times, but not admissions, this suggests that it is wait time reductions and not increased admissions that are driving the results. Of course, this set of corresponding facts do not prove this causal mechanism because there may be other factors that cause the effects to differ by diagnosis. So to further test this conclusion we consider a second source of heterogeneity.

We next turn to heterogeneity by the degree of inpatient crowding. In times where the inpatient department is more crowded, EDs may be less able to address their wait time targets by admitting patients because the inpatient wards have less spare capacity for these patients to be sent. But it is unclear that inpatient crowding would much affect the marginal wait time impacts of the target. Inpatient crowding therefore provides an opposite test of the diagnosis heterogeneity: an opportunity to observe heterogeneity that drives admission probabilities but not wait times.

To assess this, we divide the data into 50 quantiles depending on how busy the hospital inpatient department is on the day of admission. For each hospital-day, we calculate the daily number of inpatients treated by the hospital, and use this to assign each hospital-day to one of 50 groups in the hospital-specific

distribution of inpatient crowding. Patients are then assigned to each of these groups depending on their day of arrival.[45] To address differences in casemix during busy and quiet periods, we also split patients into two severity groups. 'Major' diagnoses are defined as those with a 30-day mortality rate above the overall 30-day mortality rate (1.6%). Interacting the 50 inpatient crowding groups with severity yields 100 groups. For 95 of these groups we have sufficient sample size to independently compute the effects of the target, and therefore across which to examine heterogeneity in effects.

Again, we first examine the patterns graphically. Figure E3 presents the results of this second heterogeneity test. The figure shows the results for these observations, ranked from least crowded to most crowded. Panel A shows that inpatient crowding has a weak, positive relationship with wait times. Panel B shows a strong, negative relationship between crowded inpatient departments and smaller increases in admission. So this source of heterogeneity gives the opposite results of what we saw for severity: a small effect on wait times and a large effect on admissions. Therefore, if our earlier supposition is correct that it is wait times and not admissions that drives our mortality effects, we should see little differential impact on mortality across these groups.

In fact, that is exactly what we see in Panel C in the black circles: there is no significant relationship between the degree of inpatient crowding and the estimated mortality effect. As in Figure E2c, we repeat this analysis with estimated reductions in predicted mortality (which should be unaffected by the target once we adjust for the composition of patients) to show that these results are not driven by selection. The red triangles show that the predicted mortality effects are again close to zero. There is a positive relationship between predicted mortality reductions and inpatient crowding but this is small in magnitude.[46]

Columns 4-6 of Table 3.4 formalize the results of this analysis. As for the analysis examining heterogeneity across diagnoses groups, a positive coefficient in these regressions can be interpreted as that margin being associated with a larger policy effect on 30-day mortality. Once again, we have a highly significantly relationship between the wait time reduction and mortality reduction, with a coefficient that is similar to column 1. In this case, in column 5, we do see a significant effect of the admissions effect on mortality, albeit with a wrong signed

[45] We calculate the inpatient census at the daily level as the data do not contain information on time of arrival at, or discharge from, the inpatient department.

[46] This means that our results may actually understate the mortality reductions in the most crowded periods. Given that these periods are also those with the smallest increases in admissions, this would strengthen the conclusion that mortality reductions are associated with reductions in wait times and not additional admissions.

coefficient suggesting that a larger admissions effect leads to a smaller mortality effect. But when both are included in column 6 only the wait time effect persists.

Taken together, the evidence suggests that heterogeneity associated with wait time variation appears associated with mortality variation, while heterogeneity associated with admissions variation does not. This does not prove that the wait time reductions are driving our mortality reductions, but it is highly suggestive.

3.7.2 Wait times, diagnoses and causes of death

This evidence raises the question of how reductions in wait times could lead to lower mortality rates. The most likely mechanism is that reductions in wait times lead to lower time-to-treatment for patients with severe diagnoses. An extensive medical literature makes it clear that rapid treatment is associated with better mortality outcomes for patients across a range of conditions. For example, Seymour et al. (2017) find a strong positive association between time-to-treatment and survival for ED patients with sepsis and septic shock.[47] However, it may be difficult for physicians to identify these patients as they arrive at the ED: a body of medical evidence suggests that misdiagnosis in the ED is not uncommon, while there is often disagreement between ED physician and subsequent specialist diagnosis.[48] This suggests why the target may have been successful in improving outcomes relative to an unconstrained scenario as it leads doctors to speed up treatment for all patients, which is costly but ensures that hard-to-diagnose and time-sensitive patients ultimately get the correct treatment sooner.

We explore the likelihood that this mechanism is driving our results in two ways. First, we look at how hospitals achieve the reductions in wait times by examining which parts of the treatment pathway they are compressing. This provides some evidence on whether patients start to receive treatment earlier. Second, if wait times are driving mortality reductions we would expect to see the greatest mortality reductions for patients with conditions where outcomes are known to

[47]There are also many examples from other diagnoses. For example, Saver et al. (2013) find significant improvements in mortality and post-hospital outcomes for stroke patients when cutting time-to-treatment. Cannon et al. (2000) also find substantial increases in mortality following a heart attack when patients receive angioplasty more than two hours after arriving at hospital.

[48]Shojania et al. (2003) conducted a systematic review into studies of autoposy-detected diagnostic errors over a 40 year period in the US and found a median error rate of 23.5%, although this rate was decreasing over time. Delays and misdiagnoses are particularly common for neurological and cerebrovascular patients, and many of the existing studies are in this area. For example, Newman-Toker et al. (2014) estimate that between 15,000 and 165,000 cerebrovascular events are misdiagnosed annually in US EDs, while Moulin et al. (2003) found that half of ED patient diagnoses in a large French hospital were changed after neurology consults were obtained, following access to more detailed testing equipment (e.g. CT and MRI scanners) which were not available to the original ED physician.

be time-sensitive. We therefore examine variation in mortality reductions across diagnoses and primary causes of death.

To examine how wait times are reduced, we break down the overall impacts on waits into the three separate components that the data allows: time to initial assessment; time between assessment and the beginning of treatment; and duration of ED treatment. The initial assessment is usually conducted by a triage nurse, and includes a relatively basic examination. Treatment begins when the patient is first examined by a doctor (i.e. when the first ED treatment is received, and we document the common first treatments in Table C1), and ends when the ED makes the decision to admit or discharge the patient. Admitted patients will then receive further treatment from a specialist within the hospital. As noted in Section 2.1 and shown in Appendix Tables C1 and C2, most ED treatment in England is aimed at stabilising and diagnosing patients, with more extensive treatments provided by specialists in inpatient wards. Reducing treatment time in the ED therefore means that patients start to receive this specialist treatment sooner.

We repeat the analysis in Section 4.3 using time to initial assessment, time between assessment and the start of treatment, and duration of ED treatment, as separate outcomes. The results show that the reductions are achieved both by reducing the initial wait for treatment (48% of the overall reduction), and by shortening the duration of ED treatment (45%). The remaining reduction (7%) is explained by the initial time to assessment. These results suggest that the target reduces the wait for both ED and specialist treatment to begin. Patients start to receive treatment from ED physicians sooner and spend less time receiving treatment in the ED.[49] Importantly, shorter periods spent in the ED also mean that admitted patients start to receive specialist inpatient treatment sooner. This specialist treatment often begins with further diagnostic testing (such as CT or MRI scanning, as documented in Table C3), and so reducing ED time means patients are likely to receive a detailed diagnosis sooner.

Next we turn to examining which patients benefit most (in terms of mortality reductions) from the target. If quicker treatment is responsible for improving patient outcomes, we would expect to see the greatest improvements for patients with diagnoses that can be affected by time-sensitive treatments. We therefore examine in which diagnosis groups we see the biggest impact of the target on patient outcomes. Table A3 in Appendix B.1 shows the estimated impact on mortality and wait times within each of the 40 ED diagnoses categories, ordered by the size of the mortality reduction. The largest impacts of the target on mortality

[49]The data do not include more detailed information on the amount of time actually spent with physicians. As a result we cannot test whether the target reduces time spent being treated by an ED physician. However, our results in Table 3.3 suggest that patients do not receive fewer ED treatments as a result of the target.

rates are found in patients with septicaemia, cerebo-vascular (stroke) and other vascular injuries. These are all areas, as noted above, in which medical evidence suggests benefits to patients from reduced time-to-treatment.

While the impacts are largest among these diagnoses, the total number of patients saved in each diagnosis group will also depend on the number of patients who attend ED with these diagnoses. For example, septicaemia is a relatively rare condition, while respiratory problems are more common. We therefore use the estimates to compute the number of patients for each broad ED diagnosis who survived for at least 30 days following their ED visit as a result of the target.

The last column of Table A3 reports these estimates as the share of total lives saved among patients with complete diagnosis information. The aggregate estimates indicate that in 2012/13, 17,800 patients were saved by the target, or just under one patient per hospital every three days. Among patients with complete diagnosis information, a third of the lives saved are from patients attending the ED with a respiratory problem. Gastrointestinal, cardiac and cerebro-vascular diagnoses also explain substantial shares of the lives saved. While these categories are still relatively broad, they provide reassurance that the majority of the mortality reductions come from serious conditions where timely treatment can plausibly make a difference to patient outcomes.

An alternative way of analyzing which patients are saved by the target is to examine whether we observe reductions in the specific causes of death contained in the official mortality records that are linked to our hospital records (ICD-10 codes of primary cause of death). These data provide a far more granular record of a patient's condition than the ED diagnosis data. However, while the recording of ED diagnosis should be independent of the the treatment received, mortality is an endogenous outcome. Rather than split the sample by cause of death, we therefore use indicators for specific causes of death as outcome variables and test whether the target reduced the prevalence of each cause. In this way we can further test the time-to-treatment mechanism by examining whether the target reduced deaths in time-sensitive conditions, but not in conditions that we would not expect to be time-sensitive in an acute setting.

We begin by classifying deaths into 23 categories according to the first letter of the ICD-10 code, and repeat our analysis with the 23 dummy variables as outcome variables. Table A4 in Appendix B.1 shows the results. We find that 70% of the reduction in mortality can be explained by reductions in deaths related to circulatory (30.1%), respiratory (25.7%) and digestive (15.0%) conditions. These are all categories which include specific conditions that are likely to be time-

sensitive.[50] In contrast, there is no significant reduction in mortality attributed to neoplasms (cancers), which include a number of high mortality conditions unlikely to be time-sensitive.

Undertaking the same analysis in yet more detail, we analyze more detailed causes of death using the first 2 digits from the ICD-10 codes. We examine the ten most common causes of death for ED patients, which together account for 60% of all patient deaths. Table A5 in Appendix B.1 presents the results. The estimates again show that the greatest mortality reductions are found for time-sensitive conditions. The largest effects are found in patients with cerebrovascular diseases, both as a proportion of the overall mortality reduction (column 3) and as a proportion of deaths due to the specific cause (column 4). Deaths from chronic lower respiratory diseases, influenza and pnuemonia, and ischemic and pulmonary heart diseases are also substantially reduced.

In contrast, we again observe no significant changes in mortality associated with cancers of any type. These are all conditions which we would expect to be less sensitive to time-to-treatment in an acute setting, and so act as a convincing placebo test when examining the time-to-treatment mechanism.

While these results do not provide definitive proof that wait time reductions are causing mortality reductions, they do provide reassuring evidence that many of the mortality reductions occur in diagnoses where timely treatment is known to be important, and not in areas where it is less so. Wait times, and specifically time-to-treatment, therefore do appear to play an important role in explaining patient outcomes.

3.8 Conclusion

The Emergency Department is a central node of health care delivery in developed countries around the world. It is the entry point into the hospital for a large share of patients and decisions made rapidly by ED staff have fundamental impacts on the entire course of care. Despite the complicated nature of these decisions, there remains dissatisfaction in most health care systems with the level of crowding in EDs and the speed with which cases are resolved. This has led in recent years to both open competition on ED wait times and to regulatory interventions to reduce those times.

We study one type of regulatory intervention, the four-hour wait target policy enacted in England. We find that this target had an enormous effect on wait times, as illustrated vividly by the spike in the wait times distribution at the four-hour

[50] Circulatory conditions in the ICD-10 data include strokes, which are the most common cause of death for patients with the ED diagnosis of cerebro-vascular.

mark. We use well-established bunching methodologies, which we apply in a new setting and provide support for the required assumptions, to estimate that this represents a significant reduction of around 20 minutes, or 11%, in the average wait time of impacted ED patients.

We then turn to assessing how this change in wait times impacted patient care and outcomes. We do so by introducing an econometric framework that allows us to separate the compositional impacts of individuals shifting from after to before the four-hour target from the distortionary effect of the four-hour target on medical decisions. We find this target led to a significant rise in hospital admissions. These admissions do not appear to involve much new treatment, suggesting that they may just be 'placeholders' to meet the target. But there is nonetheless a significant rise in inpatient spending of about 5% of baseline.

At the same time, we find striking evidence that the target is associated with lower patient mortality. There is a 0.4 percentage point reduction in patient mortality that emerges within the first 30 days, amounting to a large 14% reduction in mortality in that interval. This reduction fades slightly over time, so that after one year it amounts to a 3.1% mortality reduction. While modest, this effect is large relative to the extra spending, suggesting a cost of extending life by one year of $43,000 (£31,000). Finally, we exploit heterogeneity across patient types to show that this effect arises through reduced wait times, not through increased inpatient admissions, with the majority of mortality reductions occuring in diagnoses where rapid treatment is known to benefit patients.

The implications of our finding is that, unconstrained, EDs in England are not making optimal decisions on patient wait times. By reducing wait times, the four-hour target induced cost-effective mortality reductions. This is likely a lower bound on the welfare gains due to the target, as it does not value the other benefits to consumers from waiting shorter times, although there may be welfare costs from the extra admissions (Hoe, 2017).

Of course, this result only applies to the specific target studied here, and does not necessarily imply that other limits would have equal effects. It is also unclear how this result applies to other nations with different means of rewarding or incentivizing EDs. A question raised by our results is why physicians and EDs do not optimize wait times in the absence of the policy. One credible explanation is that physicians are simply imperfect agents for their patients, a longstanding concern in medical markets (Arrow, 1963). This seems especially plausible in our setting where physicians are dealing with patients prior to their full diagnosis being revealed. An alternative explanation could come from physicians lacking information on the relative benefits of timely treatment for certain patients. In

practice, however, we are unable to separate these two potential explanations in this analysis.

Importantly, in both cases, our results suggest that ED physicians working in an unconstrained setting appear to systemtically keep patients in the ED too long, such that an information-free policy (such as the four hour target) delivers better outcomes for patients. This suggests that better targeted interventions could potentially deliver further improvements for patient outcomes. More work is clearly needed to understand informational constraints and the proper set of rules and incentives necessary for delivering cost-effective ED care.

Chapter 4

The Impacts of Private Hospital Entry on the Public Market for Elective Care in England [1]

4.1 Introduction

Efforts to promote competition between providers have been a common feature across healthcare systems around the world in recent decades. In the English National Health Service (NHS), the introduction of patient choice over providers in the mid-2000s aimed to incentivise competition between hospitals in order to improve the efficiency and quality of healthcare. Previous work has found these reforms were broadly successful, with reductions in mortality as a result of increased competition (Cooper et al., 2011; Gaynor et al., 2013, 2016).

An important yet often overlooked component of these patient choice reforms was the introduction of private providers to large parts of the public market for elective healthcare. Private providers entered the market in two stages in the 2000s (Naylor and Gregory, 2009). Starting in 2005, purpose-built and privately owned surgical centres known as Independent Sector Treatment Centres (ISTCs) were introduced to boost public capacity and reduce waiting times. This was followed by the widespread entry of pre-existing private hospitals to the public elective market in the late 2000s. These hospitals were paid the same pre-determined price for providing elective care as existing public hospitals, with the aim that hospitals would instead compete for patients on the basis of quality. It was hoped

[1]This paper is joint work with Elaine Kelly, and has been published in the *Journal of Health Economics* (Kelly and Stoye, 2020). Thanks to Owen O'Donnell and two anonymous referees for helpful comments, as well as Thomas Hoe, Carol Propper, Imran Rasul, Marcos Vera-Hernandez and Ben Zaranko, and seminar participants at IFS, UCL and the University of Manchester.

that this would create competitive pressure for public hospitals to improve quality and productivity.

In this paper we examine the impact of the entry of private hospitals on the public market for elective hip replacements, where private providers delivered a fifth of all procedures by 2012/13.[2] We first consider the impact on the volumes of publicly funded procedures, considering both the total number of patients and the number of patients treated by existing public hospitals, before analysing the effects on waiting times and readmissions. Increases in total volumes would indicate that more patients benefited from publicly funded procedures due to an increase in the supply of publicly funded elective care. However, such an expansion in capacity might also dampen competitive pressures exerted by private entrants given long waiting lists for surgery: in this case, public hospitals could simply replace lost patients with those further down the list, and as a result would have little incentive to improve quality.

Changes in the size of the market would also have implications for who benefits from the reforms. We therefore also examine how provider entry changed the composition of publicly funded patients, estimating effects on the observed severity of patients, documenting heterogeneity in impacts across deprivation groups, and examining evidence of substitution away from privately financed care.

Our main analysis uses the universe of publicly funded hospital admissions to compare the changes in outcomes over an 11 year period across fixed areas which were differentially exposed to private hospital entry. This relies on the assumption that changes in outcomes would have been the same across areas with different exposure in the absence of private hospital entry. This assumption would be violated if, for example, private hospitals chose to enter markets with rising unmet demand for publicly funded care due to low public investment in healthcare. To address such concerns, we instrument private hospital entry with the pre-reform location of private hospitals. These sites were established prior to the implementation of the reform, formerly treating only privately funded patients. The reform then allowed these hospitals to treat public patients alongside their private patients. This means that while the choice to enter a specific market may be related to other factors in the area, the historical location of these hospitals should be exogenous.

To address remaining endogeneity issues, including the concern that private hospitals may be located in areas with different trends in our outcomes of interest,

[2]We focus on hip replacements for three reasons. First, it is a common procedure performed in large volumes by all large public hospitals in England. Second, private hospitals routinely conduct this procedure for private-pay patients and therefore had pre-existing capacity to carry out this surgery. Finally, unlike most other specialties in England, information on the private pay sector is available via a mandatory registry of joint replacements.

we carry out a battery of robustness checks. These include controlling for time-trends in a variety of demand factors in the local area, examining changes in areas with pre-existing private hospitals which did not enter the market, an analysis of pre-trends in outcomes, and controlling for other policy changes in the NHS at the time. In all cases, these additional analyses leave our conclusions unchanged.

We find that the introduction of a private hospital increased the total number of admissions for publicly funded hip replacements in the local area. We estimate that the entry of a private hospital in the local area by the end of the period led to an annual rise of 34 publicly funded hip replacements. This is equivalent to 11.7% of the market mean of 285 publicly funded hip replacements in 2002/03. However, there was no impact on the number of admissions at incumbent public hospitals, and therefore no change in revenues hospitals receive from these patients. This suggests that the entry of private hospitals expanded the size of the public market for elective hip replacements rather than simply reallocating patients across providers.

The estimated effects of entry on waiting times is negative but poorly defined. An additional private hospital is associated with a waiting time reduction of 11.1% (equivalent to a reduction of 27 days in 2002/03). However, this is only statistically significant at the 10% level. Estimated reductions in waiting times for patients treated at public hospitals are smaller and not statistically significant, suggesting that any reduction in waiting times were driven by shorter waits at private hospitals. The estimates also show that there were no impacts on readmission rates in any specifications or robustness tests. Together, these results are consistent with the entry of private providers increasing the capacity to deliver publicly funded hip replacements, while generating limited competitive pressure for public hospitals to improve (observable) care quality to attract new patients.

The increase in the size of the market also raises the question of who is benefiting from the policy. We explore this in several ways. First, we examine how the composition of patients changed as a result of hospital entry. Procedures are rationed partially on the basis of pain and the suitability of alternative treatments. Increases in supply could therefore lead to healthier patients receiving a hip replacement. Using different measures of secondary diagnoses, we find that private hospital entry led to a reduction in the average severity of publicly funded hip replacements. This held both for the market as a whole and for those treated by public hospitals.

Second, we explore whether the impacts of private hospital entry varied across areas with different levels of deprivation. Our estimates suggest that the impact of entry did not vary across more and less deprived areas. However, private hospitals are more likely to be based in more affluent areas. As a result, patients in these

areas are more likely to benefit from the expansion in supply. Our estimates suggest that the uneven distribution of private hospital locations can explain around 5% of the additional growth in hip replacements in the 33% least deprived areas (relative to the 33% most deprived areas) between 2002/03 and 2012/13.

Finally, we examine the extent to which the additional publicly funded hip replacements represent new procedures or substitution from the private pay sector. To do so, we use novel joint registry data that cover all hip replacements in England, including all privately and publicly funded hip replacements between 2008/09 and 2012/13, to explore the impact that private hospital entry and the associated growth in the number of procedures had on the composition of public and private hip replacement patients. Here we find that the entry of private providers to the public elective market was not associated with a reduction in the number of privately funded hip replacements conducted in the local area. This suggests that substitution between these financing streams was limited, and that the increase in publicly funded volumes represent genuinely new procedures that would not have taken place in a given year in the absence of the reform. These data again suggest that public patients were healthier on average following the entry of a private hospital, while there was no change in the severity of private patients. This reinforces our conclusion that expansions in the public market led to healthier patients receiving treatment at an earlier stage.

Our paper contributes to two literatures. First, we build upon a small body of work that has examined the impact of private provider entry to elective markets in England (Cooper et al., 2012, 2018) and in the US (Courtemanche and Plotzke, 2010; Munnich and Parente, 2018).[3] This literature has generally focused on the consequences for existing providers following the entry of purpose-built surgical centres. Existing work on reforms in England examined the impacts of ISTC entry on the efficiency of incumbent public hospitals (Cooper et al., 2018) and on the patient mix treated at new providers and public hospitals (Street et al., 2010; Bardsley and Dixon, 2011; Chard et al., 2011; Cooper et al., 2018). The subsequent reform that allowed private hospitals to enter the market has received much less attention so far, despite generating a much larger expansion in the number of providers in the market. One exception is Cooper et al. (2012), which studied the impacts of competition on the efficiency of public hospitals conducting four types of elective surgery between 2002 and 2010. They found that competition with other public hospitals led to improvements in efficiency, but these impacts

[3]Ambulatory Surgical Centres in the US play a similar role to ISTCs in England, competing with existing hospitals to provide routine procedures.

were not found when potentially facing competition from a private hospital.[4] We further contribute to this literature by examining the impact of widespread private hospital entry on the size and shape of the elective market, and the consequences for patient outcomes.

Second, our findings complement the existing evidence on the impact of hospital competition and patient choice on patient outcomes (Kessler and McClellan, 2000; Propper et al., 2004; Cooper et al., 2011; Gaynor et al., 2013, 2016). The entry of private hospitals in England played an important role in expanding the choice set of healthcare providers for patients. Understanding the impact of this change as part of the wider set of reforms to healthcare provision is therefore important. Our results suggest that, in the case of the introduction of private hospitals into the NHS, the main benefits to patients accrued from having procedures that were delivered earlier or that would not otherwise have occurred. We find no benefits to patients in terms of improved quality, which is consistent with the reform generating limited competitive pressure for public hospitals to improve observable performance. This suggests that policymakers aiming to improve quality through competition should consider restricting the size of the market when enacting any further reforms.

From a policy perspective, the role of the private sector in the NHS remains controversial more than a decade on from the original reforms. The opposition Labour Party's 2019 manifesto pledged to reduce the role of private providers, and the extent to which the NHS is included or excluded in any post-Brexit bilateral trade deals is a live political issue. This paper provides empirical evidence on the impact that the growth in private sector involvement had on patients and public hospitals in one area of NHS activity. Such evidence is often missing from current debates, and we would hope that our results could help policymakers to better understand the trade-offs when considering the role of the private sector in the NHS in future.

The rest of the paper is organised as follows. In Section 4.2 we describe the institutional setting, and set out the potential impacts of the reforms. Section 4.3 describes the data and provides descriptive evidence of the impact of private hospital entry. Section 4.4 sets out our empirical strategy. Section 4.5 presents out baseline results and a series of robustness checks. Section 4.6 examines who was affected by private hospital entry, examining variation in the impact across deprivation groups and providing suggestive evidence on the extent of substitution between public and privately funded procedures. Section 4.7 concludes.

[4]Cooper et al. (2012) found that the presence of a private hospital (which may or may not have entered the public market) was associated with a small increase in length of stay, driven by the sorting of more severe patients to public hospitals.

4.2 Institutional Background

4.2.1 Publicly funded care in England

The vast majority of health care in England is publicly funded and free at the point of use through the National Health Service (NHS). Secondary or hospital care has traditionally been delivered by publicly owned and operated NHS hospitals (henceforth 'public hospitals').[5] Patients access elective (planned) hospital services, such as hip replacements, through a referral from their primary care doctor or General Practitioner (GP). There are no self-referrals, and patients do not make any copayments. Hospitals are reimbursed by the government for the care they provide to patients, with hospitals receiving per patient payments according to a set of national tariffs.[6] NHS elective care is therefore rationed through waiting times and gatekeeping by GPs rather than prices. Patients can however choose to pay for treatment privately in a private hospital. This accounted for a fifth of hip replacements in 2002 (Arora et al., 2013).

Historically, the NHS purchased small volumes of care from the private sector on an ad-hoc basis to address short-term capacity constraints.[7] From the mid-2000s, two related reforms formalised and greatly increased the ability of privately owned providers to compete with incumbent public hospitals for publicly funded patients. The first reform introduced privately-owned surgical centres - known as Independent Sector Treatment Centres (ISTCs) - that were specifically built to initially treat only publicly funded patients. The second reform then allowed pre-existing private hospitals to enter the market to compete for publicly funded elective patients with incumbent public hospitals and ISTCs. In this paper, we examine the impacts of the second reform (the entry of pre-existing private hospitals) on the public and private market for elective hip replacements.[8]

ISTCs were initially set up as privately owned and operated facilities designed specifically to treat public patients for routine procedures. This design reflected the focus of NHS policy in the early 2000s, which aimed to reduce the very long waiting times within the NHS, initially through strict waiting time targets backed

[5]These hospitals are often grouped together to form NHS Acute Trusts. For ease of expression we will refer to these trusts as NHS or public 'hospitals' throughout.

[6]Hospital care is grouped into Healthcare Resource Groups (HRGs), which are similar to Diagnosis-related Groups in the US. Prices or tariffs are then set at a national level based on the average cost of providing the associated care. Small adjustments are made for unavoidable local differences in costs and length of stay.

[7]The 'private' or 'independent' sector include both profit-seeking and not-for-profit providers. We do not distinguish between these in our analysis.

[8]Throughout the paper, the term 'private hospitals' refers only to healthcare providers that treat privately-funded patients. Where relevant, ISTCs - that initially treated only public patients - are referred to as a separate provider.

with increases in funding.[9] The introduction of ISTCs was initially intended to allow public hospitals to focus on emergency care and elective cases that required more complex treatment in order to reduce waiting times and address NHS capacity constraints (Naylor and Gregory, 2009). The first contracts for ISTCs were signed in 2003, and public patients were treated from 2005 onwards. A second round of ISTC contracts were then signed in 2006, with patients treated from 2007 onwards in these facilities. These new contracts eased restrictions on who could be treated by ISTCs going forward.

From 2006 onwards, pre-existing private hospitals were also allowed to enter the public elective market.[10] These providers could now compete with existing providers - including both public hospitals and ISTCs - to provide care to publicly funded patients at the same nationally set fixed price that was paid to public hospitals. This reflected a shift in NHS policy in the mid 2000s towards introducing consumer choice and competition between providers. The patient choice reforms of 2006 established a requirement for GPs to offer patients a choice of hospital when referring patients for almost all elective care.[11] New private sector entrants were therefore now intended to increase competition for NHS providers and to foster innovation among providers (Naylor and Gregory, 2009).

Unlike the original ISTCs, pre-existing private hospitals were allowed to treat publicly and privately funded patients alongside one another (Cooperation and Competition Panel, 2011). Over time, they overtook ISTCs to form the majority of private provision in this public market. By 2012/13, 95 of 119 (79.8%) private providers operating in the market were pre-existing private hospitals rather than ISTCs, treating 72.5% of all publicly funded patients treated by private providers, while also continuing to treat privately funded patients.

Location decisions also varied across the two provider types. ISTCs were originally intended to be located in areas where local hospitals were lacking capacity or struggling to meet waiting time targets, and were frequently located on NHS sites (Naylor and Gregory, 2009). In contrast, virtually all of the private hospitals

[9]The first waiting times target was introduced in April 2001, with a maximum wait of 18 months between the decision to admit and inpatient admission. The target was reduced by three months each year. In December 2008 a new referral to treatment (RTT) target was introduced, with a maximum wait of 18 weeks between GP referral and inpatient admission. See Propper et al. (2010) for discussion and evidence on the waiting time targets that were implemented from 2000 onwards.

[10]Orthopaedic providers - the focus of this paper - were allowed to enter the market in 2006, but entry for other specialties was limited until 2008.

[11]Patients were initially offered a choice of 4 or 5 hospitals in 2006. The limit on the number of hospitals was then removed in 2008. This replaced a system where patients could state preferences but GPs were under no obligation to actively offer their patients a choice. These reforms were motivated by both the belief that patients valued the choice over their care, and evidence that health care competition (when prices were fixed) could improve quality (Gaynor, 2006).

that entered the market from 2006 onwards were pre-existing private hospitals that now took the decision to treat publicly funded patients alongside their private patients. The location of these hospitals pre-date the announcement of the reforms. Barriers to entry into the private healthcare market in England are high, with relatively few openings and closures of private hospitals (Competition and Markets Authority, 2014) and so the scope to build additional facilities in response to the reform was, at least in the short run, very limited. Private hospital entry was therefore determined by management choices to use spare capacity to treat public patients, but this choice was restricted by the pre-reform location of the private hospitals.

The entry of private hospitals to the NHS elective market could have important implications for public sector capacity and quality of care. However, this policy reform has received relatively little attention in the existing literature, which has instead mainly focused upon the impact of waiting time targets (Propper et al., 2008, 2010) and patient choice reforms (Cooper et al., 2011; Gaynor et al., 2013, 2016). Noticeably, the existing literature that does examine the impacts of private provider entry in England has generally focused on the impact of the ISTCs rather than the subsequent entry of pre-existing private hospitals to the market (and which now account for a much larger market share than the ISTCs). Cooper et al. (2018) examined the impact of the introduction of the first round of ISTCs on the efficiency and casemix of existing public hospitals, and found that the opening of the ISTCs led to a costlier case-mix for nearby public hospitals but also improved their efficiency as measured by pre-surgery length of stay. These findings are consistent with other existing evidence that finds that patients treated by ISTCs were healthier and wealthier than those treated by public hospitals (Street et al., 2010; Bardsley and Dixon, 2011; Chard et al., 2011).[12] We build on this literature by examining the impacts of the wider introduction of private providers on the NHS elective market.

4.2.2 Potential impacts of provider entry

The entry of new providers to the public elective market has a number of potential implications for existing public hospitals and their patients. We briefly sketch out the possible consequences below before empirically testing for these effects.

[12]Some of this sorting of less complex patients towards private providers is a consequence of government regulations on which patients were eligible, as ISTCs do not have intensive care facilities, and reflects the early objective of the policy to allow public hospitals to focus on sicker patients. However, there remain concerns about the extent that ISTCs further adjusted their eligibility criteria to 'cherry-pick' the least costly patients (Audit Commission, 2008; Bardsley and Dixon, 2011; Cooper et al., 2018).

NHS policies towards private providers in the 2000s had two partly conflicting objectives. In the early 2000s, ISTCs were introduced in order to use the private sector to provide additional capacity, to deliver more activity and reduce waiting times. From the mid 2000s, there was a shift in NHS policies towards improving the quality of care.[13] Allowing existing private hospitals to enter the public market alongside public hospitals and ISTCs aimed to improve quality by increasing compeititon in the market for elective care. This was based on the theory that when prices are fixed, providers can only compete for patients on the basis of quality. Increasing competition will therefore tend to drive quality improvements (Gaynor, 2006; Gaynor and Town, 2012).

The extent to which provider entry improves quality does however depend on the amount of additional competition generated by the new providers. Specifically, quality improvements would be unlikely to occur if competition between new and incumbent providers was weak. This would be the case if publicly funded elective care was capacity constrained prior to the reform, as seems likely given long waiting lists and the prior focus of NHS policy to increase capacity quickly. In this case, the entry of new providers would increase the supply of publicly funded elective care: patients further down the waiting list could now receive treatment at an earlier stage, increasing the total number of publicly funded patients treated in a given year. Entry would therefore increase the overall size of the market but would not reduce the numbers of patients treated by existing public hospitals, and as a result, generate little competitive pressure on incumbents that would no longer fear losing patients (or their associated revenues). Incentives to improve care quality would be muted as a result.

In addition, if the entry of private hospitals did lead to an expansion in the overall size of the market, we might also expect a change in the composition of publicly funded patients that receive treatment. Given the limited capacity of the NHS to provide elective procedures, operations are rationed. This is in large part on the basis of need: the clinical decision to perform a hip replacement is largely determined by the level of pain experienced by the patient and the ability to treat the underlying diagnosis using alternative means (such as physiotherapy).[14] This means that when supply is constrained patients with less severe needs are unlikely to receive a hip replacement. As supply constraints are relaxed the level of rationing is also reduced, with patients with less severe needs (and who are most

[13]The evolution in NHS aims from expanding capacity to improving quality of care was summarised by then-Prime Minister Gordon Brown in 2008: "If the challenge 10 years ago was capacity, the challenge today is to drive improvements in the quality of care" (Department of Health, 2008).

[14]There are no fixed national guidelines for who should be referred for a hip replacement, with some Clinical Commissioning Groups (CCGs) publishing guidance and others not. See, for example, the guidance from Ipswich and East Suffolk CCG (2017).

able to pursue non-surgical treatment) receiving a hip replacement at an earlier point in time.[15] Expansions in supply should therefore lead to a reduction in the severity of the marginal (and if large enough, the average) patient.

Establishing which of the above scenarios took place is important in understanding whether the policy was successful in improving care for patients. Given the policy aims to improve outcomes through competition, showing how and why such reforms failed would provide important lessons for future competition policy. We now examine this empirically.

4.3 Data

4.3.1 Hospital records

Data on all publicly funded care comes from the inpatient Hospital Episode Statistics (HES), and covers the period from April 2002 to March 2013. In this paper we focus on publicly funded patients undergoing elective hip replacements.[16] This covers 615,281 patients over our 11 year period of interest, and includes both procedures conducted at public hospitals and publicly funded procedures conducted at private providers (including both ISTCs and private hospitals). The inpatient data contain detailed information about the patient and the care they received, including their age, sex, GP practice, local area, admission type (emergency or elective) and dates, up to 20 diagnoses, all procedures patients receive, and a hospital identifier.

The data are linked to the precise geographic coordinates for all NHS, ISTC and private hospital sites. Patient locations are given by the centroid of their Middle Super Output Area (MSOA). MSOAs are statistical areas, similar to census tracts, with no administrative jurisdiction. There are 6,781 MSOAs in England, with an average population of 7,800 people in 2012/13.[17]

The introduction of private providers to the public market may also have impacted the private-pay market for hip replacements. As a result, in Section 4.6.2 we study whether there is any evidence of substitution between the public and private-pay markets as a result of hospital entry. However, while HES captures

[15] Over the last five years the opposite has occurred, with relatively slow growth in NHS funding raising concerns about increased rationing of services (Edwards, Nigel and Crump, Helen and Dayan, Mark, 2015) and restrictions on surgery leading to only those in the most pain receiving treatment (Iacobucci, 2017).

[16] Hip replacements include those operations with Office of Population Censuses and Surveys (OPCS) Classification of Interventions and Procedures codes (4th Edition) beginning W37, W38, W39, W93, W94 and W95. Each operation code defines a different type of hip replacement. For a full list of OPCS codes see here: http://www.surginet.org.uk/informatics/opcs.php.

[17] We use 2001 MSOAs throughout. See for National Statistics (2012) for more details.

all activity that is either provided or funded by the NHS, the data do not contain information on privately financed procedures at private hospitals. We therefore augment HES with records from the National Joint Registry (NJR), an audit of all artificial joints that are used in the procedures. These data contain information on all hip and knee replacements in England regardless of the provider type and how the procedure is funded, and enable us to study overall volumes in both the public and private market. However, the data contain much less detailed information on the patient, and in particular where the patient lives. Whereas HES contains the MSOA of patient residence, the NJR only records the patient's postal district. These postal districts are larger, with 1,993 across England, and so contain more measurement error in distances between hospitals and patients.[18] The data quality prior to 2008/09 is also somewhat poor, with missing procedures and missing information on how procedures are funded. As a result we carry out our main analysis using the more complete HES data where possible.

4.3.2 Defining hospital markets and exposure to private hospital entry

Our identification of the impact of private hospital entry on the market for publicly funded hip replacements arises from a comparison across areas or hospital 'markets' with differential levels of exposure to private hospital entry. In our baseline results, we define geographical hospital markets by assigning all (potential) patients to their nearest public hospital, as measured by the straight line distance between the centriod of the patient's MSOA and the hospital.[19] This yields a 11-year panel of 130 hospital markets. We use this definition as patients typically receive secondary care from their nearest hospital, with 72% of hip replacement patients treated by their nearest hospital in 2002/03. The capacity of the nearest hospital will therefore play an important role in whether patients received a hip replacement, and the waiting time they would face.

In our baseline analysis, we define high exposure areas as markets which contained a private hospital treating public patients in any of its MSOAs in 2012/13.[20] This measure defines high and low exposure areas that are fixed over time, and

[18]These data allow us to match patients (of any hospital) to their nearest hospital as we set out below. The lack of fine geographic identifiers do however prevent an analysis of the NJR data at very small geographic areas.

[19]We use the location of the trust headquarters site in cases where multiple hospitals within the trust conduct elective hip replacements.

[20]This does not include ISTCs. We control for separately for the presence of ISTCs in all subsequent analysis. We include private hospitals conducting at least 20 hip replacements in a year to avoid confusing provider entry with small ad-hoc purchases of care from the private sector. Results are qualitatively unaffected by this restriction.

facilitates a difference-in-difference specification that we set out in Section 4.4.[21] A map of the location of public hosptials by exposure status and private hospitals providing publicly funded hip replacements in 2012/13 is given in Figure A1.

Markets can however be defined in a number of ways that could influence our results. We therefore test our results for robustness by using other market and exposure definitions. In all cases, our conclusions are substantially unaltered by the precise definitions used.

We first examine how our results change when we define a hospital market in a different way. In our baseline results, we define markets by centering non-overlapping catchement areas around public hospitals, based on straight-line distances between the areas in which potential patients live and their nearest hospital. This contrasts with previous papers studying hospital competition in literature which create overlapping markets (Propper et al., 2008; Courtemanche and Plotzke, 2010; Cooper et al., 2011, 2018). These papers use the distribution of pre-reform distances travelled by patients for treatment at each provider to draw a radius around each hospital, and define competition based on the number of rival providers within this radius. In Section 4.5.3, we adapt this approach to define overlapping markets centred on MSOAs. We use the distribution of distance travelled by patients living in each MSOA that underwent a publicly funded hip replacement between 2001 and 2004 to define catchment areas. These areas are then included in the 'high exposure' group if a private hospital treating public patients in 2012/13 was located within this catchment area, and in the low exposure group otherwise.

This approach has advantages and disadvantages relative to our baseline definition. Using our baseline approach, a private hospital entrant can by definition only treat one area. As a result, we could misallocate markets across high and low exposure groups in two ways. First, a private hospital located near two public hospitals will by definition only treat a single hospital market (the one in which it is physically located) when in practice both hospitals will be affected. As a result, the effect of hospital entry would be understated due to the attenuation of the treatment effect. Second, we may more generally confuse low and high exposure areas if patients are most often treated by another hospital that is not their closest hospital. In this case, our estimates could overstate the true impact of provider entry. In contrast, defining overlapping markets around an MSOA allow a single entrant to affect patients in multiple markets.

[21] It also provides a geography that can be studied relatively consistently across multiple datasets. This allows us to examine effects on privately financed volumes using the NJR data on a similar geography in Section 4.6.2.

However, this second approach also risks understating the impact of provider entry. The distances travelled by patients for treatment in a given area may change if a new hospital enters in the local area. In particular, patients with an existing public hospital nearby may have very small catchment areas using the pre-reform distances for treatment. This means that an MSOA close to a public hospital that has a private hospital entering the market only slightly further away may no longer be classified in the 'high exposure' group (despite its residents being heavy users of the new hospital), whereas an MSOA far from their local hospital will be regarded as 'high exposure' if a private hospital opens slightly nearer.

We also examine how our results change when allowing entry to vary across time in different markets. Fixing areas into high and low exposure groups over time will likely understate the impact of provider entry (for a given market definition). This is because private hospitals entered the market in some areas at a later stage relative to others, and therefore would not expect to be affected by these providers throughout the entire policy period (2006/07 onwards). This is demonstrated by Table 4.1. The first column shows the total number of private hospitals conducting publicly funded hip replacements in England in each financial year. Until 2006/07, no private hospitals were operating on patients. After this, private hospitals started to enter, with the number of providers expanding much more rapidly from 2008/09. This pattern is mirrored by the percentage of markets that contained private hospitals treating public patients in each year (as shown in the second column), which increased rapidly in the late 2000s, reaching 55.4% by 2012/13. We therefore repeat our analysis using time-varying measures of exposure based on the exact years in which private hospitals treat public patients within the market.

4.3.3 Descriptive evidence of the impact of private hospital entry

Figure 4.1 shows the annual number of admissions for publicly funded hip replacements in England between 2002/03 and 2012/13. The number of procedures increased by 67.6% during this period, from 40,592 in 2002/03 to 68,031 in 2012/13. The figure also distinguishes between providers, and shows that the initial increases in admissions were driven by procedures carried out by public hospitals and then by ISTCs. After 2008/09, there was rapid growth in the number of admissions carried out by private hospitals. In 2012/13, public hospitals remained the dominant provider of publicly funded procedures, but private hospitals and ISTCs now provided 14.7% and 5.6% of procedures respectively, from a base of 0% 10 years previously.

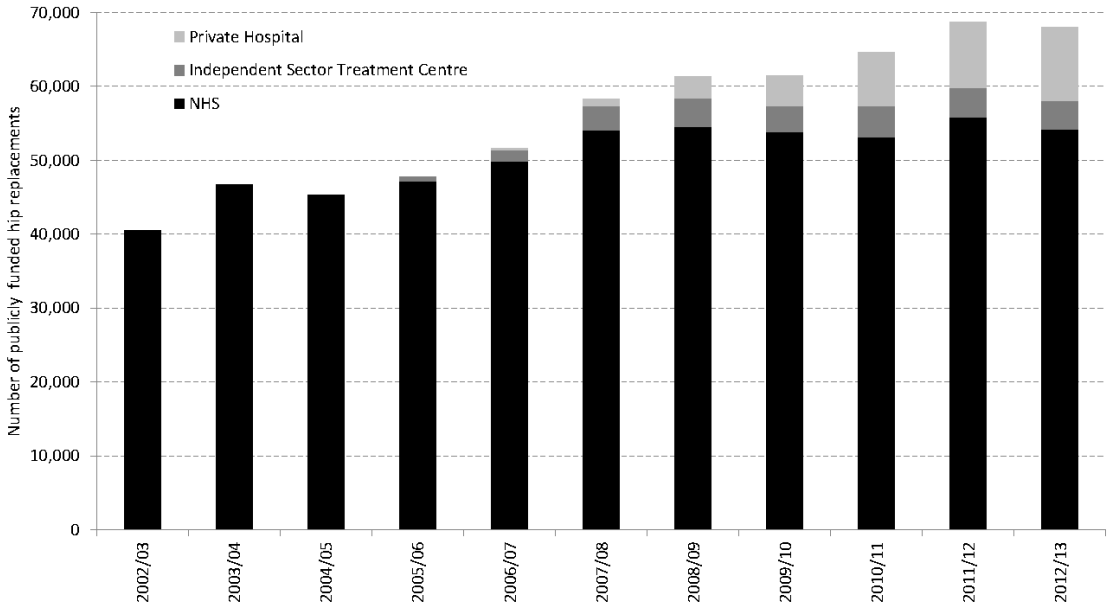
Table 4.1: *Mean hospital market exposure to private hospitals, 2002/03 - 2012/13*

Financial Year	Number of private hospitals treating public patients (1)	% of 'high exposure' markets (2)
2002/03	0	0.0%
2003/04	0	0.0%
2004/05	0	0.0%
2005/06	0	0.0%
2006/07	3	2.3%
2007/08	6	3.8%
2008/09	33	23.1%
2009/10	35	23.8%
2010/11	76	49.2%
2011/12	82	54.6%
2012/13	87	55.4%

Notes: (1) Column 1 shows the total number of private hospitals in England that conducted at least 20 publicly-funded hip replacements in a given financial year under the Any Qualified Provider scheme; (2) Column 2 shows the percentage of markets that include a private hospital treating at least 20 publicly-funded hip replacement patients within their geographic region; (3) Distances are calculated using straight line distance measures between MSOA centroid and the coordinates of the public/private hospital.

This suggests that private hospitals were, at least in part, responsible for increasing the volumes of publicly funded procedures over this period. Figure 4.2 provides further support of this hypothesis by showing the growth in mean hospital market admissions distinguishing between areas with low and high exposure to private hospitals in 2012/13. Panel A shows the growth in levels and Panel B shows indexed growth relative to 2006/07 (the first year of private hospital entry). Trends in growth appear very similar in low and high exposure areas in the pre-policy period. After 2006/07, volumes grew in high exposure areas at a much quicker rate, particularly following an expansion in private hospital entry in 2008/09. Between 2006/07 and 2012/13, admissions increased by 24.7% in areas where no private hospital treated public patients compared to 41.9% in areas with a private hospital active in the public market by the end of the period.

Figure 4.1: *The number of publicly funded hip replacements by provider type, 2002/03 to 2012/13*



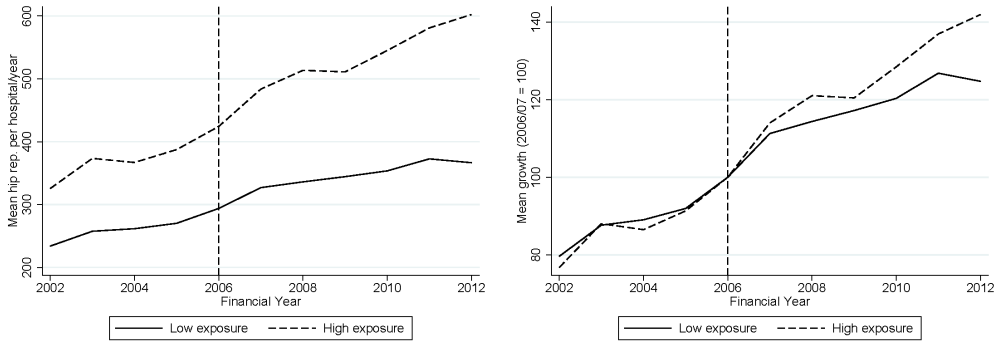
Notes: (1) Hip replacements include those operations with Office of Population Censuses and Surveys (OPCS) Classification of Interventions and Procedures codes (4th Edition) beginning W37, W38, W39, W93, W94 and W95. Each operation code defines a different type of hip replacement; (2) Independent Sector Treatment Centres are defined as providers operating under Independent Sector Treatment Centre wave 1 or 2 contracts (data provided by Monitor); (3) Private Hospitals are defined as all other providers with a site code beginning with "N".

Figure 4.3 shows the the mean number of publicly funded admissions in each financial year conducted separately by public hospitals and private hospitals across low and high exposure markets. Panel A shows mean hip replacements conducted by public hospitals, and reveals only marginally stronger growth in low exposure relative to high exposure markets. This suggests that private hospitals had only a small impact on the number of admissions at existing public hospitals. By contrast, panel B shows substantially stronger growth in admissions for hip replacements conducted by private hospitals in high exposure markets relative to low exposure markets. This suggests that private hospitals were responsible for much of the growth in hip replacements over time, and these increases were concentrated (but not exclusively) in areas where public hospitals had higher exposure to private hospitals.[22]

[22]While partly mechanical, this provides reassuring evidence that areas designated as 'high exposure' are indeed those affected by private entrants.

Figure 4.2: *Mean publicly funded hip replacements per hospital market, by private hospital exposure in 2012/13*

(a) *Levels* (b) *Growth (2006/07=100)*



Notes: (1) Volumes include all publicly funded hip replacements (as defined in Figure 1) regardless of whether they were conducted by public hospitals or private providers; (2) Patients are allocated to their nearest hospital regardless of where the surgery actually takes place; (3) High exposure areas are those with a private hospital treating public patients located within the market in 2012/13, low exposure areas are those without; (4) In Panel B growth figures are relative to 100 in 2006/07; (5) The vertical line (2006) denotes the year in which private hospitals first entered the market.

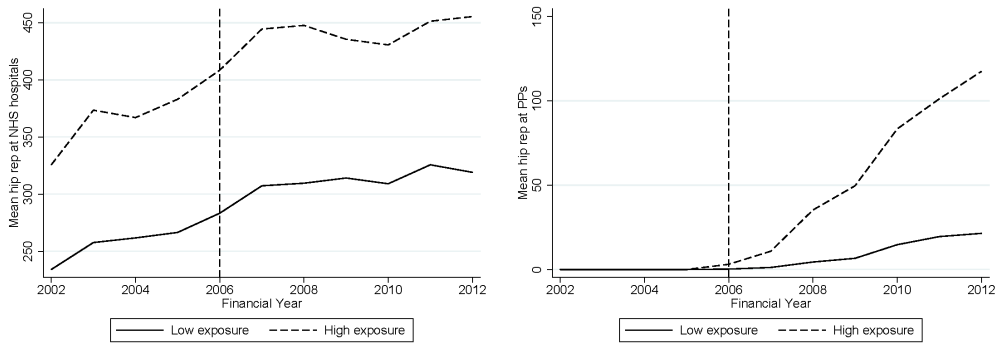
Panel A of Figure 4.4 shows a similar pattern for waiting times. It shows log median waiting times for each year of the period in low and high exposure areas. National waiting times fell considerably over this period of time in response to aggressive national waiting time standards. Importantly, we see parallel falls in low and high exposure areas prior to 2006/07. After the introduction of private providers, waiting times fell more rapidly in high exposure areas. This is consistent with the increase in capacity from these private hospitals contributing to falls in public waiting times. Panel B repeats this exercise for the log 30-day emergency readmission rate. While noisier than the other outcomes, trends prior to the reform are again similar across the high and low exposure areas and no obvious differences in overall patterns can be seen in the post-reform period.

Table 4.2 displays summary statistics for publicly funded hip replacement patients in 2002/03 and 2012/13, by provider type. Among patients treated at NHS hospitals, mean patient age has fallen slightly over the period (from 68.4 to 68.2 years old) while the percentage of patients that are male (40%) has remained unchanged. The mean number of comorbidities recorded has increased over time.[23] Length of stay and waiting times have fallen substantially, reflecting wider trends

[23]The increase in the number of diagnoses is the opposite of what we might expect if marginal patients are in better health. However, substantial increases in the number of secondary diagnoses recorded in HES over time by all (public and private) hospitals mean these changes are unlikely to be a genuine reflection of changes in patient health. Another option is to consider the Patient Reported Outcome Measures, which give detailed information about pre and post-operative health. However these data are only available from 2009 and are poorly recorded for

Figure 4.3: *Growth in mean market volume of publicly funded hip replacements (2006/07=100), by provider type and private hospital exposure in 2012/13*

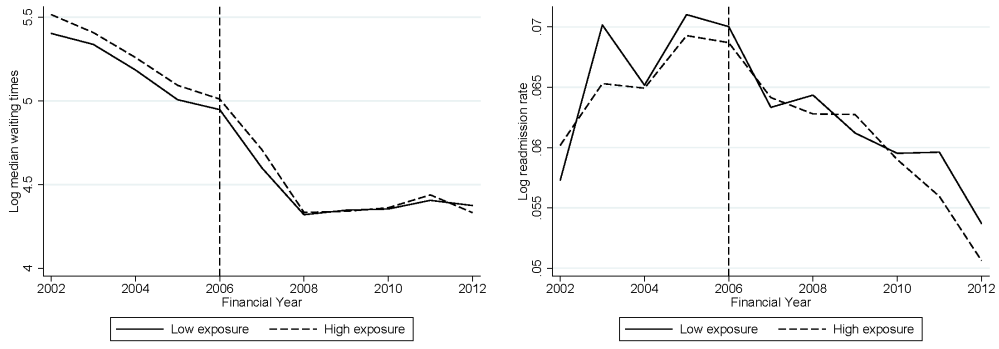
(a) *Public hospitals* (b) *Private hospitals*



Notes: (1) Growth figures relative to 100 in 2006/07; (2) Panel A shows growth in hip replacements conducted by public hospitals, Panel B shows growth in hip replacements conducted by private hospitals; (3) High exposure areas are those with a private hospital treating public patients located within the market in 2012/13, low exposure areas are those without; (4) The vertical line (2006) denotes the year in which private hospitals first entered the market.

Figure 4.4: *Log median waiting times and log 30-day emergency readmissions rates for publicly funded hip replacements, by private hospital exposure in 2012/13*

(a) *Waiting times* (b) *Emergency readmissions*



Notes: (1) In Panel A, waiting time measures the median number of days between the decision to admit a patient for a hip replacement and their admission date; (2) In Panel B, emergency readmissions measure the proportion of patients who experience an emergency inpatient readmission within 30 days of discharge after a publicly funded elective hip replacement; (3) High exposure areas are those with a private hospital treating public patients located within the market in 2012/13, low exposure areas are those without; (4) The vertical line (2006) denotes the year in which private hospitals first entered the market.

in the provision of NHS care.[24] Median length of stay fell from 9 days to 4 days over the period, and median waiting times fell from from 239 days in 2002/03 to 92 days in 2012/13. Outcomes also improved, with the 30-day emergency readmission rate falling from 6.1% to 5.6%, and the 30-day in-hospital mortality rate falling from 0.2% to 0.1%.

The table also highlights the differences in the patient mix treated by different providers by comparing the characteristics of all patients with those treated by private hospitals in 2012/13. Mean age is slightly lower at private hospitals, while patients have considerably fewer comorbidities: the mean number of comorbidities in 2012/13 was 1.8 for patients treated at private hospitals, compared to 3.1 at NHS hospitals. Length of stay is the same across providers, but waiting times are considerably lower at private hospitals. Median waiting times for private hospital patients were 63 days compared to an average of 92 days at NHS hospitals. Emergency readmission rates (to any public provider of care) are also lower at private hospitals (3.5%). These differences may reflect either differences in casemix or in the quality of the different providers.

4.4 Empirical Strategy

The descriptive evidence in the previous section suggests that the introduction of private hospitals to the public elective market had meaningful impacts on the number of admissions for publicly funded hip replacements and waiting times for these procedures. However, a variety of other changes may have taken place in different areas over time that may be conflated with the introduction of these providers. Understanding these impacts is important in understanding how private provider entry affected the overall market for publicly funded hip replacements, and its consequences for competition, public hospital performance and patient outcomes.

To estimate the impact of private hospital presence on the number of admissions, waiting times and outcomes for patients undergoing a publicly funded hip replacement we use a difference-in-difference framework, comparing changes in outcomes over time between areas with low and high exposure to these private hospitals by the end of the period. We estimate the following specification:

$$Y_{mt} = \beta_0 + \beta_1(E_m * post_t) + \beta_2 X_{mt} + \gamma_m + \lambda_t + \epsilon_{mt} \tag{4.1}$$

private hospitals for the first few years, and are therefore non-randomly missing. We therefore do not use them.

[24] Propper et al. (2008) and Propper et al. (2010) show that the majority of these falls in waiting times were due to the introduction of national waiting time targets.

Table 4.2: *Patient characteristics and outcomes in 2002/03 and 2012/13, by provider type*

	2002/03		2012/13			
	Public hospital		Public hospital		Private hospital	
	Mean	SD	Mean	SD	Mean	SD
Mean age	68.4	11.0	68.2	11.7	68.0	10.2
Male	0.40	0.49	0.40	0.49	0.40	0.49
Mean no. of comorbidities	1.20	1.65	3.10	1.86	1.83	0.6
No comorbidities (%)	48.73	49.98	16.45	37.10	30.51	46.05
2+ comorbidities (%)	31.43	46.43	68.67	46.38	51.18	49.99
Median length of stay (days)	9.0	7.5	4.0	5.3	4.0	1.5
Median wait time (days)	239.0	168.5	91.5	59.2	62.8	101.7
30-day readmission (%)	6.09	23.92	5.61	23.03	3.54	18.50
30-day mortality (%)	0.16	4.00	0.09	3.09	0.09	3.00
Observations	40,592		54,279		9,974	

Notes: (1) 'Public hospital' and 'Private hospital' refer to the type of provider (private hospital outcomes exclude hip replacements conducted by Independent Sector Treatment Centres); (2) Comorbidities outcomes include any secondary diagnoses recorded in the hospital records; (3) Wait time measures the time (in days) between the consultant's decision to admit for surgery and the admission date; (4) 30-day readmission rates measure the % of patients who had an emergency readmission in the 30 days after they were discharged following their hip replacement; (5) The 30-day mortality rate refers to in-hospital mortality only (including the initial hospital spell and any subsequent readmission).

Y_{mt} is the outcome for patients living in market m in year t, including the number of admissions for publicly funded hip replacements (including all procedures conducted at public hospitals, ISTCs and private hospitals), median waiting times and the 30-day emergency readmission rate. E_m is a binary variable that takes the value of one if a private hospital that treated publicly funded patients in 2012/13 was physically located in the market, and zero otherwise. This is interacted with $post_t$, a binary variable that takes the value of one in years when private hospitals could treat publicly funded patients (2006/07 onwards). Our coefficient of interest is β_1, the impact of private hospital exposure by the end of the period on the market outcome for publicly funded patients.[25]

We include market (γ_m) and time (λ_t) fixed effects to control for permanent differences across markets and national time trends respectively. To control for contemporaneous shocks or trends that affect the outcomes in the area, and which are correlated with private hospital exposure, we also include a rich set of area level time-varying characteristics in X_{mt}. In all specifications these include: the age composition of the local population; the number of admissions for fractured

[25]As noted above, this estimate will result in a conservative estimate of exposure as estimates may be attenuated if private hospitals only entered treated markets at the end of the period. We examine an alternative measure (exploiting the actual year of private hospital entry) along with other market definitions to test how this affects our conclusions in Section 4.5.3.

neck of femur and acute coronary syndrome to capture population need[26]; and the number of house sales and median house price to account for changes in economic conditions.[27] When examining the impact on waiting times and patient outcomes we also include direct controls for the characteristics of hip replacement patients as the introduction of new providers may have changed the attributes of patients undergoing treatment. This includes mean age, the proportion of patients who are male, and the mean Charlson score. The error term ε_{mt} is robust to heteroskedasticity and clustered at the market level.

X_{mt} also captures ISTC presence in the market. We measure this by including an analogous measure to our private hospital exposure: a binary variable that takes the value of one if an ISTC treated patients in 2012/13, and zero otherwise, interacted with a dummy variable that takes the value of one in years when ISTCs could treat publicly funded patients (2005/06 onwards, the year before private hospitals were allowed to enter). ISTCs may impact our outcomes of interest (for example, if they increase admissions) and their location may also be (negatively) correlated with private hospital entry if ISTCs were launched in areas where private hospitals were unlikely to enter the market. Controlling for their presence could therefore be important to avoid bias in our estimates. Given the aims of the policy, the coefficient on ISTC location - with ISTCs intended to be set up in areas with high waiting times - is likely to be endogenous. We therefore report these coefficients where appropriate to provide comparison with our estimated impacts of private hospital entry but do not claim these estimates capture causal impacts. We discuss these results in more detail in Appendix B.

The identifying assumption is that, conditional on our controls, exposure to private hospitals is otherwise uncorrelated with unobservable determinants of the outcomes. One threat to this assumption is any period-specific shock that differentially affected low and high private hospital exposure areas during this period. In particular, the decision of private hospitals to enter specific markets is likely to be related to other factors in the local area that may also determine the outcomes that we are interested in. This includes the decisions made by local NHS policy-

[26]Fractured neck of femur and acute coronary syndrome are emergency conditions that typically affect older people, although the average age of patients is slightly higher than for elective hip replacements. As admissions are nearly always an emergency, admission rates should reflect patient need and be uncorrelated with the introduction of private providers, which only treat elective cases. Fractured neck of femur typically results in an emergency hip replacement, which uses the same surgeons and resources as elective hip replacements. Higher rates of fractured neck of femur admissions could therefore indicate both higher need in the population, as conditions such as osteoporosis increase the need for both elective and emergency hip replacements, and greater demand on local orthopedic units from emergency patients, which could result in longer waiting times for elective patients.

[27]Population need characteristics are calculated using HES. Information on house sales and prices comes from the Office for National Statistics (http://www.ons.gov.uk/ons/rel/regional-analysis/house-price-statistics-for-small-areas/1995-2013/index.html)

makers and providers when choosing whether and how much to expand local NHS capacity. For example, private providers may have chosen to enter markets where public investment in building NHS capacity was lower as they could profit from the larger (unmet) demand for publicly funded care in these areas. If this was the case, then our estimates on the impacts of private hospital presence on our outcomes would be downward biased as we would mistakenly attribute the impact of NHS funding decisions to private provider entry. Similarly, if NHS policymakers chose to invest less in areas in which private hospitals were known to be willing to operate, then estimates would also understate these the true impacts. In both cases, we would expect our analysis to understate the impact of private hospitals on the number of admissions and waiting times reductions.[28]

We address this concern by implementing an instrumental variables strategy, using the location of pre-existing hospital sites prior to the reform to instrument for private hospital presence in the public market by the end of the period.[29] We construct this instrument in the following way. First, we create a dummy variable equal to one if a private hospital existed in the area in 2004 (prior to the policy period), and zero otherwise. Second, we interact the private hospital dummy with the $post_t$ dummy variable that takes the value of one in all years in which private hospitals were allowed to treat public elective orthopaedic patients (2006/07 onwards) and zero otherwise. This yields a time-varying variable for each financial year that indicates whether a pre-existing hospital site was located in the market and was allowed to treat public patients. We then instrument our private hospital exposure measure, $E_m * post_t$, with this variable.

For this instrument to be valid, pre-existing hospital sites should be correlated with the location of private hospital entry during the reform period (the relevance condition), and otherwise be unrelated to our outcomes of interest (the exclusion condition). Our instrument should fulfil both criteria. Private hospitals wishing to enter the public market require medical facilities in order to treat public patients. Almost all of these hospitals were built prior to the reforms, with public patients now treated alongside existing private patients. Historical presence of a private hospital should therefore be a very strong predictor of private hospital presence in the public market, and so fulfil the first criterion.

We use the location of private hospital sites in 2004 - before private hospitals were allowed to enter the NHS market for elective procedures. Our instrument should therefore fulfil the exclusion restriction, as we use only private hospitals that already existed prior to the decision to allow private entry to the public market,

[28]The potential bias for readmissions is ambiguous.

[29]We use information on the historic location of private hospitals from the Care Quality Commission (CQC), obtained through private correspondence in October 2012.

and not any that could have opened in response to the reform. Furthermore, there are high barriers to entry and expansion into the private healthcare market in the UK (Competition and Markets Authority, 2014), and the stocks and locations of hospitals are relatively fixed in the short-term. This makes it unlikely that private hospitals would open specifically in areas where NHS volumes or waiting times were changing in a specific way immediately prior to the reform.

However, it remains possible that pre-existing private hospitals were located in areas with different trends in our outcomes of interest. For example, private hospitals are typically located in less deprived areas (reflecting higher historical demand for private care in these areas). These areas may also differ along other dimensions, such as age structure, which could result in differential trends in demand for hip replacements or the composition of patients across areas with and without private hospitals. More generally, any period-specific shocks that could differently affect areas with and without pre-existing private hospitals could potentially bias our results, with our estimates falsely attributing impacts to provider entry. While our area fixed effects and time-varying controls will capture permanent and changing (observed) differences across areas, differential unobserved trends in outcomes could bias our estimates. We therefore examine this issue in detail in Section 4.5.3, and carry out a range of robustness tests to examine whether they are likely to affect our conclusions in a meaningful way. We also test for the existence of non-parallel trends in our outcomes between areas which did and did not contain private hospitals in 2004. In both cases, we show that our results are robust to these concerns.

4.5 Results

4.5.1 Impacts on market outcomes

Columns 1-3 of Table 4.3 show the results of the analysis when using the number of publicly funded admissions for hip replacements as our outcome of interest. Column 1 reports the estimates from the fixed effects specification set out in equation 4.2. It shows a positive and statistically significant relationship between admissions and exposure to a private hospital by the end of the period: the presence of at least one private hospital treating public patients in the market by 2012/13 is associated with an annual increase of 26.6 procedures. This is equivalent to 9.3% of the mean number of admissions (285) in 2002/03.

As outlined above, we would expect this coefficient to be an underestimate of the true impact of private provider presence on publicly funded admissions if private hospitals chose to enter areas where publicly funded admissions would

have increased more slowly in the absence of private provider entry. We therefore instrument the presence of private providers in 2012/13 with the location of private hospital sites in 2004. These sites were established prior to the implementation of the policy and should be independent of other policy decisions made during the reform period. They are also strong predictors of private hospital entry by 2012/13: the first stage F-stat is 87.2.[30] The strength of the instrument is not surprising given that private hospitals that entered the market were principally existing private medical facilities, with very limited scope for opening additional facilities in the short run.

Table 4.3: *Estimated impacts of private hospital exposure on volumes of publicly funded hip replacements, log median waiting times and log emergency readmissions*

	Volumes			ln(median waiting time)			ln(readmissions)		
	All	All	NHS only	All	All	NHS only	All	All	NHS only
	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pub. funded priv. hosp.	26.60***	33.75***	-0.76	-0.055	-0.111*	-0.072	0.002	0.004	0.005
	(8.96)	(13.00)	(14.85)	(0.036)	(0.066)	(0.061)	(0.002)	(0.004)	(0.004)
First stage F-stat	-	87.2	87.2	-	87.2	87.2	-	87.2	87.2
Observations	1,430	1,430	1,430	1,430	1,430	1,430	1,430	1,430	1,430
R-Squared	0.751	0.750	0.500	0.863	0.862	0.868	0.140	0.138	0.102

Notes: (1) 'Pub. funded priv. hosp.' is a dummy variable that takes the value of one in the post-reform years (2006/07) for all hospital markets where a private hospital located in the market treated public patients in 2012/13; (2) Columns 3, 6 and 9 (NHS only) use outcomes for only patients treated by NHS (public) hospitals; (3) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, an independent sector treatment centre dummy (ISTC) (equal to one if an ISTC treated public patients in 2012/13) interacted with a dummy variable that takes the value of one from 2005/06 onwards (the first year of ISTC entry), and a full set of year and hospital market fixed effects; (4) There are 130 hospitals; (5) The coefficient on private hospital presence in 2004 interacted with the post-2006/07 dummy in the first stage is 0.63 (standard error is 0.067); (6) All specifications clustered at the hospital market level, *** p< 0.01, ** p<0.05, * p<0.1.

Column 2 shows the results from the two stage least squares regression. The estimates indicate that private hospital entry by 2012/13 increased the annual number of publicly funded admissions for hip replacements by 33.8, or 11.7% of the mean number of admissions in 2002/03. This estimate is statistically significant at the 1% level and is slightly higher than the OLS estimate (although not statistically significantly different from this coefficient). This suggests that private hospitals entered markets which would have had slightly slower growth in publicly funded admissions in the absence of the reform.

Column 3 repeats this analysis using admissions for hip replacements at public hospitals only. The coefficient is negative, but is small in magnitude and not statistically significantly different from zero. This is in direct contrast to the

[30]We report the Kleibergen-Paap Wald rk F statistic in all cases.

overall increases in publicly funded admissions. Private hospitals are therefore likely to be treating new patients rather than simply taking patient numbers from incumbent hospitals. This result is very similar to the findings of Courtemanche and Plotzke (2010), where entry of ambulatory surgical centres in the US resulted only in very small reductions in volume in local incumbent hospitals, and which were nowhere close to offsetting the activity undertaken by the new centres. This is important as it suggests that private hospital entry did not result in existing public hospitals losing patients and their associated revenues. As a result, competitive effects (and the incentives to improve quality) were likely to be muted.

In columns 4-6 we repeat the analysis for waiting times, using the log of median waiting times at the hospital market level as the outcome. We might expect waiting times to fall as private hospitals enter the market, either due to increases in capacity reducing waiting lists for publicly funded patients or through public hospitals trying to lower waiting times to compete with private hospitals for patients. Column 4 indicates that the presence of a private hospital by 2012/13 was associated with a 6.0% reduction in median waiting times but this relationship is not statistically significant.

As before, if private hospitals chose to enter in areas where waiting times were not expected to fall as quickly in the absence of the reform then the OLS estimates would underestimate the impact of private hospital entry on waiting times. We therefore repeat our IV analysis. Column 5 shows the results for median waiting times for all publicly funded patients admitted for a hip replacement. The presence of a private hospital in the market by 2012/13 is now associated with a 11.1% reduction in median waiting times, and is statistically significant at the 10% level.[31] This is equivalent to a reduction of 27 days in 2002/03.

Column 6 repeats this analysis only for patients treated at a public hospital. The sign of the coefficient is again negative. However, it is smaller in magnitude than the reduction in waiting times for all publicly funded patients. It is also no longer statistically significantly different from zero. Any gains for patients in terms of reduced waiting times therefore appear to accrue to patients treated by private hospitals.

In Appendix Table B1, we also report the coefficients on ISTC presence from the same regressions described above. As for private hospitals, the OLS estimates show a negative association between ISTC presence and waiting times, but the magnitude of this coefficient is much larger than the comparable estimate for private hospitals, and is statistically significant at the 1% level. However, these results should not be treated as causal: unlike private hospital sites, ISTCs were specifically built in areas with long NHS waiting times. During this period, areas

[31] The interpretation of the estimated coefficient is $\triangle y = 100(e^{\beta 1} - 1)$.

with longer waiting times would also been likely to receive additional funding or other interventions in order to reduce waiting times. The results therefore indicate that areas where ISTCs were established did successfully reduce their waiting times faster than areas without ISTCs, but we cannot distinguish whether this is due to the introduction of an ISTC or due to other factors.[32]

Columns 7-9 of Table 4.3 consider the impacts on readmission rates. The policy aimed to improve quality of care by stimulating competition on quality between existing public hospitals and newly entered private hospitals. Previous evidence from the NHS suggests that increased competition among public hospitals led to quality improvements (Cooper et al., 2011; Gaynor et al., 2013, 2016). Cooper et al. (2018) find that the entry of ISTCs led to improvements in efficiency as measured by falls in pre-operative length of stay at nearby public hospitals. However, a number of commentators raised concerns that private providers could perform lower quality work than public hospitals and also reduce staff availability (of Surgeons of England, 2006; Pollock and Godden, 2008).[33] As a result, the entry of private hospitals could have plausibly had either a negative or positive effect on care quality. We therefore repeat the analysis for the log of 30-day emergency readmissions following a hip replacement to examine whether private hospital presence had any impact on patient outcomes. Columns 7 and 8 show the OLS and IV estimates for all publicly funded patients respectively. Column 9 repeats the IV analysis only for patients treated in public hospitals. In both cases the coefficient is not statistically significantly different from zero. We also repeated this analysis using 30 day in-hospital mortality and find no significant impacts.[34] This suggests that the introduction of private hospitals did not lead to either increases or decreases in quality on these measures. This is consistent with results for volumes and waiting times, where there is little evidence that public hospitals lost patient volumes as a result of private hospital entry.

Taken together, these results show that markets with higher exposure to private hospitals experienced stronger growth in admissions for publicly funded hip replacements than markets with lower exposure and faster reductions in waiting times. However, there was no accompanying impact on emergency readmission rates. The direction of the change in results between the OLS and IV estimates suggest that private providers entered markets that would have experienced smaller increases in capacity in the absence of the policy, although the difference

[32]This is consistent with Cooper et al. (2018), who in an appendix note that waiting times fell more quickly in areas with ISTCs but that these results cannot be interpreted as causal impacts of the ISTC reform.

[33]Most of this criticism was due to the early experience of patients treated by ISTCs rather than pre-existing private hospitals but the same concerns exist for both sets of providers

[34]Results not shown but available upon request.

in the size of the estimates is not particularly large. As a result, the OLS estimates appear to slightly underestimate the impact on admissions and waiting times. For public hospitals, results are consistent with private providers exerting very limited competitive pressure on the public incumbents. There are no statistically significant impacts on volumes at public hospitals and consequently therefore no changes in revenue. Nor is there any evidence of quality improvements or attempts to match waiting times at private hospitals, which might indicate that public hospitals had reacted to protect their existing volumes. If public hospitals did adjust care quality in response to the entry of private providers, they did so in ways that would be hard to observe to either researchers or patients.

4.5.2 Impacts on patient composition

The estimates above indicate that the entry of private hospitals to the public market did not reduce volumes at incumbent public hospitals but instead led to increases in the total number of publicly funded hip replacements. This is consistent with the hypothesis that provider entry relaxed capacity constraints in areas where supply previously failed to fully meet demand for publicly funded hip replacements.

This raises the question of who receives these additional hip replacements. Ideally we would like to identify who the 'new' patients are. This is difficult as we do not observe the identity of the marginal patient. However, we can extend our analysis in three ways to provide evidence on who is likely to be most affected by the reform. First, in this section, we examine how average patient characteristics have changed as a result of provider entry. Second, in Section 4.6.1 we examine how the estimated effect of provider entry varied across the pre-reform level of deprivation in the local area. Finally, in Section 4.6.2 we use data from the private sector to examine whether patients are substituting from privately financing their own operations to having a publicly funded procedure as a result of the reform.

We start by testing whether the entry of private hospitals affects observed health characteristics of patients. As noted in Section 4.2.2, hip replacements are rationed at least in part on the basis of need, with tighter rationing leading to longer waits for patients in less severe pain or who are better able to manage their condition with non-surgical treatment (Iacobucci, 2017).[35] As capacity expands, one might expect GPs to refer patients and specialists to operate upon patients with lower levels of need, with alternative treatments used less often or for shorter

[35]This rationing may affect either whether the GP refers the patient to see a specialist, or whether the specialist decides to operate.

periods of time.[36] Increases in supply should therefore lead to marginal patients - and if the supply increases are large enough, average patients - being healthier along these dimensions.

While ideally we would like to observe each patient's need specifically for a hip replacement through the measurement of pain levels or suitability for other treatments, such measures are unavailable in our data. We instead use measures of more general health contained in the diagnoses recorded in the hospital data to examine whether the observed severity of patients change as a result of hospital entry. Specifically, we examine three measures based on the number of comorbidities at the time of the operation: a count of the number of comorbidities; the proportion of patients with no comorbidities; and, the proportion of patients with two or more comorbidities. These characteristics may have a direct impact on hip pain and mobility, but may also affect the suitability of alternative treatments such as physiotheraphy, exercise and pain medications.[37]

Table 4.4 shows the estimated effects of provider entry on the observed mean severity of hip replacement patients. In column 1, the outcome is the mean number of comorbidties recorded for all publicly-funded hip replacement patients. Our estimates indicate that the mean number of comorbidities fell in markets with provider entry. The presence of a private hospital by the end of the period is associated with a decrease of 0.36 secondary diagnoses. This is equivalent to 18% of the mean number of comorbidities (2.0) across all publicly funded hip replacement patients.

Column 2 and 3 repeat the analysis using the mean shares of patients with no comorbidities, and two or more comorbidities, respectively. The estimates for both are consistent with the overall pattern in the number of comorbidities. The presence of a private provider in the market increases the numbers of patients with no comorbidities by 8.5%, and reduces those with two or more comorbidities by 8.9%.

Columns 4-6 repeat the specifications in columns 1-3 but restrict the sample to only patients treated by public hospitals. The aim is to assess whether the reduction in severity is restricted to patients treated by private hospitals, or whether this pattern holds across the market. In all cases, private hospital entry resulted in a reduction in the severity of patients treated by public hospitals. The coefficients are slightly smaller than in columns 1-3, but the differences are not statistically significant. This suggests that private hospital entry reduced the average severity

[36]For example, Ipswich and East Suffolk CCG (2017) provides the threshold policy for hip replacements in one area of England, which includes periods receiving alternative treatments including pain medications and physiotherapy.

[37]In Section 4.6.2 we repeat this analysis using an alternative measure of patient severity using an alternative dataset, and find results consistent with those presented in this section.

Table 4.4: *Two stage least squares estimates of the impact of private hospital exposure on average hip replacement severity*

	Comorbidities (All patients)			Comorbidities (NHS only)		
	Mean number	% none	% 2+	Mean number	% none	% 2+
	(1)	(2)	(3)	(4)	(5)	(6)
Pub. funded priv. hosp.	-0.358**	0.085**	-0.089**	-0.302**	0.077**	-0.082**
	(0.152)	(0.038)	(0.037)	(0.151)	(0.038)	(0.036)
Observations	1,430	1,430	1,430	1,430	1,430	1,430
R-squared	0.685	0.600	0.665	0.711	0.636	0.692

Notes: (1) 'Pub. funded priv. hosp.' is a dummy variable that takes the value of one in the post-reform years (2006/07) for all hospital markets where a private hospital located in the market treated public patients in 2012/13; (2) Mean number of comorbidities records the mean number of secondary diagnoses; % of patients with zero and 2+ comorbidities record the percentage of patients with this number of comorbidities; (3) Columns 4-6 (NHS only) use outcomes for only patients treated by NHS (public) hospitals; (4) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, an independent sector treatment centre dummy (ISTC) (equal to one if an ISTC treated public patients in 2012/13) interacted with a dummy variable that takes the value of one from 2005/06 onwards (the first year of ISTC entry), and a full set of year and hospital market fixed effects; (5) There are 130 hospitals; (6) First stage F-stat is 87.2; (7) All specifications clustered at the hospital market level, *** p< 0.01, ** p<0.05, * p<0.1.

of patients receiving hip replacements across the whole of the market. This is consistent with the hypothesis that an expansion in the supply for elective hip replacements enabled less severe patients to receive a hip replacement when in the absence of the reform they would not have received this treatment.

4.5.3 Robustness checks

There may still be a number of remaining threats to identification. We now examine these threats in detail and set out a range of robustness checks to test our results. We first analyse how our estimates vary when using alternative definitions of hospital markets and exposure to private hospitals, before examining a range of other remaining threats.

Alternative treatment definitions

We first examine how robust our results are to defining markets in an alternative way. As noted in Section 4.3.2, our baseline market definition assigned small areas to their nearest public hospital, and defined exposure based on these non-overlapping areas. An alternative approach would be to create overlapping markets using small geographic areas, with exposure based on whether a private hospital entered the market within a radius determined by the distances that patients in the area typically travel for treatment.

We therefore repeat our analysis using an alternative market definition, with the MSOA of patient residence now as the unit of analysis. We calculate the distribution of distance travelled by patients living in each MSOA who underwent a hip replacement between 2001 and 2004, and use the 25th, 50th and 75th percentiles of these distributions to define catchment areas around each MSOA.[38] We then assign MSOAs to the 'high exposure' group if a private hospital treated public patients in 2012/13 within these radii. This yields three measures of exposure based on each of the percentiles of interest.

Appendix Table C1 shows the results when using this definition. The estimates in columns 1 - 4 again show that private hospital presence in the market by the end of the period is associated with an increase in the volume of publicly funded hip replacements, no reductions in volumes at public hospitals, a small reduction in waiting times and no reduction in emergency readmissions within 30 days of a hip replacement. However, the magnitudes are consistently smaller: for example, when the market is defined based on the 75th percentile catchment, private hospital entry is estimated to have increased volumes by 0.3 hips per year. Compared to a mean of 5.9 hip replacements in 2002, this is equivalent to a 5% increase. This is around half the size of our initial estimates. Estimates are similar (and differences are not statistically significant) when we use the 25th and 50th percent catchment areas instead.

Columns 5-7 show that entry is again associated with a reduction in the average severity of patients. Hospital presence at the end of the period is associated with a reduction in the mean number of comorbidities and patients with two or more comorbidities, and an increase in the number of patients with no comorbidities. The magnitude of the effects are also smaller, reflecting the reduced magnitude of the estimates on volumes.

The differences in the estimated magnitudes can be explained by a number of MSOAs located very close to public hospitals being reclassified from 'high exposure' to 'low exposure' areas under the alternative definition. This occurs because, using pre-reform distances, MSOAs located very close to existing public hospitals typically have very small catchment areas. This means that even if a private hospital enters the market only slightly further away, the MSOA will not be exposed under the alternative measure. This is despite having a high number of patients being treated by private hospitals. This suggests that using pre-reform distances may not fully reflect the new choice set faced by patients, who are willing to travel slightly further for treatment at a provider that previously was not an option. For this reason, we prefer our initial definition of treatment. However, we are

[38]This is in keeping with previous work that has defined catchments based on similar percentage cutoffs (Gresenz et al., 2006; Courtemanche and Plotzke, 2010; Cooper et al., 2011)

reassured that our main conclusions that hospital entry led to an expansion in the market, no changes in observed quality, and a reduction in average patient severity, are unchanged by the use of an alternative definition.

We also examine how our results change when we exploit variation in the timing of private hospital entry in different areas. Our baseline approach uses fixed treatment and control groups for simplicity. However, this may understate the impacts of provider entry if hospitals only enter very late in the period. To examine this more closely, we estimate an augmented version of equation 1, replacing our measure of private hospital presence at the end of the period with an indicator that takes the value of one if a private hospital treated public patients in that year, and zero otherwise.[39]

The results are in given in the Appendix Table C2. Again, the estimates are consistent with our original conclusions, with increased volumes, reduced waiting times and no change in emergency readmissions. As expected, the coefficients are larger in magnitude: private hospital entry led to an increase of 68.5 hip replacements per year, twice as large as the baseline results in Table 3. The estimated reduction in waiting times is also more than twice as large as the estimated effect on waiting times estimated in Table 4, but again only statistically significant at the 10% level. There are no statistically significant effects on readmissions. Average patient severity falls, and the absolute magnitudes are again twice as large as the baseline estimates (as shown in Table 5).

As a final check of sensitivity to alternative exposure measures, we explore whether our estimated effects vary with the size of the private hospital that enters the market. Some private hospitals may only provide a small number of hip replacements to publicly funded patients and therefore may not have been available to most patients. To do this, we define large private hospitals on the basis of the number of publicly funded knee replacements they carried out in 2012/13. We do so to avoid the dependent variable (volumes) also determining the treatment variable (large private hospital defined by volume). We use knee replacements, as the procedure requires almost identical facilities and staffing. We define a large hospital as one that conducted more than 140 publicly funded knee replacements in 2012/13.[40]

Table C3 in Appendix C presents the results. The estimated coefficients indicate that the impacts on volumes are greater when a large hospital entered the market: the entry of a large hospital is associated with 59 additional hip replacements. This is much larger than our baseline estimates in Table 3, which show

[39]See Appendix C for details of the empirical specification.

[40]This is the median volume of publicly funded knee replacements among private hospitals treating knee replacement patients in this year.

entry resulted in 34 additional hip replacements. Similar results are found for the estimated impacts on waiting times. Again, we find no evidence of statistically significant changes in readmission rates. The reductions in mean severity of patients are again larger if a patient is exposed to the entry of a large private hospital, relative to our baseline results in Table 4.4 that included all private hospitals. This is consistent with a larger expansion in the market resulting in a greater reduction in patient severity.

Taken together, the results in this section demonstrate that while alternative definitions of treatment affect the magnitudes of the effects that we estimate, the pattern of results remains unchanged: the entry of private hospitals resulted in an expansion in the size of the publicly funded hip replacements, a corresponding reduction in patient severity, and no impact on emergency readmissions or public provider volumes.

Other robustness checks

Our identifying assumption that exposure to private hospitals is uncorrelated with the unobservable determinants of the outcomes in our error term could be violated by the existence of non-parallel trends in outcomes across areas with and without pre-existing private hospitals during the period prior to the reform. A visual examination of Appendix Figures A2-A4 suggests there were no obvious differences in the pre-reform period in any of our outcomes of interest. Here we examine these trends more formally, by regressing the outcomes on our time-varying controls, market and time fixed effects, and a set of interactions between private hospital presence in 2004 and year dummy variables (excluding 2006, the first year of the reform).

Table A1 in Appendix A shows the results. In all cases, there is no evidence of any pre-trends, with no statistically significant coefficients on the interaction terms prior to 2006. However, there are statistically significant differences in the period after the reform. The impacts on the number of admissions increase over time, with particularly large growth in the effect on volumes in areas with private hospitals in the last 2 years (the period when many private hospitals had entered the public market). The magnitude of waiting times reductions increased up to 2009/10 before plateauing in the final years. There are no statistically significant impacts on volumes at public hospitals or readmissions in any year, and all coefficients are very small in magnitude.

A second threat to identification is any period-specific shocks that differentially affected areas with and without pre-existing private hospitals over the period of interest. These shocks could take the form of different trends in demands in areas

with and without private hospitals, or the wider impacts of the choice reforms that took place at the same time as private providers were allowed to enter the public market.

In particular, our IV estimates would be biased if private hospitals were located in areas with differential trends in our outcomes which are unrelated to the entry of private hospitals into the public elective market. As noted in Section 4.4, and explored in more detail in Section 4.6.1, private hospitals are more likely to be located in wealthier areas. These areas are also likely to be different in other ways: for example wealthier areas also tend to have an older population on average. While permanent differences in these areas are captured by market fixed effects, health outcomes in areas with private hospitals may be changing in ways that are different to areas without private hospitals. For example, if older people are more likely to require a hip replacement over time, and areas with private hospitals have a larger older population, then we would expect volumes of hip replacements to increase in these areas even absent the reform. This would lead to an upwards bias in our results. Conversely, if disparity in health outcomes between more and less affluent areas over time led to rising demand for hip replacements in less affluent areas (which are more likely to be areas with private hospitals prior to the reform), then we might understate the impact of provider entry. As a result, the pre-reform distribution of private hospitals may be related to differential trends in demand across areas during the reform period and so our estimates could potentially either understate or overstate the true impact of private hospital entry.

To address these concerns we carry out two robustness checks. First, we repeat the IV analysis including a full interaction between our control variables and time-dummies. This controls for time trends in a variety of elements of demand for hip replacements in the local area. The inclusion of these trends should control for any changes in the relationship between the observed characteristics of the areas and our outcomes of interest. Columns 1-4 in Table 4.5 shows the results of this exercise. Our main results are qualitatively unchanged by the inclusion of these time trends.

Second, we can also consider whether there is evidence of differences in demand trends across areas with and without private hospitals by examining whether our outcomes vary across areas with pre-existing private hospitals that didn't enter the public market, and areas where there were no private hospitals in the first place. To do this, we estimate an augmented version of equation 4.2, replacing our exposure measure (E_m) with a variable that sorts markets into three categories: (i) areas without a private hospital in 2004, (ii) areas with a private hospital in

Table 4.5: *Estimates of the impact of private hospital exposure on outcomes under alternative specifications*

	Volumes	Volumes	log (med wait)	log (readmit)	Volumes	Volumes	log (med wait)	log (readmit)
	All	NHS only	All	All	All	NHS only	All	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exposure								
Pub. funded priv. hosp.	26.79*	-3.11	-0.160*	0.003	31.58**	-1.60	-0.106	0.005
	(14.48)	(17.19)	(0.084)	(0.005)	(14.47)	(15.00)	(0.068)	(0.004)
Pre-reform HHI								
HHI * Post					31.96	12.42	-0.077	-0.022***
					(22.72)	(27.52)	(0.119)	(0.007)
Patient controls	No	No	Yes	Yes	No	No	Yes	Yes
Controls x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,430	1,430	1,430	1,430	1,430	1,430	1,430	1,430
R-Squared	0.785	0.567	0.875	0.244	0.751	0.500	0.862	0.147

Notes: (1) 'Pub. funded priv. hosp.' is a dummy variable that takes the value of one in the post-reform years (2006/07) for all hospital markets where a private hospital located in the market treated public patients in 2012/13; (2) Outcomes in columns 2 and 6 refer to patients treated by NHS (public) hospitals only; (3) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, an independent sector treatment centre dummy (ISTC) (equal to one if an ISTC treated public patients in 2012/13) interacted with a dummy variable that takes the value of one from 2005/06 onwards (the first year of ISTC entry), and a full set of year and hospital market fixed effects; (4) Patient controls include the mean age, sex and Charlson Comorbidity Index score of patients undergoing an elective hip replacement; (5) There are 130 hospitals; (6) Columns 1 - 4 include Strategic Health Authority (SHA)-specific time-trends; (7) HHI in market m is the weighted average of the HHI of all MSOAs included in the market area, using the share of market m hip replacement patients who live in each MSOA as weights, for the period between 2002/03 and 2004/05. (8) All specifications clustered at the hospital market level, *** p< 0.01, ** p<0.05, * p<0.1.

2004 that had not entered the public market by 2012/13, and (iii) areas with a pre-existing private hospital treating public patients in 2012/13.[41]

Table 4.6 shows the results for each of our outcomes. If areas with pre-existing private hospitals have fundamentally different trends in outcomes which are driving our results, then we would expect outcomes in areas with private hospitals that didn't enter to change in ways that are similar to areas where private hospitals did enter. However, the results suggest that this is not the case. The estimates show a negative and statistically insignificant relationship between volumes and areas with a private hospital that had not entered the market (relative to volumes in areas with no private hospital). Similarly, there is no significant impact on waiting times or readmission rates. In contrast, the estimated impact of the entry of a private hospital entering the public market is consistent with our previous results. This suggests that the results are robust to concerns surrounding demand shocks during this period, including concerns that NHS policymakers may purposely have invested differently in areas where private hospitals existed before the reform.

[41] 30 of the 100 markets with private hospitals in 2004 had no private hospitals treating public patients in 2012/13.

We also examine the extent to which the wider patient choice reforms may explain our results. The patient choice reforms took place during an overlapping period, with elective patients offered a choice of at least four providers in 2006 and any publicly funded provider in 2008. If private hospitals were introduced in areas with greater numbers of pre-existing alternative public hospitals - and therefore greater choice - then any changes in outcomes may be caused by patient choice rather than the introduction of a private hospital into the market. We therefore examine whether our results are affected by controlling for the local pre-reform level of competition. To control for competition in the local area we calculate the Herfindahl-Hirschman Index (HHI) for each public hospital in the period between 2002 and 2004, and interact this measure with a dummy variable that takes the value of one during the reform period (2006/07 onwards) and zero otherwise.[42] If our results are driven by the choice reforms then we would expect the inclusion of a measure of potential choice to substantially attenuate our results. Columns 5-8 of Table 4.5 show the results for each of our outcomes of interest. The results are again substantially unaltered. The only difference is that the estimate on waiting times is no longer statistically significant as a result of being slightly smaller in magnitude.

Table 4.6: *OLS estimates of the impacts of private hospital exposure in 2004, by public market entry status*

	Volume	ln(med wait)	ln(readmit)
	(1)	(2)	(3)
Priv. hosp., no entry	6.66	-0.026	0.0006
	(10.60)	(0.054)	(0.0032)
Priv. hosp., entry	29.81***	-0.068	0.0018
	(10.55)	(0.045)	(0.0024)
Observations	1,430	1,430	1,430
R-Squared	0.751	0.863	0.140

Notes: (1)'Priv. hosp., no entry' is a dummy variable that takes the value of one if there was a private hospital in the area in 2004 and no private hospital treating public patients in 2012/13, interacted with the 'post' dummy (2006/07 onwards); (2) 'Priv. hosp., entry' is a dummy variable that takes the value of one if a private hospital was treating public patients in 2012/13, interacted with the 'post' dummy (2006/07 onwards); (3) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, and a full set of year and hospital market fixed effects; (4) There are 130 hospitals; (5) All specifications clustered at the hospital market level, *** p< 0.01, ** p<0.05, * p<0.1.

[42]We calculate the HHI for each MSOA in market m, and take the weighted average HHI for market m using the share of m's patients who live in each MSOA as weights. We use the pre-reform HHI data (2002/03 to 2004/05) to remove any endogenous effects of the introduction of the private provider on the level of competition in the area, and include only public hospitals.

4.6 Who benefited from hospital entry?

4.6.1 Heterogeneity in impacts across local area deprivation

Previous work has shown that ISTCs typically treated patients who were healthier and from less deprived areas than patients treated by NHS providers (Bardsley and Dixon, 2011; Chard et al., 2011; Cooper et al., 2018). If the same pattern holds for other private hospitals, benefits from hospital entry may accrue more to those living in more affluent areas, potentially exacerbating already existing socioeconomic gradients in the receipt of joint replacements (Judge et al., 2010). We therefore now examine whether hospital entry have contributed to an increase in inequality in hip replacements across areas with different levels of deprivation.

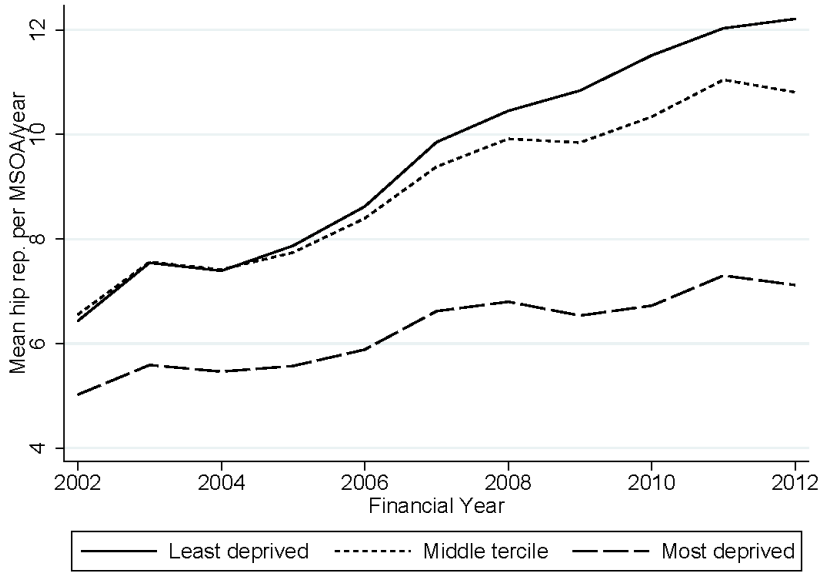
We explore impacts of hospital entry across areas with different levels of deprivation, using the 2004 Index of Multiple Deprivation (IMD) to classify different deprivation groups. In our previous analysis, we used broader hospital markets containing a number of MSOAs. We can take averages of IMD scores across MSOAs to produce market level averages of deprivation. However, this approach is likely to miss important within-market differences in deprivation, particularly for urban areas: the market average will include a range of areas, some with very low levels of deprivation and some high deprivation. To account for within market variation in deprivation, we instead study outcomes at the MSOA level (the same geography for which we observe deprivation).

Figure 4.5 shows the mean annual MSOA volume of publicly funded hip replacements between 2002/02 and 2012/13 for each deprivation tercile. As shown in Figure 4.1, the number of hip replacements has increased over time. However, this growth has been much greater in the least deprived areas. The mean number of hip replacements per MSOA grew from 6.4 to 12.2 over the period in the least deprived tercile, an increase of 90%. This compares to growth from 5.0 to 7.1 in the most deprived tercile, an increase of 42%.

The entry of private hospitals could have contributed to differential rise hip replacement volumes across areas with different levels of deprivation in two ways. First, the impacts of private hospital entry may have varied across areas with different levels of deprivation. To examine this, we estimate the following augmented version of our baseline specification:

$$Y_{mt} = \beta_0 + \beta_1(E_m * Dep_m * post_t) + \beta_2(Dep_m * post_t) + \beta_3 X_{mt} + \gamma_m + \lambda_t + \epsilon_{mt} \quad (4.2)$$

Figure 4.5: *The number of publicly funded hip replacements by provider type, 2002/03 to 2012/13*



Notes: (1) Patients are assigned to their MSOA of residence (not treatment); (2) MSOAs classified by their 2004 IMD score.

where Dep_m is the deprivation tercile of MSOA m based on its 2004 IMD score. Exposure to a private hospital in 2012/13 (E_m) is based on the measures used in our baseline at the market level: an MSOA is defined in the 'high exposure' group if a private hospital enters the market by 2012/13 in any MSOA within the trust market that the MSOA is closest to. We again instrument exposure using an indicator of whether a private hospital is located in the same area in 2004.[43]

Table 4.7 shows the results. In column 1, private hospital entry results in an increase of 0.8 hip replacements per year (or 13.5% of the sample mean in 2002) in the least deprived tercile. This closely mirrors our baseline result. The other coefficients show no statistically significant differential impacts of entry in more deprived areas. Column 2 uses volumes conducted at public hospitals only and finds no significant impacts on volumes in any deprivation tercile.

Columns 3-7 repeat this exercise for our other outcomes of interest. In all cases, the estimated impacts on the least deprived areas mirror our baseline results, with a reduction in waiting times (now statistically significant at the 1% level) and measures of patient severity. There are no additional effects associated with entry in more deprived areas. These results suggest that the impact of private hospital

[43]In this triple difference specification, β_1 now identifies the impact of private hospital presence in 2012/13 in different deprivation terciles, while β_2 accounts for other trends in outcomes during the reform period across areas with different levels of deprivation. Interactions between Dep_m and E_m are included within the MSOA fixed effect (γ_m).

entry did not vary across areas with different levels of deprivation. Instead, supply increases expanded the market equally across these groups.

However, private hospital entry may still have contributed to widening inequality in hip replacements across deprivation groups due to the location of these hospitals. Private hospitals are more commonly located in more affluent areas: 70% of MSOAs in the least deprived tercile had a private hospital treating public patients in their market in 2012/13, compared to 60% in the middle tercile and 51% in the most deprived tercile. Patients living in less deprived areas are therefore more likely to have benefited from the increased supply within their hospital market.

As a rough estimate of the share of the rising inequality explained by the location of private hospitals, we can ask how many additional hip replacements might have taken place in the most deprived areas if they had the same rate of exposure as the least deprived areas. To answer this, we first repeat our aggregate analysis at the MSOA (rather than broader hospital market level). The results are shown in Table D1 in Appendix D.[44] This suggests that private hospital entry resulted in an increase of 1.06 hip replacements each year. Increasing the likelihood of private hospital entry for those in the most deprived tercile (51%) to those in the least deprived tercile (70%) results in an increased likelihood of treatment of 19 percentage points. Multiplying this by the estimated increase suggests that the mean MSOA volume of hip replacements in the most deprived areas would be 0.2 higher. This is equivalent to just over 5% of the additional growth in hip replacements in the least deprived areas (relative to the most deprived areas) between 2002/03 and 2012/13.

Taken together, these results suggest that private hospital entry played only a small role in increasing inequality in the number of hip replacements across the deprivation distribution. Our estimates suggest that where hospitals entered the impact on volumes and other outcomes are the same across deprivation terciles. Patients living in more affluent areas are more likely to have been treated by these hospitals based on the locations of private hospitals but this only appears to explain a small proportion of the growing inequality in hip replacements by local area deprivation over this period.

4.6.2 Substitution from the private sector

One potential source of the additional publicly funded hip replacements is patients who previously would have funded their own treatment privately. This raises the question of whether the additional procedures represent an overall expansion in the

[44]The coefficients are very similar in magnitude to those in our baseline results, but now have greater statistical significance due to a larger sample size.

Table 4.7: *Two stage least squares estimates of the impact of private hospital exposure, by local area deprivation*

	Market outcomes				Mean number of comorbidities		
	Volumes (All)	Volumes (NHS only)	log (med wait)	log (readmit)	Count	% none	% 2+
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pub. funded priv. hosp.	0.808***	-0.192	-0.156***	0.002	-0.242***	0.069***	-0.067***
	(0.212)	(0.210)	(0.0287)	(0.0053)	(0.0602)	(0.0164)	(0.0150)
Pub. funded priv. hosp	0.303	0.298	0.0459	0.0025	-0.001	-0.022	0.010
* deprivation tercile 2	(0.291)	(0.289)	(0.0391)	(0.0069)	(0.0831)	(0.0210)	(0.0203)
Pub. funded priv. hosp	0.193	0.427	0.0465	0.0002	-0.018	-0.019	0.006
* deprivation tercile 3	(0.265)	(0.261)	(0.0422)	(0.0077)	(0.0893)	(0.0219)	(0.0211)
Observations	73,039	73,039	71,773	71,773	71,773	71,773	71,773
Number of MSOAs	6,640	6,640	6,640	6,640	6,640	6,640	6,640

Notes: (1) Unit of analysis is the MSOA level (all outcomes measured at this level); (2) 'Pub. funded priv. hosp.' is a dummy variable that takes the value of one in the post-reform years (2006/07) for all hospital markets where a private hospital located in the market treated public patients in 2012/13; (3) Outcomes in column 2 refers to patients treated by NHS (public) hospitals only; (4) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, an independent sector treatment centre dummy (ISTC) (equal to one if an ISTC treated public patients in 2012/13) interacted with a dummy variable that takes the value of one from 2005/06 onwards (the first year of ISTC entry), and a full set of year and hospital market fixed effects; (5) Deprivation terciles defined on the 2004 ONS Index of Multiple Deprivation at the MSOA level; (6) All specifications clustered at the MSOA level, *** p< 0.01, ** p<0.05, * p<0.1.

number of hip replacements that take place nationally each year (irrespective of funding source), or whether these procedures simply represent patients switching the source of financing for their procedures.

To examine this margin we use data from the National Joint Registry (NJR), which is a registry of all joint replacements in England, including hips, regardless of funding source.[45] This means we have information on the volumes of both privately and publicly funded procedures. Individual level patient data for privately financed care is very unusual in England, and this coverage is an important strength of the data. These data will also allow us to examine whether our baseline results are robust when using a different dataset for a similar analysis.[46]

However, the data also have two weaknesses which make them unsuitable for our main analysis. First, the geographic information for each patient is less detailed than in HES. While HES contains the MSOA of each patient, the NJR only records the patient's postal district. These postal districts are much larger

[45]The registry now covers hip, knee, ankle, elbow and shoulder joint replacements and contains more than 2.9 million records, making it the largest such database in the world. See https://www.hqip.org.uk/national-programmes/joint-replacement-surgery-the-national-joint-registry for more details.

[46]In theory, patient data for patients in HES and the NJR is linkable but we do not have permission to do this. As a result, we analyse the datasets separately.

than MSOAs, with 1,993 across England compared to 6,781 MSOAs. This means that when we define markets by assigning patients to their nearest public hospital (based on the distance between the centroid of the postal district and the hospital) we have less precise information on exactly where patients live. As a result, there is greater measurement error in the assignment of patients to their nearest hospital than there was in our original market definition when we used HES.[47]

Despite this, both the NJR and HES data produce similar volumes of publicly funded hip replacements over the period between 2008/09 and 2012/13: Appendix Figure A5 shows aggregate volumes in the NJR are slightly above those recorded in HES, but they show a similar trend in growth from 2009/10 onwards. Furthermore, we also show in Appendix Figure A6 that there is a strong positive correlation (0.87) between annual NHS volumes in the two datasets despite the differences in precise market boundaries.

Second, the NJR data quality prior to 2008/09 is poor, with a lot of missing information on how procedures are funded. As a result, the data do not include the pre-reform period. This precludes the exact empirical design that we previously used as we no longer have time variation in when private hospitals were allowed to operate. In this section we therefore instead exploit the exact timing of private hospital entry to the public market within hospital market areas to identify the impact of private hospital entry on both publicly and privately funded hip replacements (rather than studying fixed treatment and control groups as in our baseline analysis).

To do this we estimate the following equation for the period between 2008/09 and 2012/13:

$$Y_{mt} = \beta_0 + \beta_1 E_{mt} + \beta_2 X_{mt} + \gamma_m + \lambda_t + \epsilon_{mt} \tag{4.3}$$

where Y_{mt} is the volume of publicly or privately funded hip replacements for patients living in market m in year t (regardless of their actual location of treatment), and E_{mt} is a time-varying binary measure that takes the value of one if a private hospital treating publicly funded patients was located in the market in year t, and zero otherwise. We again include market and time fixed effects. The coefficient of interest β_1 now represents the association between private hospital entry and the contemporaneous number of admissions for publicly and privately

[47]Defining markets by assigning postal districts to their nearest markets will therefore yield slightly different geographic markets than in our baseline analysis (using patient MSOA in HES). However, as we show below, results from comparable regressions yield very similar estimates across the two data sources and market definitions, and there is a strong correlation in volumes in each market across the two datasets.

funded hip replacements in the local area. All standard errors are clustered at the hospital market level.

One consequence of this research design is that we can no longer use pre-existing hospital location to instrument private hospital entry. This is because the location of these hospitals does not vary over time, and would therefore be absorbed by the inclusion of area fixed effects. From our previous results, this suggests that we are likely to slightly underestimate the impact of private hospital entry on the number of publicly funded admissions in our OLS results. For privately funded admissions, we would expect the opposite effect: if private hospitals enter public markets in areas where private admissions are falling for other reasons, this would generate a negative correlation between private admission volumes and private sector entry to the public market which is not driven by entry to the public sector. As a result, we would expect to see a larger (more negative) estimated coefficient on private hospital entry than the true effect. These results should therefore be viewed as suggestive evidence rather than definitive causal impacts.

Table 4.8 shows the results. In column one the outcome is the volume of publicly funded hip replacements as recorded by HES. The estimates indicate that the presence of a private hospital in the local public market in a given year is associated with an increase of 28.9 publicly funded hip replacements. This result is statistically significant at the 1% level and is consistent with our previous estimates in Column 1 of Table 4.3 and column 1 of Table C2.[48]

[48]The results across these Tables are not directly comparable due to differences in the specifications used. The results differ from the OLS results shown in Table 4.3 as they cover a shorter period of time and exploit time-variation in exposure to private hospitals. Table C1 uses the same market definition and dataset, but presents IV (not OLS) results for a longer period of time (2002/03 onwards rather than 2008/09 onwards). However, the results across all three specifications consistently show that private hospital entry is associated with an expansion in the publicly funded market.

Table 4.8: *Estimates of the impact of private hospital exposure on hip replacement volumes, by data and funding source*

	Volume of hip replacements				% Patients 'healthy'	
	Public (HES) (1)	Public (NJR) (2)	Private (NJR) (3)	All (NJR) (4)	Public (NJR) (5)	Private (NJR) (6)
Pub. funded priv. hosp.	28.87***	22.12***	2.52	24.64***	0.0096*	-0.0113
	(8.24)	(8.45)	(2.28)	(8.53)	(0.0057)	(0.0149)
Observations	650	650	650	650	650	650
R-Squared	0.430	0.498	0.175	0.444	0.215	0.113

Notes:(1) Pub. funded priv. hosp. is a dummy variable that takes the value of one if a private hospital treats located in hospital market treats public patients in that year; (2) The outcome in column 1 is the volume of publicly-funded hip replacements as recorded by the Hospital Episode Statistics, while the outcome in columns 2-4 are the volume of publicly-funded, privately-funded and total hip replacements respectively as recorded in the National Joint Registry; (3) The outcome in columns 5 and 6 are the % of publicly funded and privately-funded patients, respectively, graded as 'healthy' on the American Society of Anesthesiologists physical status classification system scale; (4) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, and a full set of year and hospital market fixed effects; (5) All specifications clustered at the hospital market level, *** p< 0.01, ** p<0.05, * p<0.1.

In column two we repeat this analysis using the volumes of publicly funded hip replacements as recorded in the NJR as the outcome. The estimates indicate that the presence of a private hospital operating in the local public market was associated with an increased NHS volume of 22 hip replacements. The results are slightly less precise (but remain significant at the 1% level) but are consistent with the results using HES data in column one. Again, the result is consistent with our results in Section 4.5, but this time using an alternative data set and slightly different market definition (based on distances from the postal district rather than MSOA centroid).[49] Using both data sources, there is no corresponding fall in the number of procedures conducted at NHS hospitals.[50]

In column three we use the number of privately funded admissions for hip replacements as the outcome. In contrast to the strong association with public volumes, we do not find any statistically significant relationship between private hospital presence in the public market and the number of private admissions. If there was substitution between funding sources due to private hospital entry into the public market then we would expect to find a negative coefficient. However, the estimated coefficient is small and positive, and is not statistically significant

[49]These results can also be seen as an additional robustness check for our baseline results, with estimation using a separate data source providing qualitatively unchanged results.

[50]Results available upon request.

different from zero. This suggests that any substitution between public and privately funded hip replacements is minimal.

Combining the volume of public and private hip replacements gives a measure of the total number of hip replacements carried out in England in each year. In column four we use total volumes as the outcome. The results indicate that private hospital entry to the public market was associated with an increase in the total size of the hip replacement market, with the presence of a private hospital in the local area increasing total annual number of admissions for a hip replacement by 24.6 (statistically significant at the 1% level). This compares to a mean volume of 527 in 2008.

The NJR also provides some additional information on the health status of patients. This allows us to further examine changes in observed patient severity, and provides us with an additional robustness check of the results using the HES data.[51] To this end, we use the NJR data to estimate our fixed effects specification using the proportion of publicly funded patients in the area who were rated as 'healthy' on the American Society of Anesthesiologists (ASA) classification.[52] Column 5 of Table 4.8 shows the results for publicly funded patients. Again, we find evidence of improved health among patients: private hospital presence in a given year is associated with an additional 1% of publicly funded patients being rated as 'healthy' prior to the surgery. This relationship is significant at the 10% level. This means that both analyses, using two distinct data sources, indicate that patients have become observably less severe as a result of private hospital entry. We repeat this analysis for privately-financed patients in column 6. In this case, we find no statistically significant changes in average patient severity. This is consistent with no large changes in the private market taking place.

Taken together, these results suggest that the introduction of private hospitals to the public elective market led to an increase in the overall size of the market for hip replacements in England. The additional admissions for publicly funded procedures do not appear to represent financial transfers from the government to patients who would have previously financed their own treatment. Instead, the additional procedures are genuinely new procedures that would not have taken place (at least in a given year) in the absence of private sector entry. This is also consistent with the aggregate trends in Appendix Figure A5, which shows only a small decline in privately financed procedures during a period which publicly

[51] In the previous results, we address concerns that comorbidites have been recorded more accurately by all hospitals in HES over time by including year fixed effects. However, if private hospitals were slower to record comorbidities then we could overstate the impact of private hospital entry on average patient severity using this measure.

[52] The ASA scale grades patients into 6 categories based on a number of risk factors. Patients ranked as 'ASA 1' are considered to be a normal, healthy patient (Doyle, J.D. and Garmon, E.H., 2019).

funded procedures grew sharply. The results on patient severity also again point to a reduction in the average patient severity of publicly funded patients. These changes did not occur for privately financed patients. This reinforces our conclusion that the expansion in the public market enabled less severe patients to receive a hip replacement when in the absence of the reform they would not have received this treatment.

4.7 Discussion

Moves to increase the role of patient choice and promote competition between healthcare providers have been a common feature of healthcare policy across the developed world in recent years. These reforms aimed to improve efficiency among providers and to improve the quality of care provided to patients. An important component of such reforms has been the entry of new providers to compete with existing hospitals. However, despite the potential implications of such reforms relatively little is understood about the impacts of this provider entry on the structure of the elective market, incumbent providers and patient outcomes.

In this paper, we study the impacts of the entry of private hospitals on the publicly funded elective market for hip replacements. We exploit variation in the exposure to provider entry across geographic areas and the location of pre-existing private hospitals to study the impact of private hospital entry on the size and composition of the public market, the outcomes for publicly funded patients, and potential substitution from the private market.

We find that private hospital entry led to a sizeable increase in the local capacity to provide publicly funded elective care. The entry of a private hospital was associated with a 12% increase in the annual volumes of publicly funded hip replacements and importantly, no impact on the caseloads of existing public hospitals. This suggests that the competitive impacts on incumbent public providers were very weak. As a result, these hospitals faced little incentive to improve quality, with readmission rates unaffected by the reforms.

The growth in the overall size of the market, and the lack of an impact on incumbent public hospitals, is consistent with the conflicting aims of NHS policies in the 2000s. The focus on reducing waiting times and increasing activity, backed by relatively generous funding settlements for the NHS, created an environment where the markets for elective healthcare could expand quickly. This made it harder to achieve the objective of using competitive pressure from private entrants to improve the quality in public hospitals, as public hospitals could replace patients who chose private hospitals with those next on the waiting list.

The overall increase in the supply of public hip replacements also had implications for patient composition. Consistent with patients being rationed on the basis of need, our estimates from two separate datasets suggest that the increases in supply led to a reduction in the observed severity of the average patient. This means that healthier patients gained access to publicly funded elective surgery at an earlier stage than would otherwise be possible.

We examined who might potentially benefit from the reforms in two further ways. First, we examined whether private hospital entry affected areas with different levels of deprivation in the same ways. Our estimates suggested that the impacts of hospital entry did not vary along this dimension. However, the presence of private hospitals in more affluent areas did lead to a small increase in the relative number of publicly funded hip replacements in the least deprived third of areas compared to the most deprived third of areas.

Second, we examined the effects on the private market, and specifically, the evidence as to whether patients substituted from private financing to publicly funded procedures. Using a novel dataset on privately funded hip replacements, we analysed the separate impacts of private hospital entry on the size of the public, private and total market for elective hip replacements. Using these data, we corroborated our finding that private provider entry into the public market increased the number of publicly funded admissions, while having no observed impact on the size of the private market. Taken together, this evidence suggests that the reform expanded the overall market for hip replacements as opposed to reallocating patients across the public and private markets.

These findings have important policy implications. For public healthcare systems, our results show that it is possible to use the private sector to increase capacity over a relatively short period. However, the introduction of private entrants alone will not be sufficient to drive improvements in quality and efficiency in incumbent hospitals. Policymakers must think carefully about the impact that entry may or may not have on incumbent incentives. The pattern of our results, with the entry and expansion of new providers increasing the market size and relatively little impact on incumbent volumes, are very similar to ambulatory surgery centers in the US (Courtemanche and Plotzke, 2010), despite large differences in how healthcare is organised and paid for across the two countries. Changes in the location of healthcare facilities will affect the volume and pattern of use, and policymakers may wish to take this into account when making decisions.

For the UK, the role of private hospitals and the private sector more generally within the NHS remains politically controversial. We provide some empirical evidence around one area where use of the private sector has grown over the past two decades, and the implications of that for patients and public hospitals. In

an environment where the NHS budget is growing more slowly and there are ever increasing pressures from an ageing population, policymakers will need to trade off the additional capacity and lower waiting times enabled by allocating greater resources to private sector hospitals, against competing demands from within the publicly owned health system.

In any case, meeting the challenges of providing additional care in future is unlikely to be met through a large expansion in the purchase of private capacity alone. While our research examined one specific example of an expansion in the supply of publicly funded healthcare, it is unclear how the impacts of private hospital entry could differ from an expansion in supply through building new or expanding existing public hospitals. Developing further knowledge about the relative cost and benefits of these different approaches to expanding supply should therefore be a priority for future work in this area.

Chapter 5

Conclusions

Chapters 2 to 4 each contain their own conclusions. Here, I briefly set out some areas for future research.

Chapter 2 showed that there is substantial variation in the quality of individual doctors. This raises the possibility that some of the variation in patient outcomes across hospitals or geographies are driven by the average quality of doctors working in particular hospitals. For example, if hospitals located in areas with more deprived patients living nearby are staffed by doctors who on average perform worse than doctors in hospitals treating less deprived patients, these differences in average quality may be reflected in inequalities in patient outcomes across deprivation groups. Such inequalities are rarely discussed within a system of universal healthcare, and would be otherwise hard to measure. Future work should therefore consider this more explicitly. The methodology in Chapter 2 could be similarly used to calculate average doctor quality at each hospital, and to simulate scenarios where doctors are reallocated across hospitals in order to reduce health inequalities across education, deprivation and race, rather than targeting mortality reductions alone. A similar exercise already exists for understanding how much inequality in student test results across schools is explained by differences in access to teachers of different quality (Mansfield, 2015).

Chapter 2 also indicated that better patient outcomes could be achieved if existing doctors were assigned differently to patients. In practice, such reallocations are difficult to achieve. In particular, doctors may choose to work at particular hospitals for a range of personal reasons that are not observed in administrative data. Reallocating doctors across hospitals or across job roles would therefore be challenging. Understanding why doctors choose to work in certain areas, and the levers that could be used to incentivise them to move across hospitals, would therefore be a useful exercise for future work.

Chapter 3 showed that a relatively blunt regulation can be a powerful tool in changing the behaviour of doctors. Future work should seek to expand further

on the frictions that cause these improvements to be possible, and whether the target could be modified to further improve care. The results suggest that ED doctors are not well informed about the returns to quick treatment, and in an unconstrained world, would keep patients waiting for inpatient treatment longer than is optimal. One potential explanation is that ED doctors do not regularly observe what happens to their patients after they leave the ED and therefore cannot evaluate how successful treatment was. Better understanding of this mechanism would be useful to design better incentives for these doctors in future.

In recent years, hospitals in England have found it increasingly difficult to meet the ED waiting time target. The UK government also suggests that the target is outdated, and ill suited for conditions that require very quick treatment (e.g. sepsis and stroke patients). ED crowding has also intensified, and is likely to worsen patient outcomes (Woodworth, 2020). This has led to proposed changes to the existing target towards stricter targets (such as a one hour maximum wait time) for patients with particular suspected conditions, and average waiting times for all other patients in order to prevent overall waits rising too far. Prior to the coronavirus outbreak, alternative targets were being trialled in some public hospitals, but the results have not been made publicly available. An obvious extension is therefore to study how these trials (and any future wholesale changes in policy) have impacted patient outcomes and the wider production function of these hospitals.

Chapter 4 explored the impacts of the entry of private hospitals into the public elective care market in England. It showed that increased use of the private sector expanded the capacity for publicly funded care, without increasing inequalities between patients in different areas, but did not achieve the improvements in quality that the government of the time hoped to achieved. This highlighted the potential benefits and shortcomings for the NHS in interacting with the private sector.

However, more generally, the way that the private and public healthcare sectors interact in England remains poorly understood. In the face of mounting pressure from demographic changes, and recent acute shortages of elective care following widespread cancellations in the wake of the coronavirus crisis, this interplay is likely to have significant consequences for the quality and distribution of healthcare in England going forward. In particular, the private sector provides alternative opportunities for staff outside of the public hospital system. If demand for private sector care increases, this could lead to problems for the NHS in recruiting the necessary staff. Future work should therefore exploit variation across time, location, and specialties in private sector staff opportunities to examine impacts on public labour supply of doctors and the outcomes of their patients.

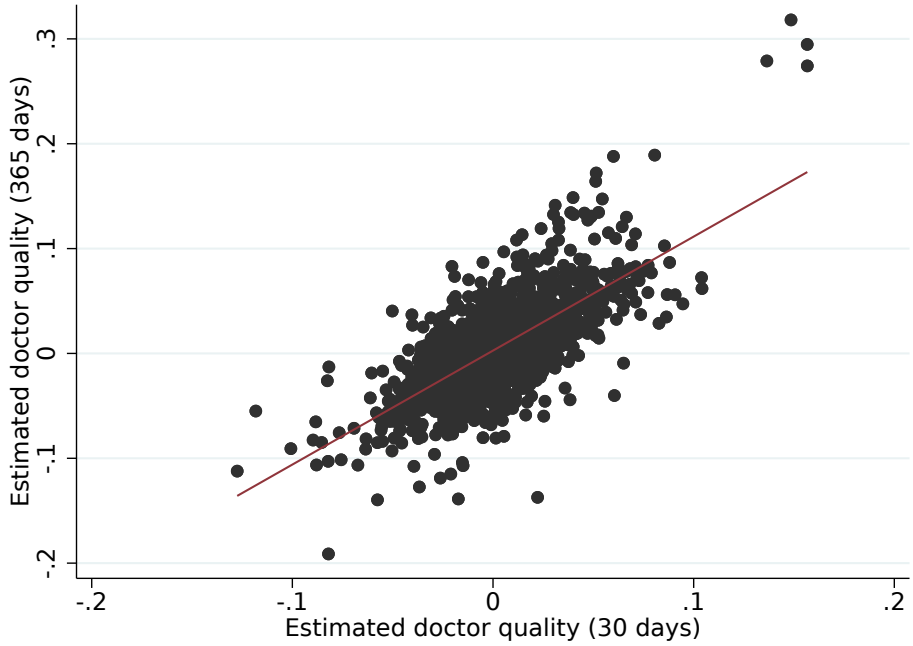
Appendices

Appendix A

Appendix to Chapter 2

A.1 Additional figures and tables

Figure A.1.1: *The correlation between estimated doctor impacts over 30 and 365 days*



Notes: (1) Doctor fixed effects from same regression as described in Table 2.4

Table A.1.1: *The estimated distribution of doctor quality under alternative hospital controls*

	365-day survival	
	Hospital fixed effect	
	Hospital (1)	Hospital-year (2)
Std Deviation	0.058	0.061
Variance	0.003	0.004
10th percentile	-0.055	-0.057
25th percentile	-0.025	-0.026
50th percentile	0.003	0.002
75th percentile	0.032	0.033
90th percentile	0.087	0.066
Number of patients	327,604	327,604
Number of doctors	1,657	1,657
Number of hospitals	140	140

Appendix B

Appendix to Chapter 3

B.1 Additional Figures and Tables

Table A1 shows the estimated impact of the target according to the lower bound of the exclusion window. In our baseline results (as shown by Table 3.3), the exclusion window began at 180 minutes. Table A1 shows two other scenarios: 170 and 190 minutes. In each case, an iterative procedure is used to automatically pick the end point of the exclusion window, as described in Section 4.2. The results show that the main estimates are robust to the choice of the lower bound of the exclusion window. While the exact magnitude of the estimates varies, the results show that the target is associated with an increased admission probability, increased number of ED investigations, increased costs and reduced 30-day mortality.

Table A2 presents the results of a second robustness check, showing how the estimated impact of the target varies by the order of polynomial used in estimation. The baseline results (shown in Table 3.3) use a polynomial of 10. Table A2 presents additional results when using a polynomial of order 6 and 8. In both cases, the results are similar to the baseline results. The final column shows the results from separately picking the polynomial for each outcomes. This is an automated

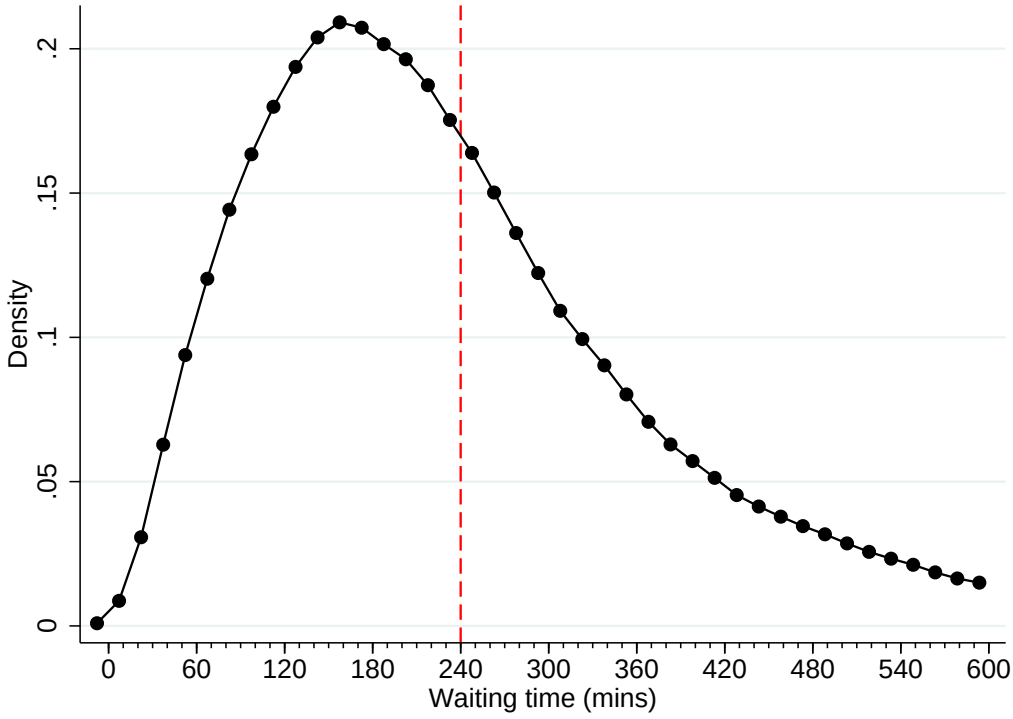
process that maximised the adjusted-R^2 statistic from estimating Equation (3.1). Again, the results are similar to the baseline estimates presented in the main text.

Table A3 shows the estimated mortality impact of the target by ED diagnosis. We created 36 outcome variables, which take the value of 1 if a patient with a specific ED diagnosis died, and 0 otherwise. We produce bootstrapped standard errors using 199 repetitions. The results show that the largest mortality reductions took place in potentially time-sensitive diagnoses with high baseline mortality rates.

Table A4 shows the estimated mortality impact of the target on broad causes of death. We created 23 outcome variables, which take the value of 1 if the ED patient dies of a specific cause of death, based on the first letter of the ICD-10 chapter recorded on their official death certificate, and 0 otherwise. We produce bootstrapped standard errors using 199 repetitions. The results show that the mortality reductions are focused among potentially time-sensitive conditions: circulatory, respiratory and digestive conditions. The results are discussed in more detail in Section 3.7.2.

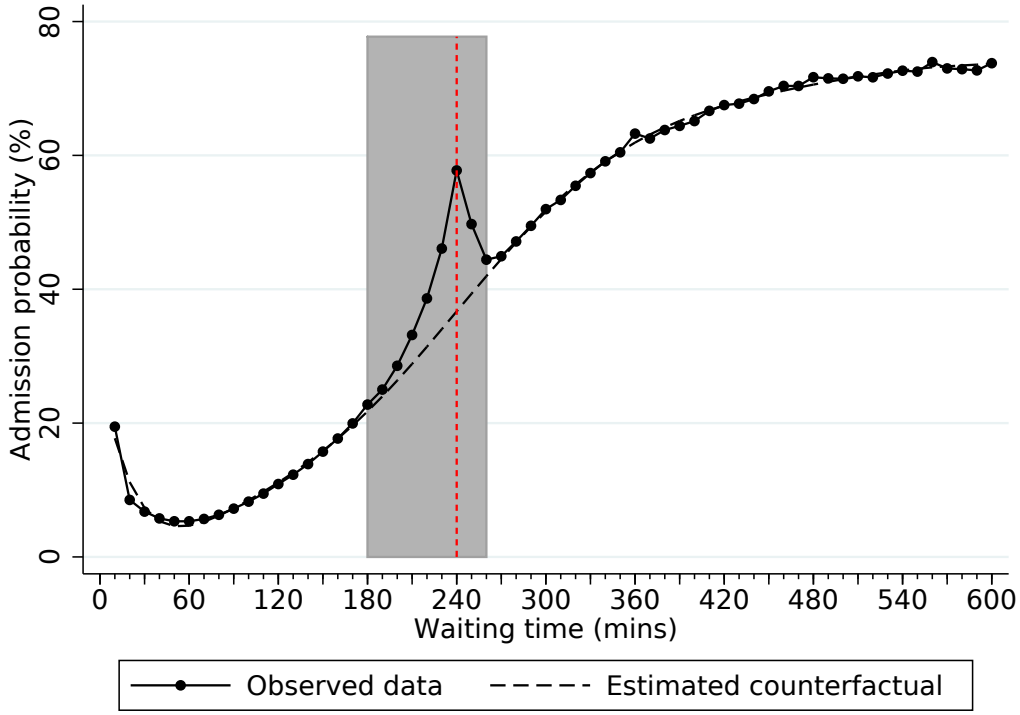
Table A5 shows the estimated mortality impact of the target on the ten most common causes of death, as defined by the first letter and first digit of the ICD-10 code recorded on official death certificates. We create 10 dummy indicators, which take the value of 1 if the ED patient dies of a specific cause within 30 days of visiting an ED, and zero otherwise. Together, these conditions account for 60% of ED patient deaths in 2011/12 and 2012/13. Again, the results show that mortality impacts are focused among the potentially time-sensitive conditions, but not among others such as cancer. The results are discussed in more detail in Section 3.7.2.

Figure A1: *Distribution of wait times at a large hospital in California*



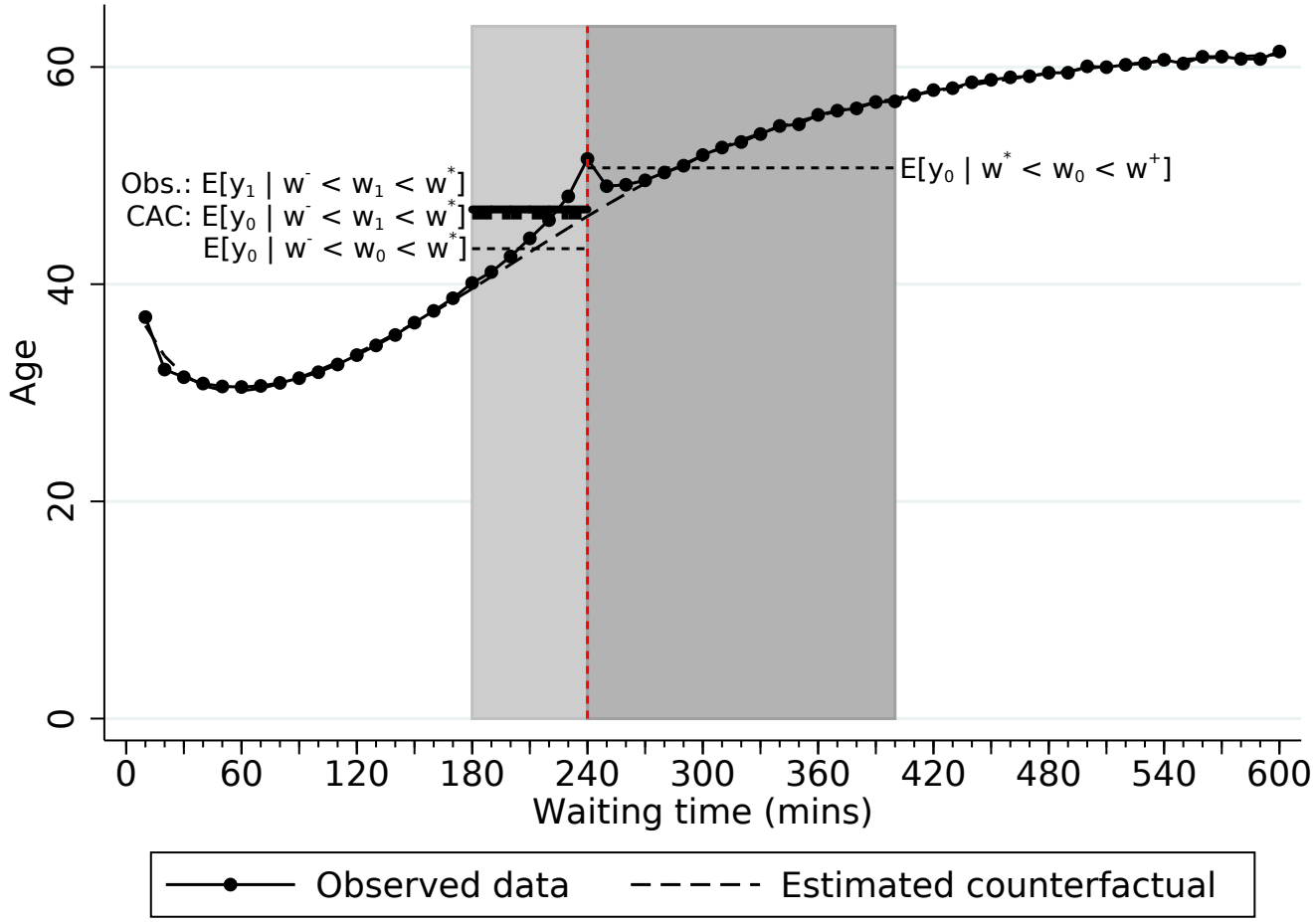
Notes: (1) The English data displays a sharp discontinuity in the wait time distribution at four hours (see Figure 3.1). Here we present the wait time distribution from a large hospital in California to illustrate that the discontinuty in the English data is unlikely to naturally occur, and is instead induced by the target; (2) We thank David Chan for providing the data for this chart.

Figure A2: *Estimated counterfactual admission probability conditional on wait times*



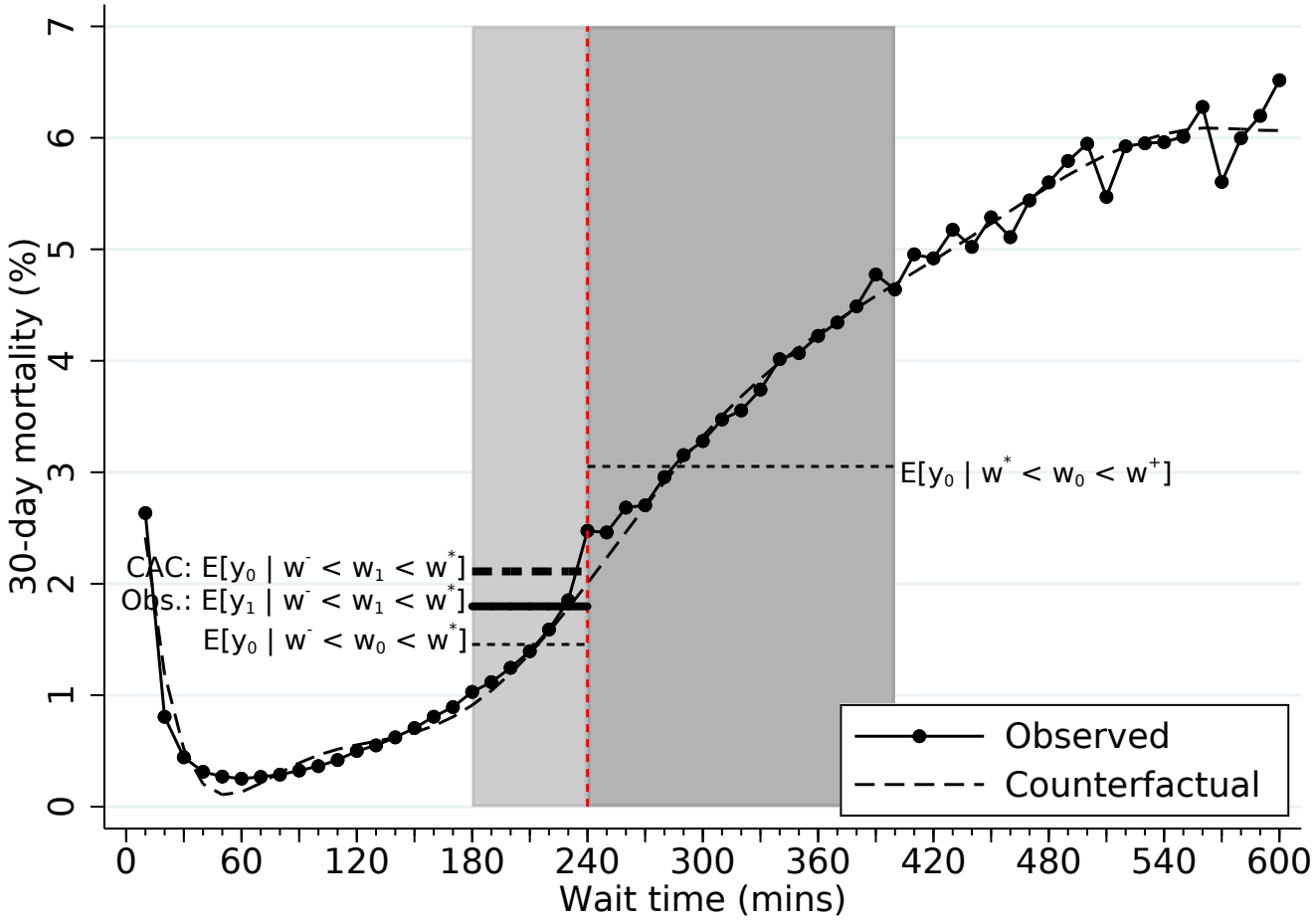
Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy; (4) The estimated counterfactual is obtained from a polynomial regression that omits the exclusion window shown in grey.

Figure A3: *Demographic test of the no-selection assumption using age*



Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy; (4) The horizontal thin dashed lines in the light grey (dark grey) region give the counterfactual outcome in the pre-threshold (post-threshold) period, $E[y_0 \mid \underline{w}_0^-]$; (5) The horizontal thick dashed line in the pre-threshold period is the composition-adjusted counterfactual, $E[y_0 \mid \underline{w}_1^-]$; (6) The horizontal thick solid line in the pre-threshold period is the observed average age, $E[y_1 \mid \underline{w}_1^-]$; (7) The distortion effect is the gap between the thick solid and dashed line, $\Delta_D = E[y_1 \mid \underline{w}_1^-] - E[y_0 \mid \underline{w}_1^-]$.

Figure A4: *Constructing the composition-adjusted counterfactual for 30-day mortality*



Notes: (1) Wait time intervals are 10-minute periods and defined as the time from arrival in the ED to leaving the ED; (2) Wait times over 600 minutes not shown; (3) 240 minutes is the four-hour threshold specified in the policy; (4) The horizontal thin dashed lines in the light grey (dark grey) region give the counterfactual outcome in the pre-threshold (post-threshold) period, $E[y_0 \mid w_0]$; (5) The horizontal thick dashed line in the pre-threshold period is the composition-adjusted counterfactual, $E[y_0 \mid \underline{w}_1^-]$; (6) The horizontal thick solid line in the pre-threshold period is the observed 30-day mortality, $E[y_1 \mid \underline{w}_1^-]$; (7) The distortion effect is the gap between the thick solid and dashed line, $\Delta_D = E[y_1 \mid \underline{w}_1^-] - E[y_0 \mid \underline{w}_1^-]$.

Table A1: *Estimated distortion effects of the target, robustness by exclusion window lower bound*

	Exclusion window lower bound (mins):				
	160	170	180 (baseline)	190	200
Panel A: ED treatment decisions					
Pr(admission)	0.033***	0.039***	0.046***	0.058***	0.074***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Pr(discharge)	−0.022***	−0.027***	−0.033***	−0.042***	−0.054***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Pr(referral)	−0.011**	−0.012***	−0.013***	−0.016***	−0.020***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
ED investigation count	0.090*	0.098**	0.108***	0.133**	0.169***
	(0.049)	(0.048)	(0.048)	(0.050)	(0.052)
ED treatment count	−0.039	−0.037	−0.033	−0.026	−0.013
	(0.029)	(0.029)	(0.028)	(0.029)	(0.029)
Panel B: Inpatient treatment decisions					
Length of stay (days)	0.022	0.022	0.035	0.087*	0.168***
	(0.047)	(0.047)	(0.048)	(0.050)	(0.052)
Inpatient procedure count	−0.003	−0.002	0.000	0.007	0.017***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Panel C: Hospital costs					
30-day ED cost	2.503***	2.786***	3.040***	3.590***	4.335***
	(0.866)	(0.876)	(0.911)	(0.972)	(1.042)
30-day inpatient cost	62.459*	93.132***	125.793***	183.993***	260.498***
	(34.028)	(33.421)	(33.992)	(35.047)	(36.550)
30-day total cost	64.962*	95.918***	128.833***	187.583***	264.833***
	(34.336)	(33.770)	(34.389)	(35.496)	(37.055)
Panel D: Patient outcomes					
30-day mortality	−0.0036***	−0.0039***	−0.0041***	−0.0040***	−0.0035***
	(0.0007)	(0.0007)	(0.0006)	(0.0006)	(0.0007)

Notes: (1) Polynomial order is set to 10 in all specifications; (2) All inpatient variables (e.g. length of stay, costs) take on the value zero for patients that are not admitted; (3) Bootstrapped standard errors clustered at the hospital trust level (199 repetitions).

Table A2: *Estimated distortion effects of the target, robustness by polynomial order*

	Polynomial order			
	6	8	10 (baseline)	Auto
Panel A: ED treatment decisions				
Pr(admission)	0.033***	0.041***	0.046***	0.046***
	(0.008)	(0.008)	(0.008)	(0.008)
Pr(discharge)	−0.013**	−0.029***	−0.033***	−0.033***
	(0.007)	(0.006)	(0.007)	(0.007)
Pr(referral)	−0.020***	−0.012***	−0.013***	−0.013***
	(0.003)	(0.003)	(0.003)	(0.003)
ED investigation count	0.101**	0.090*	0.108***	0.108***
	(0.046)	(0.048)	(0.048)	(0.048)
ED treatment count	−0.024	−0.031	−0.033	−0.026
	(0.027)	(0.027)	(0.028)	(0.030)
Panel B: Inpatient treatment decisions				
Length of stay (days)	−0.066	−0.006	0.035	0.035
	(0.051)	(0.050)	(0.048)	(0.048)
Inpatient procedure count	−0.014	−0.005	0.000	0.000
	(0.006)	(0.006)	(0.006)	(0.006)
Panel C: Hospital costs				
30-day ED cost	1.651*	2.638***	3.040***	3.080***
	(0.946)	(0.880)	(0.911)	(0.939)
30-day inpatient cost	46.035	95.305***	125.793***	125.793***
	(36.389)	(34.955)	(33.992)	(33.992)
30-day total cost	47.680	97.905***	128.833***	128.833***
	(36.857)	(35.331)	(34.389)	(34.389)
Panel D: Patient outcomes				
30-day mortality	−0.0060***	−0.0071***	−0.0041***	−0.0041***
	(0.0007)	(0.0007)	(0.0006)	(0.0006)

Notes: (1) Exclusion window begins at 180 minutes in all specifications; (2) 'Auto' selects the polynomial separately for each outcome, by selecting the polynomial that maximizes the adjusted-R^2 statistic from estimating Equation (3.1); (3) All inpatient variables (e.g. length of stay, costs) take on the value zero for patients that are not admitted; (4) Bootstrapped standard errors clustered at the hospital trust level (199 repetitions).

Table A3: *Estimated effects of the target on mortality, by ED diagnosis*

ED diagnosis	Wait time reduction (mins)	Mortality reduction (ppts)	Predicted mortality	Affected patients in 2012/13	% of total lives saved
Septicaemia	15	5.3	3.7%	14,510	6.9%
Vascular injury	7	2.2	1.6%	3,381	0.7%
Cerebro-vascular	9	2.2	3.8%	50,257	9.9%
Other vascular	7	2.1	1.9%	29,940	5.6%
Respiratory	11	1.5	2.1%	244,732	32.7%
Haematological	11	1.3	1.7%	15,094	1.7%
Central nervous system	12	0.8	1.9%	108,315	7.7%
Gynaecological	4	0.8	0.2%	38,203	2.7%
Cardiac	6	0.6	3.1%	229,047	12.2%
Gastrointestinal	10	0.5	1.3%	329,935	14.2%
Laceration	0	0.4	0.6%	99,535	3.5%
Local infection	3	0.4	0.8%	63,561	2.3%
Diabetes and endocrine	19	0.4	2.3%	28,252	1.0%
ENT	3	0.4	0.7%	51,964	1.9%
Obstetric	1	0.3	0.2%	12,728	0.3%
Soft tissue inflammation	4	0.2	0.5%	104,420	1.9%
Urological	12	0.2	2.2%	128,363	2.3%
Facio-maxillary	1	0.2	0.3%	9,715	0.2%
Social problems	12	0.2	3.2%	17,378	0.3%
Nothing abnormal detected	6	0.2	1.3%	85,682	1.5%
Head injury	5	0.1	1.1%	84,319	0.8%
Bites/stings	1	0.1	0.2%	7,692	0.0%
Infectious disease	7	0.1	0.9%	47,447	0.4%
Psychiatric	5	0.1	0.7%	46,649	0.4%
Burns and scalds	3	0	0.3%	9,913	0.0%
Poisoning (inc overdose)	12	0	0.6%	72,641	0.0%
Joint injury/fracture	6	-0.1	0.9%	217,152	-1.9%
Contusion/abrasion	4	-0.2	0.5%	69,328	-1.2%
Muscle/tendon injury	2	-0.3	0.5%	49,384	-1.3%
Dermatological	2	-0.3	0.4%	17,143	-0.5%
Allergy	1	-0.3	0.4%	14,942	-0.4%
Ophthalmological	0	-0.3	0.2%	25,701	-0.7%
Foreign body	0	-0.5	0.1%	15,314	-0.7%
Sprain/ligament injury	1	-0.6	0.2%	92,555	-4.9%

Notes: (1) Column 2 and 3 contain the estimated reduction in wait times and 30-day mortality for patients with each ED diagnosis, respectively; (2) There are 40 diagnosis categories. Non-missing data defined as patients without a missing or 'not classifiable' diagnosis.

Table A4: *Estimated mortality effects of the target by cause of death (ICD-10 chapter)*

Cause of death	Mortality reduction (percentage points)		% of overall mortality impact	% reduction in deaths
Circulatory	0.124***	(0.021)	30.1%	19.6%
Respiratory	0.105***	(0.018)	25.6%	25.5%
Digestive	0.062***	(0.010)	15.0%	35.2%
Unintentional accidents	0.024***	(0.004)	5.8%	47.9%
Mental/behavioural	0.015**	(0.007)	3.6%	15.2%
Genitourinary disease	0.014***	(0.005)	3.5%	27.1%
Infectious disease	0.013***	(0.004)	3.2%	39.0%
Neoplasms	0.011	(0.018)	2.7%	2.8%
Musculoskeletal	0.011***	(0.003)	2.6%	38.4%
Nervous system	0.009*	(0.005)	2.3%	15.2%
Endocrine/metabolic	0.005*	(0.003)	1.3%	20.5%
Disorders of the blood	0.004	(0.003)	0.9%	20.7%
Vehicle and traffic accidents	0.003**	(0.001)	0.8%	45.1%
Skin and subcutaneous tissue	0.003	(0.002)	0.7%	26.7%
Other external causes	0.002	(0.002)	0.6%	27.8%
External cause (e.g. fire, nature)	0.002	(0.003)	0.5%	7.7%
Codes for special purposes	0.002**	(0.001)	0.5%	53.9%
Congenital	0.001	(0.001)	0.2%	14.8%
Parasitic diseases	0.000	(0.001)	0.1%	9.0%
Pre-natal	0.000	(0.000)	0.1%	63.3%
Eye and ear	0.000	(0.000)	0.0%	17.9%
Pregnancy related	−0.000	(0.000)	0.0%	−109.3%
Symptoms not classified	−0.000	(0.000)	−0.1%	−3.0%

Notes: (1) Cause of death categories defined by the first letter of the ICD-10 diagnosis code; (2) Column 2 shows the estimated reduction in 30-day mortality attributed to the cause of death; (3) Column 3 shows the proportion of the overall mortality reduction that is accounted for by the cause of death; (4) Column 4 shows the proportion of deaths due to the specific cause that is avoided because of the target.

Table A5: *Estimated distortion effects of the target by cause of death (ICD-10 sub-chapter)*

Cause of death	Mortality reduction (percentage points)		% of overall mortality impact	% reduction in deaths
Cerebrovascular diseases	0.071***	(0.010)	17.2%	33.3%
Chronic lower respiratory diseases	0.055***	(0.009)	13.3%	29.0%
Influenza and pneumonia	0.034***	(0.008)	8.2%	23.0%
Ischemic and pulmonary heart diseases	0.030**	(0.012)	7.4%	11.7%
Organic mental disorders	0.013*	(0.006)	3.2%	14.1%
Malignant neoplasms (respiratory, intrathoracic)	0.005	(0.008)	1.2%	4.8%
Malignant neoplasms (lip, oral cavity and pharynx)	0.002	(0.005)	0.5%	3.8%
Malignant neoplasms (digestive)	0.001	(0.005)	0.2%	2.0%
Malignant neoplasms (male genital organs, urinary tract)	0.001	(0.004)	0.2%	1.5%
Malignant neoplasms (breast, female genital organs)	−0.000	(0.004)	−1.1%	−12.4%

Notes: (1) Cause of death categories are defined by the first letter and digit of their ICD-10 code: Cerebrovascular diseases (I6), chronic lower respiratory disease (J4), influenze and pneumonia (J1), ischemic heart diseases and pulmonary heart disease (I2), Organic, including symptomatic, mental disorders (F0), Malignant neoplasms of respiratory and intrathoracic organs (C3), Malignant neoplasms of lip, oral cavity and pharynx (C1), Malignant neoplasms of digestive organs (C2), Malignant neoplasm of breast and female genital organs (C5), and Malignant neoplasm of male genital organs and urinary tract (C6); (2) Column 2 shows the estimated reduction in 30-day mortality attributed to the cause of death; (3) Column 3 shows the proportion of the overall mortality reduction that is accounted for by the cause of death; (4) Column 4 shows the proportion of deaths due to the specific cause that is avoided because of the target.

B.2 Selection simulation

This appendix sets out the details of a simulation we conducted to evaluate the no-selection assumption. There are two stages in the simulation. The first stage is to produce a simulated 'counterfactual dataset' that is based on the counterfactual wait time and age distribution. The second stage is to use the counterfactual dataset to simulate different responses to the four-hour target, specifically in terms of how the post-threshold movers are selected. We then compare these simulated outcomes to the observed data to learn about selection and the validity of our no-selection assumption. Below we describe the two stages of the simulation and the results.

B.2.1 Constructing the counterfactual dataset

We take the following steps:

1. We compute the counterfactual wait time distribution as described in Section 3.4.1.

2. We compute the counterfactual expectation of age conditional on wait times as described in Section **??**. We use this same approach to compute the counterfactual standard deviation of age conditional on wait times.

3. Using outputs from steps 1 and 2, we create a simulated dataset of patients. This dataset has the counterfactual distribution of wait times and an age distribution that is normally distributed with its mean and standard deviation defined according to the results from step 2. As a result, the wait time distribution and the conditional expectation of age are both smooth functions through the four-hour threshold.

4. We generate a random variable, denoted ε_i, where $\varepsilon_i \sim N(0, \sigma^2_{age})$ and σ^2_{age} is the variance of the age variable created in step 3. We normalise this variance by the variance of age to help with interpretation later.

B.2.2 Simulating the selection of post-threshold movers

We take the following steps:

1. We set up the following selection equation

$$S_i = \beta age_i + \varepsilon_i, \tag{B.2.1}$$

where S_i is the selection index for patient i, β is a selection parameter to be specified, age_i is the age of patient i, and ε_i is a random term for patient i.

2. For each wait time bin in the post-threshold period, w^* to w^+, we define post-threshold movers as follows

$$M_i = 1\{S_i \geq \tau_w\}, \tag{B.2.2}$$

where M_i is a binary variable equal to one if patient i is a post-threshold mover and τ_w is a threshold specific to wait time bin w.

3. The selection thresholds τ_w are unknown but we can estimate each threshold by finding the number of post-threshold movers that equates the wait time distributions for that bin in the counterfactual and observed datasets. The post-threshold movers are then identified as those patients with $M_i = 1$.

4. We consider the following different scenarios for β:

   (a) If $\beta = 0$ then there is 'random selection' as post-threshold movers are determined purely by ε_i, which is entirely random.

   (b) If $\beta = 1$ then there is 'selection-on-observables', specifically on age. Note that age and ε_i contribute equally to variation in S_i in this scenario.

   (c) If $\beta \in (0, 1)$ then there is selection-on-observables, but age plays a smaller role than ε_i in determining S_i. Note that this scenario can also

be thought of as 'selection-on-unobservables' in the following sense: if $\beta = 0$ but ε_i contains some non-random element that is positively correlated with age, then the resulting selection equation is equivalent to Equation (B.2.1) with $\beta \in (0, 1)$.

5. To complete the simulation, we need to specify how post-threshold movers are allocated to the pre-threshold period. This allocation is unknown and so we simply adopt the simplest possible rule for the purposes of illustration. When shifting post-threshold movers we maintain their existing wait time ordering, such that those located just above w^* are moved to w^-, and those located at w^+ are moved to w^*.

B.2.3 Results

Figure B1 first shows the conditional expectation of age in the simulated dataset (Panel a), and then the results of simulating random selection of post-threshold movers (Panel b). Random selection has three main features: (i) there is a spike at 240 minutes, which is where many of the post-threshold movers are shifted to; (ii) there is an increase in the pre-threshold level which is smaller the further away it is from the 240 threshold; (iii) there is a smooth distribution after the 240 threshold.

Figure B2 compares the random selection case to the observed data. The two conditional expectation functions are very similar, with the observed data exhibiting the same three features described above. The observed data lies marginally above the simulated data, but the gap is small.

Figure B3 now introduces the different selection scenarios and in each case compares the scenario to the random selection case. The scenarios are $\beta = \{0.1, 0.5, 1\}$. As described earlier, the cases where $\beta < 1$ can be thought of as selection-on-unobservables that are correlated with age.

Looking first at the scenario with $\beta = 1$ in Panel (b). The markers of selection are clear: the spike at 240 minutes is large; the pre-threshold period is substantially

higher; and there is a pronounced drop in the post-threshold period. The post-threshold drop is more pronounced than the pre-threshold increase because, in the pre-threshold period, selection causes the post-threshold movers to be averaged with the pre-threshold non-movers while, in the post-threshold period, selection leaves behind a very select group of post-threshold non-movers.[1]

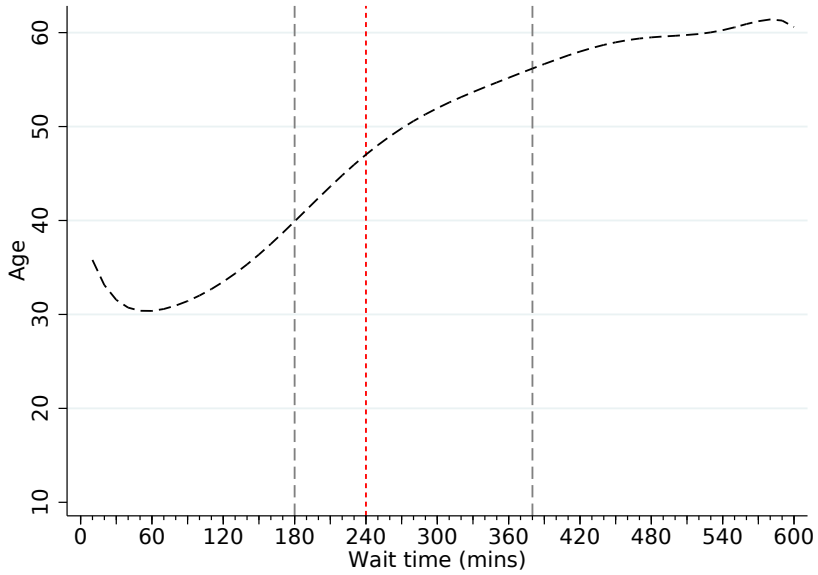
Turning now to the cases with $\beta = 0.5$ (Panel c) and $\beta = 0.1$ (Panel d). These share the same characteristics as the previous selection scenario, although the markers of selection become less pronounced as the selection mechanism is weaker. In the final case, where $\beta = 0.1$, the pre-threshold period differences are very small. Yet even in this case, the post-threshold period exhibits the pronounced drop.

These simulations highlight three key points. First, the data looks very similar to the random selection case, sharing the same three features. Second, to the extent that there is significant selection, for example in the case of $\beta = 1$, then its markers show up clearly in the data. None of these markers are present in the observed data. Third, to the extent that observable and unobservable variables are correlated, then selection-on-unobservables will manifest itself directly in the observables. As a result, concerns about selection-on-unobservables can be thought of directly in terms of this correlation. While it is difficult to quantify the correlation and link it with the power of our demographic test, there is clear evidence linking age and ambulance status to many medically relevant unobservables. For example, medical guidelines for physicians routinely incorporate age-based decision rules and, even after conditioning on age and diagnosis, the likelihood of death is more than 150% higher for ambulance patients than the average patient. These facts, which suggest that the correlation between observables and unobservables is far from negligible, are reassuring given that we find no evidence of selection on these observable variables.

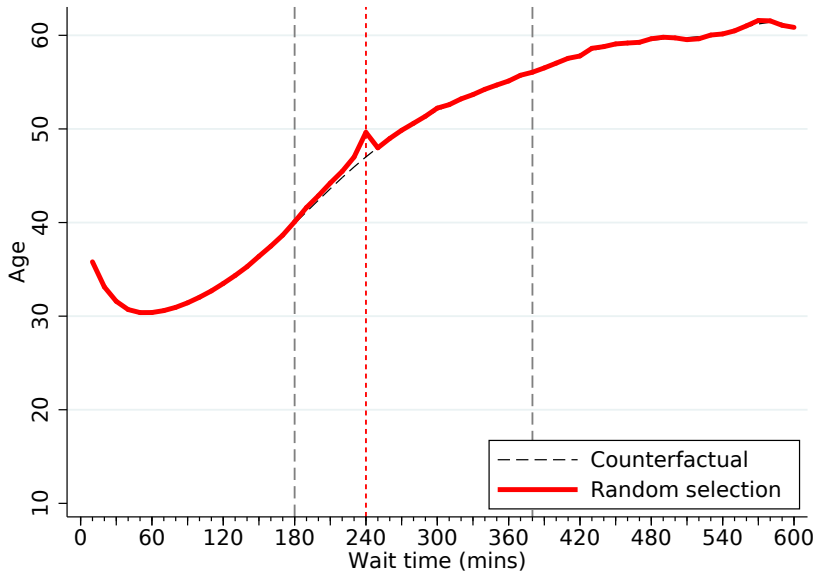
[1]Note that the specific pattern of the pre-threshold and post-threshold period is influenced by the allocation of post-threshold movers which is somewhat arbitrary in this simulation.

Figure B1: *Selection simulation using the counterfactual dataset*

(a) *Conditional expectation of age*

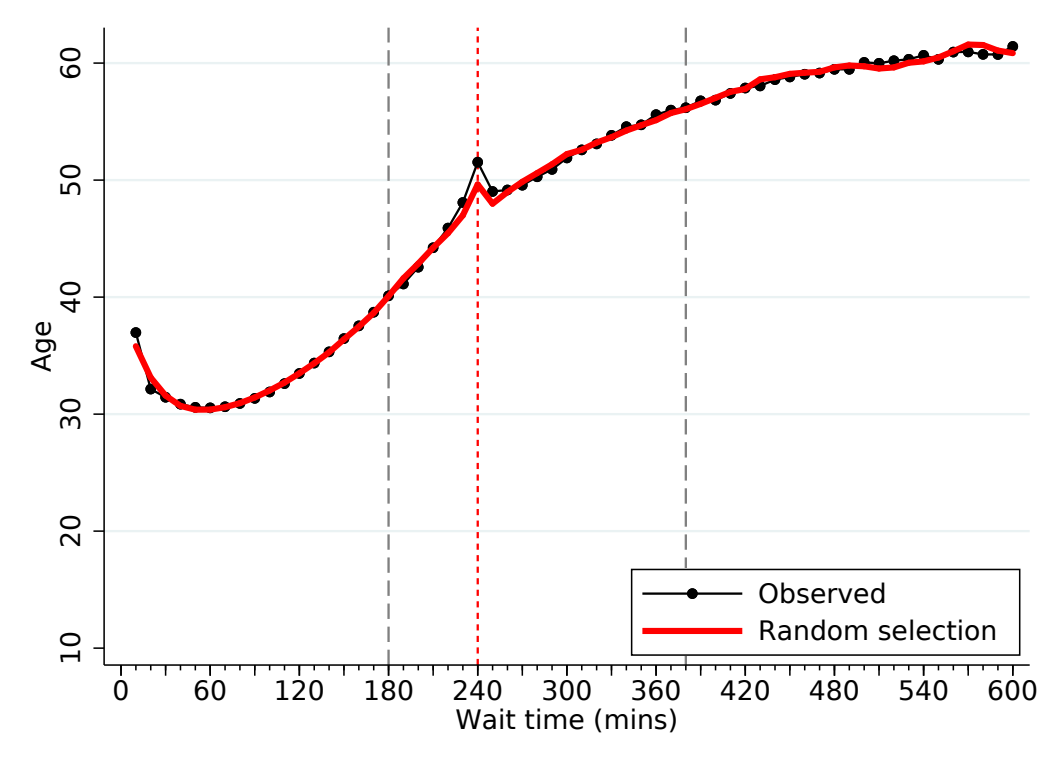


(b) *Random selection simulation of post-threshold movers*



Notes: (1) Panel (a) shows the expectation of age conditional on the wait time bin in the counterfactual dataset ('counterfactual'); (2) Panel (b) shows the simulated data when post-threshold movers are chosen at random and shifted into the pre-threshold period ('random selection'); (3) Vertical dashed lines indicate the exclusion window and the vertical dashed red line indicates the 240 minute threshold.

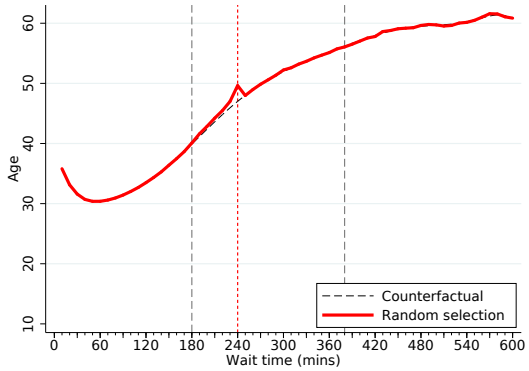
Figure B2: *Comparison of the random selection simulation and the observed data*



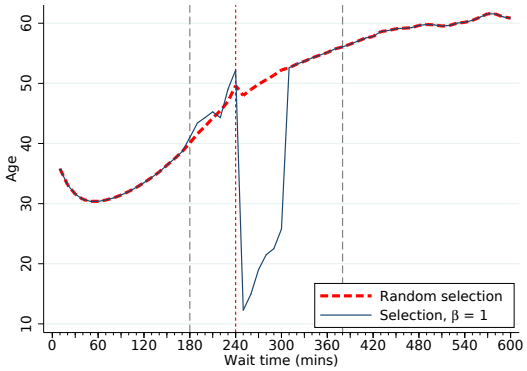
Notes: (1) This chart compares the simulated data when post-threshold movers are chosen at random and shifted into the pre-threshold period ('random selection') with the observed data ('observed'); (2) Vertical dashed lines indicate the exclusion window and the vertical dashed red line indicates the 240 minute threshold.

Figure B3: *Comparison of the random selection and selection simulations*

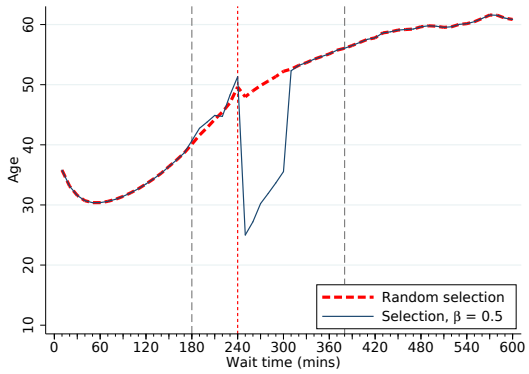
(a) *Random selection*



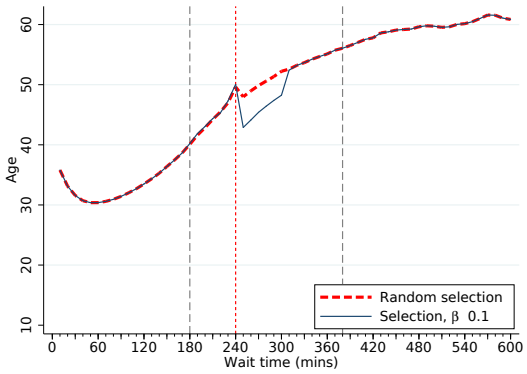
(b) *Selection:* $\beta = 1$



(c) *Selection:* $\beta = 0.5$



(d) *Selection:* $\beta = 0.1$



Notes: (1) These charts compares the simulated data when post-threshold movers are chosen at random and shifted into the pre-threshold period ('random selection') with various degrees of selection-on-observables ('selection'); (2) Vertical dashed lines indicate the exclusion window and the vertical dashed red line indicates the 240 minute threshold.

B.3 Common Treatments and Investigations in English Emergency Departments

The objective of much of the ED care provided in the ED setting is either to treat simple conditions, or to provide early diagnostic information that can be used to send patients to the correct specialist inpatient ward for future treatment (while also stabilising patients and providing basic treatment as these tests are carried out). Table C1 shows the most common ED investigations and treatments across 40 ED diagnosis categories, distinguishing between the first and subsequent investigations and treatments.[2] X-rays and blood tests are the most common primary investigations, while in a quarter of (more minor) diagnoses the majority of patients receive no specific investigations. Most treatments are simple: providing guidance or advice, or treating minor cases (e.g. wound closures for lacerations, plaster of paris for fractures). For the more serious cases, common treatments include inserting an intravenous cannula or observing patients (including taking an ECG or recording patients' pulses) while diagnostic tests such as x-rays, CT scans and blood tests are carried out. This is further reflected in Table C2, which shows the most common ED investigations and treatments for admitted (and therefore likely more serious) patients only.

Patients who require further specialist treatment are then admitted as an inpatient. Table C3 shows the most common first and subsequent inpatient procedure across each ED diagnosis. Initial inpatient treatment is also often diagnostic in nature, as shown by the frequency of the use of CT and MRI scans. More comprehensive treatment of the condition then follows.

Taken together, these tables demonstrate that the ED provides an important first stage of treatment, solving more minor problems in the department itself,

[2]There are 24 ED investigation categories, including 'none'. There are 57 ED treatments. Details of these can be found in the HES Data Dictionary (Accident and Emergency), available here: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary

while collecting important diagnostic information that is important in ensuring that more complex cases receive the correct inpatient treatment further along the treatment pathway.

Table C1: *The most common ED invesigations and treatments, all patients*

	Most common ED investigations				Most common ED treatments			
ED diagnosis	First investigation	% of patients	Subsequent investigation	% of patients	First treatment	% of patients	Subsequent treatment	% of patients
Laceration	None	65.8%	Biochemistry	2.6%	Wound closure	20.1%	Wound closure	18.8%
Contusion/abrasion	X-ray plain film	55.0%	Biochemistry	3.0%	Guidance/advice only	36.8%	None	6.8%
Soft tissue inflammation	X-ray plain film	51.9%	Biochemistry	6.2%	Guidance/advice only	37.2%	Medication administered	7.0%
Head injury	None	61.6%	Biochemistry	6.9%	Guidance/advice only	30.9%	Observation (ECG, pulse oximetry etc)	12.8%
Joint injury/fracture	X-ray plain film	81.6%	Haematology	8.0%	Plaster of paris	17.9%	Plaster of paris	9.4%
Sprain/ligament injury	X-ray plain film	65.0%	Biochemistry	1.7%	Guidance/advice only	38.8%	Medication administered	6.4%
Muscle/tendon injury	X-ray plain film	44.9%	Biochemistry	7.7%	Guidance/advice only	30.5%	Medication administered	9.1%
Nerve injury	None	42.9%	Biochemistry	19.2%	Recording vital signs	36.4%	Medication administered	17.2%
Vascular injury	None	31.9%	Biochemistry	21.5%	Guidance/advice only	16.7%	Observation (ECG, pulse oximetry etc)	16.0%
Burns and scalds	None	84.4%	ECG	1.8%	Dressing	42.4%	Dressing	17.8%
Electric shock	X-ray plain film	50.7%	Urinalysis	2.9%	Guidance/advice only	44.2%	Medication administered	15.1%
Foreign body	None	62.3%	Biochemistry	1.5%	Removal foreign body	26.4%	Removal foreign body	12.2%
Bites/stings	None	78.3%	Biochemistry	2.1%	Medicines prepared to take away	18.9%	Medicines prepared to take away	19.3%
Poisoning (inc overdose)	None	28.4%	Biochemistry	31.0%	Observation (ECG, pulse oximetry etc)	22.5%	Observation (ECG, pulse oximetry etc)	19.1%
Near drowning	None	73.4%	Biochemistry	5.8%	Dressing	28.2%	Dressing	36.9%
Visceral injury	None	38.1%	Biochemistry	20.8%	Guidance/advice only	16.0%	Observation (ECG, pulse oximetry etc)	22.0%
Infectious disease	None	43.2%	Biochemistry	22.3%	Guidance/advice only	18.4%	Observation (ECG, pulse oximetry etc)	15.1%
Local infection	None	49.5%	Biochemistry	16.7%	Medicines prepared to take away	18.5%	Observation (ECG, pulse oximetry etc)	13.8%
Septicaemia	X-ray plain film	44.3%	Haematology	47.3%	Intravenous cannula	21.3%	Observation (ECG, pulse oximetry etc)	27.7%
Cardiac	X-ray plain film	38.7%	Haematology	49.4%	Intravenous cannula	19.2%	Observation (ECG, pulse oximetry etc)	21.5%
Cerebro-vascular	CT scan	26.4%	Biochemistry	43.7%	Intravenous cannula	21.8%	Observation (ECG, pulse oximetry etc)	21.6%
Other vascular	Haematology	20.5%	Biochemistry	33.5%	Medication administered	18.7%	Observation (ECG, pulse oximetry etc)	17.5%
Haematological	X-ray plain film	29.0%	Biochemistry	36.3%	Guidance/advice only	18.7%	Observation (ECG, pulse oximetry etc)	16.8%
Central nervous system	None	24.6%	Biochemistry	33.1%	Observation (ECG, pulse oximetry etc)	20.3%	Observation (ECG, pulse oximetry etc)	18.8%
Respiratory	X-ray plain film	40.7%	Haematology	33.1%	Observation (ECG, pulse oximetry etc)	15.2%	Observation (ECG, pulse oximetry etc)	19.5%
Gastrointestinal	X-ray plain film	21.8%	Biochemistry	36.8%	Observation (ECG, pulse oximetry etc)	15.2%	Observation (ECG, pulse oximetry etc)	18.4%
Urological	Urinalysis	19.3%	Biochemistry	34.7%	Observation (ECG, pulse oximetry etc)	13.5%	Observation (ECG, pulse oximetry etc)	19.3%
Obstetric	None	21.8%	Biochemistry	24.9%	Observation (ECG, pulse oximetry etc)	19.6%	Observation (ECG, pulse oximetry etc)	18.0%
Gynaecological	Haematology	26.4%	Biochemistry	32.1%	Guidance/advice only	19.2%	Observation (ECG, pulse oximetry etc)	14.3%
Diabetes and endocrine	Haematology	24.1%	Biochemistry	44.2%	Intravenous cannula	20.6%	Observation (ECG, pulse oximetry etc)	19.4%
Dermatological	None	68.4%	Biochemistry	9.3%	Guidance/advice only	23.1%	Recording vital signs	11.8%
Allergy (inc anaphylaxis)	None	65.9%	Biochemistry	11.2%	Medication administered	18.6%	Observation (ECG, pulse oximetry etc)	15.7%
Facio-maxillary conditions	None	66.0%	Biochemistry	6.8%	Guidance/advice only	20.1%	Observation (ECG, pulse oximetry etc)	11.9%
ENT	None	61.5%	Biochemistry	11.9%	Guidance/advice only	20.8%	Observation (ECG, pulse oximetry etc)	11.4%
Psychiatric	None	55.8%	Biochemistry	16.5%	Guidance/advice only	28.0%	Observation (ECG, pulse oximetry etc)	11.6%
Ophthalmological	None	58.3%	Biochemistry	1.6%	Guidance/advice only	26.4%	Medication administered	11.9%
Social problems	None	30.8%	Biochemistry	29.5%	Observation (ECG, pulse oximetry etc)	21.1%	Observation (ECG, pulse oximetry etc)	17.1%
Diagnosis not classigiable	None	41.1%	Biochemistry	21.7%	None	18.5%	Observation (ECG, pulse oximetry etc)	12.6%
Nothing abnormal detected	None	53.0%	Biochemistry	15.1%	None	33.1%	None	12.1%
Diagnosis missing	None	39.2%	Biochemistry	19.1%	Guidance/advice only	19.3%	Observation (ECG, pulse oximetry etc)	11.7%

Notes: (1) A full list of investigations and treatments are available from the NHS Digital HES Data Dictionary (Accident and Emergency): https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary; (2) First ED investigation/treatment contains the first recorded investigation/treatment code for a specific ED visit; (3) Subsequent investigations/treatments combined information across all other investigation/treatment codes in a specific ED visit (up to 12 investigations and 8 treatments).

Table C2: *The most common ED invesigations and treatments, admitted patients only*

	Most common ED investigations				Most common ED treatments			
ED diagnosis	First investigation	% of patients	Subsequent investigation	% of patients	First treatment	% of patients	Subsequent treatment	% of patients
Laceration	X-ray plain film	36.7%	Haematology	25.9%	Observation (ECG, pulse oximetry etc)	11.6%	Observation (ECG, pulse oximetry etc)	16.2%
Contusion/abrasion	X-ray plain film	50.0%	Haematology	37.1%	Observation (ECG, pulse oximetry etc)	14.1%	Observation (ECG, pulse oximetry etc)	19.3%
Soft tissue inflammation	X-ray plain film	40.8%	Biochemistry	40.0%	Intravenous cannula	15.7%	Observation (ECG, pulse oximetry etc)	18.5%
Head injury	CT scan	26.7%	Haematology	27.9%	Observation (ECG, pulse oximetry etc)	25.1%	Observation (ECG, pulse oximetry etc)	22.2%
Joint injury/fracture	X-ray plain film	72.3%	Haematology	39.4%	Intravenous cannula	16.6%	Intravenous cannula	18.8%
Sprain/ligament injury	X-ray plain film	56.8%	Biochemistry	30.6%	Medication administered	19.3%	Recording vital signs	17.0%
Muscle/tendon injury	X-ray plain film	56.5%	Haematology	34.7%	Medication administered	20.5%	Recording vital signs	19.8%
Nerve injury	ECG	36.1%	Biochemistry	58.1%	Recording vital signs	59.1%	Intravenous cannula	47.7%
Vascular injury	X-ray plain film	18.9%	Biochemistry	40.8%	Intravenous cannula	21.0%	Observation (ECG, pulse oximetry etc)	22.9%
Burns and scalds	Haematology	48.5%	ECG	37.0%	Other parenteral drugs	33.0%	Medication administered	9.6%
Electric shock	Haematology	27.0%	Haematology	19.4%	Guidance/advice only	16.9%	Medication administered	16.4%
Foreign body	X-ray plain film	54.2%	Haematology	19.6%	Observation (ECG, pulse oximetry etc)	16.4%	Observation (ECG, pulse oximetry etc)	15.6%
Bites/stings	X-ray plain film	34.5%	Haematology	24.2%	Intravenous cannula	14.4%	Intravenous cannula	16.4%
Poisoning (inc overdose)	Haematology	29.4%	Biochemistry	41.5%	Observation (ECG, pulse oximetry etc)	23.1%	Observation (ECG, pulse oximetry etc)	22.0%
Near drowning	X-ray plain film	56.0%	Haematology	42.0%	Intravenous cannula	18.8%	Observation (ECG, pulse oximetry etc)	22.8%
Visceral injury	X-ray plain film	28.9%	Biochemistry	50.8%	Intravenous cannula	18.2%	Observation (ECG, pulse oximetry etc)	35.6%
Infectious disease	X-ray plain film	33.7%	Biochemistry	45.4%	Other parenteral drugs	17.4%	Observation (ECG, pulse oximetry etc)	23.3%
Local infection	X-ray plain film	32.2%	Biochemistry	41.1%	Intravenous cannula	21.4%	Observation (ECG, pulse oximetry etc)	24.1%
Septicaemia	X-ray plain film	50.8%	Haematology	57.2%	Intravenous cannula	25.8%	Intravenous cannula	31.6%
Cardiac	X-ray plain film	47.4%	Haematology	57.7%	Intravenous cannula	24.9%	Observation (ECG, pulse oximetry etc)	24.3%
Cerebro-vascular	CT scan	36.4%	Haematology	54.0%	Intravenous cannula	29.4%	Observation (ECG, pulse oximetry etc)	26.0%
Other vascular	X-ray plain film	27.9%	Haematology	44.8%	Intravenous cannula	19.0%	Observation (ECG, pulse oximetry etc)	25.0%
Haematological	X-ray plain film	32.5%	Haematology	48.6%	Intravenous cannula	24.5%	Observation (ECG, pulse oximetry etc)	25.3%
Central nervous system	X-ray plain film	21.8%	Haematology	46.9%	Observation (ECG, pulse oximetry etc)	22.0%	Observation (ECG, pulse oximetry etc)	22.8%
Respiratory	X-ray plain film	51.3%	Haematology	48.7%	Intravenous cannula	18.4%	Observation (ECG, pulse oximetry etc)	23.8%
Gastrointestinal	X-ray plain film	32.6%	Biochemistry	46.7%	Intravenous cannula	23.9%	Intravenous cannula	23.6%
Urological	X-ray plain film	24.1%	Biochemistry	48.7%	Intravenous cannula	21.3%	Observation (ECG, pulse oximetry etc)	24.8%
Obstetric	Haematology	28.0%	Biochemistry	35.7%	Observation (ECG, pulse oximetry etc)	24.4%	Observation (ECG, pulse oximetry etc)	23.0%
Gynaecological	Haematology	33.8%	Biochemistry	44.0%	Intravenous cannula	21.5%	Observation (ECG, pulse oximetry etc)	18.1%
Diabetes and endocrine	X-ray plain film	33.6%	Biochemistry	51.0%	Intravenous cannula	27.2%	Intravenous cannula	24.1%
Dermatological	None	25.9%	Biochemistry	34.2%	Observation (ECG, pulse oximetry etc)	17.4%	Recording vital signs	23.1%
Allergy (inc anaphylaxis)	None	31.9%	Biochemistry	30.2%	Observation (ECG, pulse oximetry etc)	16.6%	Observation (ECG, pulse oximetry etc)	24.1%
Facio-maxillary conditions	X-ray plain film	37.7%	Biochemistry	32.8%	Intravenous cannula	17.9%	Observation (ECG, pulse oximetry etc)	29.0%
ENT	Haematology	29.8%	Biochemistry	34.9%	Intravenous cannula	18.6%	Observation (ECG, pulse oximetry etc)	20.4%
Psychiatric	None	22.7%	Biochemistry	40.3%	Observation (ECG, pulse oximetry etc)	23.4%	Observation (ECG, pulse oximetry etc)	19.3%
Ophthalmological	None	23.7%	Biochemistry	30.3%	Observation (ECG, pulse oximetry etc)	17.1%	Observation (ECG, pulse oximetry etc)	27.1%
Social problems	X-ray plain film	34.1%	Biochemistry	46.2%	Observation (ECG, pulse oximetry etc)	23.1%	Observation (ECG, pulse oximetry etc)	20.2%
Diagnosis not classigiable	X-ray plain film	35.4%	Biochemistry	45.0%	Observation (ECG, pulse oximetry etc)	21.5%	Observation (ECG, pulse oximetry etc)	22.1%
Nothing abnormal detected	X-ray plain film	29.8%	Biochemistry	38.2%	None	31.8%	Recording vital signs	21.1%
Diagnosis missing	X-ray plain film	34.5%	Biochemistry	48.1%	Recording vital signs	18.6%	Intravenous cannula	21.9%

Notes: (1) Includes only ED visits which resulted in an inpatient admission; (2) A full list of investigations and treatments are available from the NHS Digital HES Data Dictionary (Accident and Emergency): https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary; (3) First ED investigation/treatment contains the first recorded investigation/treatment code for a specific ED visit; (4) Subsequent investigations/treatments combined information across all other investigation/treatment codes in a specific ED visit (up to 12 investigations and 8 treatments).

Table C3: *The most common inpatient procedures*

ED diagnosis	Most common inpatient procedures			
	First procedure	% of patients	Subsequent procedure	% of patients
Laceration	Suture of skin	17.4%	Debridement/cleaning of skin/wound	48.6%
Contusion/abrasion	CT / MRI scan of head, spine or CNT	33.2%	Treatment/examination of pelvis or spine	36.6%
Soft tissue inflammation	CT / MRI scan (site not specified)	12.3%	Treatment/examination of pelvis or spine	22.4%
Head injury	CT / MRI scan of head, spine or CNT	67.8%	CT / MRI scan of head, spine or CNT	49.2%
Joint injury/fracture	Closed reduction of fracture	17.3%	Debridement/cleaning of skin/wound	13.9%
Sprain/ligament injury	CT / MRI scan of head, spine or CNT	23.8%	Treatment/examination of pelvis or spine	23.9%
Muscle/tendon injury	Primary repair of tendon	21.4%	Debridement/cleaning of skin/wound	29.8%
Nerve injury	CT / MRI scan of head, spine or CNT	22.0%	Treatment/examination of pelvis or spine	21.0%
Vascular injury	CT / MRI scan (site not specified)	12.5%	CT / MRI scan (site not specified)	22.2%
Burns and scalds	CT / MRI scan of head, spine or CNT	18.9%	Debridement/cleaning of skin/wound	21.5%
Electric shock	CT / MRI scan of head, spine or CNT	28.8%	Debridement/cleaning of skin/wound	23.1%
Foreign body	Removal of inorganic substance from the skin	17.0%	Debridement/cleaning of skin/wound	28.4%
Bites/stings	Debridement/cleaning of skin/wound	28.1%	Debridement/cleaning of skin/wound	72.3%
Poisoning (inc overdose)	CT / MRI scan of head, spine or CNT	31.3%	CT / MRI scan of head, spine or CNT	35.6%
Near drowning	Ventilation	19.0%	CT / MRI scan of head, spine or CNT	50.0%
Visceral injury	CT / MRI scan (site not specified)	26.5%	Treatment/examination of pelvis or spine	62.3%
Infectious disease	CT / MRI scan of head, spine or CNT	22.3%	CT / MRI scan of head, spine or CNT	32.1%
Local infection	Drainage/incision of lesion or skin	23.1%	Debridement/cleaning of skin/wound	21.8%
Septicaemia	CT / MRI scan of head, spine or CNT	19.0%	Treatment/examination of pelvis or spine	34.5%
Cardiac	Echocardiography	25.3%	Echocardiography	46.6%
Cerebro-vascular	CT / MRI scan of head, spine or CNT	76.5%	CT / MRI scan of head, spine or CNT	73.0%
Other vascular	CT / MRI scan of head, spine or CNT	31.2%	CT / MRI scan of head, spine or CNT	23.0%
Haematological	Blood transfusion (inc blood stem cell transplant)	15.1%	Treatment/examination of pelvis or spine	41.2%
Central nervous system	CT / MRI scan of head, spine or CNT	70.9%	CT / MRI scan of head, spine or CNT	67.8%
Respiratory	Ventilation	23.1%	Treatment/examination of pelvis or spine	24.5%
Gastrointestinal	CT / MRI scan (site not specified)	29.2%	Treatment/examination of pelvis or spine	68.3%
Urological	CT / MRI scan (site not specified)	30.7%	Treatment/examination of pelvis or spine	40.3%
Obstetric	Aspiration/extraction of products of conception from uterus	21.5%	Examination of female genital tract	25.6%
Gynaecological	Examination of female genital tract	22.9%	Treatment/examination of pelvis or spine	28.5%
Diabetes and endocrine	CT / MRI scan of head, spine or CNT	26.0%	Treatment/examination of pelvis or spine	39.4%
Dermatological	Drainage/incision of lesion or skin	17.7%	Treatment/examination of pelvis or spine	29.5%
Allergy (inc anaphylaxis)	CT / MRI scan of head, spine or CNT	16.4%	Treatment/examination of pelvis or spine	22.9%
Facio-maxillary conditions	Operations on tooth and surrounding area	20.9%	Extraction of teeth	30.5%
ENT	Packing of cavity of nose	28.2%	Packing of cavity of nose	70.4%
Psychiatric	CT / MRI scan of head, spine or CNT	32.9%	CT / MRI scan of head, spine or CNT	25.6%
Ophthalmological	CT / MRI scan of head, spine or CNT	35.8%	CT / MRI scan of head, spine or CNT	24.6%
Social problems	CT / MRI scan of head, spine or CNT	52.4%	CT / MRI scan of head, spine or CNT	34.4%
Diagnosis not classigiable	CT / MRI scan of head, spine or CNT	26.6%	Treatment/examination of pelvis or spine	32.0%
Nothing abnormal detected	CT / MRI scan of head, spine or CNT	27.0%	Treatment/examination of pelvis or spine	36.0%
Diagnosis missing	CT / MRI scan of head, spine or CNT	23.3%	Treatment/examination of pelvis or spine	31.5%

Notes: (1) Inpatient procedures are recorded using OPCS4.8 codes. For a mapping of OPCS4.8 codes to procedures,

B.4 Further details of the empirical methodology

In Section 3.4, we set out our approach for estimating the impact on treatment decisions and mortality outcomes. In this appendix, we provide further details of this methodology. We first introduce some notation to define the different channels through which the target can affect outcomes and then show how we identify and estimate the 'distortion effects' of the target.

B.4.1 Composition and distortion effects

In a potential outcomes framework, let y_t be an outcome (treatment decision or mortality outcome) and w_t be the wait time in regime $t \in \{0, 1\}$. We then define two conditional expectation functions. The first is $E[y_t \mid w_t]$, which is the expected outcome conditional on the wait time. This allows us to express average outcomes (either in the targeted or non-targeted regime) for groups of patients located in different parts of the wait time distribution. For example, the observed data can be written as $E[y_1 \mid w_1]$. It is also possible to think about $E[y_0 \mid w_0]$, outcomes in the absence of the target, and combinations such as $E[y_0 \mid w_1]$ which are the outcomes in the non-targeted regime for patients at certain points of the wait time distribution in the targeted regime.

We also define $E[y_t \mid w_1, w_0]$, which is the expected outcome for patients with wait time w_1 in the targeted regime and wait time w_0 in the non-targeted regime. This notation allows us to denote outcomes for groups of individuals that have had a change in wait time due to the target. For example, $E[y_t \mid w^- < w_1 \leq w^*, w^* < w_0 < w^+]$ is the expected outcome for post-threshold movers. Since we will repeatedly refer to this and other related groups, we abbreviate these conditioning inequalities in the following way: $E[y_t \mid \underline{w}_1^-, \underline{w}_0^+]$.

Using this notation we can decompose the observed outcomes in the pre-threshold period.[3] Note that, from the wait time analysis, we know that the target causes a number of patients to shift from the post-threshold to the pre-threshold period ('post-threshold movers'). So with the target, outcomes in the pre-threshold period are a weighted-average of pre-threshold non-movers and post-threshold movers. Abbreviating the pre-threshold period as $\underline{w}_1^-$, outcomes can be written as

$$E[y_1 \mid \underline{w}_1^-] = \rho E[y_1 \mid \underline{w}_1^-, \underline{w}_0^-] + (1-\rho)E[y_1 \mid \underline{w}_1^-, \underline{w}_0^+], \tag{B.4.1}$$

where $\rho \equiv \big[F_0(w^*) - F_0(w^-)\big] / \big[F_1(w^*) - F_1(w^-)\big]$ and F_t is the *cdf* of wait times. The parameter ρ is defined by the observed and counterfactual wait time distributions, where ρ is the proportion of pre-threshold non-movers and $1-\rho$ is the proportion of post-threshold movers.

The composition and distortion effects are then defined as follows.

Definition 1 (Composition effect)**.** *The composition effect is the change in expected outcomes conditional on the wait time that occurs in the pre-threshold period because the target shifts some patients into this period from the post-threshold period:*

$$\Delta_C \equiv \rho\big(E[y_0 \mid \underline{w}_1^-, \underline{w}_0^-] - E[y_0 \mid \underline{w}_1^-, \underline{w}_0^-]\big) + (1-\rho)\big(E[y_0 \mid \underline{w}_1^-, \underline{w}_0^+] - E[y_0 \mid \underline{w}_1^-, \underline{w}_0^-]\big) \tag{B.4.2}$$

$$= (1-\rho)\big(E[y_0 \mid \underline{w}_1^-, \underline{w}_0^+] - E[y_0 \mid \underline{w}_1^-, \underline{w}_0^-]\big). \tag{B.4.3}$$

Definition 2 (Distortion effect)**.** *The distortion effect is the change in expected outcomes conditional on the wait time that occurs in the pre-threshold period be-*

[3]We refer to parts of the wait time distributions as 'periods'. The pre-threshold period includes all patients with waiting times between 180 (w^-) and 240 (w^*) minutes. The post-threshold period includes all patients with waiting times between 240 minutes and 400 minutes (w^+).

cause the target has a direct effect on the outcomes in each regime:

$$\Delta_D \equiv \rho\big(E[y_1 \mid \underline{w}_1^-, \underline{w}_0^-] - E[y_0 \mid \underline{w}_1^-, \underline{w}_0^-]\big) + (1-\rho)\big(E[y_1 \mid \underline{w}_1^-, \underline{w}_0^+] - E[y_0 \mid \underline{w}_1^-, \underline{w}_0^+]\big). \tag{B.4.4}$$

Note that the distortion effects may impact both pre-target non-movers (first term) or post-target movers (second term). With these definitions the observed outcomes in the pre-threshold period can be written as

$$\underbrace{E[y_1 \mid \underline{w}_1^-]}_{\text{Targeted regime (observed)}} = \underbrace{E[y_0 \mid \underline{w}_0^-]}_{\text{Non-targeted regime}} + \underbrace{\Delta_C}_{\text{Composition effect}} + \underbrace{\Delta_D}_{\text{Distortion effect}} \tag{B.4.5}$$

which can be verified by substituting in Equations (B.4.1), (B.4.3) and (B.4.4) and rewriting the non-targeted regime outcome as $E[y_0 \mid \underline{w}_1^-, \underline{w}_0^-]$.

B.4.2 Identification of the distortion effect

To identify the distortion effect we make use of the following definition.

Definition 3 (Composition-adjusted counterfactual)**.** *The composition-adjusted counterfactual (CAC) is the outcomes from the non-targeted regime in the pre-threshold period that would occur in the presence of the composition effect only:*

$$E[y_0 \mid \underline{w}_1^-] \equiv E[y_0 \mid \underline{w}_0^-] + \Delta_C \tag{B.4.6}$$

$$= \rho E[y_0 \mid \underline{w}_1^-, \underline{w}_0^-] + (1-\rho) E[y_0 \mid \underline{w}_1^-, \underline{w}_0^+]. \tag{B.4.7}$$

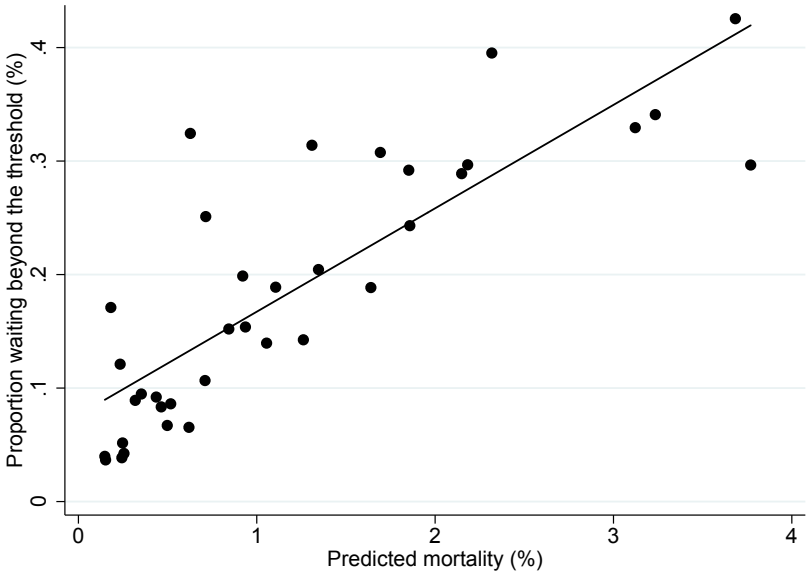
where the second line follows from the definition of Δ_C*.*

With this definition it is straightforward to show that the distortion effect is identified as the difference between the observed data and the CAC: $\Delta_D = E[y_1 \mid \underline{w}_1^-] - E[y_0 \mid \underline{w}_1^-]$. Moreover, Equation (B.4.7) shows the CAC can be constructed as a weighted average of the counterfactual outcomes for two groups, the pre-threshold non-movers and the post-threshold movers, where the weights can be constructed from the observed and counterfactual wait time distributions.

B.5 Further exploration of heterogeneity to identify mechanisms

Figure E1 graphs the proportion of patients hitting the wait time target (in the counterfactual wait time distribution) against the severity of the diagnosis. Severity is measured by mean predicted 30-day mortality for patients within each diagnosis. It shows that that the probability of hitting 240 minutes is much higher for the most severe diagnoses.

Figure E1: *Proportion wait beyond the threshold vs. predicted mortality by diagnosis groups*



Notes: (1) Each data point corresponds to a diagnosis group average; (2) Proportion waiting beyond the threshold defined using the counterfactual distribution of wait times; (3) Predicted mortality defined using a regression of 30-day in-hospital mortality on past-CCI and a fully interacted set of age, gender, and ambulance arrival fixed effects.

Figure E2 graphically examines the relationship between wait time effects and distortion effects for admissions and 30-day mortality across diagnoses groups. Panel A shows that higher severity diagnoses have larger wait time effects, while Panel B shows that the effects of the target on hospital admissions is no higher for more severe diagnoses.

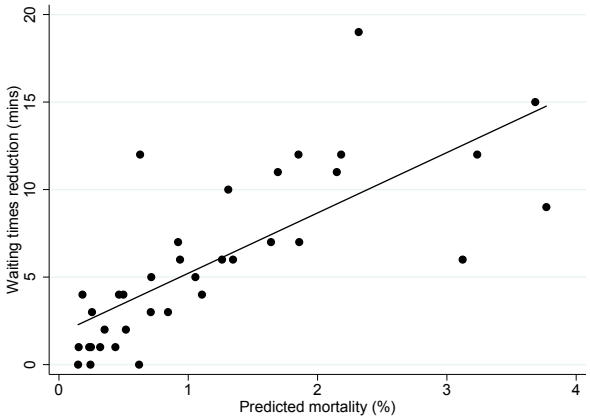
Panel C shows the differential treatment effect on mortality by diagnosis category, where black circles correspond to actual mortality outcomes and red triangles correspond to predicted mortality outcomes. The y-axis shows the absolute value of mortality reduction, so that a larger value means a larger mortality reduction. Looking at the black circles, there is a clear upward slope showing that the mortality effect of the four-hour target is strongest for the most severe diagnoses. To ensure that selection is not driving our result, the graph also repeats this exercise for predicted mortality. If our assumption of no-selection (Assumption 5) holds, these effects should not be statistically different from zero. The red triangles shows that this is indeed the case, with all estimates clustered around zero and no systematic relationship between the effects of the target on predicted mortality and the severity of the diagnosis.

Figure E3 presents the results of the second heterogeneity test examining the relationship between wait time reductions and distortion effects for admissions and 30-day mortality across crowding groups. The figure shows the results for these observations, ranked from least crowded to most crowded. Panel A shows that inpatient crowding has a weak, positive relationship with wait times. Panel B shows a strong, negative relationship between crowded inpatient departments and smaller increases in admission. So this source of heterogeneity gives the opposite results of what we saw for severity: a small effect on wait times and a large effect on admissions. Therefore, if our earlier supposition is correct that it is wait times and not admissions that drives our mortality effects, we should see little differential impact on mortality across these groups.

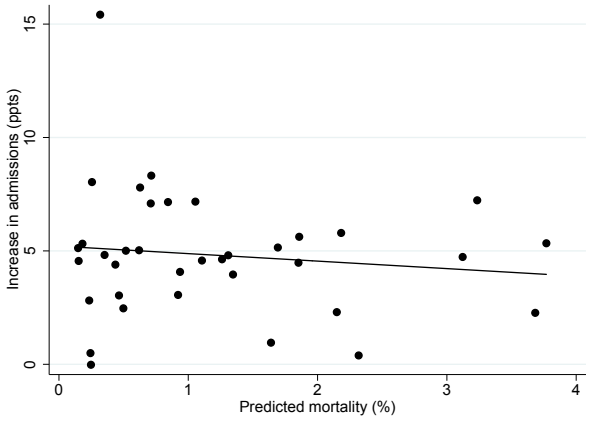
In fact, that is exactly what we see in Panel C in the black circles: there is no significant relationship between the degree of inpatient crowding and the estimated mortality effect. As in Figure E2c, we repeat this analysis with estimated reductions in predicted mortality (which should be unaffected by the target once we adjust for the composition of patients) to show that these results are not driven by selection. The red triangles show that the predicted mortality effects are

Figure E2: *Estimated effects of the target vs. predicted mortality by diagnosis groups*

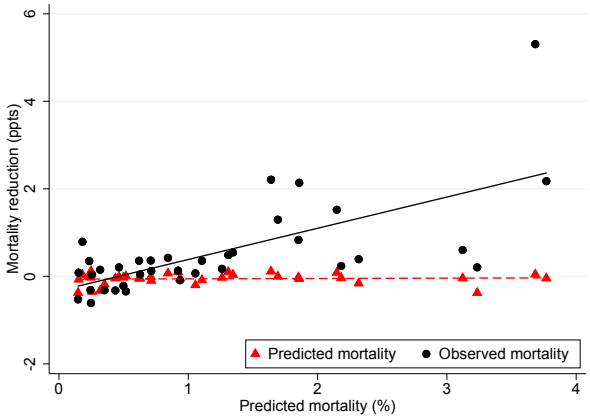
(a) *Wait times reductions*



(b) *Admissions increases*



(c) *Mortality reductions*



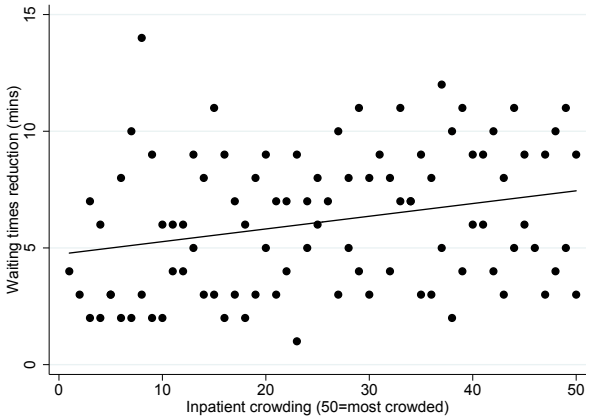
Notes: (1) Each data point corresponds to a diagnosis group average; (2) Predicted mortality defined using a regression of 30-day in-hospital mortality on past-CCI and a fully interacted set of age, gender, and ambulance arrival fixed effects.

again close to zero. There is a positive relationship between predicted mortality reductions and inpatient crowding but this is small in magnitude.[4]

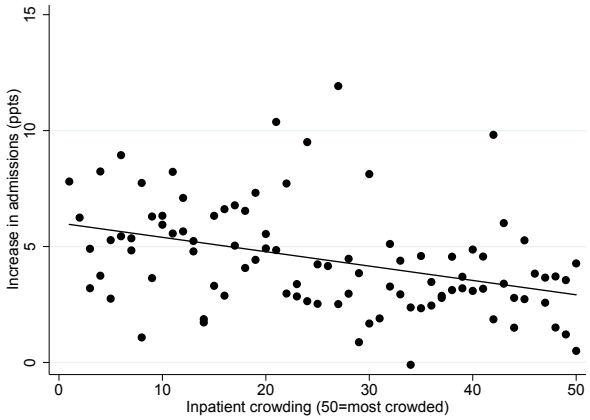
[4]This means that our results may actually understate the mortality reductions in the most crowded periods. Given that these periods are also those with the smallest increases in admissions, this would strengthen the conclusion that mortality reductions are associated with reductions in wait times and not additional admissions.

Figure E3: *Estimated effects of the target vs. inpatient crowding by crowding-severity groups*

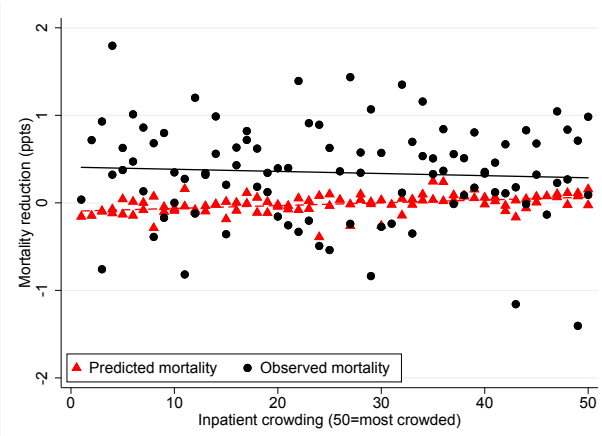
(a) *Wait times reductions*



(b) *Admissions increases*



(c) *Mortality reductions*



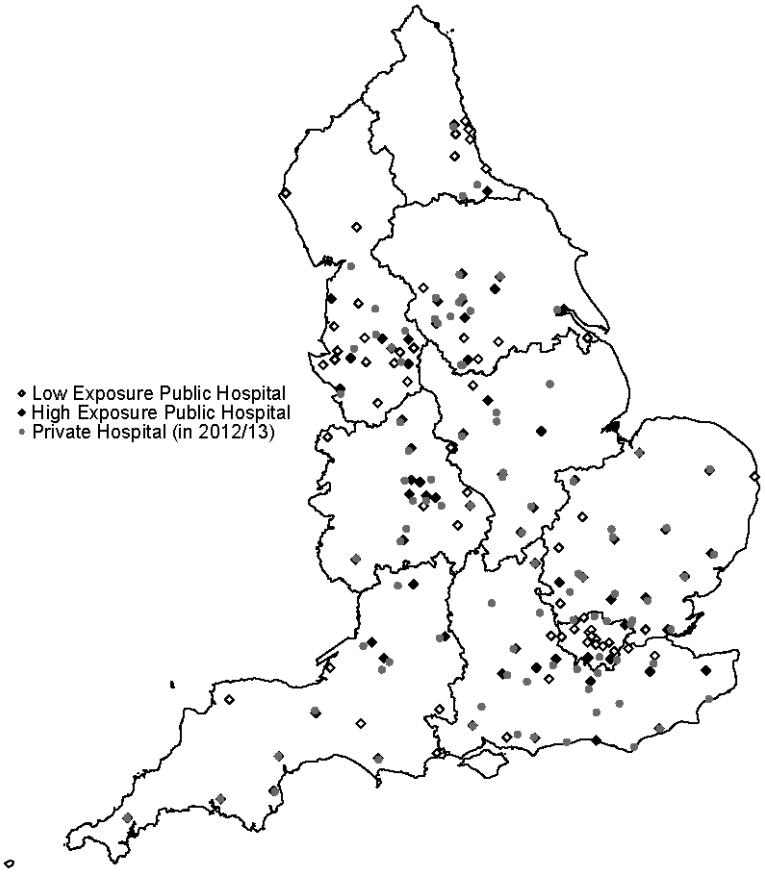
Notes: (1) Each data point corresponds to an inpatient crowding-severity group average; (2) Inpatient crowding groups defined according to the number of inpatients treated per hospital-day, which we then use to split into 50 quantiles; (3) Severity is defined as diagnoses with a mean 30-day mortality rate above the mean overall 30-day mortality rate; (4) Predicted mortality defined using a regression of 30-day in-hospital mortality on past-CCI and a fully interacted set of age, gender, and ambulance arrival fixed effects.

Appendix C

Appendix to Chapter 4

C.1 Additional Figures and Tables

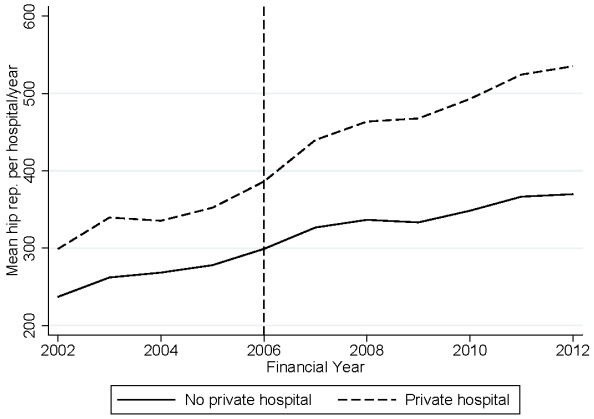
Figure A1: *Public and private hospitals treating publicly-funded hip replacements in 2012/13, by exposure status*



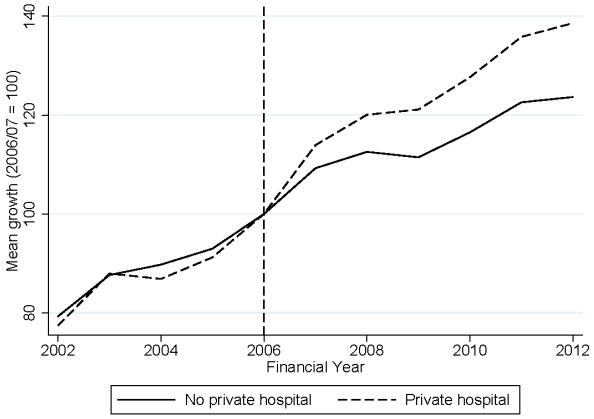
Notes: (1) Unfilled diamonds show public hospitals located in low exposure markets; (2) Filled diamonds show public hospitals located in high exposure markets; (3) Circles show private hospitals that treat publicly funded hip replacement patients in 2012/13.

Figure A2: *Mean publicly funded hip replacements per hospital market, by pre-reform private hospital exposure*

(a) *Levels*



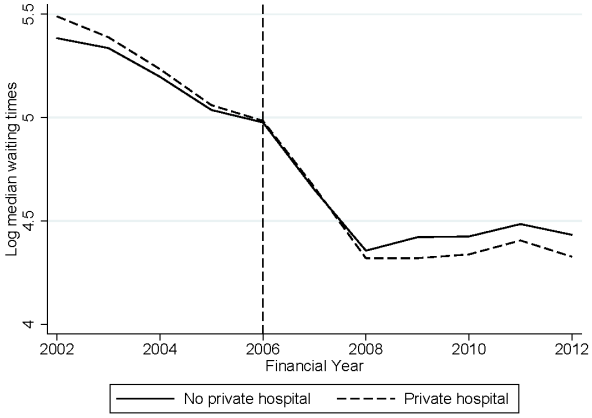
(b) *Growth (2006/07=100)*



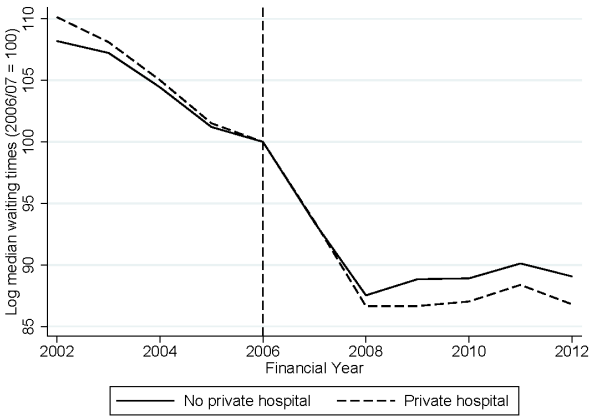
Notes: (1) Volumes include all publicly funded hip replacements (as defined in Figure 1) regardless of whether they were conducted by public hospitals or private providers (private hospitals or Independent Sector Treatment Centres); (2) Patients are allocated to their nearest hospital regardless of where the surgery actually takes place; (3) Private hospital areas are those which contained a private hospital in 2004; (4) In panel B growth figures are relative to 100 in 2006/07; (5) The vertical line (2006) denotes the year in which private hospitals first entered the market.

Figure A3: *Log median waiting times for publicly funded hip replacements, by pre-reform private hospital exposure*

(a) *Levels*



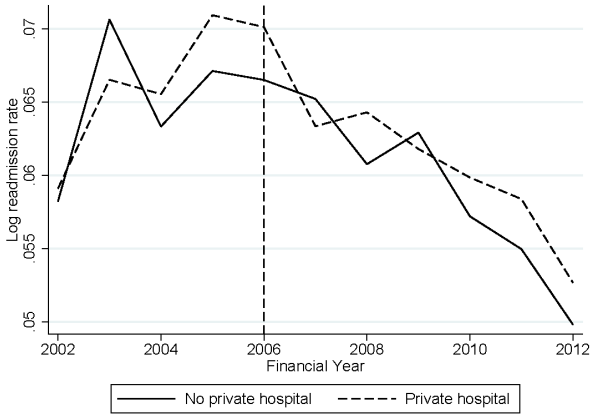
(b) *Growth (2006/07=100)*



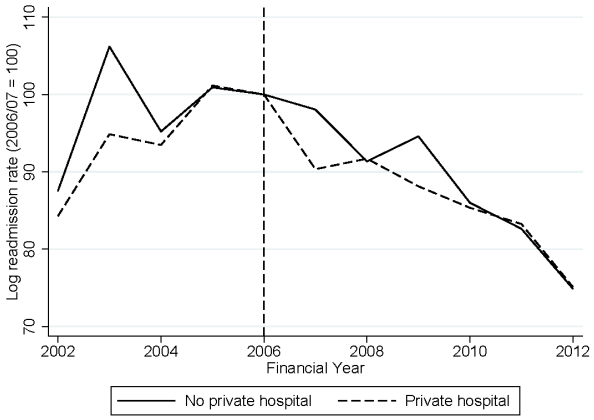
Notes: (1) Changes relative to 100 in 2006/07; (2) Waiting times measures the median number of days between the decision to admit a patient for a hip replacement and their admission date; (3) Private hospital areas are those which contained a private hospital in 2004; (4) The vertical line (2006) denotes the year in which private hospitals first entered the market.

Figure A4: *Log 30-day emergency readmissions rates for publicly funded hip replacements, by pre-reform private hospital exposure*

(a) *Levels*



(b) *Growth (2006/07=100)*



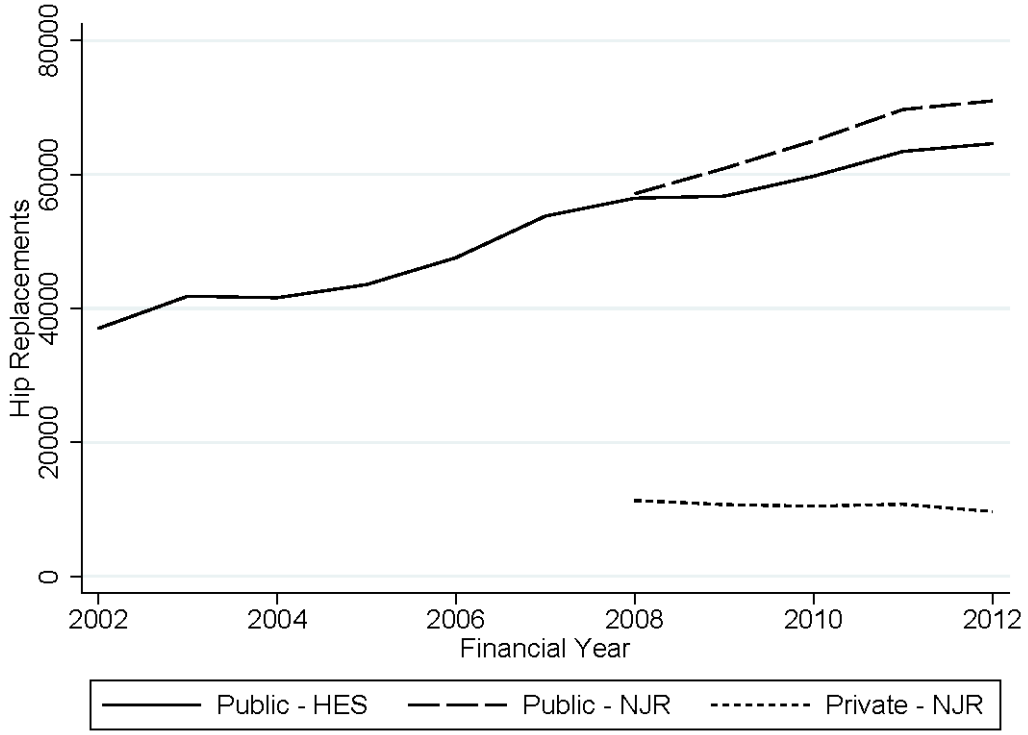
Notes: (1) Changes relative to 100 in 2006/07; (2) Emergency readmissions measures the proportion of patients who experience an emergency inpatient readmission within 30 days of discharge after a publicly funded elective hip replacement; (3) Private hospital areas are those which contained a private hospital in 2004; (4) The vertical line (2006) denotes the year in which private hospitals first entered the market.

Table A1: *Estimates of trends in volumes, waiting times and readmissions, by pre-reform private hospital exposure*

	Volumes		log (med wait)	log (readmit)
	All	NHS only	All	All
	(1)	(2)	(3)	(4)
Trends in pre-reform period				
Priv. hospital in 2004 * 2002/03	7.87	-4.71	0.08	-0.002
	(11.23)	(11.72)	(0.06)	(0.005)
Priv. hospital in 2004 * 2003/04	12.93	6.68	0.03	-0.007
	(10.27)	(9.26)	(0.05)	(0.006)
Priv. hospital in 2004 * 2004/05	2.47	-2.29	0.01	-0.001
	(9.47	(9.57)	(0.04)	(0.004)
Priv. hospital in 2004 * 2005/06	-9.91	-6.98	0.01	0.001
	(8.15)	(7.64)	(0.02)	(0.004)
Trends in post-reform period				
Priv. hospital in 2004 * 2007/08	15.12*	2.02	0.01	-0.005
	(9.04)	(8.63)	(0.03)	(0.004)
Priv. hospital in 2004 * 2008/09	20.65*	0.74	-0.02	0.002
	(11.06)	(10.41)	(0.05)	(0.005)
Priv. hospital in 2004 * 2009/10	22.10	-6.11	-0.09*	-0.003
	(14.29)	(13.54)	(0.05)	(0.005)
Priv. hospital in 2004 * 2010/11	27.27	-10.31	-0.07	0.001
	(16.67)	(16.63)	(0.05)	(0.005)
Priv. hospital in 2004 * 2011/12	44.39**	3.98	-0.06	0.002
	(17.57)	(19.78)	(0.05)	(0.004)
Priv. hospital in 2004 * 2012/13	42.42**	-1.25	-0.07	0.004
	(18.65)	(21.22)	(0.06)	(0.004)
Patient controls	No	No	Yes	Yes
Observations	1,430	1,430	1,430	1,430
R-Squared	0.753	0.504	0.864	0.146

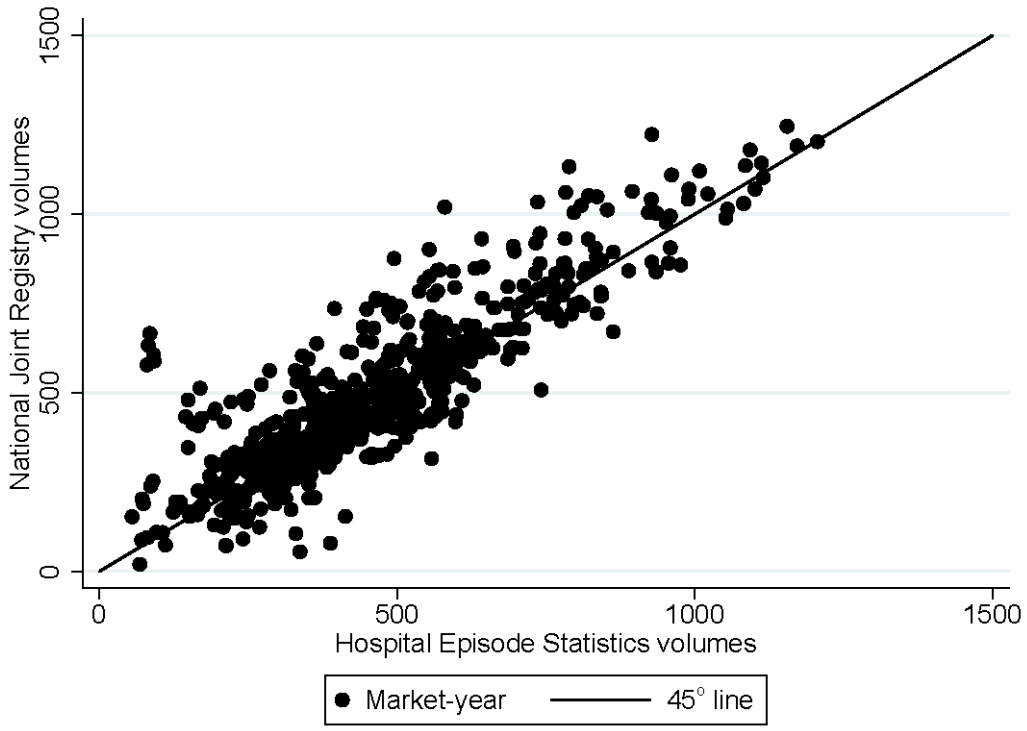
Notes: (1) 'Priv. hospital in 2004' is a dummy variable that takes the value of one if a private hospital was located in the hospital market in 2004; (2) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, and a full set of year and hospital market fixed effects; (3) Patient controls include the mean age, gender and Charlson Comorbidity Index score of patients undergoing an elective hip replacement; (4) There are 130 hospitals; (5) All specifications clustered at the hospital market level, *** p< 0.01, ** p<0.05, * p<0.1.

Figure A5: *Hospital market annual volumes of hip replacements, by dataset and funding stream*



Notes: (1) Public-HES reports all publicly funded hip replacements recorded in the Hospital Episode Statistics; (2) Public-NJR reports all publicly funded hip replacements recorded in the National Joint Registry; (3) Private-NJR reports all privately funded hip replacements recorded in the National Joint Registry.

Figure A6: *Hospital market annual volumes of publicly funded hip replacements recorded in the National Joint Registry and Hospital Episode Statistics, 2008/09 to 2012/13*



Notes: (1) Each observation is a hospital market and year combination (N=650).

C.2 The relationship between ISTCs and public patient outcomes

Appendix Table B1 shows the coefficients associated with the interaction between a dummy variable that takes the value of one when an ISTC treated public patients in the market in 2012/13, and a dummy variable that takes the value of one in years when ISTCs were allowed to operate in the market (2005/06) onwards. These coefficients are from the same regressions displayed in columns 2, 5 and 8 of Table 4.3.

Table B1: *Two stage least squares estimates of the association between Independent Sector Treatment Centre exposure and selected outcomes, 2002/03 to 2012/13*

	Admissions	ln(med wait)	ln(readmissions)
	(1)	(2)	(3)
Ind. Sec. Treat. Centre	28.29**	-0.238***	-0.000
	(12.36)	(0.055)	(0.004)
Observations	1,430	1,430	1,430
R-squared	0.687	0.129	0.215

Notes: (1) 'Ind Sec Treat Centre' is a dummy variable that takes the value of one if an Independent Sector Treatment Centre located in the hospital market treats public funded hip replacement patients in 2012/13, interacted with another dummy variable that takes the value of one in all years from 2005/06 (the period when Independent Sector Treatment Centres could operate in the public market); (2) Reported coefficients in columns 1,2 and 3 are from the corresponding regressions in columns 2,5 and 8 of Table 4.3 respectively (see Table 4.3 for details of the specifications); (3) All specifications clustered at the hospital market level, *** p< 0.01, ** p<0.05, * p<0.1.

Column 1 shows the estimated association between ISTC presence and the number of admissions for publicly funded hip replacements. The coefficient is positive and statistically significantly different from zero. It is slightly smaller than the estimated impact of private hospital presence on publicly funded volumes. Column 2 shows the estimated association with waiting times. The coefficient is negative and statistically significant at the 1% level. This suggests that waiting times fell more quickly in areas where ISTCs were established than in areas where

they were not. However, ISTCs were intended to be located in areas with high waiting times at the beginning of the period, and so this coefficient may partly include the impact of other measures taken to reduce waiting times in the local area. In column 3, the dependent variable is 30-day emergency readmission rate. As with private hospital entry, the coefficient is small in size and not statistically significantly different from zero.

C.3 Alternative definitions of markets and exposure

Table C1 shows the results from a specification using markets defined at the MSOA level, and are discussed in the main text.

For each MSOA, we measure exposure based on whether a private hospital based within a certain distance from the MSOA centroid treated public patients in 2012/13. This is again interacted with the *post* dummy variable to create a time-varying measure of treatment. These catchment areas are defined using the distribution of distance travelled by publicly funded hip replacement patients living in the MSOA between 2001 and 2004. We calculate the 25th, 50th and 75th percentile of the distribution of these distances and create exposure measures based on these. In all cases, the unit of observation is now the MSOA level, with outcomes recorded for patients living in each MSOA in each year.

We also build corresponding instruments using this method. Specifically, we create an indicator of whether a private hospital was located within the catchment area of each MSOA (using each percentile) in 2004. We then interact this indicator with the *post* dummy. Private hospital entry into the public market is then instrumented using pre-existing private hospital presence in a similar approach to our baseline specification.

This specification includes all the same control variables as in our baseline analysis. We also include an interaction between year dummies and the nearest NHS hospital trust to control for changing trends in outcomes in the wider local area.

Table C1: *Two stage least squares estimates of the impact of private hospital exposure using overlapping MSOA markets*

	Market outcomes				Mean number of comorbidities		
	Volumes	NHS Volumes	log(med wait)	log(readmit)	Count	% none	% 2+
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A: 25th percentile							
Pub. funded priv. hosp.	0.307***	0.0502	-0.0430***	0.00211	-0.0429**	0.0109**	-0.0149***
	(0.0722)	(0.0685)	(0.00914)	(0.00217)	(0.0218)	(0.00478)	(0.00502)
Observations	73,039	73,039	71,773	71,773	71,773	71,773	71,773
B: 50th percentile							
Pub. funded priv. hosp.	0.307***	0.0733	-0.0234***	0.00318	-0.0469**	0.0121**	-0.0162***
	(0.0715)	(0.0678)	(0.00905)	(0.00215)	(0.0216)	(0.00474)	(0.00497)
Observations	73,039	73,039	71,773	71,773	71,773	71,773	71,773
C: 75th percentile							
Pub. funded priv. hosp.	0.296***	0.0603	-0.0191**	0.00199	-0.0564***	0.0149***	-0.0164***
	(0.0705)	(0.0669)	(0.00893)	(0.00212)	(0.0213)	(0.00467)	(0.00490)
Observations	73,039	73,039	71,773	71,773	71,773	71,773	71,773

Notes: (1) Unit of observation is the MSOA (all outcomes measured at this level); (2) 'Pub. funded priv. hosp.' is a dummy variable equal to 1 if a private hospital located within the market radius treated public patients in 2012/13, interacted with a dummy variable equal to 1 in the post-reform period; (3) Panel A uses the 25th percentile of distance travelled to define market radius (B uses the 50th and C uses the 75th percentiles respectively); (4) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, Independent Sector Treatment Centre market presence; a full set of year and MSOA fixed effects; and a time-trend for the nearest public hospital; (5) There are 6,640 MSOAs; (6) All specifications clustered at the MSOA level, *** p< 0.01, ** p<0.05, * p<0.1.

In Table C2 we allow exposure to a private hospital (at the market level) to vary across each year. Using our original market definition as the unit of observation, we estimate the following specification:

$$Y_{mt} = \beta_0 + \beta_1 E_{mt} + \beta_2 X_{mt} + \gamma_m + \lambda_t + \epsilon_{mt} \tag{C1}$$

where E_{mt} takes the value of one if a private hospital that treated public patients was located in market m in year t, and zero otherwise. All other variables are specified as in equation 4.2. This means that, relative to our baseline specification, markets can switch between being high and low exposure areas during the post-reform period. As discussed in the main text, we would expect the estimates to be larger in magnitude using this specification, as the fixed geography method would understate the impact of hospital entry if entry only occurred at the end of the period. This is indeed what Table C2 shows, with the magnitude of estimates roughly doubling in value as compared to our baseline results.

Table C2: *Two stage least squares estimates of the impact of time-varying private hospital exposure*

	Market outcomes				Mean number of comorbidities		
	Volumes	NHS volumes	log(med wait)	log(readmit)	Count	% none	% 2+
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pub. funded priv. hosp. in year t	68.52**	1.619	-0.244*	0.00782	-0.659**	0.154*	-0.164**
	(30.70)	(31.23)	(0.139)	(0.00769)	(0.316)	(0.0785)	(0.0766)
Observations	1,430	1,430	1,430	1,430	1,430	1,430	1,430
R-squared	0.749	0.502	0.848	0.125	0.647	0.541	0.607

Notes: (1) Unit of observation is the hospital market level (all outcomes measured at this level); (2) 'Pub. funded priv. hosp in year t' is a dummy variable that takes the value of one if a private hospital located in the hospital market treats public funded hip replacement patients in that year; (3) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, Indendent Sector Treatment Centre presence; and a full set of year and hospital market fixed effects; (4) First-stage F-stat is 67.6; (5) The number of hospital markets is 130; (6) All specifications clustered at the market level, *** p< 0.01, ** p<0.05, * p<0.1.

In Table C3 we re-estimate the baseline specification with 'high exposure' areas defined as only areas with private hospitals conducting large numbers of publicly funded procedures. We define a 'large' hospital using the volume of publicly-funded knee replacements in 2012/13. Hospitals with knee replacement volumes above the median for in that year (more than 140 procedures) are classified as 'large' hospitals. We then create our 'high exposure' measure using only these large hospitals (markets with small private hospitals are classified as 'low exposure' areas). We then instrument this exposure measure using our original instrument.

Table C3: *Two stage least squares estimates of the impact of large private hospital exposure*

	Market outcomes				Mean number of comorbidities		
	Volumes	NHS volumes	log(med wait)	log(readmit)	Count	% none	% 2+
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Large pub. funded priv. hosp.	58.77**	-1.317	-0.193*	0.00665	-0.585**	0.133**	-0.143**
	(25.53)	(25.83)	(0.112)	(0.00641)	(0.264)	(0.0659)	(0.0642)
First stage F-stat	44.8	44.8	44.8	44.8	44.8	44.8	44.8
Observations	1,430	1,430	1,430	1,430	1,430	1,430	1,430
R-squared	0.748	0.500	0.861	0.124	0.667	0.579	0.642

Notes: (1) Large pub. funded priv. hosp. is a dummy variable that takes the value of 1 if a 'large' private hospital located in the hospital market treats public funded hip replacement patients in 2012/13, interacted with a dummy variable that takes the value of 1 in the post-reform period (2006/07 onwards); (2) Large private hospital is defined as a private hospital conducting more than 140 publicly funded knee replacements in 2012/13; (3) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, Independent Sector Treatment Centre presence, and a full set of year and hospital market fixed effects; (4) First stage F-stat is 44.8; (5) There are 130 hospital markets; (6) All specifications clustered at the hospital market level, *** p< 0.01, ** p<0.05, * p<0.1.

C.4 Alternative specifications exploring heterogeneity by local area deprivation

Table D1 repeats our baseline analysis using the MSOA as the unit of analysis (rather than the constructed markets that we use in Table 4.3). We therefore estimate a version of equation (1) where m is now the MSOA rather than the wider market. High exposure is defined by assigning the treatment status of the wider market to the MSOA (i.e. if the MSOA was in a market where a private hospital entered by 2012/13, then the MSOA is defined as 'high exposure'). We then instrument this with an analogous measure of private hospital presence in 2004.

As expected, the results in Table D1 closely reflect our baseline results (Tables 4.3 and 4.4). This shows an increase in the volume of hip replacements of around 15% of the baseline (2002) level. There are no reductions in NHS volumes (not displayed). There is a reduction of 13% in waiting times. This is now statisticallty significant at the 1% level (rather than the 10% level in our baseline results). There are similarly sized effects on the measures of patient severity.

Table D1: *Two stages least squares estimates of the impact of private hospital exposure, at MSOA level*

	Market outcomes			Mean number of comorbidities		
	Volumes	log (med wait)	log (readmit)	Count	% none	% 2+
	(1)	(2)	(3)	(4)	(5)	(6)
Pub. funded priv. hosp.	1.060***	-0.128***	0.00273	-0.269***	0.0576***	-0.0651***
	(0.107)	(0.0161)	(0.00288)	(0.0348)	(0.00812)	(0.00815)
Number of MSOAs	6,640	6,640	6,640	6,640	6,640	6,640
Observations	73,039	71,773	71,773	73,039	71,773	71,773

Notes: (1) Unit of analysis at the MSOA level; (2) 'Pub. funded priv. hosp.' is a dummy variable that takes the value of one in the post-reform years (2006/07) for all hospital markets where a private hospital located in the market treated public patients in 2012/13; (3) All specifications control for the age-sex profile of the local population, numbers of emergency fractured neck of femur and acute coronary syndrome admissions of residents in the area, house sales and prices, an independent sector treatment centre dummy (ISTC) (equal to one if an ISTC treated public patients in 2012/13) interacted with a dummy variable that takes the value of one from 2005/06 onwards (the first year of ISTC entry), and a full set of year and MSOA fixed effects; (4) There are 6,640 MSOAs hospitals; (5) All specifications clustered at the hospital market level, *** p< 0.01, ** p<0.05, * p<0.1.

Bibliography

Aaronson, D., L. Barrow, and W. Sander (2007). Teachers and student achievement in the chicago public high schools. *Journal of Labor Economics 25*(1), 95–135.

Abowd, J. M. and F. Kramarz (1999). The analysis of labor markets using matched employer-employee data. In O. C. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3, pp. 749–804. Amsterdam: North-Holland.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). High wage workers and low wage firms: negative assorting matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society) 171*(3).

Arora, S., A. Charlesworth, E. Kelly, and G. Stoye (2013, May). Public pay and private provision: the changing landscape of healthcare in the 2000s. Nuffield Trust and Institute for Fiscal Studies Research Report.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review 53*(5), 941–973.

Audit Commission (2008, June). Is the treatment working? progress with the nhs system reform programme. Audit Commission, London.

Bardsley, M. and J. Dixon (2011). Quality of care in independent sector treatment centres. British Medical Journal 343. d6936.

Best, M., J. Cloyne, E. Ilzetzki, and H. J. Kleven (2017, November). Estimating the elasticity of intertemporal substitution using mortgage notches. Working Paper.

Best, M. and H. J. Kleven (2018). Housing market responses to transaction taxes: Evidence from notches and stimulus in the uk. *Review of Economic Studies* (85), 157–193.

Bitler, M., S. Corcoran, T. Domina, and E. Penner (2019). Teacher effects on student achievement and height: A cautionary tale. *NBER Working Paper 26480*.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer employee data. *Econometrica 87*(3), 699–738.

Cannon, C. P., C. M. Gibson, C. T. Lambrew, D. A. Shoultz, D. Levy, W. J. French, J. M. Gore, W. D. Weaver, W. J. Rogers, and A. J. Tiefenbrunn (2000). Relationship of symptom-onset-to-balloon time and door-to-balloon time with mortality in patients undergoing angioplasty for acute myocardial infarction. *Journal of the American Medical Assosciation 283*(22), 2941–2947.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *Quarterly Journal of Economics 128*(3).

Chan, D. (2016). Teamwork and moral hazard: Evidence from the emergency department. *Journal of Political Economy 124*(3).

Chan, D. (2017). The efficiency of slacking off: Evidence from the emergency department. *Econometrica*. Forthcoming.

Chan, D., M. Gentzkow, and C. Yu (2019). Selection with variation in diagnostic skill: Evidence from radiologists. *NBER Working Paper 26467*.

Chandra, A. and D. O. Staiger (2007). Productivity spillovers in healthcare: Evidence from the treatment of heart attacks. *Journal of Political Economy 115*, 103–140.

Chard, J., M. Kuczawski, N. Black, and J. van der Meulen (2011). Outcomes of elective surgery undertaken in indepdnent sector treatment centres adn NHS providers in England: audit of patient outcomes in surgery. British Medical Journal 343.

Chetty, R., J. Friedman, and J. Rockoff (2014a). Measuring the impacts of teachers i: Evaluation bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

Chetty, R., J. Friedman, and J. Rockoff (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 2633–2679.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2013). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *Quarterly Journal of Economics 126*(2), 749–804.

Competition and Markets Authority (2014, April). Private health care market investigation: Final report. Competition and Markets Authority case report, HMSO.

Cooper, Z., S. Gibbons, S. Jones, and A. McGuire (2011). Does Hospital Competition Save Lives? Evidence from the English NHS Patient Choice Reforms. *Economic Journal 121*(554), F228–F260.

Cooper, Z., S. Gibbons, S. Jones, and A. McGuire (2012). Does competition improve public hospitals' efficiency? evidence from a quasi-wxperiment in the english national health service. CEP DIscussion Paper No 1125.

Cooper, Z., S. Gibbons, and M. Skellern (2018). Does competition from private survival centres improve public hospital' performance? evidence from the english national health service. *Journal of Public Economics 166*, 63–80.

Cooperation and Competition Panel (2011, July). Review of the operation of 'any willing provider' for routine elective care. CCP Report.

Courtemanche, C. and M. Plotzke (2010). Does competition from Ambulatory Surgical Centres affect hospital surgical output? *Journal of Health Economics 29*(5), 765–773.

Currie, J. and W. B. Macleod (2020). Understanding doctor decision making: The case of depression treatment. *Econometrica 88*(3), 847–878.

Currie, J., W. B. Macleod, and J. V. Parys (2016). Practice style and patient health outcomes: The case of heart attacks. *Journal of Health Economics 47*, 64–80.

Cutler, D. (2003). *Your Money Or Your Life: Strong Medicine for America's Health Care Sytem*. Oxford University Press.

Cutler, D., J. S. Skinner, A. D. Stern, and D. Wennberg (2019). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *American Economic Journal Economic Policy 11*(1), 192–221.

Diamond, R. and P. Persson (2016, June). The long-term consequences of teacher discretion in grading of high-stakes tests. Working Paper.

Doyle, J. J., S. M. Ewer, and T. H. Wagner (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of Health Economics 29*(6), 866–882.

Doyle, J. J., J. Graves, J. Gruber, and S. Kleiner (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy 123*(1), 170–214.

Doyle, J.D. and Garmon, E.H. (2019). American society of anesthesiologists classification (asa class). StatPearls [Internet].

Dranove, D., K. Daniel, M. McCellan, and M. Satterthwaite (2003). Is more information better? The effects of "Report Card" on health care providers. *Journal of Political Economy 111*(2), 555–588.

Edwards, Nigel and Crump, Helen and Dayan, Mark (2015). Rationing in the nhs: Policy briefing 2 february 2015. The Nuffield Trust, London.

Eeckhout, J. and P. Kircher (2011). Identifying sorting: In theory. *Review of Economic Studies 78*(3).

Einav, L., A. Finkelstein, and M. Polyakova (2018). Private provision of social insurance: Drug-specific price elasticities and cost sharing in medicare part d. *American Economic Journal: Economic Policy*. Forthcoming.

Einav, L., A. Finkelstein, and P. Schrimpf (2015). The response of drug expenditure to non-linear contract design: Evidence from medicare part d. *Quarterly Journal of Economics 130*(2), 841–899.

Einav, L., A. Finkelstein, and P. Schrimpf (2017). Bunching at the kink: implications for spending responses to health insurance contracts. *Journal of Public Economics 146*, 27–40.

for National Statistics, O. (2012). 2011 Census: Population and household estimates for small areas in England and Wales, March 2011.

Gaynor, M. (2006, December). What do we know about competition and quality in health care markets? *Foundations and Trends in Microeconomics 2*(6), 441–508.

Gaynor, M., R. Moreno-Serra, and C. Propper (2013). Death by Market Power: Reform, competition and patient outcomes in the National Health Service. *American Economic Journal: Economic Policy 5*(4), 134–166.

Gaynor, M., C. Propper, and S. Seiler (2016). Free to Choose? Reform and demand responses in the English National Health Service. *American Economic Review 106*(11), 3521–3527.

Gaynor, M. and R. Town (2012). Competition in health care markets. In *Handbook of Health Economics*, Volume 2, Chapter Competition in health care markets. Elsevier.

Gerard, F., C. Rothe, and M. Rokkanen (2018, May). Bounds on treatment effects in regression discontinuity designs with a manipulated running variable, with an application to unemployment insurance in brazil. *NBER Working Paper No. 22892*.

Geweke, J., G. Gowrisankaran, and R. Town (2003). Bayesian inference for hospital quality in a selection model. *Econometrica 71*(4), 1215–1238.

Glance, L., A. Dick, D. Mukamel, Y. Li, and T. Osler (2008). Are high-quality cardiac surgeons less likely to operate on high-risk patients compared to low-quality surgeons? evidence from new york state. *Health Services Research 43*(1), 300–312.

Gowrisankaran, G. and R. Town (1999). Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics 18*(6), 747–767.

Gresenz, C., J. Rogowski, and J. Escarce (2006). Updated variable-radius measures of hospital competition. *Health Services Research 39*, 417–430.

Hanushek, E. (2009). Teacher deselection. In D. Goldhaber and J. Hannaway (Eds.), *Creating a New Teaching Profession*, pp. 165–180. Washington, DC: Urban Institute Press.

Hoe, T. P. (2017, November). Are public hospitals overcrowded? evidence from trauma and orthopaedics in england. Working Paper.

Hoot, N. R. and D. Aronsky (2008). Systematic review of emergency department crowding: Causes, effects and solutions. *Annals of Emergency Medicine 52*(2), 126–137.

Hull, P. (2020, January). Estimating hospital quality with quasi-experimental data. Working Paper.

Iacobucci, G. (2017). Exceptional requests for care surge as rationing deepens. *BMJ 358:j3188*.

Ipswich and East Suffolk CCG (2017). Threshold policy - y18a hip replacement (formerly t9). Ipswich and East Suffolk CCG.

Judge, A., N. Welton, J. Sandhu, and Y. Ben-Shlomo (2010). Equity in access to total joint replacement of the hip and knee in England: cross sectional study. *BMJ 341*.

Kane, T., J. Rockoff, and D. Staiger (2008). What does certification tell us about teacher effectiveness? evidence from new york city. *Economics of Education Review 27*, 615–631.

Kane, T. and D. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *NBER Working Paper 14607*.

Kelly, E. and G. Stoye (2020). The impacts of private hospital entry on the public market for elective care in england. *Journal of Health Economics 73*.

Kessler, D. P. and M. B. McClellan (2000, May). Is hospital competition socially wasteful? *Quarterly Journal of Economics 115*(2), 577–615.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics* (8), 435–464.

Kleven, H. J. and M. Waseem (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *Quarterly Journal of Economics 128*(2), 669–723.

Locker, T. and S. M. Mason (2006). Are these emergency department performance data real? *Emergency Medicine Journal 23*(7), 558–559.

Mansfield, R. K. (2015). Teacher quality and student inequality. *Journal of Labor Economics 33*(3), 751–788.

Marshall, M., P. Shekelle, S. Leatherman, and R. Brook (2000). The public release of performance data: what do we expect to gain? a review of the evidence. *Journal of the American Medical Association 281*, 2098–2105.

Maxwell, S. (1999). Emergency management of acute myocardial infarction. *Br J Clin Pharmacol 48*(3), 284–298.

McCabe, C., K. Claxton, and A. J. Culyer (2008). The nice cost-effectiveness threshold: What it is and what that means. *Pharmacoeconomics 26*(9), 733–744.

McCaffrey, D. F., T. R. Sass, J. R. Lockwood, and K. Mihaly (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy 4*, 572–606.

Molitor, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy 10*(1), 326–356.

Moulin, T., D. Sablot, E. Vidry, F. Belahsen, E. Berger, P. Lemounaud, L. Tatu, F. Vuillier, A. Cosson, E. Revenco, G. Capellier, and L. Rumbach (2003). Impact of emergency room neurologists on patient management and outcome. *European neurology 50*(4), 207–214.

Munnich, E. and S. Parente (2018). Returns to specialization: Evidence from the outpatient surgery market. *Journal of Health Economics 57*, 147–167.

National Audit Office (2004). Improving emergency care in england. HC 1075 Session 2003-2004.

Naylor, C. and S. Gregory (2009). Briefing{:} Independent Sector Treatment Centres. Kings Fund http://www.kingsfund.org.uk/publications/briefing-independent-sector-treatment-centres.

Newman-Toker, D. E., E. Moy, E. Valente, R. Coffey, and A. L. Hines (2014). Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. *Diagnosis 1*(2), 155–166.

of Surgeons of England, R. C. (2006). Inquiry into independent sector treatment centres. *Ann R Coll Surg Engl (Suppl) 88*, 116–119.

Pollock, A. and S. Godden (2008). Independent sector treatment centres: evidence so far. *BMJ 336*, 421–424.

Propper, C., S. Burgess, and K. Green (2004). Does competition between hospitals improve the quality of care? Hospital death rates and the NHS internal market. *Journal of Public Economics 88*(7–8), 1247–1272.

Propper, C., M. Sutton, C. Whitnall, and F. Windmeijer (2008). Did tarhets and terror reduce waiting times for hospital care in England? *BE Journal of Economic Analysis and Policy 8*(2), 318–335.

Propper, C., M. Sutton, C. Whitnall, and F. Windmeijer (2010). Incentives and Targets in Hospital Care: Evidence from a natural experiment. *Journal of Public Economics 94*, 318–335.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Saver, J. L., G. C. Fonarow, E. E. Smith, M. J. Reeves, M. V. Grau-Sepulveda, W. Pan, D. M. Olson, A. F. Hernandez, E. D. Peterson, and L. H. (2013). Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke. *JAMA 309*(23), 2480–2488.

Schochet, P. Z. and H. S. Chiang (2013). What are error rates for classifying teacher and school performance using value-added models? *Journal of Educational and Behavioral Statistics 38*, 142–171.

Seymour, C. W., F. Gesten, H. C. Prescott, M. E. Friedrich, T. J. Iwashyna, G. S. Phillips, S. Lemeshow, T. Osborn, K. M. Terry, and M. M. Levy (2017). Time to treatment and mortality during mandated emergency care for sepsis. *New England Journal of Medicine 376*(23), 2235–2244.

Shojania, K. G., E. C. Burton, K. M. McDonald, and L. Goldman (2003). Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *Journal of the American Medical Association 289*(21), 2849–56.

Silver, D. (2016, November). Haste or waste? peer pressure and the distribution of marginal returns to health care. Working Paper.

Skinner, J. (2011). Causes and consequences of regional variations in health care. In *Handbook of Health Economics*, Volume 2, Chapter 2, pp. 45–93. Elsevier.

Street, A., P. Sivey, A. Mason, M. Miraldo, and L. Siciliani (2010). Are English treatment centres treating less complex patients? *Health Policy 94*(2), 150–157.

Van Parys, J. (2016). Variation in physician practice styles withinand across emergency departments. *PLoS ONE 11, number =*.

Windecker, S. (2014). 2014 esc/eacts guidelines on myocardial revascularization: The task force on myocardial revascularization of the european society of cardiology (esc) and the european association for cardio-thoracic surgery (eacts) developed with the special contribution of the european association of percutaneous cardiovascular interventions (eapci). *European Heart Journal 35*(37), 2541–2619.

Woodcock, S. (2015). Match effects. *Research in Economics 69*(1).

Woodworth, L. (2020). Swamped: Emergency department crowding and patient mortality. *Journal of Health Economics 70*, 102279.