Essays on the Organisation of Hospital Care

Laurie Rachet Jacquet

PhD

University of York

Economics and Related Studies

May 2021

Abstract

This thesis focuses on the organisational determinants of quality of care in the hospital sector. It consists of three chapters, of which the first two investigate determinants of quality at the hospital (*economies of scale*) or surgeon level (*surgical skills*), while the third examines the trade-off between hospital quality and costs. Using patient-level data from hospitals in the English National Health Service, these essays contribute to the understanding of the optimal organisation of hospital care, related to the consolidation of hospital activity (Chapter 1), surgeons' work schedules (Chapter 2) or the efficient use of hospital resources (Chapter 3).

Chapter 1 investigates the existence of hospital economies of scale in quality for planned hip replacement. It makes use of rich condition-specific patient-reported outcome measures which increase the scope for risk-adjustment. It shows that hospital volume, though positively correlated with health outcomes, does not have a causal effect on patient health after accounting for volume endogeneity.

Chapter 2 focuses on a key actor of quality of care – surgeons – by exploring how breaks in surgical practice impact health outcomes for patients admitted after a hip fracture. Using a large panel of surgeons in England, it finds that short breaks of four to six days reduce 30-day mortality rates by around six percent relative to surgeons who had no prior breaks. Further, results show that short breaks alter the choice of treatment, holding other patient characteristics fixed.

Chapter 3 estimates the effect of reducing inpatient length of stay, a possible cost-containment strategy for hospitals, on 28-day readmission rates for emergency chest pain patients. Patients who are discharged on the same day as admission have lower readmission rates. However, the effect disappears after instrumenting for patients' length of stay, indicating that there is no causal effect of same-day discharge on health outcomes in this context.

Contents

List of Figures

List of Tables

Acknowledgements

Funding for the thesis was provided by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement number 721402.

I wish to express here my sincere thanks to a number of persons for their help during the PhD. First, I am very grateful to my main supervisor, Luigi Siciliani, for his support, implication and supervisory guidance, which have made the thesis journey stimulating and enjoyable throughout. I also thank Martin Chalkley, my co-supervisor and thesis advisor, Nils Gutacker, thesis advisor and co-author, James Gaughan, co-author and colleague, together with everyone at CHE for their help and advice. They all contributed to an incredibly supportive work environment. My thanks also go to all supervisors, organisers and fellows of the Marie Curie PhD programme for providing regular feedback and creating such a nice and unique environment to do one's thesis.

I am grateful to my family for being always supportive, showing interest in my research and kind enough to put up with my anglicisms. My thanks also go to Luís and Joana, my PhD friends, for our numerous conversations and not less numerous coffees; to Alessandro, Anika, David, Finn, Francesco, Gloria, Idaira, James, Laia, Maria José, Vigyan, Wiktoria, my friends in York and in the office for making life sweeter, and certainly more sugary. On the other side of the channel, Antoine, Léa and Setti, by their knowledge of the research world, have always been very helpful.

Finally, this is the opportunity to formally thank Yong Jie for his support during all these years; his presence has made everything easier. I am also indebted to his rigorous proofreading and patient chasing of traces of French-sounding English wordings.

Declaration

I declare that this thesis represents original work, of which I am the primary author. Chapter 1 "Scale Economies in the Health Sector: The Effect of Hospital Volume on Health Gains from Hip Replacement Surgery" is co-authored with Nils Gutacker and Luigi Siciliani. I prepared the data set, performed the analyses and wrote the first draft of the manuscript. All authors assisted in the refinement of the text. Earlier versions of this paper were presented and discussed at the European Training Network research in progress workshop in Rotterdam, the European Health Economics Association (EuHEA) conference in Maastricht, the EuHEA PhD student-supervisor conference in Catania, the Health Econometrics Data Group seminar at the University of York, the Health Economists' Study Group meeting in York and at the Irdes-Dauphine workshop in Paris. An early version was published in the Centre for Health Economics Research paper series, No.168, 2019.

Chapter 2 "The Effect of Short Breaks on Performance: Evidence from the Medical Workforce" is single-authored. Previous versions of this paper were presented and benefitted from suggestions of discussants and participants at the European Training Network research in progress workshop in Odense, the Health Economists' Study Group meeting in Newcastle, the Health Econometrics and Data Group seminar in York, the American Society of Health Economists conference (online) and at the Journées des Economistes de la Santé Francais (online).

Chapter 3 "Does Containing Costs Reduce Hospital Quality?" is co-authored with James Gaughan, Nils Gutacker and Luigi Siciliani. An early version of the data set was produced by James Gaughan as part of a previous project. I constructed the final data set and carried out the empirical analyses. All co-authors contributed to the refinement of the empirical strategy. I wrote the first draft of the chapter, which all authors helped improve.

Introduction

Health care represents a considerable sector of the economy. In 2018, health expenditure averaged 8.8% of GDP in Organisation for Economic Cooperation and Development (OECD) countries and is forecast to further increase to 10.2% by 2030 (OECD, 2019). At the same time, health gains have slowed across OECD countries, in part due to an ageing population and the rising prevalence of chronic diseases (OECD, 2019). In the U.K. for instance, life expectancy has stalled over the past decade and health inequalities have widened (Elwell-Sutton et al., 2019). Although not the only determinant of population health, the performance of healthcare systems is an important driver of health, influencing whether patients stay well (preventive care), recover quickly when ill (acute care), live well with co-morbidities (chronic care) or receive appropriate care at the end of life (palliative care). A major policy focus is therefore on improving the quality of care, such that the care provided is effective, safe and patient-centred (OECD, 2017a).

The literature has documented the existence of substantial geographic variation in quality and utilisation of care (Baicker and Chandra, 2004; Chandra and Staiger, 2007; Skinner, 2011). The evidence suggests that this variation is less due to differences in demand-side factors, such as patient severity or demand for care, than to supply-side factors (Finkelstein et al., 2016; Molitor, 2018; Cutler et al., 2019). Among supply-side factors, differences in surgeons' preferences appear to only partly explain variation in treatment across areas (Molitor, 2018; Cutler et al., 2019), suggesting that a high share of the variation in quality of care may be due to differences in hospitals' production function or choice of inputs (Castelli et al., 2015; Skinner and Staiger, 2015; Chandra et al., 2016b).

This thesis focuses on quality of care in the hospital sector. Quality of care is a core policy focus given its importance in patient care pathway and the share of health spending it represents. In the U.K. alone, hospitals accounted for 40.7% of total health care expenditures in 2018 (Eurostat, 2020). Understanding the drivers of hospital quality would help inform qualityenhancing strategies and reduce the variation in quality observed across hospitals (Castelli et al., 2015; Ali et al., 2018). Previous research has examined the role of external determinants of hospital quality, such as the impact of hospital market forces, as well as internal aspects pertaining to e.g., hospital managerial quality or technology utilisation (Gaynor et al., 2015; Bloom et al., 2020; Barrenho et al., 2021). In relation with the literature outlined above, this thesis investigates several determinants of quality related to the organisation of hospital care in the English National Health Service (NHS).

The potential for economies of scale in quality of care has received considerable attention in the policy world (Luft et al., 1987; Gaynor et al., 2005; Ho, 2014). If higher volumes permit quality gains via learning effects, there is a rationale for concentrating hospital care or closing healthcare providers that operate below some safety volume threshold. Using patient-reported outcome measures, Chapter 1 explores the existence of hospital economies of scale in quality for a planned orthopaedic surgery, providing causal empirical evidence for England. More recently, the availability of finer micro-level data has permitted to investigate the role of workforce in hospitals. Recent evidence suggests for instance that the staffing levels and composition of clinical teams in hospitals can be important drivers of quality (Bartel et al., 2014; Friedrich and Hackmann, 2017; Chan, 2021). Chapter 2 contributes to the understanding of on-the-job performance by investigating the role of breaks in activity on surgical performance.

Raising the quality of care may however come at the expense of higher costs (Hussey et al., 2013; Jamalabadi et al., 2020). Healthcare systems have been under substantial cost pressure, due to increased demand driven in part by an ageing population and the rising cost of new technologies. In recent years, health expenditure has outpaced economic growth across OECD countries, challenging the sustainability of health systems in the long term (OECD, 2019). In this context, identifying sources of expensive but low-value care could reduce wasteful expenses and help deliver more efficient hospital services (Skinner and Staiger, 2015; OECD, 2017b). Chapter 3 contributes to this strand of literature by reviewing the effect of a cost-containment strategy, which involves shifting more patients to same-day discharge care, on quality of care. The remainder of this section gives an overview of each chapter in more detail.

Chapter 1 explores whether hospitals benefit from economies of scale in quality by investigating the causal effect of hospital volume on patient health gains from planned hip replacement surgery. It focuses on a common orthopaedic surgery which is possibly amenable to returns to scale given the importance of peri-operative and follow-up care. This chapter contributes to the limited causal literature on volume-outcome effects and the policy debate around the opportunity to concentrate the provision of care as a means of raising the standards of care (Luft et al., 1987; Gaynor et al., 2005; Hentschker and Mennicken, 2018; Avdic et al., 2019).

The analysis uses a panel data set of all public hospitals in England between 2011 and 2015. The data set links routine hospital records and hip-specific patient-reported outcome measures (PROMs), which assess patients' hip pain and mobility shortly before and after the surgery. This chapter employs both a pooled OLS model and specifications with hospital fixed effects to account for unobserved heterogeneity in hospital quality. Differences in hospital case-mix are accounted for through pre-surgery hip-specific PROMs, medical and socioeconomic indicators. Nevertheless, higher-quality hospitals may also attract higher volumes of patients (Luft et al., 1987), if hospital demand responds to quality (Brekke et al., 2014; Chandra et al., 2016a; Gutacker et al., 2016). To address this possible reverse-causality bias, the analyses use a measure of predicted hospital volumes obtained from a patient model of hospital choice, as done in the hospital competition literature (Kessler and McClellan, 2000; Gowrisankaran and Town, 2003; Gaynor et al., 2013).

Results from the pooled OLS model indicate that the effect of volume on health outcomes in hip replacement surgery is positive but clinically small, but no longer significant after accounting for the endogeneity of volume with the predicted volumes. Results from an alternative specification with hospital fixed effects also show a non statistically significant effect of volume on health outcomes. Together, these results indicate that hospital volume does not have a causal impact on health outcomes, thus rejecting the hypothesis of positive scale economies in quality in this context. While a positive causal effect of volume has been sometimes found for other procedures and settings (Gaynor et al., 2005; Hentschker and Mennicken, 2018; Avdic et al., 2019), the state of the literature calls for more causal evidence (Sheldon, 2004).

Chapter 2 investigates the role of short breaks in surgical activity on the performance of surgeons. It models the causal effect of surgeons' breaks, defined as the number of days between two surgeries, on health outcomes for patients admitted for an emergency hip fracture. It relates to the wider literature on the effect of work schedule or within-team organisation on performance for the medical workforce (Cook et al., 2012; Caruso, 2014; Chan, 2021). It contributes to the scarce evidence on the effect of interruptions in practice on patient health outcomes for surgeons or teams, mostly focused on cardiac procedures (David and Brachet, 2011; Hockenberry and Helmchen, 2014; Huesch, 2014; Van Gestel et al., 2017).

The data set uses hospital records for emergency hip fracture patients admitted in England between 2009 and 2016, consisting of an unbalanced panel data set of around 2,000 orthopaedic surgeons. Using a linear probability model, patient health outcomes, measured by 30-day mortality, are regressed against surgeons' time breaks, a set of patient controls and surgeon fixed effects to account for unobserved time-invariant heterogeneity in surgeon ability. Controls include a rich set of patients' medical and socio-economic factors (age, gender, ethnicity, comorbidities, socio-economic deprivation, day of the week dummies, pre-surgery length of stay and type of hip fracture) and surgeons' volume of practice. The empirical strategy exploits the quasi-exogenous variation in surgeons' time breaks that arises from unanticipated emergency hip fracture admissions conditional on the large set of controls.

Results show that short breaks of four to six days reduce mortality rates by around six percent, relative to surgeons who had no prior break in surgical practice. Heterogeneity analyses further suggest that the beneficial effect of short breaks may be more pronounced for surgeons with lower volume of hip fracture patients. Short breaks also lead to a different choice of the type of surgery carried out, by increasing the probability of carrying out the less intensive surgery. Overall, the results indicate that surgeons perform better after short breaks possibly because they are more alert and apt to choose a more appropriate treatment. The findings differ from previous small evidence on cardiac surgeries, which suggest a null or detrimental effect of short breaks, concluding to fast depreciation of surgical skills (Hockenberry et al., 2008; Hockenberry and Helmchen, 2014).

Chapter 3 reviews the impact of a possible cost-containment strategy on hospital quality. It investigates the causal effect of being discharged from hospital on the same day as admission, therefore reducing patient length of stay in hospital, on quality of care. Discharging patients on the same day as admission when clinically safe is a potential policy lever to contain costs in hospitals (British Association for Ambulatory Emergency Care, 2014; OECD, 2017a). However, discharging patients too early may result in poorer patient health if not done appropriately. This chapter contributes to the wider literature on the effect of austerity measures on quality of care (Borra et al., 2019; Arcà et al., 2020; Bordignon et al., 2020).

The analysis uses patient hospital records from English NHS hospitals between 2010 and 2014. It focuses on emergency patients presenting with chest pain, which is a common reason for attendance at the Emergency Department. In a linear probability model, an indicator for whether a patient had an emergency readmission is regressed against the key independent variable, i.e. being discharged on the same day as admission, with patient medical and socioeconomic controls as well as hospital fixed effects. Patients who are discharged on the same day as admission are likely to be less severely ill, possibly resulting in unobserved patient severity. To account for this possible omitted variable bias, the chapter employs an instrument variable strategy where being discharged on the same day as admission is instrumented by patient exposure to a major policy shift. Introduced in 2012, the policy financially incentivised hospitals to discharge low-severity chest pain patients on the same day as admission (Allen et al., 2016; Gaughan et al., 2019).

OLS results show that being discharged on the same day is associated with lower 28-day emergency readmission rates by around 10%, while the IV results indicate no significant effect. Taken together, the results point to a null causal effect of same-day discharge treatment on health outcomes, once we account for the endogeneity of patient treatment. The results suggest that reductions in inpatient length of stay for medically approved conditions can be a valid strategy to contain costs without harming quality of care. Results are in line with the scarce existing evidence for longer inpatient stays which finds little systematic effect of shortening length of stay on health outcomes (Picone et al., 2003; Hauck and Zhao, 2011).

Chapter 1

Scale economies in the health sector: the effect of hospital volume on health gains from hip replacement surgery

1 Introduction

Improving quality of care is a key policy objective in health systems across high income countries (OECD, 2017a). Policy initiatives commonly rely on the premise that high-volume healthcare providers are able to deliver better care, by exploiting economies of scale and learning-by-doing effects often cited in the economics literature (Argote and Epple, 1990; Benkard, 2000; Mukoyama, 2006; Thompson, 2010; Ho, 2014). For instance, the Leapfrog group, a coalition of healthcare purchasers in the United States, has set minimum volume standards for hospital referrals since the early 2000s (Birkmeyer and Dimick, 2004). Similarly, France, Germany and the Netherlands have introduced minimum volume regulations for certain surgeries (Com-Ruelle et al., 2008; Bauer and Honselmann, 2017; Mesman et al., 2017).

Despite a large literature investigating the relation between volume and health outcomes across a range of procedures and countries (Ferguson et al., 1997; Halm et al., 2002; Gutacker et al., 2017), evidence of a causal effect of volume on quality remains limited due to the potential endogeneity of volume (Luft et al., 1987). Specifically, volume-outcome studies are prone to a reverse causality bias if patients' choice of hospital responds to quality via reputation or public reporting (Brekke et al., 2014; Gutacker et al., 2016). Yet, understanding the causal mechanisms behind the volume-outcome association is essential in the context of policies seeking to improve quality of care by concentrating the provision of care.

If the volume-outcome association is driven by demand's responsiveness to quality, sometimes referred to as 'selective-referral', concentrating surgical activity will not improve quality and may have adverse effects on patient accessibility to care (Blanco et al., 2017). Alternatively, a higher volume of operations can lead to better outcomes, by increasing surgeons' technical proficiency via repetition (learning-by-doing) or by fostering coordination within clinical teams (Bartel et al., 2014; Chan, 2016). Higher volume may also make it economically viable for hospitals to invest time and resources in more streamlined production processes that follow patient pathway and to invest in better infrastructure. In contrast to the selective-referral hypothesis, these mechanisms capture different forms of economies of scale[1]. In this instance, more concentrated hospital markets may lead to improved patient outcomes (Gaynor and Town, 2011; Brekke et al., 2017).

This study investigates the effect of hospital volume on the health gains of patients receiving a primary (i.e. non-revision) planned hip replacement procedure in the English NHS between 2011/12 and 2015/16. Hip replacement surgery is a common planned procedure, which involves replacing the damaged part of a hip joint by an artificial one. Hip replacement is well suited to studying economies of scale given the importance of peri-operative, rehabilitation and follow-up care[2] (Reagans et al., 2005). Hospitals with higher volumes of hip replacement patients may exert effort to design better pre-surgery and discharge protocols or build up relationships with healthcare or other providers during patient care pathway (Kizer, 2003; Ho, 2014). Further, unexplained variations in patient-reported outcomes after hip replacement have been reported at the hospital (Street et al., 2014) and surgeon level (Varagunam et al., 2015a), while fixation methods and implant types are associated with differences in revision rates (Healthcare Quality Improvement Partnership, 2018), suggesting that there is room for quality improvements. Even

[1]Rapid increases in volume may however also lead to lower quality of care if less clinical time is spent with each patient (i.e. congestion effect).

[2]Patients who undergo a planned hip replacement procedure are typically referred to the hospital by their family physician (called general practitioner in the U.K.) or after being assessed by a musculoskeletal clinic. They have a pre-surgery assessment with nurse practitioners before being seen by an anaesthetist and operated by an orthopaedic surgeon or one of the team members. After care can be supervised by an occupational therapist or a physiotherapist (Healthcare Quality Improvement Partnership, 2018).

modest improvements in the health gains for individual patients would sum to important gains at the health system level given the high incidence of hip surgeries in an ageing population.

To test for the presence of economies of scale in health outcomes, we use two alternative strategies. We first run a pooled OLS model, exploiting variation in volume across all public hospitals in England and over time (2011-2015). Variation in volumes can be driven by geographical differences in population density or in the organisation of hospital services. In addition, we implement a hospital fixed effects model to account for remaining unobserved time-invariant hospital factors beyond our set of hospital control variables. This specification tests for the effect of increasing volumes within hospitals over time on patient health outcomes. The results from a pooled OLS model show that the observed effect of volume on health outcomes in hip replacement surgery is positive and clinically small, but no longer statistically significant once we account for the endogeneity of volume. Results from an alternative specification with hospital fixed effects further confirm that hospital volume does not have a causal effect on health outcomes. We therefore conclude that we do not find evidence that economies of scale affect quality to support the argument for concentrating the provision of care in this setting.

Our contribution to the previous literature (reviewed briefly in Section 1.1) is threefold. First, we use patient-reported outcome measures (PROMs) to capture the effect of volume in terms of improvements in patients' health status. The English NHS is one of the first healthcare systems to routinely collect these novel data, which permit an examination of the benefit of treatment as perceived by the patient. In contrast, post-operative mortality is very low for planned hip replacement patients (0.06% in our data), rendering commonly used measures (mortality or complication rates) insensitive to finer variations in quality (Shojania and Forster, 2008; Varagunam et al., 2015a). Second, the availability of rich patient-reported data on functional status collected just before the surgery ensures that we thoroughly control for patient severity and minimise the risk of omitted variable bias through unobserved severity (Tsai et al., 2006; Kahn et al., 2009).

Third, we address the reverse causality bias by employing a measure of *predicted* volumes, rather than *actual* volumes. Predicted volumes are derived from a conditional logit choice model where patients' choice of hospital is a function of exogenous determinants, including the distance between patient residence and each hospital. In doing so, we apply a method commonly used in the literature on hospital competition following the seminal study by Kessler and McClellan (2000), which uses predicted volumes (patient flows) to construct Herfindahl-Hirschman indices based on hospital market shares (Gaynor and Town, 2011; Gaynor et al., 2013; Cooper et al., 2018). We also control for a rich set of hospital variables and characteristics of the catchment area around the hospital to minimise the risk of omitted variable due to hospital related variables.

By addressing these two sources of endogeneity, i.e. insufficient risk-adjustment and reverse causality, and controlling for a rich set of hospital characteristics including hospital fixed effects, we obtain causal estimates of the effect of hospital volume on patients' health benefits following planned hip replacement surgery. Our findings suggest that failing to account for hospital volume endogeneity can generate a spurious positive relationship whereby hospitals of higher quality also face a higher demand and thus higher volumes. We also show that controlling for surgeon volumes does not change our results at the hospital level, suggesting that the relation between hospital volumes and outcomes does not reflect surgeon effects.

In the remainder of this section we give a brief account of the literature. Section 2 introduces the data. Section 3 lays out the methods and Section 4 presents the results. Section 5 concludes.

1.1 Related literature

Quality improvements driven by volume of practice may occur through different channels. At the hospital level, economies of scale may take place through better collaboration between surgeons and nursing staff, familiarity with the operating theatre, the presence of specialists and technology-based services or more standardised processes of care (Kizer, 2003; Ho, 2014). A recent literature has stressed the importance of teamwork and peer effects in surgical settings in increasing productivity and efficiency of care. These studies suggest that high volumes of patients can contribute to better quality of care, via more effective work routines, through better coordination between nursing staff and surgeons, a better allocation of patients across surgeons or through more frequent learning opportunities from senior colleagues in the surgical team (Reagans et al., 2005; Chan, 2016). At the surgeon level, the volume-outcome effect is more readily understood as a learning-by-doing effect in surgical skills or a better choice of treatment, with higher volumes leading to better outcomes through improved surgical technique or ability to detect and prevent complications (Chowdhury et al., 2007).

Previous causal studies have used hospital fixed effects to control for time-invariant unobserved hospital quality, using within-hospital variation in volumes over time to estimate learning effects (Hamilton and Hamilton, 1997; Ho, 2002; Sfekas, 2009). Using a quasi-natural experiment in bariatric surgery in the U.S., several studies exploit the increase in patient referrals to high-volume hospitals after a policy restricted coverage to hospitals with certain minimum volume for Medicare patients (Livingston, 2009; Nguyen et al., 2010; Dimick et al., 2013). Alternatively, the previous literature has exploited the geographical distribution of patients in an IV setting (Gaynor et al., 2005, for cardiac care in the U.S.), or used the variation in volume caused by the closure/opening of surrounding clinics (Avdic et al., 2019, for cancer care in Sweden).

The limited causal literature on orthopaedic surgery suggests mixed results. Luft et al. (1987) find evidence of both demand's response to quality ('selective-referral') and practice-makes-perfect effect in hip replacement using simultaneous equation methods with U.S. data. Hamilton and Hamilton (1997) find no effect of volume on in-hospital mortality for hip fracture patients in Canada, after controlling for unobserved time-invariant hospital quality with hospital fixed effects. Hentschker and Mennicken (2018) use the distribution of patients and hospital competitors around the hospital as an instrument for hospital volumes in Germany. They find that hospital volume reduces in-hospital mortality for emergency hip replacement after hip fracture.

Medical studies in orthopaedic surgery find a positive association between hospital volume and health outcomes after primary hip replacement surgery. Hospitals with high volume of planned and emergency hip replacement patients are associated with lower mortality rates or complication rates in England, in the U.S. and in the Netherlands (Judge et al., 2006; De Vries et al., 2011; Singh et al., 2011). Previous studies find a negative relation between surgeon volumes and the rate of revisions or complications after primary planned hip replacement, using data from the U.S. (Losina et al., 2004) or Canada (Paterson et al., 2010; Ravi et al., 2014). These studies further show that patients treated by low-volume surgeons are associated with higher rates of revision, at both low and high-volume hospitals (Losina et al., 2004; Ravi et al., 2014). In a setting close to ours, Varagunam et al. (2015b) use Hospital Episode Statistics from England for 2011/12 and patient-reported outcome measures for planned hip replacement. They report no association between hospital volume and outcome[3] but find a significant and positive association between surgeon volume and PROMs scores. The authors, however, implicitly assume volume to be exogenously determined.

2 Data

2.1 Sample

We extract data from the Hospital Episodes Statistics (HES) on all planned (i.e. non-emergency) hip replacement surgeries in England performed between April 2011 and March 2016. HES is an administrative data set on hospital admissions in England, which includes detailed patient demographic and medical information. The original sample consists of about 360,000 patients. To ensure sample homogeneity, we exclude revision surgeries, which are less common and more complex procedures[4]. Patients who are younger than 50 years are also excluded from the sample as they represent infrequent (i.e. approximately five percent of planned hip replacements) and atypical medical cases who require replacement of a damaged

[3] Our results here differ from theirs, potentially because we exclude private hospitals as their reported volumes do not include all treated patients (see Section 2.1 for more detail).

[4] For example, revision surgeries represented around 10% of total planned hip patient admissions in our data.

hip joint much earlier than expected given usual wear. We further exclude uncommon types of hip replacement (e.g. total prosthetic replacements of the head of the femur or resurfacing arthroplasties of the joint, which account for less than 0.01% of the sample). Hospitals reporting an unusually low number of cases (below 20 annual hip replacements) are also excluded to attenuate the risk of coding errors. Relaxing these sample restrictions based on patient age or hospital size does not affect our results (Table A.1 in Appendix).

Since April 2009, a national programme requires hospitals in England to collect patient-reported outcome measures (PROMs) from patients who undergo certain planned surgeries (hip or knee replacement, varicose vein and groin hernia repair). Participation in this programme is voluntary for patients but mandatory for all hospitals that treat NHS-funded patients. All eligible orthopaedic patients are asked to report through a paper-based survey their health status, functioning and health-related quality of life immediately before and six months after surgery. We use data collected via the Oxford Hip Score (OHS) questionnaire, which is a hip-specific instrument that has been clinically validated as an accurate measure of health status for patients with problems of the hip joint (Dawson et al., 1996; Ostendorf et al., 2004) (see Section 2.2 below for more detail). The PROMs and HES data are linked based on a number of identifying characteristics, including their unique NHS number (NHS Digital, 2017). 67% of all hip patient admissions are successfully matched for at least one PROM record, which corresponds to about 235,000 admission records. We discuss potential risks of attrition bias in Section 4.5.

We further exclude private hospitals from our sample. Our data set includes all patient admissions (privately and NHS-funded) in NHS hospitals, but only admissions for NHS-funded patients in private hospitals. The observed volumes for private hospitals would therefore underestimate their actual volume of activity. The degree of measurement error depends on each hospital's unobserved volume of private patients, such that the relative distribution of the observed volumes for private hospitals will also differ from the distribution of their actual volumes. Results from a volume-outcome analysis based on the observed volumes will therefore be biased. After sample cleaning, restriction to complete PROMs records and exclusion of private hospitals, our final sample includes 105,229 patients.

Table A.2 in Appendix indicates that patients in the final sample are slightly older, have more comorbidities and lower pre-surgery Oxford Hip Score than patients in the initial sample with private providers. Though the difference is small, this suggests that private providers in England treat a healthier population overall (Moscelli et al., 2018a). This would impact the external validity of our results only if the potential for scale economies varies across levels of pre-operative health. However, this means that we focus on the upper end of the severity distribution of hip replacement patients for which health gains are more likely to occur.

2.2 Dependent variable

Our measure of patient health, the OHS, contains 12 items relating to functional status (mobility) and pain, each of which is evaluated on a scale from zero to four. Patients are asked to rate the degree or frequency of pain felt ("During the past four weeks, have you had any sudden severe pain (shooting, stabbing, or spasms) from your affected hip?"), their ability to walk ("During the past four weeks, have you been limping when walking because of your hip?"), use public transportation ("Have you had any trouble getting in and out of a car or using public transportation because of your hip?"), climb stairs or do household shopping autonomously, among other items[5]. The OHS is the sum of the scores obtained for each item and goes from zero (worst) to 48 (best health status). The same OHS questions are distributed to the patients shortly before and six-month after surgery. We use patients' post-surgery OHS as our dependent variable but control for the pre-surgery OHS, thereby assessing patient's *health gain* from the surgery.

2.3 Independent variables

Our key independent variables are the annual hospital volumes, measured as the number of patients who have undergone a planned primary hip replacement at a given hospital during each financial year from 2011/12 to 2015/16. In the English NHS, hospitals are organised

[5]The full questionnaire can be found online at: `http://www.orthopaedicscore.com/scorepages/oxford_hip_score.html` [accessed 02.04.2020].

into legal entities, formally called NHS trusts. We measure volume at the more disaggregated hospital (site) level rather than at the trust level to obtain the physical concentration of activity in a facility, which we assume to be more relevant to economies of scale.

We control for patients' demographic characteristics (age, gender and ethnic group) and socio-economic deprivation, where the latter is based on the quintiles of the 2010 or 2015 index of multiple deprivation (IMD)[6] measured at the small residence area level (lower-level super output area, LSOA) of the patients. Our model includes pre-surgery OHS grouped in narrow bands to capture potential non-linear effects. Our model also controls for the patient's self-assessed disability status prior to the surgery, symptom duration and living arrangements, as well as self-reported depression and assistance in filling the questionnaire (Department of Health, 2012). We count the Elixhauser comorbidities reported in a patient's hospital stays up to one year prior to the admission for hip replacement (Elixhauser et al., 1998; Gutacker et al., 2016). We also control for whether the patient has previously undergone hip surgery on the other hip in the past year, the primary diagnosis (e.g. osteoarthritis) (Losina et al., 2004) and the type of surgery (i.e. total hip replacement vs hybrid prosthetic replacement).

We control for a large set of hospital characteristics that may be associated with higher quality (e.g. via medical expertise or better resources) independently of volumes. We include controls for hospitals' teaching status, whether the hospital is a specialist (orthopaedic) hospital[7] or a NHS foundation trust (FT) as the latter have greater financial autonomy (Gravelle et al., 2014). Hospitals located in more affluent areas may enjoy better facilities or find it easier to recruit healthcare staff. We proxy for these exogenous geographical differences by using the market forces factor (MFF) which reflects unavoidable differences in hospital costs of labour or capital and is used to adjust hospital reimbursement tariffs.

To ensure that we isolate the effect of hip replacement volumes on health outcomes from potential confounders, we also control for the average socio-economic and demographic characteristics of the population in hospitals' catchment area. Hospitals that serve a more frail or

[6] The index of multiple deprivation measures deprivation across seven domains, including income, employment and education.

[7] We extract information on teaching and specialist status from the Estates Returns Information Collection collated by NHS Digital (NHS Digital, 2016).

deprived population may have poorer outcomes, independently of volume, and hospitals may face higher demand pressure. We define the hospital catchment area as the area within 30 km of the hospital (in line with the competition literature mentioned above) and measure the proportion of over 65-year-olds and the average deprivation score of the population in that area. Poorer access and availability of primary care may result in lower coordination of care in the community and put more strain on the hospital services as a whole. We therefore include the mean distance to the closest family physician practice, the General Practitioner (GP), for the population who lives in the hospital catchment area[8].

Robustness check analyses also include controls for the degree of competition in the hospital catchment area, proxied by the number of equivalent public hospitals whose headquarters lie within 30 km of the hospital, for the overall size of the hospital and for hospital staff composition. Hospital size is measured by the total number of beds for general acute care (including overnight and day-only beds) at the trust level. We construct dummies corresponding to seven categories of hospital size: less than 400 beds, 400-549 beds, 550-699 beds, 700-849 beds, 850-999 beds, 1000-1149 beds and over 1150 beds. Data are published quarterly by NHS England and averaged across quarters to obtain hospitals' yearly mean number of beds. Data on hospital staff are reported monthly through the Electronic Staff Records and published quarterly by NHS digital. We construct the proportion of hospital staff who are consultants (i.e. senior NHS doctors), the ratio of nursing staff to doctors in full time equivalent (FTE) and the ratio of nursing staff to beds as the yearly mean across quarters.

[8]We construct these variables based on population statistics from the Office for National Statistics for small homogenous geographic areas called Lower Super Output Areas (LSOAs). The hospital catchment area comprises all the LSOAs whose centroid falls within 30 km of the hospital's headquarters.

3 Methods

3.1 Baseline model with observed volumes

We study the effect of hospital volume on health gains after hip replacement surgery. Our econometric model is specified as follows:

$$y_{iht} = \alpha + vol^{'}_{ht}\beta_1 + x^{'}_{iht}\beta_2 + k^{'}_{ht}\beta_3 + \delta_t + \epsilon_{iht}, \tag{1.1}$$

where y_{iht} is the post-surgical OHS of patient i in hospital h at time of admission t, x_{iht} is a vector of patient characteristics (age in 10-year bands, gender, comorbidities, the pre-surgery OHS, socio-economic status) to adjust for differences in case-mix across hospitals. k_{ht} is a vector of time-varying controls for hospital characteristics (i.e. NHS foundation trust, specialist orthopaedic, teaching hospitals and market forces factor factor in given year t)[9] and control for characteristics of hospitals' catchment area (proportion of population over 65 years old, mean deprivation and distance to closest GP). δ_t is a vector of year dummies which account for aggregate change in quality over time. ϵ_{iht} is a random error term.

Our main interest is in the effect of hospital volume on patients' post-surgery health status. Hospital volume vol_{ht} is entered as a vector of four dummy variables corresponding to volume categories: vol_{ht}<150, 150≥ vol_{ht}>200, 200≥ vol_{ht}>300 and vol_{ht} ≥300. This allows for a non-linear relationship due to decreasing marginal returns to scale; especially at the lower end of the volume distribution where scale economies are likely to occur. To our knowledge, there is no evidence on the safety threshold for planned hip replacement using PROMs. Using volume quartiles would not permit comparability of the results across specifications given that observed and predicted volumes follow different distributions. We therefore define category thresholds that allow for more weight to be placed on the lower volume categories given expected diseconomies of scale, whilst ensuring that we have enough hospitals in each category

[9]These characteristics are defined at the trust level. For simplicity, we use the same subscript h for hospitals and trusts.

and both volume distributions for consistent estimation[10]. We also present results using the log of hospital volume. We estimate Equation (1.1) with pooled OLS. We also estimate a second model where we add hospital fixed effects, denoted with γ_h, to control for unobserved time-invariant hospital factors. We adjust standard errors for clustering at the hospital level[11].

3.2 Endogeneity concerns

Regression models, such as that defined in Equation (1.1), may provide a biased estimate of the volume-outcome relationship in the presence of reverse causality from quality to volume, or omitted variables linked to unobserved patient severity or hospital characteristics.

First, low (high) quality hospitals will face a lower (higher) demand, thus inducing a positive correlation in our estimates of Equation (1.1). To address this, we borrow from the literature on the effect of competition on hospital quality. A similar challenge in this literature arises since hospitals' market share, measuring the hospital market structure via the Herfindahl-Hirschman Index (HHI), may be potentially (endogenously) determined by the quality of the hospital and of its competitors (Gowrisankaran and Town, 2003; Gaynor et al., 2013). These studies use discrete patient choice models, based on patients' distance to the hospital and hospital characteristics, to obtain predicted patient volumes and thus predicted market shares to derive an exogenous measure of hospital market structure (i.e. exogenous HHI).

We follow a similar approach but focus on predicted hospital volumes rather than market shares. This amounts to constructing the volumes that would be observed if patients were choosing hospitals based on geographical proximity[12]. Our identification strategy is therefore based on assumptions commonly made in the literature that i) patients' residential choices are not based on the quality of the surgical interventions provided by the surrounding hospitals,

[10]The smallest volume category accounts for 10% of the volume distribution and a minimum of 21 hospitals.

[11]Technically, we cluster at the trust (i.e. legal entity) level, given possible correlation across hospitals within a trust.

[12]There is an analogy between our method and previous instrumental variable strategies (Gaynor et al., 2005; Hentschker and Mennicken, 2018) because both rely on the exogeneity of patients' distance to the hospital. However, the conditional logit model allows for non-linear effects whereas the first stage in an IV strategy is estimated by OLS and thus assumes linearity in the parameters.

and ii) that patients derive higher disutility and costs from travelling further (Kessler and McClellan, 2000; Gaynor et al., 2005; Gutacker et al., 2016; Hentschker and Mennicken, 2018; Moscelli et al., 2018a). Whilst residential sorting is plausible in the context of education, as families may choose to live close to a charter school for instance (Horowitz et al., 2009; Chung, 2015), this is unlikely in the context of acute hospital care in high-income countries. Residential sorting in our case would imply that patients anticipate their future need for a specific healthcare procedure, here orthopaedic surgery, when choosing where to live, and have a good knowledge of this specific aspect of hospital quality. This is even less plausible in the context of a hip replacement, which is a one-off acute treatment for relatively healthy individuals, as opposed to patients with chronic conditions who may require repeated treatments for the rest of their lives. Further, hospital quality has been shown to be only weakly correlated across conditions (low-risk vs high-risk conditions) and types of care (emergency vs planned care) (Gravelle et al., 2014; Skellern, 2017) and is likely to vary over time. In particular, Gravelle et al. (2014) test for correlation between different measures of quality for a sample of English hospitals, and find that hospitals' overall mortality rates are not correlated with any measures of quality related to planned orthopaedic activity (i.e., readmission or revisions after hip or knee replacements). A formal presentation of the choice model is given in Section 3.3.

Second, family physicians may refer their most severely ill patients to hospitals with better quality and higher volumes (Geweke et al., 2003; Hentschker and Mennicken, 2018). We control for differences in hospitals' case-mix with patients' self-reported pre-surgery health and a comprehensive set of comorbid conditions. Pre-surgery measures of functional status and pain allow us to adjust more thoroughly for differences in patients' ability to benefit from surgery than has been possible in previous studies. Any remaining differences between hospitals in terms of unobserved patient severity should be limited.

Finally, hospitals may be able to provide higher quality through unobserved determinants of quality that also correlate with volume. By failing to control for these, parameter estimates in Equation (1.1) will suffer from omitted variable bias. We address these concerns by running specifications with hospital fixed effects. Results from a hospital fixed effects model estimate the effect of change in volume within hospitals over time on patient health outcomes. While using hospital fixed effects may also curtail relevant variation in volume, e.g. across hospitals, it allows for a thorough control of potential unobserved time-invariant hospital factors, and thus mitigates further the risk of omitted variable bias. In addition, we test the sensitivity of our results to the inclusion of additional time-varying hospital control variables that may be potentially correlated with volume (degree of competition) or endogenous to volume (overall size of the hospital, or hospital staff composition) in the robustness checks reported in Section 4.5.

3.3 A model of patient choice of hospital

To implement the empirical strategy outlined above, we estimate a conditional logit model of patient choice of hospital (McFadden, 1974). We include in the choice set all public and private hospitals that treat NHS patients. The sample includes the whole population of planned hip replacement patients who had surgery, regardless of whether they participated in the PROMs survey. The utility of patient i choosing hospital h at time t can be written as:

$$u_{iht} = V_{iht} + v_{iht}, \tag{1.2}$$

where V_{iht} is the utility of patient i derived from observed characteristics of hospital h and v_{iht} is the unobserved utility. We specify V_{iht} as:

$$V_{iht} = \gamma_1 d_{iht} + \gamma_2 d_{iht}^2 + \gamma_3 d_{iht}^3 + \gamma_4 close_{iht} + z_{ht}^{'}\gamma_z + \sum_{k=1}^{K} x_{ikt}(\gamma_1 d_{iht} + \gamma_2 d_{iht}^2 + \gamma_3 d_{iht}^3), \tag{1.3}$$

where d_{iht} represents the distance between patient i and hospital h at time t, measured as the straight-line distance between hospital's postcode and the centroid of patient's LSOA of residence, and γ_1 is the associated (dis)utility of travel. We include quadratic and cubic terms of distance to allow for a non-linear effect on patient's choice utility. We add a dummy variable, $close_{iht}$, to capture the utility of avoiding any excess travel past the closest hospital. The vector z_{ht} consists of dummy variables for hospital characteristics (i.e. NHS foundation trusts, specialist (orthopaedic) hospital, teaching hospital, private hospital) as well as the number of hospitals (sites) within a trust and whether the hospital is a treatment centre. Hospital groups (trusts) may direct their patients to a specific hospital (site). Treatment centres typically do not admit complex patients. We therefore control for these two admission restrictions. We add interaction terms between all the distance terms and $x = (x_{ikt}, k = 1, .., K)$, a vector of K patient characteristics (age, sex, socio-economic status, Elixhauser comorbidities, and whether the patient lives in a rural area[13]) as the impact of distance on hospital choice also depends on patients' socioeconomic and clinical factors[14]. Standard errors are clustered at the family physician practice level to account for correlation in hospital choice across patients of the same practice.

Assuming that the unobserved utility terms v_{iht} are iid extreme-value (Train, 2009), the probability that patient i chooses hospital h at time t can be estimated by maximum likelihood and is given by:

$$\hat{p}_{iht} = \frac{exp(\hat{V}_{iht})}{\sum_{h' \in M_{it}} exp(\hat{V}_{iht'})}, \tag{1.4}$$

where M_{it} is the patient choice set containing patient i's 50 closest hospitals. The predicted volume of hospital h is equal to the sum of the estimated probabilities $\hat{p}_{iht}$ across all patients of choosing hospital h:

$$\widehat{vol}_{ht} = \sum_{i=1}^{N} \hat{p}_{iht} = \sum_{i=1}^{N} \frac{exp(\hat{V}_{iht})}{\sum_{h' \in M_{it}} exp(\hat{V}_{iht'})}, \tag{1.5}$$

We estimate Equation (1.4) for the whole sample of planned primary hip replacement patients in England for all years between 2011/12 and 2015/16, after exclusion of patients under

[13]The geographical information for lower super output areas (LSOAs) comes from the Office of National Statistics.

[14]In a sensitivity analysis, we included an indicator variable for patients who had a hip replacement surgery in the previous year (slightly under four percent of the sample), to account for the fact they will likely return to the same hospital. Predicted volumes under this alternative specification were highly correlated with our baseline predicted volumes (Pearson correlation coefficient = 0.99).

50 years old who are atypical medical cases for planned hip replacement and hospitals with less than 20 hip replacement cases per year. In the robustness check presented in Table A.1 in Appendix, we also relax these sample restrictions when estimating the choice model. Preferences in hospital choice may vary across years. We therefore estimate the choice model separately for each year before merging all years to obtain the final sample.

Appendix Tables A.3, A.4 and Figure A.1 present summary statistics for the choice model sample, consisting of 261,743 patients[15]. Predicted hospital volumes are less dispersed than observed volumes (Figure A.2, Appendix). We use the same sample restrictions to compute observed volumes to ensure that both predicted and observed volumes sum up to the same total patient population[16]. The correlation coefficient between both measures of volumes in our estimation sample is 0.61 ($p<0.01$).

Conditional logit models have the advantage that they are tractable and computationally simple. These properties however rely on the assumption of independent error terms. If this holds, estimated coefficients are invariant to which alternatives/choices are available (independence of irrelevant alternatives, IIA). We *omit* hospital quality from our model specification[17], thus creating potential correlation in the error terms. The IIA property of logit models is problematic in forecasting exercises (i.e. when forecasting the demand for a new alternative) as it imposes strong restrictions on substitution behaviours. However, it is considered less crucial when estimating average aggregate preferences (Train, 2009, p.36). Our model is therefore an approximation of patients' demand for hospitals, if they were to ignore hospital quality considerations. We re-estimate our model with varying sets of alternatives (comprising the 30 and 10 closest hospitals in patients' choice set). Hospitals' predicted volumes under both specifications are highly correlated (Pearson correlation coefficient= 0.97), suggesting that any potential violation of the IIA assumption does not affect our results.

[15]Estimated coefficients from the choice model are available in Table A.5 in Appendix.

[16]The observed volumes with and without these sample restrictions have a correlation coefficient of 0.98.

[17]As a sensitivity analysis, we also estimate our choice model with hospital quality using hospitals' average lagged risk-adjusted Oxford Hip Score gain and standardized overall mortality rates, before removing these effects for the computation of predicted volumes. Predicted volumes under this alternative specification of the choice model are highly correlated with our baseline choice model (Pearson correlation coefficient =0.98) and results are unchanged.

4 Results

4.1 Summary statistics

In our sample, hospitals treat on average 222 hip replacements patients annually, ranging from 20 to 1,238 surgeries. Table 1.1 also reports the total, between and within-hospital standard deviation for all hospital characteristics. The within-hospital standard deviation is much smaller (40.31), about 27%, than the between-hospital standard variation (147.69) as hospitals are less likely to experience dramatic changes in volumes over time. 57% of hospitals are NHS Foundation Trusts (Table 1.1). Teaching hospitals and specialist orthopaedic hospitals account respectively for 21% and two percent of hospitals. The average hospital has a market forces factor of 1.08 and a (total) standard deviation of 0.07. On average, 16.93% of the population in the hospital catchment area is over 65 years old. Population in the hospital catchment area has a mean deprivation rank of 15,971 (relative to the catchment area with highest deprivation with a rank of 24,208), and lives on average 1.48 km away from the closest GP practice.

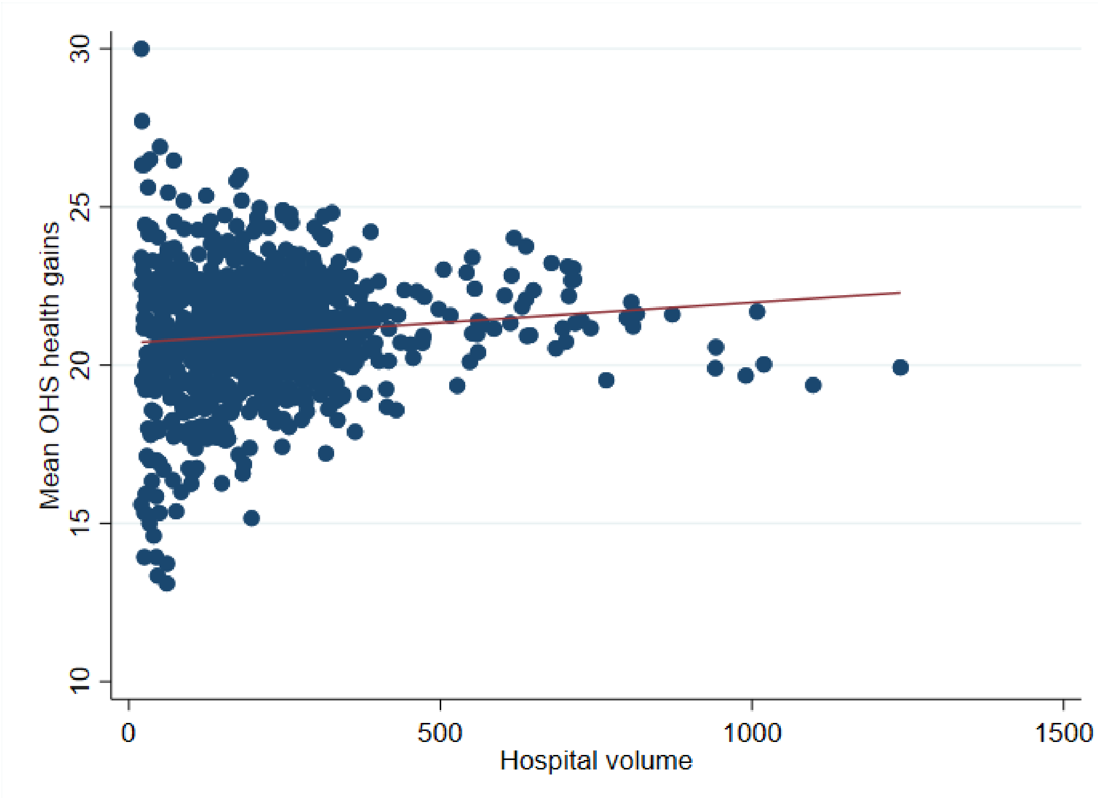
Table 1.1: Summary statistics for hospital characteristics

	Mean	Std. Dev.			Min.	Max.
		Total	Between	Within		
Observed volume	222.54	156.45	147.69	40.31	20.00	1238.00
Foundation Trust	0.57	0.49	0.49	0.07	0.00	1.00
Teaching hospital	0.21	0.41	0.41	0.06	0.00	1.00
Specialist hospital	0.02	0.13	0.12	0.00	0.00	1.00
Market forces factor	1.08	0.07	0.07	0.03	1.00	1.30
Hospital catchment area						
% of pop. over 65 years	16.93	3.10	3.05	0.49	11.23	24.55
Mean deprivation rank	15717.74	3143.93	3224.20	393.15	11383.87	24665.11
Mean distance to GP (km)	1.48	0.54	0.52	0.04	0.78	3.44
Hospital-years	892					

Notes: Volume is the number of annual planned primary hip replacements per hospital. The market forces factor index adjusts hospital resource allocation for unavoidable geographical differences in the costs of labour and capital. A hospital catchment area comprises all the small homogenous geographic areas (Lower Super Output Areas, LSOAs) whose centroid falls within 30 km of the hospital's headquarters. The index of multiple deprivation ranks each LSOA according to their level of deprivation, from 0 (the most deprived) to 32,844 (the least deprived).

Figure 1.1 shows the unadjusted relationship between OHS health gains and hospital volumes, suggesting a small positive association between hospital volumes and surgery health gain.

Figure 1.1: Association between hospital volume and Oxford Hip Score (OHS) health gain



Notes: Plot of the average OHS health gains (post-surgery minus pre-surgery OHS, unadjusted for other patient characteristics) per hospital against hospital volumes, for all hospital-years between 2011/12 and 2015/16.

Table 1.2 presents descriptive statistics of patient characteristics. The average pre-surgery OHS is 17.52 points, and patients gain on average 22 points (from 17.52 to 38.69) six months after surgery. Patients are on average 70 years old and 40% of our sample are male. On average, patients report slightly over one (1.37) Elixhauser comorbidity. The large majority of patients (70%) report having hip-related symptoms for between one and five years.

Table 1.2: Summary statistics for patient characteristics

	Mean	Std. Dev.	Min.	Max.
Post-surgery OHS	38.69	9.03	0	48
Pre-surgery OHS	17.52	8.07	0	48
Age	69.87	8.90	50	99
Male patient	0.40	0.49	0	1
Elix. comorbidity count	0.25	0.73	0	8
Index of multiple deprivation (IMD):				
1st quintile	0.24	0.43	0	1
2nd quintile	0.25	0.43	0	1
3rd quintile	0.22	0.42	0	1
4th quintile	0.17	0.38	0	1
5th quintile - Most deprived	0.12	0.32	0	1
Ethnicity: white	0.91	0.28	0	1
Surgery on the other hip	0.18	0.39	0	1
Diagnosed with osteoarthritis	0.97	0.18	0	1
Hybrid prosthetic replacement	0.19	0.39	0	1
PROMs questions:				
Self-reported disability	0.61	0.49	0	1
Self-reported depression	0.08	0.27	0	1
Received assistance in filling questionnaire	0.07	0.26	0	1
Symptoms duration:				
$<$1year	0.13	0.34	0	1
1-5 years	0.70	0.46	0	1
6-10 years	0.11	0.31	0	1
$>$10 years	0.06	0.23	0	1
Living arrangements:				
Lives alone	0.28	0.45	0	1
Lives with family	0.71	0.45	0	1
Other	0.01	0.08	0	1
Observations	105229			

Notes: The IMD is calculated for small residence areas (LSOAs) in England. The Oxford Hip Scores (OHS) range from the worst reported health state (=0) to the best (=48) and are collected for each patient shortly before and six months after the operation.

Table 1.3 presents summary statistics for surgeon characteristics used in Section 4.3. On average, a surgeon treats around 56 planned hip replacement patients per year. 99% of orthopaedic surgeons are male. 63% of the surgeons are trained in the U.K. and have on average 24 years since their primary medical qualification. We exclude surgeons who report less than 10 cases a year (N=1,130 surgeons, more than half of which only treat one hip replacement patient per year), and four surgeon outliers who report more than 300 annual cases.

Table 1.3: Summary statistics for surgeon characteristics

	Mean	Std. Dev.			Min.	Max.
		Total	Between	Within		
Surgeon yearly volume	56.67	40.69	37.31	14.59	10	270
Male surgeon	0.99	0.11	0.12	0.00	0	1
Years since qualification	24.73	7.52	7.79	1.29	9	45
Qualified in the UK	0.63	0.48	0.49	0.00	0	1
Surgeon-years	4619					

Notes: Surgeon volume comprises all patients treated in all hospitals (if the surgeon holds multiple appointments). We exclude surgeons with less than 10 annual cases or above 300 annual cases who are probable volume outliers. Years since qualification is the time since primary medical qualification, after which surgeons follow additional specialty training.

Table 1.4 presents summary statistics for the additional hospital controls used in robustness checks in Section 4.5. On average, a hospital has slightly under six equivalent rival hospitals in its catchment area. Hospital bed categories are relatively evenly distributed. Hospitals with less than 400 beds, the smallest category, stands for eight percent of the hospital sample, and hospitals with over 1150 beds, the largest category, representing 24% of the hospital sample. On average, doctors account for 12.68% of hospital staff in full-time equivalent. There are on average 2.28 nursing staff for one doctor, and 1.53 nurses per bed in our sample of hospitals.

Table 1.4: Summary statistics for additional hospital characteristics

	Mean	Std. Dev.			Min.	Max.
		Total	Between	Within		
Number of rivals	5.89	6.79	6.96	0.67	0.00	25.00
Overall number of beds						
<400 beds	0.08	0.27	0.25	0.11	0.00	1.00
400-549 beds	0.13	0.34	0.25	0.11	0.00	1.00
550-699 beds	0.15	0.36	0.29	0.17	0.00	1.00
700-849 beds	0.16	0.37	0.32	0.21	0.00	1.00
850-999 beds	0.15	0.36	0.29	0.21	0.00	1.00
1000-1149 beds	0.09	0.29	0.25	0.17	0.00	1.00
>1150 beds	0.24	0.43	0.41	0.16	0.00	1.00
% of doctors in hospital staff	12.68	2.37	2.24	0.76	5.39	25.87
Nurses-doctors ratio	2.28	0.54	0.47	0.23	0.82	7.45
Nurses-beds ratio	1.53	0.34	0.32	0.16	0.90	3.46
Hospital-years	892					

Notes: The number of equivalent hospital rivals in the hospital's catchment area comprises all public hospitals (trusts) whose headquarters lie within 30 km of the hospital. The number of beds (total of overnight and day only beds, published quarterly by NHS England) and staff data (the proportion of hospital staff who are doctors, the nurses-to-doctors ratio and nurses-to-beds ratio) are yearly average for hospital trusts and are lagged by one year. Data on hospital staff are reported monthly through the Electronic Staff Records and published quarterly by NHS digital.

4.2 Main results

Table 1.5 provides the results for the pooled OLS regression with observed hospital volumes, indicating a positive and statistically significant association between hospital volume and health. Relative to patients treated in hospitals with more than 300 hip replacement cases per year (the reference group with highest volume), patients treated in hospitals with less than 150 cases (with lowest volume) are estimated to gain 0.72 fewer OHS points. The estimate is statistically significant at the 0.1% level. Patients treated in hospitals performing between 150 and 200 cases are estimated to gain 0.48 fewer OHS points, and the coefficient is statistically significant at the one percent level. There is no statistically significant effect of hospital volumes of 200-300 patients relative to the base category. The volume-outcome association is therefore weakly monotonic. A change in OHS is considered clinically meaningful if above four points (Varagunam et al., 2015b). Therefore, the estimated association is quantitatively small, at around 20% (0.7 points) of a clinically meaningful change[18].

The coefficients on patient characteristics are all statistically significant, though are not substantial in clinical terms. Patients with higher pre-surgery health (OHS score) also have higher health after surgery. Coefficients on pre-surgery scores are close to or above four OHS points, suggesting that the difference is clinically important. Older patients tend to have worse outcomes and male patients tend to report slightly better outcomes. Our results suggest the existence of a socioeconomic gradient. More deprived patients have worse outcome than less deprived ones. Having one more Elixhauser comorbidity leads to a reduced post-surgery OHS score by about 0.47 points (12% of a minimally clinically important difference). Self-reported depression, disability, or help in filling questionnaires are negatively associated with post-surgery outcomes by two OHS points or above. Only specialist orthopaedic hospitals are associated with slightly better health outcomes, though the difference is not clinically important.

Table 1.5: Results from OLS regression with observed volumes

	Post-surgery OHS	
	Coefficient	SE
Volume [Ref.≥300]		
<150 cases	-0.720***	(0.191)
150-200 cases	-0.484**	(0.177)
200-300	-0.242	(0.130)
Pre-surgery OHS [Ref. 0-6 pts]		
6-12 pre-surgery OHS	2.996***	(0.155)
12-18 pre-surgery OHS	5.007***	(0.163)
18-24 pre-surgery OHS	6.314***	(0.168)
24-30 pre-surgery OHS	7.174***	(0.175)
30-36 pre-surgery OHS	7.987***	(0.180)
36-42 pre-surgery OHS	8.730***	(0.226)
42-48 pre-surgery OHS	9.250***	(0.402)
Age [Ref. 50-59 years]		
60-69 years	0.380***	(0.097)
70-79 years	-0.653***	(0.109)
80-89 years	-1.463***	(0.126)
90-105 years	-1.196***	(0.294)

[18]Note however that the total OHS gains at the aggregate level can amount to more substantial health gains.

Male patient	0.969***	(0.056)
Ethnic group [Ref. white]		
Other ethnic group	-0.934**	(0.332)
Ethnicity not coded	1.083***	(0.134)
Deprivation Index IMD [Ref. 1st quintile]		
2nd quintile	-0.307***	(0.075)
3rd quintile	-0.579***	(0.093)
4th quintile	-1.499***	(0.095)
5th quintile - Most deprived	-2.732***	(0.127)
Surgery on the other hip	-0.677***	(0.075)
Elix. comorbidity count	-0.470***	(0.055)
Diagnosed with osteoarthritis	1.468***	(0.173)
Hybrid prosthetic replacement	0.209*	(0.084)
Self-reported disability	-2.414***	(0.066)
Self-reported depression	-2.995***	(0.132)
Received assistance in filing questionnaire	-2.749***	(0.137)
Symptoms duration [Ref. <1 year]		
1 to 5 years	-0.767***	(0.082)
6 to 10 years	-1.504***	(0.106)
More than 10 years	-1.815***	(0.154)
Living arrangements [Ref. alone]		
Lives with family	0.393***	(0.059)
Other	-0.835*	(0.374)
Teaching hospital	-0.069	(0.137)
Specialist hospital	0.632*	(0.281)
Foundation Trust	-0.007	(0.126)
Market forces factor	0.271**	(0.101)
Mean deprivation rank (catchment area)	0.000	(0.000)
Mean distance to GP (catchment area)	0.307	(0.189)
Year dummies [Ref. 2011]	ref.	
2012	0.638***	(0.102)
2013	0.613***	(0.119)
2014	0.758***	(0.111)
2015	1.626***	(0.132)
R^2	0.175	
Observations	105229	

Notes: In parentheses, robust standard errors are clustered on hospitals.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In panel A (top panel) of Table 1.6, we report the regression results for observed hospital volumes and predicted hospital volumes using pooled OLS. The covariate coefficients are similar for the specifications with observed and predicted volumes. We therefore only present the coefficients for hospital volume, under different functional forms: using volume categories or continuous volume in log form. Both functional forms allow for a nonlinear effect of volume on health outcomes.

Table 1.6: Effect of observed and predicted hospital volumes on patient post-surgery OHS

	Observed hospital volumes		Predicted hospital volumes	
	(1)	(2)	(3)	(4)
Panel A: Pooled OLS				
Volume [Ref.≥300]				
<150 cases	-0.720***		0.012	
	(0.191)		(0.250)	
150-200 cases	-0.484**		0.011	
	(0.177)		(0.152)	
200-300	-0.242		-0.199	
	(0.130)		(0.140)	
Log(volume)		0.404***		-0.028
		(0.089)		(0.178)
R^2	0.175	0.175	0.174	0.174
Panel B: Fixed Effects				
Volume [Ref.≥300]				
<150 cases	-0.316		-0.106	
	(0.239)		(0.274)	
150-200 cases	-0.092		-0.151	
	(0.174)		(0.204)	
200-300	-0.146		-0.101	
	(0.098)		(0.118)	
Log(volume)		0.202		0.157
		(0.195)		(0.263)
R^2	0.182	0.182	0.182	0.182
Observations	105229	105229	105229	105229

Notes: The same covariates (patient controls, hospital time-varying controls and year dummies) as in Table 1.5 are included. In parentheses, robust standard errors clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.6, panel A, shows that observed hospital volume is associated with higher patient post-surgery OHS scores, irrespectively of the functional form of volume chosen. However, when using the predicted (exogenous) hospital volumes, the volume coefficients reported in columns 3-4 are smaller and no longer statistically significant. This suggests the presence of a spurious positive relation between health outcomes and volumes. After accounting for reverse causality, due to hospitals with higher quality attracting more patients, hospital volumes are no longer associated with improved health outcomes (Table 1.6, Panel A, columns 3-4)[19].

In panel B (bottom panel) of Table 1.6, we report the same results but estimate a hospital fixed-effects model. Our results are essentially similar: predicted hospital volumes are not statistically significant, indicating no causal effect of volume on patient health (panel B, columns 3-4). However, observed hospital volumes are no longer significantly associated with better patient health outcomes in this context (panel B, columns 1-2). This may be because the fixed effects specification solely relies on the within-hospital variation in volume over time, which is limited in this context as shown in summary statistics in Table 1.1[20]. Results are also robust to the exclusion of (observed) volume outliers (i.e., hospitals treating over 1,000 annual hip replacement cases).

4.3 Testing for the effect of surgeon volume

Volume-outcome effect may be due to hospital factors and/or personnel effects. Healthcare personnel, and chiefly, surgeons, may experience positive learning effects as their volumes of activity increase. They may gain technical proficiency and become more apt at detecting complications. A higher volume of patients, if it entails a more regular practice, may also ensure that operating skills are maintained over time (Ramanarayanan, 2008; Hockenberry and Helmchen, 2014). To test that hospital volume is not simply a proxy for individual surgeon effects, we run the same models but additionally control for individual surgeons' yearly volume and characteristics, such as their gender, years since graduation as a proxy for seniority and being trained in the U.K.

[19]Note that with predicted volumes, bootstrapping standard errors would be the appropriate technique to account for the fact that the predicted volumes are generated in a first stage choice model. Model-based standard errors do not account for sampling variation in the predicted volumes, which may lead to downward-biased standard errors (Murphy and Topel, 1985). However, because the procedure is computationally intensive and because larger standard errors would not affect our results as we find a null effect for predicted hospital volumes, we do not bootstrap the standard errors throughout the study.

[20]More substantial changes in hospital volumes over time may arise after the closure of nearby hospitals or hospital mergers. Such quasi-exogenous shocks in volume have been used elsewhere in the literature (e.g. Avdic et al. (2019))

Our strategy of predicting hospital volumes based on distance cannot be extended to surgeon volumes because patients would travel the same distance to surgeons within the same hospital. We argue that selective referral to high-quality surgeons should be limited, given that little information was available on surgeon performance during our study period. In 2015/16, online statistics were limited to a surgeon's 90-day mortality rate after primary hip replacement (NHS Commissioning Board, 2012), and reported that *all* surgeons were in line with expectations[21] (Varagunam et al., 2015a). However, we cannot completely exclude that surgeon volumes are endogenous, e.g., if allocation of patients to surgeons within a hospital is not random. We therefore do not claim causality of the surgeon volume effect.

Table 1.7 shows the results of a specification where we allow for surgeon effects. As in Table 1.6, we report results from two specifications: first, we run a pooled OLS model (panel A) and second, we estimate a surgeon fixed effects model (panel B) to account for unobserved surgeon effects. Hospital fixed effects will be highly collinear with surgeon fixed effects, therefore we only include surgeon fixed effects. We alternatively use continuous surgeon volumes in logs to allow for a nonlinear effect of surgeon volumes on health outcomes (columns 1 and 3) and a categorical variable (columns 2 and 4) for surgeon volume above the safety threshold (35 annual cases) identified by Ravi et al. (2014) for surgeons performing total hip arthroplasty in the U.S.

Results from the pooled OLS model indicate that there is a significant effect of surgeon volume on patient health. The quantitative effect however is small. A 10% increase in surgeon volume is associated with around 0.05 additional OHS points, equivalent to 1.25% of a clinically minimal important difference (four OHS points). The effect is stronger (coefficient around 0.6, i.e. approximately 15% of a minimally important difference) when we compare surgeons who perform less than 35 annual cases with surgeons above that threshold. The coefficient for surgeon volumes is stable across specifications.

[21]This is also stated on the National Joint Registry website to inform patient choice for current data: `http://www.njrsurgeonhospitalprofile.org.uk/FAQ#10`.

Table 1.7: Results with surgeon volumes, pooled OLS and surgeon fixed effects

	Observed hospital volumes		Predicted hospital volumes	
	(1)	(2)	(3)	(4)
Panel A: Pooled OLS				
Hospital volume [Ref.≥300]				
<150 cases	-0.412*	-0.534**	0.103	0.065
	(0.172)	(0.174)	(0.226)	(0.237)
150-200 cases	-0.243	-0.347*	0.153	0.080
	(0.165)	(0.168)	(0.129)	(0.142)
200-300	-0.122	-0.194	-0.099	-0.140
	(0.117)	(0.126)	(0.122)	(0.136)
Log(surgeon volume)	0.487***		0.537***	
	(0.071)		(0.071)	
Surgeon volume ≥35		0.570***		0.637***
		(0.108)		(0.108)
Male surgeon	0.608	0.683*	0.592	0.689*
	(0.310)	(0.309)	(0.313)	(0.311)
Qualified in the UK	0.281*	0.359**	0.303*	0.397**
	(0.118)	(0.120)	(0.120)	(0.124)
Years since qualification	-0.030***	-0.027***	-0.030***	-0.027***
	(0.006)	(0.006)	(0.006)	(0.006)
R^2	0.177	0.177	0.177	0.177
Panel B: Surgeon Fixed Effects				
Hospital volume [Ref.≥300]				
<150 cases	-0.507**	-0.542**	0.348	0.309
	(0.183)	(0.184)	(0.243)	(0.248)
150-200 cases	-0.142	-0.173	0.111	0.073
	(0.171)	(0.168)	(0.174)	(0.176)
200-300	-0.088	-0.106	0.033	0.005
	(0.098)	(0.097)	(0.108)	(0.110)
Log(surgeon volume)	0.282*		0.324**	
	(0.126)		(0.124)	
Surgeon volume ≥35		0.203		0.216
		(0.155)		(0.153)
R^2	0.200	0.200	0.200	0.200
Observations	101304	101304	101304	101304

Notes: The same covariates (patient controls, hospital time-varying controls and year dummies) as in Table 1.5 are included. In parentheses, robust standard errors are clustered on hospitals. In panel B, surgeon characteristics (being male, qualification in the U.K. or years since qualification) are not included as they would be collinear with the surgeon fixed effects or the year dummies. Sample size is slightly smaller due to missing surgeon characteristics or exclusion of surgeon outliers. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Hospitals with small volumes of hip replacement patients, i.e. under 150 cases a year, are associated with worse health outcomes, even though the effect is smaller once we control for surgeon volumes and surgeon characteristics. Surgeons who qualified in the U.K. are associated with slightly better health outcomes, while the number of years since graduation is associated with slightly worse outcomes, possibly because older surgeons are less familiar with the medical state-of-the-art knowledge. Results from the specification with surgeon fixed effects (panel B) however show that surgeon volume is associated with differences in patient health but the magnitude of the association is smaller once we control for time-invariant surgeon effects with surgeon fixed effects.

Overall, the results in Table 1.7 confirm that observed hospital volume is significantly associated with patient health outcomes, even after controlling for surgeon effects. Predicted hospital volume has no effect on health gains even after controlling for surgeon effects (Table 1.7, columns 3 and 4), even when we include surgeon fixed effects (Table 1.7, panel B).

Some surgeons in public hospitals may also treat privately funded patients in private (independent sector) hospitals. For these surgeons, our measure of volume is smaller than the total carried out across public and private hospitals. We run the same analysis on the subsample of surgeons whom we observe to work only in NHS hospitals (around 55%, N=689)[22]. Our results (Table A.6 in Appendix) are unchanged.

The results show that observed hospital volume is associated with health outcomes, even after controlling for individual surgeon volume and surgeon fixed effects. In a robustness check in Section 4.5, we control for other measures of personnel effects, such as the proportion of staff who are doctors or the nurses-to-beds ratio both measured at the hospital level.

[22]While private hospitals represent an important share of all NHS funded care (i.e. in 2015/16, close to one third of NHS-funded planned hip replacement patients were treated in private hospitals in our data), privately-funded hip replacements across all hospitals accounted for less than 13% of the total hip replacement volume in 2010/11 according to Kelly and Stoye (2016). The likelihood that surgeons working in private hospitals only treat private patients is therefore low, and any remaining unobserved volumes would be small in magnitude.

4.4 Broader measures of volume

Scale economies may also arise from performing treatments that are different but related to planned hip replacements, and therefore contribute to improvement in health outcomes for these patients through ameliorated processes or learning-by-doing effects (Schilling et al., 2003). To address this, in this section, we employ more comprehensive measures of orthopaedic volumes. We first include emergency hip replacements in our measure of hospital and surgeon volumes, as this relate to the same procedure but in an emergency rather than an elective setting. Second, we further add knee replacements, which are also performed in an elective setting and involve similar surgeon skills. Table A.7 in Appendix present summary statistics for these different definitions of volumes.

Table 1.8 compares the results when using different measures of hospital volumes. Given that different definitions of volume have a different support and distribution, we use the log of volume, rather than the previously defined volume categories, to compare the results. Columns (1) and (4) correspond to our baseline measure of volume comprising all planned hip replacements. The coefficient on observed volume is significant in most specifications in columns 1-3 but diminishes in size as we include additional activity, first by adding emergency hip replacements (column 2), and then, planned knee replacements (column 3). These results suggest that these additional surgeries are less relevant to returns of scale and including them potentially introduces some measurement error. Coefficients for predicted hospital volumes are not statistically significant for either definitions of volumes (columns 4-6). When we introduce hospital fixed effects (panel B), as with our baseline results, no measure of volume is statistically significant.

Table 1.8: Results with broader measures of hospital volumes

	Observed hospital volumes			Predicted hospital volumes		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pooled OLS						
Log(volume)	0.404***			-0.028		
	(0.089)			(0.178)		
Log(volume) - with emergency hip replacements		0.325**			-0.173	
		(0.105)			(0.177)	
Log(volume) - with emergency hip and planned knee replacements			0.295**			-0.205
			(0.106)			(0.188)
R^2	0.175	0.175	0.174	0.174	0.174	0.174
Panel B: Fixed effects						
Log(volume)	0.202			0.157		
	(0.195)			(0.263)		
Log(volume) - with emergency hip replacements		0.070			-0.225	
		(0.236)			(0.266)	
Log(volume) - with emergency hip and planned knee replacements			0.061			-0.151
			(0.234)			(0.296)
R^2	0.183	0.183	0.183	0.183	0.183	0.183
Observations	105229	105229	105229	105229	105229	105229

Notes: The same covariates (patient controls, hospital time-varying controls and year dummies) as in Table 1.5 are included. In parentheses, robust standard errors are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Similarly, Table 1.9 shows the results when adding surgeon volume using these different definitions. We keep our baseline measure of hospital volume (planned hip replacements) as results from Table 1.8 show these seem the most relevant to measure returns to scale. Surgeon volume is associated with better health post-surgery OHS, though the magnitude of the association is similar or lower when we include related orthopaedic activity, as with hospital volumes in Table 1.8. The overall results are unchanged: observed hospital volumes are associated with patient post-surgery health, even after controlling separately for individual surgeon volume and characteristics.

Table 1.9: Results with broader measures of surgeon volumes (pooled OLS)

	Observed hospital volumes			Predicted hospital volumes		
	(1)	(2)	(3)	(4)	(5)	(6)
Hospital volume [Ref.≥300]						
<150 cases	-0.412*	-0.439*	-0.538**	0.103	0.103	0.044
	(0.172)	(0.172)	(0.179)	(0.226)	(0.228)	(0.235)
150-200 cases	-0.243	-0.271	-0.346*	0.153	0.152	0.092
	(0.165)	(0.164)	(0.170)	(0.129)	(0.129)	(0.142)
200-300	-0.122	-0.142	-0.186	-0.099	-0.107	-0.137
	(0.117)	(0.117)	(0.128)	(0.122)	(0.123)	(0.136)
Log(volume)	0.487***			0.537***		
	(0.071)			(0.071)		
Log(volume) - with emergency hip replacements		0.510***			0.566***	
		(0.078)			(0.078)	
Log(volume) - with emergency hip and planned knee replacements			0.295***			0.358***
			(0.079)			(0.076)
Male surgeon	0.608	0.626*	0.682*	0.592	0.613*	0.675*
	(0.310)	(0.305)	(0.309)	(0.313)	(0.309)	(0.312)
Qualified in the U.K.	0.281*	0.281*	0.360**	0.303*	0.304*	0.393**
	(0.118)	(0.118)	(0.123)	(0.120)	(0.121)	(0.127)
Years since qualification	-0.030***	-0.027***	-0.027***	-0.030***	-0.027***	-0.027***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
R^2	0.177	0.177	0.177	0.177	0.177	0.176
Observations	101304	101304	101304	101304	101304	101304

Notes: The same covariates (patient controls, hospital time-varying controls and year dummies) as in Table 1.5 are included. In parentheses, robust standard errors are clustered on hospitals. Sample size is slightly smaller due to missing surgeon characteristics or exclusion of surgeon outliers. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.5 Robustness checks

Our estimation of predicted volumes based on the patient choice model relies on patients' distance to the hospital being exogenous to hospital quality, conditional on our set of controls. This assumes that there are no unobserved patient confounders that relate both to patient distance to the hospital and to health outcomes. Though untestable, this is plausible considering our large set of patient controls. Unlike for certain acute conditions that require immediate treatment, e.g. heart attacks or strokes, where delays in access to care may impact health outcomes, timely access to care does not affect health outcomes in the case of planned hip surgery (Tuominen et al., 2010; Brealey et al., 2012). Further, we show in Appendix, Table A.8, that patient distance to the hospital is not correlated with patient pre-surgery OHS or with the number of comorbidities after controlling for main patient socio-economic characteristics[23], suggesting that remaining unobserved confounders are unlikely.

Our dependent variable, post-surgery Oxford Hip Score, is a subjective measure of health. Patients who care about hospital quality, and select actively into better quality hospitals, may also answer the outcome questionnaire differently. As a robustness check, we use the probability of having a revision surgery within three years of the index surgery as an alternative and objective measure of patient health outcome[24]. Results, in Appendix Table A.9, confirm our baseline results with PROMs: there is no statistically significant effect of predicted hospital volume on mortality rates. Like with PROMs, patients treated in hospitals with smaller observed volumes are associated with a higher risk of having a revision surgery, though the effect is only statistically significant at the 10% level.

Furthermore, patient participation in the PROM survey is voluntary and attrition may happen for different reasons. If attrition is systematically correlated with health outcomes and hospital volumes, our estimates will be biased. If non-response to the PROMs questionnaires is driven by poorer underlying patient health, our rich set of risk-adjustment variables (including the pre-surgery OHS) ensures that differences in hospitals' case mix are accounted for. In addition, we regress the rate of PROMs participation per hospital, corresponding to the number of patients who answered the questionnaires out of the total number of eligible patients, on hospital volumes, patient case-mix and hospital status. Results in Table A.10 in Appendix indicate no systematic correlation between hospitals' rate of participation to PROMs and hospital volumes. Overall, this suggests that bias linked to attrition is unlikely.

Hospitals may provide better quality through unobserved time-varying determinants that correlate with hospital volume. Our large set of hospital-level controls together with the

[23]Results from Table A.8 show that patient distance to the closest hospital is weakly correlated with the number of comorbidities (i.e. increasing patient distance to the closest hospital by 100 km is associated with around 0.6 lower comorbidities) but the effect is driven by patients with the highest distance to the closest hospital. The correlation disappears when we remove the top percentile of patients with the largest distance to their closest hospital.

[24]Post-surgical mortality is another objective measure of health outcomes but it is very low for planned hip replacement surgery (around 0.06% in our initial sample), as opposed to around 1.3% for three-year hip revision surgery.

exclusion of private hospitals mitigates the risk of systematic quality differences (e.g., linked to ownership type). In addition, we run the following robustness checks. First, previous studies have found that hospitals in more competitive areas respond to the competition by increasing quality (Gaynor et al., 2013; Bloom et al., 2015; Cooper et al., 2018), though the effect varies across countries and procedures[25]. We therefore control for the degree of competition in the hospital market, proxied by the number of equivalent rival hospitals in the hospital catchment area (Bloom et al., 2015; Moscelli et al., 2018b). This is because competition could be correlated with hospital volume, if higher aggregate supply causes lower volume for each provider, for a given demand, or if more competitive areas face proportionally larger demand than less competitive areas even accounting for higher supply.

Second, we check that our results are due to the effect of the volume of hip replacement patients on health outcomes, rather than the overall size of the hospital. For instance, larger hospitals may benefit from economies of scope across clinical departments, by pooling resources or skills, which may benefit the quality of care. We measure hospital size by the total number of hospital beds for general acute care and include controls for seven categories: less than 400 beds, 400-549 beds, 550-699 beds, 700-849 beds, 850-999 beds, 1000-1149 beds and over 1150 beds. We do not control for hospitals' size in our baseline model because, similarly to hip replacement volume, it is prone to a reverse causality effect, whereby hospitals with high-quality reputation will attract more patients, thus driving hospitals' bed capacity upward.

Third, hospital staff composition may impact the quality of care, independently of volume. The presence of experienced colleagues may have a positive effect on the team (Ayoubi et al., 2017). We additionally control for the proportion of hospital staff who are doctors, the nurses-to-doctors ratio and the nurses-to-beds ratio across the hospital[26]. Again, we do not include this variable in our baseline regressions because staff composition may also reflect hospital quality, if high-quality hospitals are more successful in recruiting more qualified personnel.

[25] Feng et al. (2015) find no association between market competition of hospitals and patient-reported health outcomes for planned hip replacements in England, while Skellern (2017) finds that a pro-competition reform in the English NHS had a negative effect on PROMs for hip and knee surgeries.

[26] Variables that are directly under hospital control, such as staff composition and the hospital's total number of beds, are lagged by one year, ensuring that they were measured *before* our dependent variable and thus could not have been affected by the contemporaneous level of quality.

The regression results, shown in Table 1.10, do not differ from our main results. The positive relationship between the observed hospital volumes (lowest volume category) and the patient outcomes is unchanged and statistically significant. The predicted volumes show no causal effect on patient health gains in any of the specifications. The coefficients on the pre-surgery Oxford Hip Scores are stable across specifications. None of the additional hospital characteristics shows a statistically significant relationship with patient outcomes. The sample size in Table 1.10 is smaller because of some missing staff characteristics for certain years for certain hospitals.

Other unobserved mechanisms might be at work within hospitals. For instance, the presence of physical therapists, whom we do not have data on, could improve patient rehabilitation. Overall, however, our set of additional controls suggest that our results successfully isolate the potential for economies of scale in hip replacement from a range of potential confounders.

Table 1.10: Results with additional hospital control variables

	Observed volumes	Predicted volumes
Volume [Ref.≥300]		
<150 cases	-0.690***	-0.023
	(0.188)	(0.246)
150-200 cases	-0.432*	0.053
	(0.178)	(0.180)
200-300	-0.276*	-0.200
	(0.140)	(0.154)
Measure of hospital competition		
Number of rivals	-0.012	-0.016
	(0.030)	(0.031)
Total number of beds (Ref. >1500 beds)		
<400 beds	-0.267	-0.275
	(0.336)	(0.338)
400-549 beds	-0.368	-0.452
	(0.303)	(0.308)
550-699 beds	-0.304	-0.339
	(0.286)	(0.287)
700-849 beds	-0.071	-0.061
	(0.281)	(0.276)
850-999 beds	0.043	0.042
	(0.278)	(0.275)
1000-1149 beds	-0.126	-0.078
	(0.253)	(0.249)
Staff composition		
% of doctors in hospital staff	-0.004	0.026
	(0.071)	(0.074)
Nurses-doctors ratio	0.178	0.196
	(0.139)	(0.149)
Nurses-beds ratio	-0.131	-0.140
	(0.089)	(0.093)
R^2	0.175	0.175
Observations	85918	85918

Notes: In parentheses, robust standard errors are clustered on hospitals. The same covariates (patient controls, hospital time-varying controls and year dummies) as in Table 1.5 are included. A hospital catchment area comprises all the small homogenous geographic areas (Lower Super Output Areas, LSOAs) whose centroids fall within 30 km of the hospital's headquarters. The index of multiple deprivation ranks each LSOA according to their level of deprivation, from 0 (the most deprived) to 32,844 (the least deprived). The number of equivalent hospital rivals in the hospital's catchment area comprises all public hospitals (trusts) whose headquarters lie within 30 km of the hospital. The number of beds and staff data are lagged yearly average statistics at the trust level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5 Conclusions

This study investigates the effect of hospital volumes on health gains, as measured by patient-reported outcomes, for planned hip replacement surgery in public hospitals in England. Our key finding is that there is a clinically small and positive association between observed hospital volume and health outcomes, but this disappears once we adjust for volume endogeneity due to reverse causality (i.e. hospitals with higher quality attract more patients). Our results differ from the study by Hentschker and Mennicken (2018) who find a positive causal effect of hospital volume on patient outcomes after emergency hip fracture in Germany. Results may differ because the authors investigate a more complex procedure typically involving frail patients, with a high post-surgery mortality rate at around six percent.

Our pooled OLS results overall suggest that the hospital-level association can be driven by hospital demand's responsiveness to quality rather than economies of scale. This shows the importance of accounting for volume endogeneity in volume-outcome studies for planned procedures whose results may otherwise be biased. In the absence of a causal effect of volume on patient outcomes, increasing the provision of planned hip replacements at any hospital would not result in improvements in health outcomes in that hospital. We conclude that we do not find evidence that economies of scale affect quality to support the argument for concentrating the provision of care in this setting.

Nevertheless, we find a small positive correlation between volume and outcomes. Transferring patients from low-volume hospitals with lower-performance to better quality hospitals could result in health improvements for these patients. However, the potential gains are unlikely to offset the potential adverse effects of concentrating hip replacement activity. The hospital market in England is already concentrated, and further concentration may have adverse effects on patient access to care, as proximity remains a key determinant of patient choice to the provider, especially for relatively older patients in need of a hip replacement (Gutacker et al., 2016). Policies that concentrate provision of health care may also risk shifting some of the NHS costs onto patients and carers by increasing travel times or transportation costs, which may be particularly problematic for patients from disadvantaged socio-economic backgrounds (Ferguson et al., 1997). Further concentration of care may also affect hospital competition, which could have knock-on effects on quality efforts (Cooper et al., 2011; Gaynor et al., 2013; Moscelli et al., 2018b).

The study has some limitations. First, we cannot exclude the possibility that most hospitals may already be operating at the flat end of the volume-outcome curve, despite being able to observe hospitals with low volumes (the lowest volume category starts at 20 cases per year). However, the fact that we do find a positive association between outcomes and observed volumes suggests that there remains room for improvement. Second, we cannot perfectly disentangle the various channels behind higher hospital volume. We have focused on a procedure which is likely to benefit most from economies of scale and for which concentration of care would be a possible policy option given the planned nature of the surgery. We use hospital volumes to measure hospital economies of scale (e.g. improvement in processes of care), but hospital volume may also be a proxy for increased team experience. However, we include analyses where we control for a range of personnel factors (such as individual surgeon characteristics and volume, and hospital staff composition). The results are robust to the inclusion of these personnel effects, which indicates that attenuation bias linked to measurement error should not be a concern in this case. Future research could explore empirical strategies that disentangle the causal contributions of personnel and hospital factors. Volume effects that are driven primarily by individual surgical learning-by-doing would have different policy implications, and may be best addressed during surgeons' medical residency, or via regular trainings or work schedules that allow for regular practice (Hockenberry and Helmchen, 2014). Provided that data become available, future research could also investigate the effect of volume on other dimensions of quality, such as care responsiveness or facility features.

A Appendix

Table A.1: Results after relaxing restriction to hospitals above 20 annual cases (1) or to patients above 50 years old (2), for observed or predicted hospital volumes.

	(1)		(2)	
	Observed	Predicted	Observed	Predicted
Volume [Ref.$\geq$300]				
<150 cases	-0.743***	-0.042	-0.744***	-0.175
	(0.203)	(0.271)	(0.192)	(0.219)
150-200 cases	-0.449**	-0.018	-0.440*	-0.008
	(0.170)	(0.151)	(0.169)	(0.143)
200-300	-0.231	-0.135	-0.273	-0.193
	(0.128)	(0.132)	(0.139)	(0.143)
R^2	0.178	0.177	0.178	0.177
Observations	110559	110559	110669	110669

Notes: Results with different sample restrictions. In parentheses, robust standard errors are clustered on hospitals. Patient controls include, besides the pre-surgery Oxford Hip Scores (OHS), the patient age, sex, ethnicity, index of multiple deprivation, the number of Elixhauser comorbidities, diagnosis with osteoarthritis, control for the type of surgery carried out (hybrid prosthetic versus total hip replacement) and for having had a surgery on the other hip, self-reported disability, depression, assistance in filling questionnaires, the symptoms duration and the patient' living arrangements. Hospital controls include hospital status (teaching, specialist orthopaedic and Foundation Trust hospitals), for the market forces factor index, and for socio-demographic characteristics of population in hospitals' catchment area (proportion of over 65-year-olds, mean deprivation rank and mean distance to the closest GP). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.2: Patient characteristics for sample with private providers (1) and final sample (2)

	(1)		(2)	
	Mean	Std. Dev.	Mean	Std. Dev.
Post-surgery OHS	39.44	8.69	38.69	9.03
Pre-surgery OHS	18.11	8.11	17.52	8.07
Age	69.68	8.73	69.87	8.90
Male patient	0.40	0.49	0.40	0.49
Elix. comorbidity count	0.23	0.68	0.23	0.71
Index of multiple deprivation (IMD):				
1st quintile	0.26	0.44	0.24	0.43
2nd quintile	0.26	0.44	0.25	0.43
3rd quintile	0.22	0.42	0.22	0.42
4th quintile	0.16	0.37	0.17	0.38
5th quintile - Most deprived	0.10	0.30	0.12	0.32
Ethnicity: white	0.88	0.32	0.91	0.28
Surgery on the other hip	0.19	0.40	0.18	0.39
Diagnosed with osteoarthritis	0.96	0.19	0.97	0.18
Hybrid prosthetic replacement	0.18	0.39	0.19	0.39
Self-reported disability	0.57	0.49	0.61	0.49
Self-reported depression	0.08	0.26	0.08	0.27
Received assistance in filling questionnaire	0.06	0.24	0.07	0.26
Symptoms duration:				
<1 year	0.14	0.35	0.13	0.34
1-5 years	0.69	0.46	0.70	0.46
6-10 years	0.11	0.31	0.11	0.31
>10 years	0.06	0.23	0.06	0.23
Living arrangements:				
Lives alone	0.27	0.44	0.28	0.45
Lives with family	0.72	0.45	0.71	0.45
Other	0.01	0.07	0.01	0.08
Observations	148617		105229	

Notes: The index of multiple deprivation (IMD) measures deprivation across seven domains, including income, employment and education, and is calculated for small residence areas (LSOAs) in England. The Oxford Hip Scores (OHS) range from the worst reported health state (=0) to the best (=48) and are collected for each patient shortly before and six months after surgery.

Table A.3: Patient summary statistics for the choice model

	Mean	Std. Dev.	Min.	Max.
Patient age	70.01	9.14	50.00	102.00
Male patient	0.39	0.49	0.00	1.00
Elixhauser conditions	0.27	0.76	0.00	8.00
Index of multiple deprivation (IMD):				
1st quintile	0.24	0.43	0.00	1.00
2nd quintile	0.25	0.43	0.00	1.00
3rd quintile	0.22	0.42	0.00	1.00
4th quintile	0.17	0.38	0.00	1.00
5th quintile - Most deprived	0.12	0.33	0.00	1.00
Patient living in rural area	0.27	0.45	0.00	1.00
Distance to chosen hospital (km)	12.94	12.97	0.00	326.35
Patient choosing closest hospital	0.42	0.49	0.00	1.00
Observations	261743			

Notes: Statistics are calculated for the total sample of planned hip replacements, after exclusion of hip admissions for revision surgeries, patients below 50 and providers with less than 20 cases a year.

Table A.4: Hospital summary statistics for the choice model

	Mean	Std. Dev.	Min.	Max.
NHS Treatment Centre (TC) site	0.03	0.17	0	1
Teaching trust	0.16	0.37	0	1
Specialist trust	0.01	0.09	0	1
#hospital sites within trust	1.89	1.21	1	6
Provider type:				
Independent Sector TC	0.07	0.26	0	1
Independent Sector non-TC	0.23	0.42	0	1
NHS Foundation Trust (FT)	0.38	0.49	0	1
NHS non-FT	0.31	0.46	0	1
Observations	465			

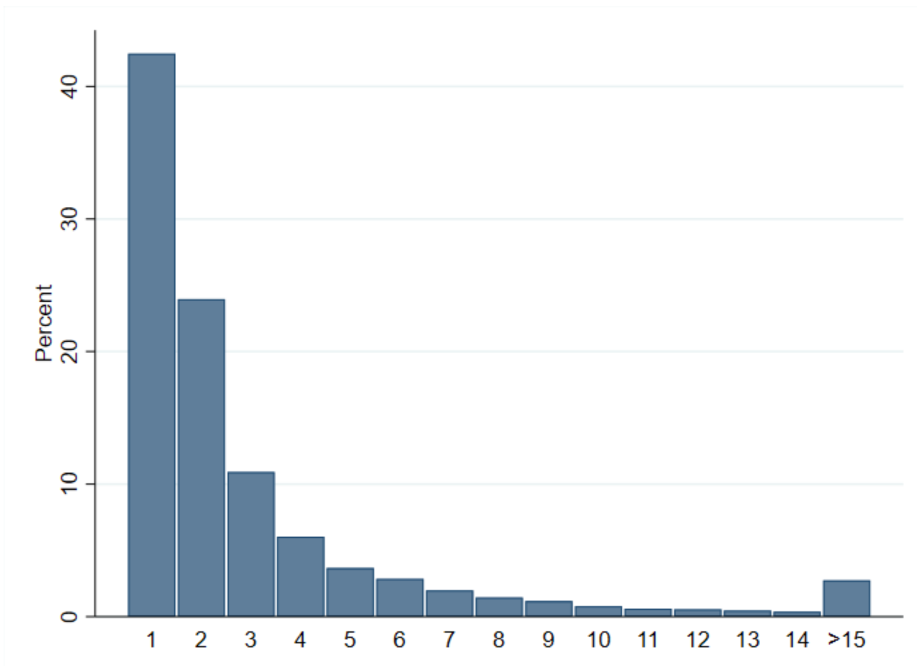
Table A.5: Estimated coefficients of the choice model

	Est.	SE
Distance (km)	-0.171***	(0.013)
Distance^2	0.002***	(0.000)
Distance^3	-0.000**	(0.000)
Closest	-0.025	(0.021)
NHS Trust [Ref. category]		
Foundation Trust	0.270***	(0.030)
Independent Sector (IS)	-0.660***	(0.029)
IS Treatment Centre	-0.050	(0.037)
Teaching Trust	0.028	(0.030)
Specialist Trust	1.344***	(0.072)
1 hospital site [Ref. category]		
2 hospital sites	-0.359***	(0.032)
3 hospital sites	-0.369***	(0.042)
4 hospital sites	-1.127***	(0.084)
5 hospital sites	-0.596***	(0.075)
NHS treatment centre	0.969***	(0.058)
Interaction with distance		
x Male	-0.005	(0.003)
x Patient age	-0.001***	(0.000)
x Income deprivation [Ref. 1st quintile]		
x Deprivation (2nd quintile)	0.010*	(0.005)
x Deprivation (3rd quintile)	0.009	(0.005)
x Deprivation (4th quintile)	-0.012	(0.007)
x Deprivation (5th quintile)	-0.052***	(0.009)
x Comorbidity count	-0.001	(0.001)
x Rural residence	0.050***	(0.007)
Interaction with distance^2		
x Male	0.000*	(0.000)
x Patient age	0.000**	(0.000)
x Income deprivation [Ref. 1st quintile]		
x Deprivation (2nd quintile)	-0.000	(0.000)
x Deprivation (3rd quintile)	0.000	(0.000)
x Deprivation (4th quintile)	0.001***	(0.000)
x Deprivation (5th quintile)	0.001***	(0.000)
x Comorbidity count	0.000	(0.000)
x Rural residence	-0.001***	(0.000)
Interaction with distance^3		
x Male	-0.000*	(0.000)
x Patient age	-0.000	(0.000)
x Income deprivation [Ref. 1st quintile]		
x Deprivation (2nd quintile)	-0.000	(0.000)
x Deprivation (3rd quintile)	-0.000*	(0.000)

x Deprivation (4th quintile)	-0.000***	(0.000)
x Deprivation (5th quintile)	-0.000***	(0.000)
x Comorbidity count	-0.000	(0.000)
x Rural residence	0.000***	(0.000)
N patients	59833	
N providers	312	

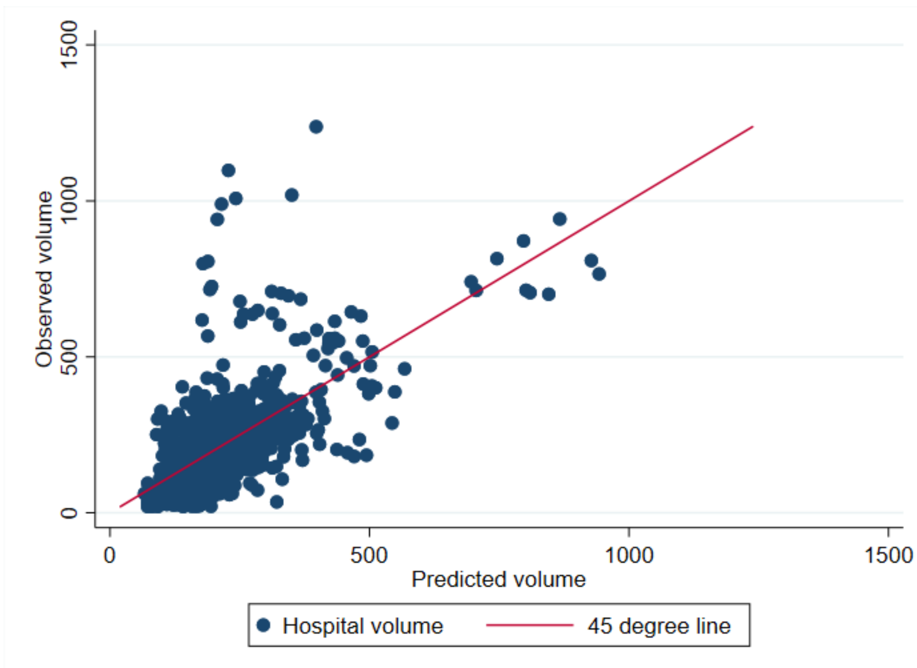
Notes: Conditional logit model of choice of hospital for elective hip replacement patients treated in England in the financial year 2015/16. Coefficients are marginal utilities. Standard errors clustered on family physician practice level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure A.1: Percentage of planned hip patients who went to their Nth nearest provider



Notes: This corresponds to the total sample of 261,743 planned primary hip replacements in England between 2011/12 and 2015/16 over which the multinomial logit model of hospital choice is run.

Figure A.2: Scatterplot of the observed against the predicted hospital volumes



Notes: Hospital volumes are the number of planned primary (non-revision) patients treated by a hospital in between 2011/12 and 2015/16 in the English NHS. The predicted volumes are constructed using a conditional logit model of hospital choice. The data points are for each hospital-year in the final sample.

Table A.6: Results on subsample of surgeons who work for public hospitals only

	Observed hospital volumes		Predicted hospital volumes	
	(1)	(2)	(3)	(4)
Panel A: Pooled OLS				
Hospital volume [Ref. ≥300]				
<150 cases	-0.441*	-0.650**	0.247	0.129
	(0.201)	(0.202)	(0.334)	(0.355)
150-200 cases	-0.156	-0.310	0.079	-0.054
	(0.206)	(0.209)	(0.148)	(0.169)
200-300	-0.047	-0.162	-0.177	-0.254
	(0.147)	(0.163)	(0.141)	(0.163)
Log(surgeon volume)	0.568***		0.624***	
	(0.091)		(0.087)	
Surgeon volume ≥35		0.544***		0.626***
		(0.130)		(0.131)
Male surgeon	0.645	0.711*	0.625	0.723*
	(0.342)	(0.341)	(0.347)	(0.343)
Qualified in the UK	0.256	0.352*	0.281	0.401*
	(0.148)	(0.154)	(0.147)	(0.154)
Years since qualification	-0.030***	-0.025**	-0.031***	-0.026**
	(0.008)	(0.008)	(0.008)	(0.008)
R^2	0.179	0.178	0.179	0.178
Panel B: Surgeon Fixed Effects				
Hospital volume [Ref. ≥300]				
<150 cases	-0.799**	-0.854**	0.585	0.530
	(0.266)	(0.262)	(0.357)	(0.365)
150-200 cases	-0.163	-0.205	-0.022	-0.071
	(0.253)	(0.248)	(0.243)	(0.245)
200-300	-0.140	-0.166	-0.102	-0.136
	(0.171)	(0.169)	(0.185)	(0.188)
Log(surgeon volume)	0.288		0.362*	
	(0.161)		(0.152)	
Surgeon volume ≥35		0.193		0.220
		(0.187)		(0.184)
R^2	0.205	0.205	0.205	0.205
Observations	61668	61668	61668	61668

Notes: Results for subsample of surgeons working for NHS hospitals only. In parentheses, robust standard errors are clustered on hospitals. Patient controls include, besides the pre-surgery Oxford Hip Scores (OHS), the patient age, sex, ethnicity, index of multiple deprivation, the number of Elixhauser comorbidities, diagnosis with osteoarthritis, control for the type of surgery carried out (hybrid prosthetic versus total hip replacement) and for having had a surgery on the other hip, self-reported disability, depression, assistance in filling questionnaires, the symptoms duration and the patient' living arrangements. Hospital controls include hospital status (teaching, specialist orthopaedic and Foundation Trust hospitals), for the market forces factor index, and for socio-demographic characteristics of population in hospitals' catchment area (proportion of over 65-year-olds, mean deprivation rank and mean distance to the closest GP). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.7: Summary statistics for broader definitions of hospital and surgeon volumes

	Mean	Std. Dev.	Min.	Max.
Hospital volume:				
Planned hip replacements (baseline)	222.54	156.45	20	1238
With emergency hip replacements	307.32	176.46	20	1238
With emergency hip + planned knee replacements	553.61	317.57	24	2422
Hospital-years	892			
Surgeon volume:				
Planned hip replacements (baseline)	56.67	40.69	10	270
With emergency hip replacements	66.63	44.24	10	589
With emergency hip + planned knee replacements	121.83	68.22	10	589
Surgeon-years	4619			

Table A.8: Correlation between distance to closest hospital and pre-surgery severity

	Pre-surgery OHS		Elixhauser comorbidities		
	(1)	(2)	(3)	(4)	(5)
Distance to hospital	0.011	-0.007	-0.002	-0.007**	-0.003
	(0.010)	(0.018)	(0.002)	(0.002)	(0.003)
Distance to hospital (squared)		0.001		0.000*	0.000
		(0.000)		(0.000)	(0.000)
Patient age	-0.042***	-0.042***	0.028***	0.028***	0.028***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)
Male patient	2.333***	2.333***	0.015*	0.015*	0.015*
	(0.057)	(0.057)	(0.007)	(0.007)	(0.007)
Deprivation index (rank)	0.000***	0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ethnicity: white	-0.595**	-0.595**	0.147***	0.147***	0.148***
	(0.186)	(0.186)	(0.020)	(0.020)	(0.020)
Constant	17.623***	17.684***	-0.742***	-0.728***	-0.738***
	(0.354)	(0.367)	(0.043)	(0.044)	(0.045)
R^2	0.041	0.041	0.060	0.060	0.060
Observations	100930	100930	100930	100930	100347

Notes: Distance to the hospital is patient's distance to their closest hospital. Controls also include year dummies. In column (5), the top percentile of distances in our sample were dropped. The deprivation index ranks all small geographical areas in England (LSOAs) and goes from one (the most deprived) to 32,844 (the least deprived). In parentheses, robust standard errors clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.9: Results of the effect of hospital volume on 3-year revision rates

	Pooled OLS		Fixed Effects	
	(1) Observed	(2) Predicted	(3) Observed	(4) Predicted
Volume [Ref.≥300]				
<150 cases	0.241^{+}	0.116	0.572^{+}	-0.160
	(0.143)	(0.154)	(0.332)	(0.318)
150-200 cases	-0.043	0.090	0.242	-0.000
	(0.159)	(0.145)	(0.258)	(0.232)
200-300	0.001	0.003	0.240	0.039
	(0.111)	(0.116)	(0.170)	(0.167)
R^2	0.003	0.003	0.006	0.006
Observations	105229	105229	105229	105229

Notes: In parentheses, robust standard errors are clustered on hospitals. Patient controls include the pre-surgery OHS score, patient age, sex, ethnicity, index of multiple deprivation, the number of Elixhauser comorbidities, osteoarthritis diagnosis, control for the type of surgery carried out (hybrid prosthetic versus total hip replacement) and for having had a surgery on the other hip, self-reported disability, depression, assistance in filling questionnaires, the symptoms duration and the patient' living arrangements. Hospital controls include hospital status (teaching, specialist orthopaedic and Foundation Trust hospitals), for the market forces factor index, and for socio-demographic characteristics of population in hospitals' catchment area (proportion of over 65-year-olds, mean deprivation rank and mean distance to the closest GP). + p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001

Table A.10: Determinants of participation to PROM questionnaires for Oxford Hip Score

	PROMs participation rate		
	(1)	(2)	(3)
Hospital volume	0.017*	0.013	0.012
	(0.007)	(0.007)	(0.008)
Average patient age		-0.899**	-1.054*
		(0.341)	(0.458)
% of male patients		0.151	-0.009
		(0.166)	(0.175)
% of patients of white ethnicity		0.099	-0.066
		(0.098)	(0.114)
Average deprivation rank		0.000	0.000
		(0.000)	(0.000)
% of patients with surgery on other hip		0.406**	0.343**
		(0.131)	(0.116)
Average Elixhauser comorbidity count		1.074	1.232
		(2.620)	(2.625)
% of patients with osteoarthritis		0.367***	0.268**
		(0.083)	(0.093)
% of patients with hybrid prosthetic surgery		0.044	0.100
		(0.055)	(0.052)
% of patients who died after surgery			-0.487
			(2.556)
Teaching hospital			-3.534
			(2.347)
Specialist hospital			-4.114
			(6.443)
Foundation trust			4.006*
			(1.980)
Market forces factor			-2.978
			(1.834)
% of pop. over 65 years old (catchment area)			-1.717**
			(0.609)
Mean deprivation rank (catchment area)			-0.000
			(0.000)
Mean distance to GP (catchment area)			7.633**
			(2.555)
R^2	0.080	0.150	0.197
Observations	892	892	892

Notes: Results from a Linear Probability Model with 0-100 response variable. The rate of participation to PROMs questionnaire is the number of patients who answered the PROMs questionnaire (one or both pre- and post-surgery questionnaires were answered) out of the total number of eligible patients per hospital, expressed as a percentage. Patient level variables are aggregated at the hospital level and expressed as a percentage, so that a one unit increase corresponds to one percentage point increase. The market forces factor is standardized by its sample standard deviation (0.07). The index of multiple deprivation ranks all small geographical areas in England (LSOAs) and goes from one (the most deprived) to 32,844 (the least deprived). A hospital catchment area comprises all the small homogenous geographic areas (Lower Super Output Areas, LSOAs) whose centroids fall within 30 km of the hospital's headquarters. In parentheses, robust standard errors clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 2

The effect of short breaks on performance: Evidence from the medical workforce

1 Introduction

The organisation and management of human resources matter for firms' performance (Bloom and Van Reenen, 2007; Lazear and Shaw, 2007). Within firms, workforce turnover, levels of staffing or team composition can impact workers' productivity, by affecting the way workers learn from task repetition, learn from each other or coordinate (Cook et al., 2012; Bartel et al., 2014; Chan, 2016, 2021). The labour and medical literature have also explored the role of individuals' work schedules, to understand how working hours and shift patterns impact workers' productivity and wellbeing (Caruso et al., 2006; Brachet et al., 2012; Caruso, 2014; Pencavel, 2015). While long working hours are generally associated with poorer performance, regular breaks are beneficial for workers' mental well-being, physical health and safety (Spurgeon, 2003), explaining why most national and international labour regulations mandate regular rest periods[27].

However, regular breaks may also impact workers' performance, if individual performance benefits from repeated practice (Ho, 2014). The economics literature posits that experience can depreciate with interruptions in practice, such that breaks of various lengths may lead to skill depreciation for certain tasks (Besanko et al., 2010). Empirically, breaks in activity induced

[27] In the European Union for instance, Working Time Directive 2003/88/EC sets a limit to the weekly hours of workers and commands regular periods of rest.

by drops in production levels have been shown to affect worker productivity across a variety of production settings, such as franchise shops, assembly lines, aircraft or ship production (Argote and Epple, 1990; Darr et al., 1995; Benkard, 2000; Shafer et al., 2001; Thompson, 2007), though the effect varies across sectors. Alternatively, breaks in activity may reduce decision fatigue. The psychological literature has shown that repeated decisions being made in a row can lead to worse judgement making by judges, voters or healthcare workers (Danziger et al., 2011; Flinn and Armstrong, 2011; Augenblick and Nicholson, 2016; Hunt et al., 2021), potentially impacting performance. For instance, nurses are more likely to refer patients and doctors are more likely to prescribe antibiotics as the day passes, reflecting a shift to more conservative or less appropriate clinical decisions (Linder et al., 2014; Allan et al., 2019). In the healthcare sector, even short breaks in activity can have important implications for patients.

This paper therefore explores whether surgeons' time breaks, defined as the number of days since their last surgery, have a causal impact on 30-day mortality rates after hip fracture surgery. To this end, I construct a panel of over 2,000 orthopaedic surgeons in the English National Health Service (2009-2016). Hip fracture is the most common reason for admission into emergency orthopaedic wards, with around 70,000 to 75,000 hip fractures occurring annually in the United Kingdom (National Clinical Guidelines Centre, 2011, p.6). Post-surgical mortality is high, at around seven percent one month after hip fracture in 2017 (Healthcare Quality Improvement Partnership, 2018), suggesting potential for improvement at the aggregate level given the high incidence of hip fractures.

To mitigate concerns of endogeneity linked to surgeon selection, I estimate a fixed effects model where surgeon fixed effects allow for unobserved time-invariant heterogeneity in surgeon ability. Interacted hospital-year fixed effects further control for flexible hospital-specific time trends in health outcomes over the sample period. Results show that, for breaks between four and six days, average 30-day mortality rates decrease by about six percent (around 0.4 percentage points), compared to surgeons who were in the operating room the day before. A possible interpretation of the findings is that surgeons who treat a hip fracture patient after a short break perform better because they are less fatigued.

Only a few studies have explicitly modelled the effect of time breaks in surgical practice on quality of care. In a study on a cardiac procedure (transcatheter aortic valve implantations, TAVI) for one Belgian hospital, Van Gestel et al. (2017) find no effect of time breaks on 24-month mortality. They however estimate that an additional day between TAVI practices is associated with a higher probability of adverse events (renal failure and stroke) but show that this effect is driven by the most extreme values of time breaks, i.e. above 100 days. Using a panel data on trauma-related ambulance runs in the U.S., David and Brachet (2011) find evidence of skill decay among paramedics, showing that additional days of inactivity adversely impact paramedics' performance, measured by ambulances' out-of-hospital time. Fluctuations in performance measured at the hospital or team level may however also arise from labour turnover, therefore confounding the effect of time breaks for individual surgeons (David and Brachet, 2011).

The small evidence available on the effect of time breaks for individual surgeons suggests, in most cases, an adverse effect of breaks on outcomes for cardiac procedures. Hockenberry and Helmchen (2014)'s study on coronary artery bypass grafting (CABG) in the United States finds that an additional day away from the operating room raises in-hospital mortality by 2.4%. A previous study by Hockenberry et al. (2008) suggests similar effect of time breaks for CABG and percutaneous coronary intervention (PTCA), a less invasive cardiac surgery, using Taiwanese data. The authors indicate that the one-month likelihood of death after CABG is 17% higher for surgeons after a 3-14 day breaks relative to a 0-2-day break. However, another U.S. study by Huesch (2014) compares patient outcomes for cardiac surgeons who have performed a CABG in the previous month and those who have not, and finds no association with post-surgical length of stay or mortality.

There is limited evidence on the potential mechanisms behind the effect of time between surgeries on health outcomes. Previous studies have interpreted the adverse effect of time breaks on patient health outcomes as evidence of depreciation of surgical skills. Hockenberry and Helmchen (2014) find that time breaks both increase patient mortality and reduce hospitalisation costs, which they suggest indicates that surgeons' lower performance after time breaks is due to inattention and insufficient care being provided. Hockenberry et al. (2008) test whether the level of task repetition prior to the break, measured by the volume of activity prior to the break, may explain the adverse effect but find little evidence of this. David and Brachet (2011)'s findings suggest that overall work inactivity and tasks interference (i.e. performing a large range of tasks) contribute to paramedics' lower performance, measured by out-of-hospital ambulance time, but do not investigate the impact on patient outcomes.

This study makes several contributions to the literature on the effect of time breaks on patient health outcomes. First, in most settings, time breaks are not exogenously determined. Changes in practice are likely to respond to changes in performance, resulting in a reverse causality bias. Surgeons who become worse at treating patients may take breaks more often, which will introduce a spurious negative relationship between surgical ability and breaks length. Following David and Brachet (2011), this endogeneity concern is mitigated by focusing on an emergency condition[28]. Patients break their proximal femur, often after a fall, and need to be treated within 48 hours of admission to the hospital (National Clinical Guidelines Centre, 2011). Variation in surgeons' time breaks is driven by unanticipated emergency admissions in pre-determined schedules, and thus arguably exogenous conditional on the large set of controls.

Second, this study adds to the understanding of the effect of time breaks on performance, by testing for heterogeneity in the effect of time breaks along different dimensions of surgical practice. Breaks are defined here as the time elapsed since any orthopaedic surgery; orthopaedic surgeons tend to perform a range of bone-related surgeries, from hips to knees, ankles or upper limbs. Heterogeneity analyses indicate that the effect of breaks may be more important for surgeons with less experience in hip fracture care, as measured by their annual volume of hip fracture patients, but is not affected by surgeons' degree of specialisation in hip fracture care. In addition, I investigate potential mechanisms behind the effect of time breaks. I find no evidence that surgeon skills deteriorate with time breaks as was suggested in previous studies on cardiac procedures. Allowing time breaks to be task-specific, thereby

[28]Related studies commonly include surgeon fixed effects to account for the correlation of ability and average time breaks across surgeons. However, within-surgeon variations in activity and performance over time may be spuriously correlated, especially for planned conditions when surgeons choose when and whom to operate.

measuring the effect of the number of days since last hip fracture surgery and excluding related orthopaedic surgical activity, does not change the overall results: breaks in hip fracture activity, like general time breaks, have a null or positive effect on patient health outcomes.

Further, I explore whether surgeons who return to the operating ward after some days off choose a different type of surgery (namely, total/partial hip replacement versus reduction or internal fixation of hip fracture using nails or screws), holding patient characteristics and the type of fracture fixed. If the choice of surgery type explains the effect of time breaks, specific treatment can be enforced by stricter enforcement of clinical guidelines (Chowdhury et al., 2007). Findings show that short time breaks also impact the type of surgery carried out, holding patient characteristics fixed, by decreasing the probability of receiving a full or partial hip replacement surgery (i.e. increasing the probability of fracture reduction using nails or screws). Importantly, I provide some evidence that these changes in treatment choices are not due to differences in the type of patients treated after a break, but rather point to differences in surgeons' decision after a break.

The rest of this paper is structured as follows. Section 2 describes the institutional context for surgeons in England. Next, Section 3 introduces the data used and documents the occurrence of time breaks in surgeons' activity. Section 4 details the econometric strategy. Section 5 presents the results and robustness checks, while Section 6 concludes.

2 Institutional setting

Health care in England is free at the point of use for residents and primarily funded by general taxation. Most hospital care is provided in public (National Health Service, NHS) hospitals. Patients are seen by surgeons in hospitals either after being referred by their family doctor (called general practitioner, GP) who act as gatekeepers to planned hospital care, or after being admitted directly to the hospital through the Emergency Department (Ikenwilo and Scott, 2007).

After four to six years of undergraduate medical training, medical students, called junior doctors, undertake postgraduate training consisting in two years of general training (foundation years), followed by specialty training (British Medical Association, 2020). Upon successful completion of the specialty training programme, surgeons can become consultants, which is the most senior status for surgeons in the NHS and refers to surgeons who lead on clinical decisions with a team including nurses and junior doctors. An estimated 40,395 consultants were working for the NHS in 2012 (Department of Health, 2013). Consultants' basic salary is negotiated nationally at the U.K. level and evolves with career progression along a national scale (Ikenwilo and Scott, 2007). The basic payment rate can however be modified to account for extra work or work undertaken on call during unsocial hours. Additional payments can arise if surgeons receive distinction awards, linked to excellence in clinical practice.

Consultants' activity is regulated under a consultancy contract which was reformed in 2003 to improve accountability of surgeons' activity and increase their starting salary[29]. The contract stipulates that full-time consultants must take on minimum 10 programmed activities per week for the NHS, where one programme activity corresponds to around 4 hours during normal working hours. A typical programmed activity may involve a) direct clinical care comprising patient care, ward rounds, theatre sessions and emergency work duties and on-call work; b) supporting professional activities such as training or teaching, NHS responsibilities such as clinical tutor or director; or c) external duties (e.g. trade union or work for the royal colleges) (Williams and Buchan, 2006). Emergency work includes both scheduled activities, as part of surgeons' rotas in programmed activities, as well as additional work arising during on-call duties. Consultants may also work for the private sector on the side, though they are required to undertake an additional programmed activity for the NHS first. A typical consultant contract balances programmed activities to around 7.5 programmes activities of direct clinical care and 2.5 of supporting professional activities, though that may vary. In reality, contracts may often exceed the minimum of 10 programmed activities, especially in acute specialities (British Medical Association, 2009).

In compliance with the European Working Time Directive, since 1998 the working hours of

[29]Consultants who started after 2003 were automatically covered by the contract while consultants who started before could choose to switch to the new contract. In 2012, 97% of consultants were under the 2003 contract (Department of Health, 2013).

consultants are limited to a maximum of 48 hours on average per week under the Working Time Regulation. Junior doctors in training were later included in the regulation in 2004, though they retain the choice to opt out, e.g., to pursue more training for instance. In addition, the directive stipulates that consultants are entitled to 11-hour rest a day out of work, a day off each week, a rest break for working days longer than six hours and 5.6 weeks paid leave each year (British Medical Association, 2021). Consultants may also refuse non-emergency work that arise during unsocial hours, ie., outside of 7am to 7pm during the week and anytime during the weekend. However, consultants may exceed the maximum average weekly working hours if they also work for the private sector as the directive only covers time for employees (Ikenwilo and Scott, 2007).

Despite efforts to reconsider consultants' contracts, the NHS has been under substantial staffing issues. Staff shortages were estimated at 100,000 vacancies in 2018 (The King's Fund, 2018), and healthcare staff report chronic excessive workload (Wilkinson, 2015; West, 2020). The number of doctors and nurses rose by slightly under 10% between 2010 and 2016 but at a lower rate than hospital activity, with total inpatient admissions soaring by around 17% during the same period (Propper et al., 2020).

3 Data

3.1 Sample

Using a detailed patient-level hospital administrative data set for all National Health Service hospitals in England called Hospital Episode Statistics (HES), I extract hospital records for all patients admitted as emergency patients with a diagnosis of hip fracture between 2009 and 2016[30]. The patient hospital records are then merged with a data set on surgeon characteristics (surgeon sex, year and country of qualification, training and status on the medical register) provided by the General Medical Council.

Proximal femoral fracture, or hip fracture, usually occurs among elderly patients after a

[30] Hip fracture diagnosis is based on the International Classification of Diseases (ICD)-10th revision codes: S720, S721, S722 and S729.

fall, often as a result of underlying poor health (National Clinical Guidelines Centre, 2011). Hip fracture patients who are deemed too frail to undergo surgery are treated non-surgically, often by non-orthopaedic surgeons (e.g. ortho-geriatrists), though conservative treatment is deemed rarely appropriate[31]. The analysis focuses on hip fracture patients who are treated surgically by orthopaedic surgeons, to ensure that the surgeon population and surgical tasks are homogenously defined in the sample. Surgeons who treat less than 30 hip fracture patients in the overall study period (2009-2016) are excluded (N=752 surgeons). This ensures that health outcomes are estimated on a sufficiently large number of cases, whilst removing reporting errors in the data or surgeons who very occasionally treat hip fracture cases[32]. Further, the final sample excludes surgeons who have received a fitness to practice warning or have been suspended or erased from the medical register, most likely after a medical error (N=25). The assumption that time breaks are uncorrelated with a surgeon's ability would not hold for these surgeons. The final data set is an unbalanced panel of 2,124 orthopaedic surgeons over eight years.

The surgeons identified in the administrative hospital data are the doctors who are responsible for a patient's episode of care, called consultants in the NHS. Though clinically responsible for the patient's care, consultants may not be carrying out the surgery themselves, e.g. if they are supervising a junior trainee surgeon. Observations where surgeons report more than four hip fracture cases in a given day are dropped from the sample (i.e. one percent of the initial sample) as surgeons may not be carrying all surgeries themselves[33], which would introduce measurement error in the definition of time breaks. This potential data limitation is further addressed in Section 5.4.

Hip fractures are more frequent among an older population, though traumas or accidents leading to a hip fracture may happen for younger patients. Therefore, only patients below 30-year-old, for whom the risk of having a hip fracture is substantially lower, accounting for less

[31] Patients with very short life expectancy, complete immobility or who refuse surgery would receive conservative treatment for their hip fracture (British Orthopaedic Association, 2007, p.20).

[32] 30 is used as a cut-off as it is generally considered the minimum sample size required for the central limit theorem to hold. Note that this corresponds to 30 patients over the whole study period (ie, some surgeons may treat fewer patients per year) and thus remains a conservative sample restriction.

[33] A hip replacement surgery takes around one to two hours, according to the NHS (NHS, 2020). On average, a surgeon will be able to complete at most four surgeries in an eight-hour work shift.

than one percent of emergency hip fracture patients, are removed from the sample[34]. The final sample excludes admissions for revision surgeries, which is a more complex and rare surgery consisting in removing and replacing an existing artificial hip implant (i.e. around one percent of the sample). Patients who are transferred from another hospital are also removed from the sample (i.e. around five percent of the initial hip fracture admissions), as transfers may be decided based on a patient's particular surgical needs, and thus will be endogenous to hospital quality. Extreme values for pre-surgery length of stay above 30 days representing less than one percent of hip fracture admissions and probable coding errors or outliers are dropped from the sample.

This paper focuses on the effect of surgeons' time breaks, exploiting observed breaks in surgeons' activity. To this end, observations for which surgeon time breaks between surgeries lies above 31 days are excluded, in order to remove outliers or potential errors[35]. These are rare (i.e. top one percentile of the distribution) and may be due to misreporting or exceptionally long work leaves (e.g. severe illness or incapacity to work). Restricting the sample ensures that extreme values of breaks are not driving the results. After sample exclusions, the final data set counts 371,271 observations over eight years (2009-2016).

3.2 Dependent variable

Post-surgical health outcomes are measured using the 30-day mortality rate, which is supplied by the Office of National Statistics and linked to the hospital records data. The 30-day mortality rate was on average seven percent for hip fracture patients in our sample, suggesting that post- surgical death is not negligible (Healthcare Quality Improvement Partnership, 2018). Patient mortality is measured for 30 days starting from the date of patient admission, also following patients after being discharged from the hospital.

[34]Young patients are more likely to break their femur as a result of an important trauma (such as an accident), as opposed to older patients who are more likely to suffer from osteoporosis and could break their femur after a fall. This would result in unobservable differences in the severity of the fracture beyond what can be accounted for via controls for the type of fracture (ie, neck of femur, pertrochanteric, subtrochanteric or else). Removing very young patients therefore ensures better homogeneity in the profile of patients treated for a hip fracture. In addition however, I also show that this sample restriction does not affect the results. Appendix Table B.1 presents results after inclusion of patients under 30 years old and controls for smaller age groups (5-year age bands).

[35]Relaxing this sample restriction however does not affect the results.

The administrative data set only reports all-cause deaths, therefore the exact cause of patient mortality is not known. In all likelihood, deaths occurring within one month of a major surgery after hip fracture are linked to the surgery. Further, this data limitation would introduce bias in the results only if unrelated deaths are systematically correlated with surgeons' variation in time breaks. There are little theoretical reasons for this to be the case.

3.3 Independent variables

The key independent variable is the number of days, including weekend days, bank holidays or holiday breaks, since the operating surgeon's last orthopaedic surgery. The latter is not limited to emergency hip fracture surgery and includes planned surgeries on bones or joints identified with OPCS chapters, mostly involving joint replacement surgeries for upper or lower limbs. The data set contains information on the date of surgery but not on the exact time of the surgery. Therefore, I do not know the order in which surgeries took place in a given day. The variable for time breaks thus only considers whether the surgeon performed a surgery in the previous days, but do not take into account surgeries which took place during the same day. As a result, the minimum value for the time breaks is one and varies by surgeon and day of surgery.

The control variables include a full set of surgeon dummies (surgeon fixed effects) and interacted hospital-year dummies (hospital-year fixed effects). Surgeon performance may be influenced by their volume of hip fracture activity, e.g. if there are positive learning effects from treating more hip fracture patients. To separate the effect of time breaks from the potential effect of experience, I control for surgeons' yearly volume of hip fracture cases, equal to the number of hip fracture cases treated in the 365 days leading up to the patient admission.

The model includes a comprehensive vector of clinical and socioeconomic patient characteristics, which may be associated with poorer health outcomes. Namely, patient sex, age (via dummies for 10-year age categories to reflect potential non-linearity in the effect of age), ethnicity, diagnosis of osteoporosis and the number of emergency and total hospital admissions in the year leading up to the hip fracture episode, are controlled for. The control variables also include patients' socio-economic status, proxied by the percentage of population who experience economic deprivation in the patient's residence area level (known as lower layer super output area, LSOA[36]) and produced by the Office for National Statistics.

Hip fracture often signals underlying poor patient health and comorbidities. To thoroughly account for the potential confounding effect of other comorbidities on health outcomes, the vector of control variables includes a full set of indicator variables for each Elixhauser condition (Elixhauser et al., 1998) (see Appendix, Table B.2 for a full list and associated prevalence in the sample). Importantly, I also control for the type of hip fracture which are associated with different levels of severity and complexity, by including indicator variables corresponding to each type of hip fracture (neck of femur, pertrochanteric, subtrochanteric, or unspecified). In particular, intracapsular fractures, i.e. involving the femoral head or neck of the femur, may disrupt the blood supply of the femoral head (National Clinical Guidelines Centre, 2011). Medical guidelines recommend early treatment of the fracture (Healthcare Quality Improvement Partnership, 2018). Health outcomes may be impacted by a long time between patient fall leading to the fracture, and surgery. To proxy this, I control for patients' pre-surgery length of stay, defined as the number of days between patient admission and surgery. Patients who are admitted on a weekend day may be unobservably more ill (Meacock et al., 2019). Similarly, patients treated on the weekend may receive a lower level of care, if there are lower levels of staffing or available services at the weekend. Indicator variables for admission and surgery during a weekend day are included to account for this possibility.

3.4 Summary statistics

Table 2.1 shows that the 30-day mortality rate decreases over the period studied, reflecting overall improvement in hip fracture care. The overall number of hospitals and surgeons varies over time, following hospital mergers or openings, and influx of new surgeons or surgeons' retirement from practice.

[36]There are over 32,000 LSOAs in England with an average population of 1,500.

Table 2.1: Aggregate mortality rates after hip fracture, by year of sample

	2009	2010	2011	2012	2013	2014	2015	2016
30-day mortality (mean)	7.16	6.85	6.76	7.01	6.36	6.38	5.95	6.07
N hospitals	144	143	144	141	140	139	135	133
N surgeons	1450	1538	1586	1623	1650	1683	1656	1579
Observations	42295	45452	46348	46915	48343	48009	47510	46399

Notes: Aggregate statistics per year of panel, for all patients admitted emergently with a diagnosis of hip fracture, after exclusion of uncommon patient characteristics, and low-volume surgeons

Patients in the sample are 81 years old on average and 28% are male (Table 2.2). The majority of the sample is white. 69% of the admissions for hip fracture involve a fracture of the neck of the femur. The average patient lives in areas where 14% of the population is economically deprived. 15% of the sample have a diagnosis of osteoporosis. On average, patients report slightly over two Elixhauser comorbidities, 0.55 emergency admissions and one hospital admission in the year leading up to current admission. Pre-surgery length of stay is two days on average. 27% of the sample are admitted (treated) during a weekend day. Around half of the sample (48%) receives a hip replacement surgery, as opposed to a reduction of fracture using nails or screws. On average, patients stay in hospital for 21 days after the surgery. Summary statistics in Table 2.2 show that some patients have very high values of pre and post length of stay or number of hospital admissions. Long pre-surgery lengths of stay may occur if patients break their femur while already in hospital. Though plausible, these outliers may signal unobserved patient severity. However, because these are emergency patients, unobserved severity is unlikely to be systematically related to surgeons' break length and therefore unlikely to introduce bias in my estimates. To ensure that extreme values are not influencing the results, I run again the main specification after applying 99% winsorisation, where (remaining) extreme values are set to the 1 percentile values for all covariates. Results, shown in Appendix Table B.3, are unchanged.

Table 2.2: Summary statistics for patient characteristics

	Mean	Std. Dev.	Min.	Max.
30-day mortality	0.07	0.25	0	1
Patient age	81.14	10.95	30	113
Male patient	0.28	0.45	0	1
White ethnicity	0.92	0.28	0	1
Income deprivation	0.14	0.10	0	1
Diagnosis of osteoporosis	0.15	0.35	0	1
Elixhauser index	2.28	1.77	0	19
Past emergency admissions	0.55	1.11	0	38
Past all admissions	1.03	4.47	0	207
Pre-surgical length of stay	1.44	2.47	0	30
Surgery during weekend	0.26	0.44	0	1
Admission during weekend	0.27	0.44	0	1
Type of fracture				
Neck of femur	0.69	0.46	0	1
Pertrochanteric	0.26	0.44	0	1
Subtrochanteric	0.04	0.20	0	1
Unspecified fracture	0.00	0.05	0	1
Partial or complete hip replacement	0.48	0.50	0	1
Post-surgical length of stay	21.26	19.00	0	111
Observations	371271			

Notes: Income deprivation measures the proportion of the population suffering from income deprivation in the area of residence of the patient (i.e. Lower Super Output Area). Pre-surgery length of stay is defined as the number of days between patient admission and surgery.

Table 2.3 indicates that 96% of surgeons in the sample are male. Surgeons obtained their medical degree between 1968 and 2005 and the majority (68%) received their medical qualification in the United Kingdom. The data set is an unbalanced panel. The average surgeon is observed for six years and works for slightly more than one hospital (1.20) over the period. The average time break in the sample lies slightly above one, indicating that surgeons are in the operating ward almost every day (Table 2.4). Surgeons may perform different orthopaedic surgeries, such as (planned) hip or knee replacements along with hip fracture surgeries. Table 2.4 therefore also reports specifically the number of days since a surgeon's last hip fracture surgery. On average, around 11 days elapsed between surgeon's hip fracture cases, though the median value is much lower, at only three days. The yearly surgeon volume averages 50 hip fracture surgeries, ranging from one[37] to 763 yearly cases[38]. Overall, individual surgeons' volumes of hip fractures represent around 14% of a surgeon's activity in a year.

Table 2.3: Summary statistics for surgeon characteristics

	Mean	Std. Dev.	Min.	Max.
Male	0.96	0.20	0	1
Year of qualification	1990.68	7.62	1968	2005
Junior surgeons (i.e. qualified in >1998)	0.21	0.40	0	1
Trained in the UK	0.68	0.47	0	1
#Years in panel	6.01	2.20	1	8
#Hospitals where surgeons work	1.20	0.54	1	6
Observations	2124			

Notes: Data on surgeon characteristics were obtained from the General Medical Consult (GMC).

Because my empirical strategy relies on within-surgeon variation in time breaks (via the surgeon fixed effects), I also report the within-surgeon standard deviation for the main surgeon characteristics. The statistics in Table 2.4 confirm that most of the variation in time breaks occurs within rather than between surgeons.

Table 2.4: Distribution of time breaks and surgeon activity

	Mean	Median	Std. Dev.			Min.	Max.
			Total	Between	Within		
Days since last orthopaedic surgery	1.77	1.00	1.97	0.58	1.91	1.00	31.00
Days since last hip fracture	11.38	3.00	15.83	4.67	15.40	1.00	100.00
Yearly volume of hip fractures	50.68	41.00	66.21	19.71	14.56	1.00	763.00
% of activity in hip fractures	13.85	11.68	10.51	5.92	3.52	0.19	100.00
Observations	371271						

Notes: The number of days between two cases varies by surgeon and day of surgery. The days since last surgery includes any type of orthopaedic activity, identified by the broad OPCS chapters on bones and joints (starting by letters V, W or O). The proportion of activity in hip fractures corresponds to the proportion of a surgeon activity spent on hip fracture patients over a year.

[37] Surgeons with less than 30 hip fractures over the whole sample period are excluded. However their yearly volume can be lower than 30 cases.

[38] As a hip replacement lasts around two hours, a surgeon can treat around four hip fractures in 8-hour workday. Multiplied by the total number of working days, the maximum possible surgeon yearly volume is estimated at around 800 cases.

4 Methods

4.1 Baseline model

The aim of this study is to understand whether patients' health outcomes are affected by the time elapsed between a surgeon's orthopaedic surgery cases. A fixed-effects model is specified as follows:

$$y_{isht} = \beta_0 + D_{isht}^{'}\beta_1 + \beta_2 ln(volume)_{st} + X_{isht}^{'}\beta_3 + \alpha_s + \delta_{ht} + \epsilon_{isht}, \tag{2.1}$$

where y_{isht} is the post-surgical 30-day mortality for patient i treated by surgeon s in hospital h and admitted to hospital at time t. X_{isht} is a vector of patient characteristics which may be associated with poorer health outcomes, namely, patient sex, age in 10-year categories, ethnicity, economic deprivation[39], past hospital utilisation, indicator variables for each Elixhauser comorbidity, pre-surgery length of stay, admission or treatment during a weekend day and the type of hip fracture.

Surgeons' time breaks, D_{isht}, are a set of indicator variables corresponding to 1 day, 2-3, 4-6, and above 7 days since the last surgery for the operating surgeon s. This allows for a non-linear effect of breaks on health outcomes; without assuming a specific functional form. Though arbitrarily defined, the categories broadly correspond to observed patterns of the data, from a few days breaks to a week or more (above 7 days)[40]. The categories are defined to span the distribution of time breaks but increase in magnitude along the distribution to ensure sufficient statistical power. Figure B.1 in the Appendix shows the distribution of the categories in the sample. The effect of interest is in the vector of coefficients β_1 for the effect of surgeons' time breaks on patients' health outcomes. In addition, I report results using linear, linear-quadratic and log of surgeons' breaks[41].

[39]The data used measure economic deprivation in 2010 and was not updated after that. However, changes in deprivation level over time across areas will be accounted for by the hospital-year fixed effects.

[40]These broadly correspond to a weekend, a long weekend with bank holidays or a longer holiday break. Furthermore, I report in the Appendix the association between a dummy variable for each value of day break (from one to a 30 day break) and health outcomes which inform the grouping into categories (Table B.4).

[41]Results using these alternative specifications indicate a significant effect of time breaks for non-linear func-

Surgeon fixed effects, α_s, account for time-invariant heterogeneity in surgical ability. Changes in surgeons' volume of hip fractures over the years may be associated with improvement in health outcomes as well as shorter time breaks. To isolate the effect of time breaks on health outcomes from potential confounding surgeon factors, Equation (2.1) includes the surgeon's yearly lagged volume of hip fracture patients, vol_{st}. Hospital-year fixed effects, δ_{ht}, account for aggregate changes in health outcomes over time. While year dummies assume homogenous time trend across hospitals, interacted hospital-year fixed effects allow for a hospital-specific flexible time trend[42] (Gormley and Matsa, 2014). This controls for unobserved hospital heterogeneity in outcomes over time, linked to e.g. differences in the adoption of medical guidelines (i.e. for teaching vs general hospitals), one-off investments in technical equipment or facilities, changes in the medical team or in the supply of long-term care around the hospital (Gaughan et al., 2017). These factors, if they correlate with changes in surgeons' frequency of activity, would be important confounders. ϵ_{isht} is an idiosyncratic error term.

Equation (2.1) is estimated by a linear probability model (LPM) which is often used to model health outcomes with fixed effects. Probit or logit models suffer from the incidental parameters bias with fixed effects (Greene, 2004; Cameron, 2009). The user-written Stata command reghdfe (Correia, 2016) is used given high dimension of the fixed effects. Standard errors are adjusted for clustering at the hospital level to account for geographical correlation across patients of a given hospital[43].

4.2 Endogeneity concerns

Causal identification of the effect of time breaks on health outcomes relies on the assumption that surgeons' variations in time breaks are exogenous, conditional on the set of fixed effects and controls.

The model with surgeon fixed effects specified in Equation (2.1) accounts for unobserved surgeon ability as long as it is time-invariant. Endogeneity issues may however arise if changes

tional form (linear-quadratic and log), though the effect is significant at the 10% level.

[42] Note that hospital-year fixed effects are not perfectly collinear with surgeon fixed effects as surgeons may work for several hospitals.

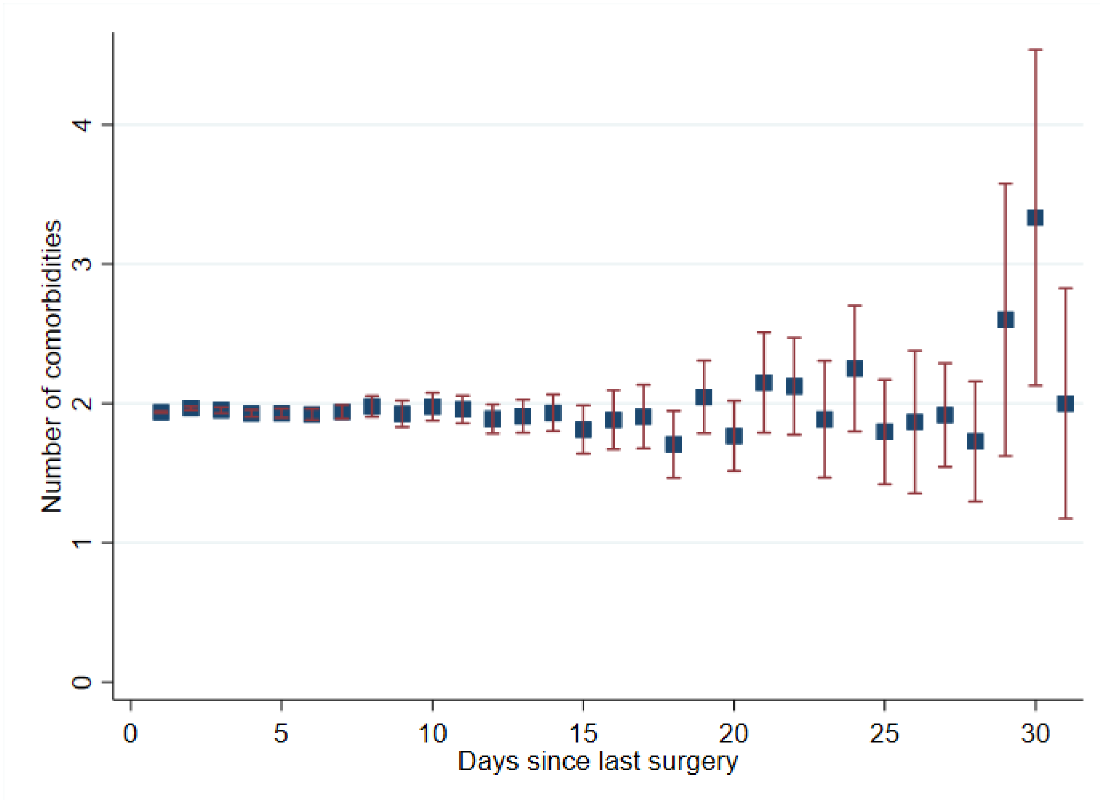
[43] Clustering standard errors at the surgeon level alongside hospital level, to account for serial correlation due to some surgeons changing hospitals, did not affect the results.

in surgeons' frequency of practice responds to changes in surgeons' ability. In particular, surgeons whose surgical ability improves may start treating patients more often and thus, have shorter time breaks. Equally, patients may choose to go more to hospitals where quality has improved. This potential reverse causality will introduce a positive bias in the estimates of time breaks on surgeons' ability: lower patient mortality will be associated with shorter time breaks.

This endogeneity concern is mitigated here in the context of an emergency procedure. Due to the unpredictable nature of hip fracture, the volumes and timings of patient admissions are not anticipated by surgeons. Further, patients need to be treated within 48 hours after the hip fracture, which leaves little room for patient choice of hospital or surgeon based on reputation of quality. Figure B.2 in the Appendix shows that the large majority of emergency hip fracture patients (78% of the sample) do not bypass their closest hospital.

Similarly, in the emergency setting, surgeons have little control on which patients to treat, thus alleviating concerns that surgeons would choose to treat less complex patients after a longer time break. Figure 2.1 plots the average number of Elixhauser comorbidities along the distribution of time breaks. The number of comorbidities varies only slightly along the distribution of time breaks, indicating that observed patient severity does not differ greatly across time breaks. The greater variation in number of comorbidities towards the higher values of breaks reflects the lower number of cases. The same absence of association is observed for patient age, another important proxy of severity, and length of time breaks (see Appendix Figure B.3), further suggesting that bias due to remaining unobserved severity is unlikely.

Figure 2.1: Mean number of Elixhauser comorbidities, by break length



Notes: Mean and 95% confidence intervals for the number of Elixhauser comorbities per number of days since last surgery.

In the rest of this section, I provide additional evidence that surgeons' time breaks are mostly driven by exogenous factors, such as seasonality effects in the incidence of hip fractures or surgeons' fixed work schedule. Figure B.4 in Appendix shows that volumes of admissions are perceptibly higher over the winter months[44]. Whilst the volume of hip fracture admissions remains stable across the days of the week (Figure B.5, in Appendix), surgeons' time breaks vary depending on the day of admission, reflecting surgeons' pre-determined work schedule or on-call duties over the weekend (Figure B.5, bottom plot). As expected, the average number of days of breaks without surgical practice is highest after the weekend and decreases afterwards.

[44]Whilst monthly variations in hip fracture admissions provide interesting within-surgeon variation, the quality of care may also vary across months, due to, e.g., differences in staffing or higher demand for hospital care in winter. In additional analyses, I account for potential seasonality in quality of care, by including month dummies. The results (Table B.6 in Appendix) are fully robust to these controls.

I further test whether having had an adverse event (i.e. patient death) influences the surgeon's frequency of activity. I create an indicator variable equal to one if a surgeon has had a patient die in the hospital in the past 30 days. Table B.7 in Appendix shows that having had a recent patient death is not associated with longer time breaks, indicating that surgeons do not stay out of practice longer after an adverse event.

5 Results

5.1 Main results

Regression results for the effect of time breaks in surgical practice are reported in Table 2.5, for the main coefficients of interest. Column (1) first shows the results with the minimum set of controls, and column (2) presents results from the full set of patient controls. Coefficients are stable across specifications[45]., and I therefore focus on the full specification. The full table of coefficients for the regression with all controls is available in the Appendix (Table B.8). Results from Table 2.5 show that mortality rates decrease for surgeons returning to surgery after breaks of four to six days. Patients treated by surgeons after a time break of four to six days have a 0.402 percentage points lower probability of dying, relative to patients treated by surgeons who were also in the operating room the day before. The coefficient is statistically significant at the one percent level. Results indicate no statistically significant effect on patient health outcomes for very short time breaks, of two to three days, nor for longer time breaks, of above seven days. Surgeon volume is not statistically associated with mortality[46]. The coefficient is close to zero and precisely estimated.

Relative to the average 30-day mortality rate of 6.55% (i.e. the sample mean), the coef-

[45]Coefficient stability usually indicates that there is limited risk of bias due to unobserved patient severity. Tests have been developed to measure the potential bias from unobserved severity based on coefficient stability and variance in the outcome when controls are included (Altonji et al., 2005; Oster, 2019). I further discuss risk of potential unobserved severity and present a series of robustness checks in Section 5.4

[46]Note that some studies have reported a positive association between surgeon volume and outcomes for orthopaedic surgeries, though these have focused on different outcomes (eg. measures of functional status as in Rachet-Jacquet et al. (2021)) and on a planned procedure which may be more prone to economies of scale than an emergency procedure.

ficient for breaks of four to six days corresponds to a relative effect of around six percent[47]. This suggests that treating a hip fracture patient after some time off, of between four to six days, may improve outcomes as surgeons may be more alert and less fatigued.

Table 2.5: Effect of surgeons' time breaks on 30-day mortality

	30-day mortality	
	(1)	(2)
Days since last surgery (Ref: 1 day)		
2-3 days	0.069	0.008
	(0.105)	(0.102)
4-6 days	-0.602***	-0.402**
	(0.148)	(0.145)
$\geq$ 7 days	-0.441	-0.258
	(0.240)	(0.228)
Yearly surgeon volume (ln)	0.187	0.136
	(0.123)	(0.118)
Surgeon FE	*Yes*	*Yes*
Patient controls	*No*	*Yes*
Hospital-year FE	*Yes*	*Yes*
R^2	0.010	0.074
Observations	371271	371271

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission and surgeon's annual lagged volume. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2 Heterogeneity

In this section, I explore whether the effect of time breaks depends on surgeons' volume of hip fracture patients or degree of specialisation in hip fracture surgery. Surgeons with relatively less experience in hip fracture care, proxied by lower annual volume, may benefit more from short breaks. Alternatively, a higher level of practice may make surgeons more resilient to fatigue. Surgeon volume is centred around its median values such that the top panel of

[47]However this represents a smaller effect in magnitude than certain risk factors or patient demographics. For instance, the effect of a four to six day break corresponds to around 12% of the effect of being in their 70s or of being male, both age and sex having been reported to be important factors of mortality in this context (Liu et al., 2017).

Table 2.6 corresponds to the effect of time breaks for a surgeon with median volume of hip fracture patients (i.e. around 41 annual surgeries).

Results in the top panel of Table 2.6 are unchanged. Results in the top panel of Table 2.6 are unchanged. The interaction terms are positive, indicating a higher mortality, though the effect is only statistically significant at the 10% level. This suggests that the beneficial effect of time breaks may be offset by higher volumes, suggesting that surgeons with higher yearly volume of practice and thus more experience can be more resilient to fatigue.

Table 2.6: Effect of surgeons' time breaks, by yearly volume of hip fracture patients

	30-day mortality
Days since last surgery (Ref: 1 day) for median volume	
2-3 days	-0.006
	(0.103)
4-6 days	-0.410**
	(0.144)
$\geq$ 7 days	-0.247
	(0.230)
Yearly volumes of hip fractures (median)	0.355
	(0.428)
Days since last surgery X yearly volume	
2-3 days	0.175^{+}
	(0.102)
4-6 days	0.518^{+}
	(0.277)
$\geq$ 7 days	0.573
	(1.157)
R^2	0.074
Observations	371271

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission and surgeon's annual lagged volume. Standard errors (in parentheses) are clustered on hospitals. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

I also test for heterogeneity in the effect of time breaks across surgeons' degree of specialisation in hip fracture, proxied by the share of past annual surgical activity in hip fracture. Surgeons with lower share of recent activity in hip fracture may benefit more from breaks to be fully alert. Alternatively, they may suffer more from days out of practice, if interference with other tasks has a detrimental effect on focus (Wixted, 2004).

The top panel in Table 2.7 shows the effect of time breaks for surgeons with median share of surgical activity in hip fracture. Results correspond to the baseline results. The interaction terms are not statistically significant, indicating that the effect of time breaks does not depend on surgeons' degree of specialisation in hip fracture surgery. This may be because the surgeries considered require broadly similar skills (eg. hip vs knee replacement), such that the relative share of activity in hip fracture does not make a substantial difference.

Table 2.7: Effect of surgeons' time breaks, by yearly share of activity in hip fracture

	30-day mortality
Days since last surgery (Ref: 1 day) for median % of activity	
2-3 days	-0.012
	(0.104)
4-6 days	-0.444**
	(0.144)
$\geq$ 7 days	-0.267
	(0.244)
% of surgical activity in hip fracture (median)	0.359
	(1.159)
Days since last surgery X yearly % hip fracture activity	
2-3 days	0.919
	(0.800)
4-6 days	1.712
	(1.208)
$\geq$ 7 days	0.021
	(1.985)
R^2	0.074
Observations	371271

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission and surgeon's annual lagged volume. Standard errors (in parentheses) are clustered on hospitals. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The within-surgeon variation in volume and specialisation may be limited. Running regressions without the surgeon fixed effects to reduce the potential collinearity between the fixed effects and the interacted variables does not change the results, and also show statistically insignificant interaction terms.

5.3 Potential mechanisms

This section investigates potential mechanisms behind surgeons' time breaks, by first exploring the role of surgeons' time breaks on the type of surgery carried out. Second, I consider the effect of the number of days since last hip fracture surgery on patient health. While time breaks in general surgical practice capture the effect of time off, breaks in hip fracture surgery may more accurately measure the potential for skill depreciation, by focusing on the task-specific dimension of time breaks.

In what follows, I test whether surgeons who return from a break choose a different type of surgery to carry out, controlling for other patient characteristics. A strand of literature suggests that surgeons vary in their diagnostic skills and treatment choices (Epstein and Nicholson, 2009; Abaluck et al., 2016; Currie and MacLeod, 2017), and that the latter can also vary for surgeons over time[48]. Clinical guidelines produced by the British Orthopaedic Association indicate that the best surgical treatment depends on the range of factors, such as the type and stability of the fracture, patients' frailty and surgical dexterity with implant type. While internal fixation of the fracture may be preferred to replacement of the joint for the younger or very frail elderly, complications such as displacement of the fracture or fixation failure may arise. Partial (hemiarthroplasty) or full replacement of the joint is a more invasive surgery with higher risks of post-operative complications and subsequent need of revision procedures (British Orthopaedic Association, 2007, p21-25).

Table 2.8 shows that surgeons returning from a short break are less likely to perform a partial or complete hip replacement, as opposed to a fixation or reduction of fracture using nails or screws, holding other patient characteristics fixed. Short time breaks, between four to six days, reduce the probability of having a full or partial hip replacement surgery by 0.826 percentage points. The effect is statistically significant at the one percent level but quantitatively small (around 1.7% relative effect).

[48] Changes in treatment choice can be linked to surgeons' work environment (Molitor, 2018; Chan, 2021), or depend on the time of patient admission, as surgeons may choose a less intensive treatment option for patients admitted near the end of their work shift in order to preserve leisure time (Halla et al., 2016; Costa-Ramón et al., 2018; Persson et al., 2019; Costa-Ramón et al., 2020).

To ensure that differences in the choice of surgery do not reflect differences in the severity of patients treated after a break, Table B.9 in Appendix presents the results of a regression of the type of fracture on time breaks, conditional on patient and surgeon controls. The results indicate that surgeons' time breaks and the probability of treating a patient with a fracture of the neck of femur (i.e., the most common type of fracture), are not correlated.

Table 2.8: Effect of surgeons' time breaks on type of surgery

	Full or partial hip replacement surgery
Days since last surgery (Ref: 1 day)	
2-3 days	0.003
	(0.102)
4-6 days	-0.826**
	(0.283)
$\geq$ 7 days	-0.570
	(0.435)
R^2	0.074
Observations	371271

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission and surgeon's annual lagged volume. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Previous studies have interpreted the detrimental effect of time breaks as evidence of skill depreciation (Hockenberry et al., 2008; Hockenberry and Helmchen, 2014). This rest of this section provides a partial test for this by focusing specifically on the effect of breaks in hip fracture surgery, as opposed to general breaks in surgical practice. The distribution of hip fracture-specific breaks is less concentrated, as surgeons will work in the operating ward most days but may not treat hip fracture patients every day. I re-define time breaks categories to span the entire distribution of hip fracture breaks and present the results in Table 2.9 (column 2).

Defining the time breaks to hip fracture practice does not change the overall results. Breaks of between 15-45 days have a positive effect on health outcomes, by reducing mortality by 0.40 percentage points, whilst shorter and longer time breaks have no effect. Surgeons who return to hip fracture surgery after a while perform better possibly because they are more alert or apt to choose a more appropriate surgery.

Table 2.9: Effect of general time breaks and hip fracture-specific time breaks on mortality

	30-day mortality	
	(1)	(2)
Days since last surgery (Ref: 1 day)		
2-3 days	0.008	
	(0.102)	
4-6 days	-0.402**	
	(0.145)	
$\geq$ 7 days	-0.258	
	(0.228)	
Days since last hip fracture surgery (Ref: 1 day)		
2-3 days		-0.055
		(0.124)
4-6 days		-0.025
		(0.164)
7-14 days		-0.223
		(0.128)
15-29 days		-0.409**
		(0.138)
30-44 days		-0.474*
		(0.193)
$\geq$45 days		-0.050
		(0.189)
R^2	0.074	0.074
Observations	371271	371271

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission and surgeon's annual lagged volume. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.4 Robustness checks

This section further tests that the results are not impacted by error measurement in the data or endogeneity concerns. First, in the NHS hospital administrative database (Hospital Episodes Statistics), surgical activity is only reported for consultants, whereas patient care involves a larger clinical team. Reassuringly, the literature shows that senior surgeons have a considerable influence on medical teams' decision (Chan, 2021). Nevertheless, consultants may not always be the ones carrying out the surgery in practice if they delegate it to the junior doctors that they supervise. In such case, the operating surgeon may have a different (unobserved) number of days since their last surgery, resulting in unsystematic measurement error. This may introduce attenuation bias in the effect of interest. To mitigate this concern, the same analysis is therefore run on the subsample of consultants who are the least likely to take on supervisees, defined as the consultants at the start of their consultancy period (N=436 surgeons). This assumes that more senior consultants are more likely to have supervisees, which seems credible. Results in Table B.10, in Appendix, show that the results are unaffected by the sample restriction though the point estimate is now slightly larger, consistent with the hypothesis that there may be some attenuation effect.

In the next robustness checks, I check that the results are not driven by the selective allocation of patients to hospitals or to surgeons in the emergency ward. As surgeons start practising more often resulting in shorter time breaks, they may also treat more complex cases, which will affect patient outcomes. Any remaining unobserved patient severity systematically correlated with surgeon ability may introduce a negative bias in the estimate of the effect of practice frequency. Selective patient allocation based on unobserved severity should be limited for emergency conditions. However, ambulances may direct the most complex patients to the best hospitals. In a first robustness check, the sample is restricted to the patients who are treated in their closest hospital, therefore excluding patients who bypassed their closest hospital or were brought to another hospital by the ambulance. Results shown in Appendix Table B.11 are unchanged.

Alternatively, patients may be allocated to specific surgeons within the clinical team in the emergency ward, based on unobserved patient needs. In a first robustness check, I include a surgeon-specific time trend to account for (linear) changes in surgeons' outcomes over time. The time trend would account for changes in surgeons' patient case mix over time, reflecting the fact that unobservably more complex cases may be allocated to the most able surgeon within the clinical team. Linear time trends assume that any selective allocation of patients to surgeons is linearly increasing over time, which here seems appropriate to reflect the effect of surgeon seniority over the sample (i.e. 8 years). Table B.12, in Appendix, shows that the results are unchanged. In a second robustness check, I run the same specifications only for the sample of surgeries that took place during a bank holiday or a weekend when surgeons are on call. The lower levels of staffing and the presence of fewer surgeons on the wards during these days means there is limited possibility to selectively allocate patients across surgeons. The results, shown in Table B.13 are unchanged.

A potential limitation of this study is that the reason why surgeons go on breaks is not known. For instance, an alternative interpretation of the results could be that surgeons, while observed to be 'on break' in the data, perform surgeries elsewhere. This would mean additional surgical experience which may translate into better health outcomes after the break[49]. The HES data report activity for all patients in NHS hospitals, but only activity for NHS patients in private sector hospitals[50]. I run again the baseline specification after excluding observations where surgeons also work in a private hospital (called independent sector treatment centre) that year[51]. The sample size is smaller but the results are unchanged (Table B.14). Alternatively, surgeons may appear to be on break when they go for training courses, during which they would gain knowledge and technical proficiency. It is unlikely that all observed breaks are due to surgeons going on training courses. However, results should be interpreted as an upper bound of the actual effect of breaks.

[49]Note however that the results of this study indicate no statistically significant association between surgeon volume and mortality rates.

[50]The HES data cover all admissions paid for by the NHS, which is estimated to represent around 98% of hospital activity in England (Herbert et al., 2017).

[51]Surgeons are considered to work for a private (independent sector) hospital if they have treated at least one NHS patient during that year. Private (independent sector) hospitals do not treat emergency hip fracture patients. However, surgeons there may treat planned hip replacement surgeries.

6 Conclusions

This study investigates the causal effect of surgeons' breaks in surgical practice on 30-day mortality, for a large panel of surgeons in the English NHS. The key finding is that time breaks of between four to six days improve 30-day survival probability by around six percent. Findings are in line with the previous literature which, with the notable exception of Hockenberry et al. (2008) and Hockenberry and Helmchen (2014) for coronary bypass surgeries, has found little compelling evidence that surgeons' time breaks have a detrimental effect on patient health outcomes (Huesch, 2014; Pearce et al., 2015; Van Gestel et al., 2017). Hockenberry et al. (2008) and Hockenberry and Helmchen (2014) find a fast and salient negative effect of breaks on patient health outcomes. Results potentially differ here because hip fracture surgery requires a lower range of complex skills (Nembhard, 2000), or because cardiac surgeries involve a larger surgical team in which case frequent surgeries may also help coordination across tasks. Emergency departments are known to be strenuous work environments, which could explain the positive effect for short breaks in this setting. A limitation of this study is that it relies on all surgeons' activity to measure breaks, resulting in limited variation in terms of days since last surgery. Most breaks are concentrated on short breaks of one to three days. Provided survey data on surgeons' holidays is available, future research could investigate the effect of longer breaks.

Nevertheless, the current findings have policy implications for the organisation of activity in hospitals. Understanding the impact of short-run changes in surgeons' activity on performance would allow for better-targeted and more effective policies to increase the quality of care. Possible policy interventions include regulating surgeons' work schedules to accommodate more regular short breaks, without necessarily reducing the overall workload. This also ties in with the research on the impact of work shift length on performance, which suggests an adverse effect of long working hours (Brachet et al., 2012; Collewet and Sauermann, 2017). Work schedules could allow for more regular short breaks to maintain focus and increase alertness. Given the importance of the health workforce in the production of healthcare, optimizing work schedules to reflect these potential effects could translate into substantial effects for healthcare systems (McConnell et al., 2013; Bloom et al., 2015, 2020).

More generally, the findings raise awareness about non-financial determinants of improvements in quality of care (Lagarde et al., 2019). While a previous literature has stressed the importance of workforce management quality, reputation, diffusion of clinical information, team composition or levels of staffing on quality of care, the organisation of activity and type of work shift are other potential levers of quality (Phelps, 2000; Brachet et al., 2012; Kolstad, 2013; Bartel et al., 2014; Friedrich and Hackmann, 2017; Bloom et al., 2020).

B Appendix

Table B.1: Results without sample restriction based on patient age

	30-day mortality
Days since last surgery (Ref: 1 day)	ref.
2-3 days	0.003
	(0.101)
4-6 days	-0.404**
	(0.142)
$>$ 7 days	-0.272
	(0.222)
Surgeon yearly volume (ln)	0.137
	(0.114)
Surgeon FE	Yes
Patient controls	Yes
Hospital-year FE	Yes
R2	0.076
Observations	373273

Notes: Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.2: List of Elixhauser comorbidities and associated sample prevalence

	Mean	Std. Dev.	Min.	Max.
Congestive heart failure	0.10	0.31	0	1
Cardiac arrhythmias	0.25	0.43	0	1
Valvular disease	0.08	0.27	0	1
Pulmonary circulation disorders	0.02	0.14	0	1
Peripheral vascular disorders	0.04	0.21	0	1
Hypertension, uncomplicated	0.53	0.50	0	1
Paralysis	0.02	0.14	0	1
Other neurological disorders	0.09	0.29	0	1
Chronic pulmonary disease	0.20	0.40	0	1
Diabetes, uncomplicated	0.15	0.36	0	1
Diabetes, complicated	0.01	0.12	0	1
Hypothyroidism	0.11	0.31	0	1
Renal failure	0.13	0.34	0	1
Liver disease	0.02	0.13	0	1
Peptic ulcer diseases, excl. bleeding	0.01	0.10	0	1
AIDS/HIV	0.00	0.01	0	1
Lymphoma	0.01	0.08	0	1
Metastatic cancer	0.02	0.13	0	1
Solid tumour without metastasis	0.05	0.22	0	1
Rheumatoid arthritis/collagen vascular	0.05	0.22	0	1
Coagulopathy	0.01	0.09	0	1
Obesity	0.01	0.11	0	1
Weight loss	0.02	0.14	0	1
Fluid and electrolyte disorders	0.14	0.35	0	1
Blood loss anemia	0.00	0.04	0	1
Deficiency anemia	0.05	0.21	0	1
Alcohol abuse	0.04	0.21	0	1
Drug abuse	0.00	0.05	0	1
Psychoses	0.01	0.10	0	1
Depression	0.08	0.28	0	1
Hypertension, complicated	0.02	0.15	0	1
Observations	371271			

Table B.3: Effect of surgeons' time breaks on 30-day mortality (99% winsorisation)

	30-day mortality	
	Coefficient	SE
Days since last surgery (Ref: 1 day)	ref.	
2-3 days	0.004	(0.102)
4-6 days	-0.408**	(0.145)
> 7 days	-0.255	(0.229)
Surgeon yearly volume (ln)	0.133	(0.118)
40-49 years	0.484**	(0.184)
50-59 years	1.488***	(0.166)
60-69 years	2.155***	(0.165)
70-79 years	3.301***	(0.175)
80-89 years	5.612***	(0.173)
90-99 years	10.366***	(0.235)
>100 years	18.673***	(1.094)
=1 if male	2.999***	(0.108)
Other ethnic group	0.889***	(0.241)
Ethnicity not coded	0.914***	(0.175)
Pre-surgery length of stay	-0.068**	(0.024)
Pertrochanteric	0.284*	(0.124)
Subtrochanteric	0.267	(0.213)
Unspecified	-0.535	(0.655)
Congestive heart failure	8.125***	(0.242)
Cardiac arrhythmias	2.760***	(0.137)
Valvular disease	-0.160	(0.166)
Pulmonary circulation disorders	4.855***	(0.442)
Peripheral vascular disorders	1.692***	(0.224)
Hypertension, uncomplicated	-1.628***	(0.082)
Paralysis	-0.657*	(0.269)
Other neurological disorders	0.318	(0.164)
Chronic pulmonary disease	2.428***	(0.125)
Diabetes, uncomplicated	-0.030	(0.128)
Diabetes, complicated	-1.303**	(0.389)
Hypothyroidism	-0.546***	(0.133)
Renal failure	3.213***	(0.170)
Liver disease	4.581***	(0.437)
Peptic ulcer diseases, excl. bleeding	-0.980*	(0.417)
AIDS/HIV	-18.005***	(2.134)
Lymphoma	1.217	(0.625)
Metastatic cancer	9.808***	(0.551)
Solid tumour without metastasis	2.651***	(0.287)
Rheumatoid arthritis/collagen vascular	-0.717***	(0.163)
Coagulopathy	2.979***	(0.575)
Obesity	-2.094***	(0.331)

Weight loss	2.375***	(0.344)
Fluid and electrolyte disorders	5.307***	(0.203)
Blood loss anemia	-1.113	(1.051)
Deficiency anemia	-0.963***	(0.224)
Alcohol abuse	-1.035***	(0.189)
Drug abuse	-1.871**	(0.600)
Psychoses	0.497	(0.424)
Depression	-0.607***	(0.147)
Hypertension, complicated	2.262***	(0.439)
Past emergency admissions	0.507***	(0.084)
Past all hospital admissions	-0.363***	(0.052)
Diagnosed with osteoporosis	-0.256	(0.198)
Operated during a weekend	0.173	(0.101)
Admitted during a weekend	0.131	(0.101)
R^2	0.074	
Observations	371271	

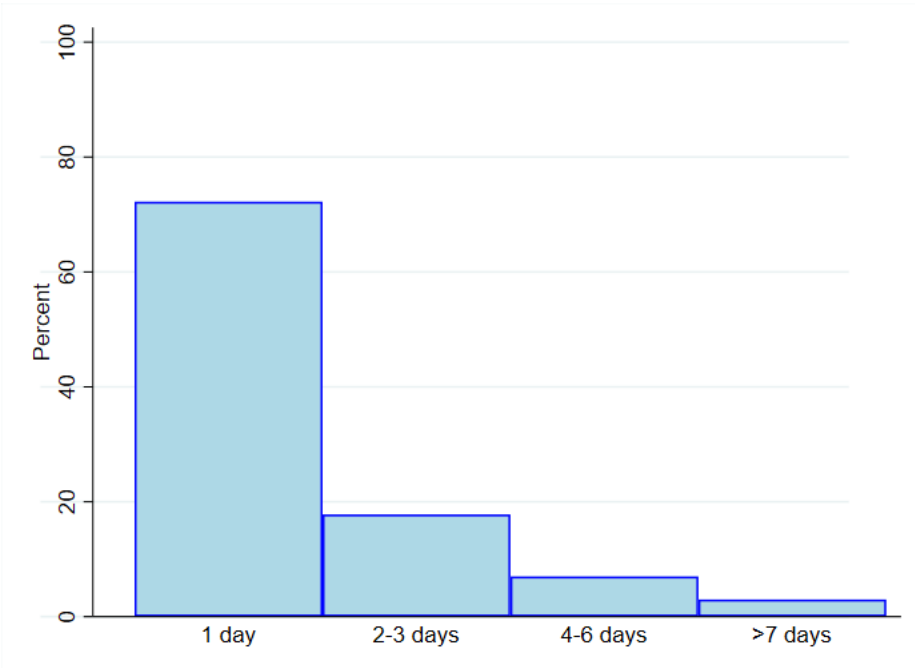
Notes: Coefficients are expressed in percentage points. Time breaks corresponds to the number of days since the operating surgeon's last surgery. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.4: Association between each day of break and 30-day mortality rates

	Coef.	SE
Days since last surgery (Ref: 1 day)		
2 days	0.025	(0.120)
3 days	0.170	(0.174)
4 days	-0.557**	(0.200)
5 days	-0.562*	(0.270)
6 days	-0.775*	(0.336)
7 days	-0.396	(0.433)
8 days	-0.001	(0.566)
9 days	-1.197	(0.741)
10 days	-0.972	(0.755)
11 days	-0.790	(0.722)
12 days	-0.171	(0.879)
13 days	-0.594	(1.000)
14 days	-0.033	(1.199)
15 days	-3.279**	(1.061)
16 days	3.254	(2.043)
17 days	-2.473	(1.816)
18 days	-1.778	(1.880)
19 days	2.281	(2.227)
20 days	1.440	(2.285)
21 days	-1.712	(2.027)
22 days	3.022	(3.862)
23 days	-3.869	(2.481)
24 days	0.744	(3.644)
25 days	-1.924	(2.962)
26 days	2.831	(5.598)
27 days	-3.078	(3.431)
28 days	-1.035	(3.740)
29 days	23.150	(17.310)
30 days	-6.043***	(1.060)
31 days	2.445	(9.333)
R^2	0.010	
Observations	371271	

Notes: Coefficients are expressed in percentage points. Controls include surgeon fixed effects and hospital-year fixed effects. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure B.1: Distribution of the number of days since a surgeon's last surgery



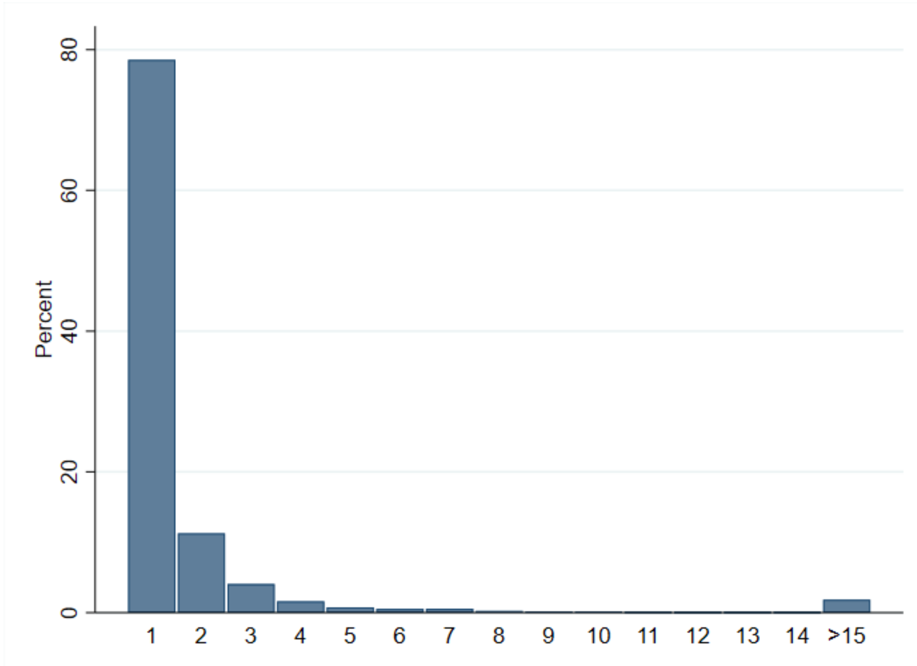
Notes: The number of days between two surgeries varies by surgeon and day of surgery.

Table B.5: Results using different functional forms of time breaks

	30-day mortality		
	(1)	(2)	(3)
Time breaks	-0.030	-0.095^{+}	
	(0.020)	(0.049)	
Time breaks^2		0.005	
		(0.04)	
Time breaks (ln)			-0.125^{+}
			(0.067)
R^2	0.074	0.074	0.074
Observations	371271	371271	371271

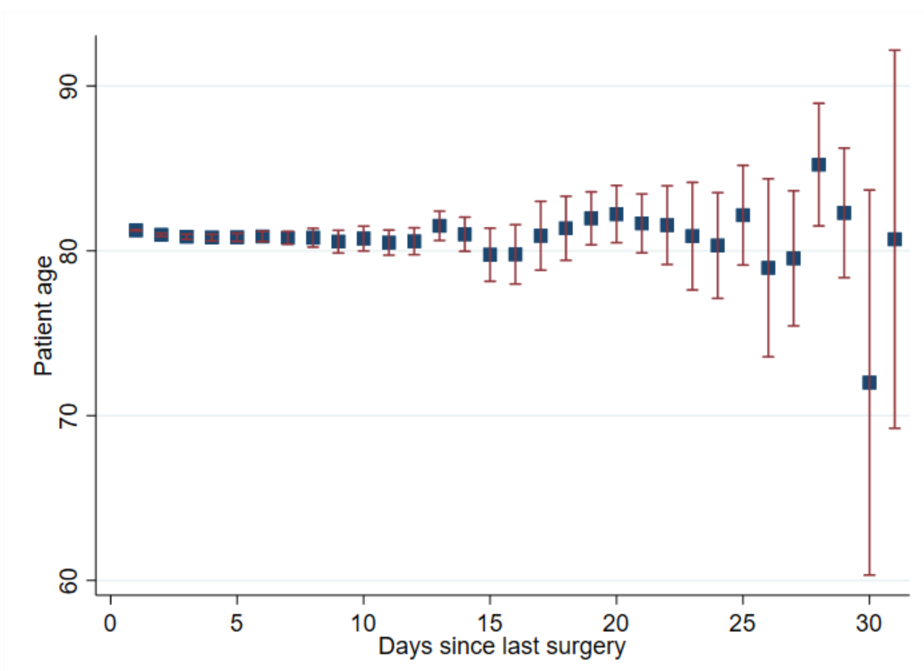
Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission and surgeon's annual lagged volume. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure B.2: Percentage of hip fracture patients who go to their Nth closest hospital



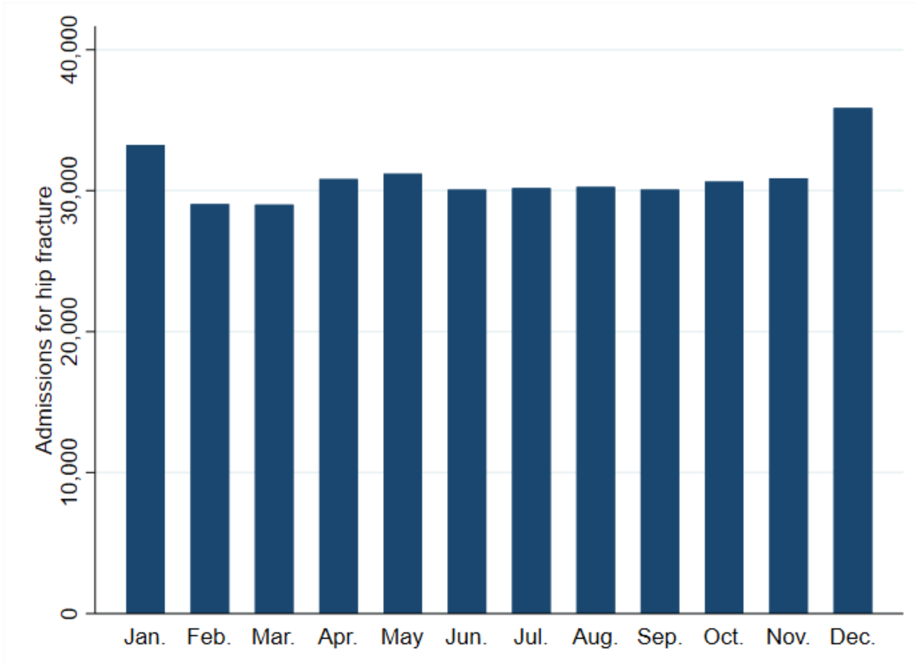
Notes: 78% of the patients go to their closest hospital (trust). This is calculated on the sample of patients for which the information on patient's residence is non-missing (N=289,514).

Figure B.3: Mean patient age, by break length



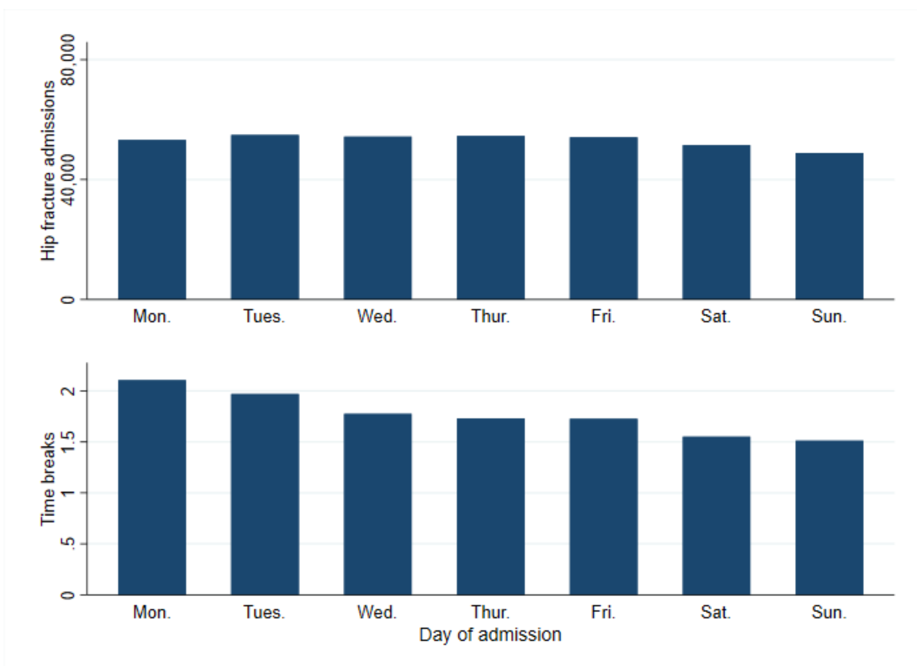
Notes: Mean and 95% confidence intervals for average patient age per number of days since last surgery.

Figure B.4: Admissions for hip fracture, by month of admission



Notes: Aggregate admissions for hip fracture, by month of admission (final sample).

Figure B.5: Variation in hip fracture admissions and time breaks, by day of the week



Notes: Top plot: Aggregate admissions for hip fracture, by day of admission. Bottom plot: Average surgeons' time breaks (i.e. days since last surgery), by day of admission.

Table B.6: Effect of surgeons' time breaks on mortality, with month dummies

	30-day mortality
Days since last surgery (Ref: 1 day)	
2-3 days	-0.011
	(0.103)
4-6 days	-0.441**
	(0.145)
$\geq$ 7 days	-0.326
	(0.229)
Month dummies (Ref. January)	
February	-0.153
	(0.206)
March	-0.543**
	(0.196)
April	-0.875**
	(0.204)
May	-1.198***
	(0.191)
June	-1.375***
	(0.217)
July	-1.502***
	(0.211)
August	-1.240***
	(0.192)
September	-0.963***
	(0.202)
October	-1.077***
	(0.201)
November	-0.650**
	(0.204)
December	-0.202*
	(0.211)
R^2	0.075
Observations	317271

Notes: Coefficients are expressed in percentage points. Time breaks corresponds to the number of days since the operating surgeon's last surgery. Only controls that may influence time breaks are included (i.e. month and day of surgery). Standard errors (in parentheses) are adjusted for clustering at the hospital level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.7: Effect of a recent patient death on surgeons' time breaks

	Time breaks
Patient died in previous month	-0.020
	(0.018)
Surgeon volume in previous month	-0.075***
	(0.008)
Day of surgery	*Yes*
Month of surgery	*Yes*
Surgeon FE	*Yes*
Hospital-year FE	*Yes*
R^2	0.094
Observations	371271

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission, and surgeon's annual lagged volume. Standard errors (in parentheses) are adjusted for clustering at the hospital level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.8: Effect of surgeons' time breaks on 30-day mortality (full table)

	30-day mortality	
	Coefficient	SE
Days since last surgery (Ref: 1 day)		
2-3 days	0.008	(0.102)
4-6 days	-0.402**	(0.145)
$\geq$ 7 days	-0.258	(0.228)
Surgeon yearly volume (ln)	0.136	(0.118)
40-49 years	0.468*	(0.183)
50-59 years	1.457***	(0.164)
60-69 years	2.123***	(0.163)
70-79 years	3.283***	(0.174)
80-89 years	5.651***	(0.172)
90-99 years	10.452***	(0.233)
>100 years	18.776***	(1.097)
=1 if male	2.985***	(0.108)
Other ethnic group	0.898***	(0.240)
Ethnicity not coded	0.946***	(0.176)
Pre-surgery length of stay	-0.089***	(0.019)
Pertrochanteric	0.286*	(0.124)
Subtrochanteric	0.253	(0.213)
Unspecified	-0.564	(0.657)
Congestive heart failure	8.141***	(0.241)
Cardiac arrhythmias	2.759***	(0.136)

Valvular disease	-0.184	(0.166)
Pulmonary circulation disorders	4.839***	(0.441)
Peripheral vascular disorders	1.623***	(0.224)
Hypertension, uncomplicated	-1.665***	(0.082)
Paralysis	-0.641*	(0.269)
Other neurological disorders	0.338*	(0.164)
Chronic pulmonary disease	2.411***	(0.125)
Diabetes, uncomplicated	-0.038	(0.128)
Diabetes, complicated	-1.468***	(0.386)
Hypothyroidism	-0.559***	(0.134)
Renal failure	3.164***	(0.169)
Liver disease	4.484***	(0.436)
Peptic ulcer diseases, excl. bleeding	-1.069*	(0.419)
AIDS/HIV	-17.941***	(2.138)
Lymphoma	0.580	(0.619)
Metastatic cancer	9.445***	(0.549)
Solid tumour without metastasis	2.476***	(0.286)
Rheumatoid arthritis/collagen vascular	-0.795***	(0.163)
Coagulopathy	2.875***	(0.574)
Obesity	-2.151***	(0.331)
Weight loss	2.299***	(0.347)
Fluid and electrolyte disorders	5.328***	(0.203)
Blood loss anemia	-1.184	(1.050)
Deficiency anemia	-1.009***	(0.224)
Alcohol abuse	-1.000***	(0.188)
Drug abuse	-1.881**	(0.602)
Psychoses	0.544	(0.422)
Depression	-0.608***	(0.147)
Hypertension, complicated	2.153***	(0.439)
Past emergency admissions	0.088	(0.055)
Past all hospital admissions	0.021	(0.013)
Diagnosed with osteoporosis	-0.264	(0.197)
Operated during a weekend	0.168	(0.102)
Admitted during a weekend	0.132	(0.101)
R^2	0.074	
Observations	371271	

Notes: Coefficients are expressed in percentage points. Time breaks corresponds to the number of days since the operating surgeon's last surgery. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.9: Association between time breaks and type of fracture

	Fracture of the neck of femur
Days since last surgery (Ref: 1 day)	
2-3 days	0.248
	(0.204)
4-6 days	0.419
	(0.289)
$\geq$ 7 days	-0.687
	(0.463)
R^2	0.048
Observations	371271

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, indicator for weekend admission and surgeon's annual lagged volume. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.10: Effect of time breaks on mortality, for surgeons starting their consultancy

	30-day mortality
Days since last surgery (Ref: 1 day)	
2-3 days	-0.425
	(0.270)
4-6 days	-1.230***
	(0.354)
$\geq$ 7 days	-0.393
	(0.584)
R^2	0.089
Observations	51295

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission, and surgeon's annual lagged volume. Sample of consultants who started their consultancy in 2008 or after. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.11: Effect of time breaks on mortality, without patients who bypass closest hospital

	30-day mortality
Days since last surgery (Ref: 1 day)	
2-3 days	0.026
	(0.121)
4-6 days	-0.436**
	(0.166)
$\geq$ 7 days	-0.250
	(0.248)
R^2	0.077
Observations	289508

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission, and surgeon's annual lagged volume. Junior surgeons are consultants who start their consultancy in 2008 or after. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.12: Effect of time breaks on mortality, with a surgeon-specific time trend

	30-day mortality
Days since last surgery (Ref: 1 day)	
2-3 days	0.010
	(0.103)
4-6 days	-0.421**
	(0.146)
$\geq$ 7 days	-0.259
	(0.233)
Surgeon FE	*Yes*
Patient controls	*Yes*
Surgeon time trend	*Yes*
R^2	0.077
Observations	371271

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission, and surgeon's annual lagged volume. Junior surgeons are consultants who start their consultancy in 2008 or after. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.13: Effect of time breaks on mortality, on bank holidays or weekend days

	30-day mortality
Days since last surgery (Ref: 1 day)	
2-3 days	-0.117
	(0.205)
4-6 days	-0.816*
	(0.348)
$\geq$ 7 days	-0.455
	(0.609)
R^2	0.098
Observations	104144

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission, and surgeon's annual lagged volume. Junior surgeons are consultants who start their consultancy in 2008 or after. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.14: Effect of time breaks on mortality, without surgeons who work in private sector

	30-day mortality
Days since last surgery (Ref: 1 day)	
2-3 days	-0.068
	(0.172)
4-6 days	-0.507*
	(0.237)
$\geq$ 7 days	-0.406
	(0.276)
R^2	0.083
Observations	164670

Notes: Coefficients are expressed in percentage points. Controls include patient age, sex, ethnicity, economic deprivation, pre-surgical length of stay, Elixhauser comorbidities, number of past hospital admissions, diagnosis of osteoporosis, type of hip fracture, indicator for weekend admission, and surgeon's annual lagged volume. Junior surgeons are consultants who start their consultancy in 2008 or after. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Chapter 3

Does containing costs reduce hospital quality?

1 Introduction

Rising healthcare expenditures, due to an increasing demand for health services driven by demographic changes and technological innovation, pose a challenge to the sustainability of healthcare systems (OECD, 2015). To limit the growth in healthcare spending, policymakers must find effective ways of containing healthcare costs without affecting quality of care. The hospital sector in particular accounts for an important share of healthcare spending. In 2017, in the U.K., half of government-funded healthcare expenditure was spent on hospitals (Office for National Statistics, 2017). There may however be scope to reduce unnecessary spending. A recent report from the OECD suggests that as much as 20% of healthcare expenditure makes no or only minimal contribution to improving health outcomes (OECD, 2017b).

In this context, reducing patients' length of stay in hospital when clinically appropriate is a potential policy lever to reduce costs across many treatments. Shorter inpatient lengths of stay can also improve patient safety and comfort, by reducing the risk of hospital-acquired infections and allowing patients to recover in the familiar environment of their home (OECD, 2017a). For this reason, strategies to discharge patients sooner, whereby patients are admitted to a hospital bed, receive the necessary care and are discharged on the same calendar day, have been incentivised for a list of low-risk emergency conditions in the English NHS (British Association for Ambulatory Emergency Care, 2014). However, despite the potential for cost reduction across hospitals and treatments, real-world evidence around the safety of discharging patients early rather than admitting them overnight remains weak, especially in emergency settings (Miani et al., 2014). Discharging patients too early may result in poorer patient health and even increase costs if patients have to seek care again later (Chen et al., 2010).

In this study, we investigate the causal effect of being discharged from hospital on the same day as admission, rather than staying overnight, on hospital quality for emergency patients admitted with chest pain symptoms in England. Chest pain is one of the most common reasons for emergency admission (Forberg et al., 2006), with around 200,000 patients admitted to a hospital each year in England. Although chest pain often resolves itself within a short timeframe, it can be a symptom of more severe cardiac conditions, such as coronary heart disease or a myocardial infarction (i.e., heart attack). Therefore, discharging patients too quickly may come at the cost of poorer quality of care. We proxy hospital quality by patients' risk of having an emergency admission within 28 days of hospital discharge. Emergency readmission is a widely used metric of hospital quality, often reported in the U.S. or England to monitor quality of care. It fits our context particularly well as we may be concerned that unduly discharging patients on the same day as admission will cause them to be readmitted later[52].

Patients who are discharged on the same day may be unobservably healthier than patients admitted for an overnight hospital stay, and thus will have better health outcomes independently of their inpatient length of stay. Therefore, we control for a wide range of observed clinical and socio-economic dimensions of patient health available in hospital records (e.g., patient age, sex, ethnicity, income deprivation, comorbidities, risk factors and past hospital utilisation). We further account for possible remaining unobserved patient severity by using an instrumental-variable approach. In particular, we instrument for whether a patient was discharged on the same day of admission by using differences in patient exposure to a major bonus policy in England that incentivises early patient discharge for certain conditions (Allen et al., 2016; Gaughan et al., 2019). We utilise the fact that patients who attended the Emergency Department during daytime were more impacted by the reform than patients who attended at night (see Section 5 on Methods). In line with a growing IV literature (Duflo,

[52]While mortality is a commonly used quality metric, it is too rare in this context to be informative: in our sample only 0.5% of patients died within 30 days of hospital discharge.

2001; Lundborg and Majlesi, 2018; Aouad et al., 2019; Ma, 2019), our instrument therefore relies on exogenous variation in the intensity of the effect of the policy, induced by the time at which patients arrive at an Emergency Department. Importantly, any systematic unobserved differences in hospital quality or case-mix or differences in patient severity linked to the time of attendance are accounted for by a set of hospital fixed effects and hour of arrival dummies.

Our OLS results indicate that, after controlling for a rich set of patient characteristics, being admitted and discharged on the same day is associated with a lower probability of being readmitted within 28 days. However, our instrumental variable approach suggests that there is no causal effect of being discharged on the same day as admission on the probability of being readmitted.

This study makes several contributions to the literature, which we review in more detail in the next section. First, we contribute to the literature on the trade-off between costs and quality in the hospital sector, by providing novel causal evidence on the effect of a cost-reducing strategy on hospital quality. Importantly, the policy considered here pays hospitals a higher tariff for a lower-cost treatment (ie., same-day discharge treatment rather than overnight stays). A large literature has sought to assess the relationship between costs and quality of care (Gutacker et al., 2013; Hussey et al., 2013; Häkkinen et al., 2014; Jamalabadi et al., 2020), relying on variation in costs across space and/or time. A major endogeneity concern in this context is omitted variable bias due to possible unobserved patient severity, which may correlate both with hospital costs and patient health outcomes. By design, because our instrument relies on the time at which patients attend the Emergency Department (ED), it varies across patients within hospitals rather than solely across hospitals or regions. Our setting and instrument therefore mitigate the risk of remaining unobserved variables at both patient and hospital level.

Second, our study adds to the broader literature on healthcare providers' behaviour. Understanding hospitals' behaviour provides important insights for policymakers. For instance, Dafny (2005) shows that U.S. hospitals responded to a change in prices by 'upcoding' patients to diagnoses with largest price increases but did not alter the intensity or quality of care provided. Dranove et al. (2003) provide evidence that following the public reporting of hospital performance for cardiac surgery in the U.S., hospitals shift to treating healthier patients, while Kolstad (2013) finds that surgeons responded to report cards by increasing quality, suggesting that surgeons are also motivated by reputational concerns. Closer to our setting, Arcà et al. (2020) find that hospitals responded to austerity measures implemented in certain Italian regions by cutting beds and staff capacity, which led to lower patient health outcomes. In this study, we consider how reductions in inpatient length of stay may have impacted the care provided in hospitals. Our results indicate that cost reductions based on inpatient length of stay can be achieved without harming quality.

Finally, this study contributes to the limited but growing literature using the introduction of policies as instrumental variables in a panel data framework (Hudson et al., 2017). In education economics, years of schooling have been instrumented using variation in exposure to schooling reforms across time or space. Acemoglu and Angrist (2000) estimate returns to schooling on future earnings for pupils in the U.S., using differences in compulsory schooling across states and cohorts of students. Duflo (2001) relies on variation in the intensity of a school construction programme across regions of Indonesia to estimate individuals' returns to schooling in terms of wages[53]. Ma (2019) and Lundborg and Majlesi (2018) exploit variation in the rolling out of reforms to extend compulsory schooling to estimate the effect of extra years of schooling on parents' health and longevity in China and Sweden respectively. In the healthcare context, we are only aware of few studies with similar IV designs. Aouad et al. (2019) study the effect of receiving treatment in an ambulatory surgery centre on patient health outcomes after a colonoscopy in the U.S. Being treated in an ambulatory surgery centre, rather than in a hospital, is instrumented for by exposure to a change in patients' cost-sharing for hospital treatments for certain insurers, which incentivised these patients to receive care in ambulatory surgery centres. Frimmel and Pruckner (2020) use variation in the legal retirement age driven by pension reforms in Austria to examine the effect of retirement on healthcare use.

[53] In this influential study, Duflo uses a similar instrumental design where the instrument is the effect of the policy, exploiting differences in the intensity of a school construction programme (the policy) across regions and cohorts of pupils. While the school construction programme was a major national reform, regions and pupils were impacted to a different degree: more schools were built in some regions while younger pupils were exposed to the reform for a longer period of time

In the remainder of this section, we give a brief account of the literature (Section 1.1). Section 2 describes our institutional context, while Section 3 lays out our conceptual framework. Section 4 introduces the data. Section 5 details our instrumental variable strategy. Section 6 presents the results while Section 7 concludes.

1.1 Related literature

A major strand of literature has sought to understand the relationship between costs and quality, by exploiting geographic variations in costs (see Hussey et al. (2013) and Jamalabadi et al. (2020) for two recent systematic reviews of the literature). Despite generally mixed findings, the literature suggests that the direction and significance of the cost-quality association depend on the medical conditions investigated, the measure of outcomes and the ability to account for patient severity. A positive relationship between costs or prices and quality is more commonly reported when studies focus on process measures, rather than health outcomes, and when potential omitted variable bias is accounted for. Some studies also find that the relationship between costs and quality is non-linear (Steven, 1991; Chen et al., 2010; Gutacker et al., 2013; Jamalabadi et al., 2020). This may also explain why findings across studies vary if they are effectively estimating the cost-quality relation at different points of the cost-quality curve. McKay and Deily (2008) further indicate that the effect may depend on the source of cost reduction, suggesting that rather than attempting to reduce overall costs, policy makers should target efficiency efforts on costs that are deemed wasteful. Doyle et al. (2017) for instance find that high spending on inpatient care leads to better chances of survival, but that higher outpatient spending reduces survival. While there is substantial literature on the trade-off between costs and quality, there are fewer studies which attempt to account for potential omitted variable bias linked to unobserved patient severity in regressions with hospital costs or prices.

Several causal studies investigate the effect of costs on quality and implement instrumental variable strategies. Stargardt et al. (2014) investigate the effect of hospital costs on 1-year mortality for acute myocardial infarction (AMI) patients in Germany, and instrument hospital costs alternatively with the average costs in the federal state or the average price per square meter in hospitals' county. The authors find that lower costs lead to increased mortality. Another study by Schreyögg and Stargardt (2010) assesses the relationship between hospital costs and patient health outcomes after AMI in the U.S., where they instrument for hospital costs by an adjustment index of wages across regions in the U.S. and general overhead costs per day at the hospital. Their results also suggest that increased costs lead to higher mortality and readmission rates. Overall, there are concerns that instruments which use variation across hospitals or regions may not fully account for unobserved severity, as there may remain unobserved differences in population health across hospitals or regions. Doyle et al. (2015) investigate whether patients with an emergency condition who are treated in high-spending hospitals have a lower risk of mortality in the U.S. After documenting that ambulance companies have different preferences as to which hospital to refer patients, they instrument patients' probability of being treated in a high-spending hospital by the ambulance service dispatched to patients. They find that being treated in a higher-spending hospital leads to reduced mortality.

Some studies focus specifically on the effect of inpatient length of stay, as a driver of costs, on patient health outcomes, which is also our focus. Harrison et al. (1995) find that the decreasing inpatient length of stay over time in a region of Canada was not associated with increases in 30-day admission rates or in physician visits after patient discharge from the hospital for a range of conditions, which include acute myocardial infarction. Picone et al. (2003) estimate the effect of length of inpatient stay on mortality and functional status for Medicare patients admitted after a health shock (hip fracture, stroke, coronary heart disease) in the U.S. They fit a quasi-maximum likelihood discrete factor model to account for unobserved severity, and find that length of stay does not lead to better patient health outcomes. Hauck and Zhao (2011) use Australian hospital data to study the causal effect of an additional day in hospital on patients' risk of experiencing an adverse event. Instrumenting patient length of stay by the days and month of patient discharge, the authors show that an additional day in hospital increases the risk of infections and ulcers. While these studies look at length of stay in general, we focus on the lower end of the distribution of length of stay, for the conditions where it is considered safe to not keep patients overnight. The reduction in hospital length of stay goes therefore from one to zero hospital days.

A recent strand of literature leverages variation in austerity measures across regions or over time in Spain or Italy to estimate the effect of cutting costs on hospital quality (Borra et al., 2019; Arcà et al., 2020). In an austerity context however, other factors, such as economic hardship and delayed access to care, can directly affect population health, potentially confounding the direct role of hospital quality on patient heath. Similarly, enforced austerity measures may affect hospital quality in many different ways, via cutbacks in medical staff, reduced hospital capacity or availability of technology and equipment, making it difficult to identify the relevant mechanisms causing changes in quality (Borra et al., 2019).

More broadly, there is a large literature exploring hospital behaviour in relation to e.g., financial incentives, hospital market structure, reputational concerns, ownership status or institutional context. A large literature has studied the effect of changes in payment systems, in the context of the widespread adoption of activity-based hospital payment (Hodgkin and McGuire, 1994; Farrar et al., 2009). Changes in prices prompt hospitals to increase activity for incentivised surgeries or upcode patient severity (Dafny, 2005; Januleviciute et al., 2016; Verzulli et al., 2017). The empirical literature on the effect of marketplace competition indicate mixed effects on quality of care depending on the particular features of the pro-competition policies (Kessler and McClellan, 2000; Gowrisankaran and Town, 2003; Cooper et al., 2011; Gaynor et al., 2015), but suggests positive effects on efficiency (Cooper et al., 2018; Longo et al., 2019) and management quality (Bloom et al., 2015). Other studies consider the role of intrinsic motivation and reputational concerns by reviewing the effect of reporting public information on hospital performance. The evidence, mostly from the U.S., on quality efforts is mixed (Dranove et al., 2003; Lindenauer et al., 2007; Lagarde et al., 2019; Yoon, 2019), while there are some concerns that providers engage in patient screening. Ownership status may also impact the care provided. The existing evidence for Europe suggests that private hospitals are not providing better nor more efficient care (Kruse et al., 2018; Moscelli et al., 2018a).

Evidence from France indicates that private for-profit hospitals provide similar quality of care as public hospitals due to a higher capacity to use innovative procedures (Gobillon and Milcent, 2016). The institutional context may also influence hospitals' provision of care (Francese et al., 2014; De Luca et al., 2021). De Luca et al. (2021) show that hospitals operating in regions where the institutional quality is lower, tend to provide less appropriate care, measured by higher rates of unscheduled c-sections.

2 Institutional context

Since the creation of the National Health Service (hereafter, NHS) in 1948, access to health care services in England has been universal and funded primarily through general taxation. Health care is free at the point of use, and this includes both planned (elective) and emergency care (Dunn et al., 2016)[54]. However, to access planned hospital care patients need a referral from their family doctor (general practitioner), who act as gatekeepers. Most hospital care for NHS patients is provided by public hospitals (NHS Trusts). These are public bodies subject to financial and regulatory control and are expected to break even[55].

Since 2003 in England, hospitals have been funded primarily through a prospective payment system, moving away from the previous system of block grants (fixed budget). The prospective payment system, common across OECD countries to fund hospital care, maps the activity of care provided to pre-determined tariffs. Each patient is mapped to a Healthcare Resource Group (HRG), the English equivalent of the Diagnosis-Related Groups (DRGs) in the U.S., which consists in a series of categories defined by patient characteristics such as diagnoses, age and procedures carried out. The payment made for a given HRG is calculated from national costs in recent previous years (Grašič et al., 2015).

Since 2008 and the austerity period that followed the financial crisis, spending on health care has considerably slowed down, growing by 1.4% annually on average, as opposed to a long-term annual average growth rate of 3.7% (The King's Fund, 2020). In this context, the

[54] Only prescriptions of certain pharmaceuticals, dental and optical care are subject to user charges.

[55] NHS providers that have been granted a Foundation Trust status have more financial flexibility and are for instance able to retain surpluses to reinvest in later years.

healthcare sector has been under substantial pressure to make efficiency savings of between £15 to £20 billion between 2011 and 2014, resulting in numerous spending cuts and declining financial strength of hospitals (NHS, 2009; Lafond et al., 2014). The unprecedented slowdown in health funding over the past decade together with a rising demand for healthcare services have put financial pressure on the NHS. In 2018/19, nearly half of the NHS providers in England were in financial deficit at the end of the year (Anandaciva, 2020).

Over time, several pay-for-performance schemes have been introduced with the objective of linking payment more closely to providers' performance, both in primary and secondary care. In the latter, some of these schemes, labelled "best practice tariffs", incentivise higher quality of care for a range of conditions (e.g., cataract surgery, gallbladder removal, stroke, or fragility hip fracture) by inducing hospitals to adopt a range of process measures corresponding to the best practice (Meacock et al., 2014).

Some schemes also focused on incentivising efficiency of care. Between 2010 and 2014, a pay-for-performance scheme was introduced and gradually expanded to incentivise hospitals to increase the number of patients discharged on the same day as admission, and therefore reduce costs. The scheme was informed by the recommendations from medical associations who identified a list of 13 planned and 19 emergency conditions that could be safely treated on the same day as admission rather than involving an overnight admission (British Association of Day Surgery, 2006; British Association for Ambulatory Emergency Care, 2014). Under the scheme, a bonus is paid to hospitals for each patient discharged on the same calendar day as admission, with the aim of progressively changing care practices in hospitals (e.g., allowing for re-organization of the hospital ward) and reducing medically unnecessary stays (see Gaughan et al. (2019) for a comprehensive overview of the scheme).

Chest pain, which is the focus of this study, is part of the list of emergency conditions that where included in this pay-for-performance scheme encouraging hospitals to discharge patients on the same day as admission. Prior to the policy, around 40% of chest pain patients were discharged on the same day as their admission, well below the national recommended rate of 60%. Since 2012, hospitals have been financially incentivised to admit and discharge patients presenting at the Emergency Department with mild chest pain on the same calendar day. The financial bonus was significant: hospitals received a tariff for each patient discharged within the same day which was 37% higher than the tariff for overnight stays (£748 vs £543), despite it being more costly to keep patients overnight (Monitor and NHS England, 2014).

3 Conceptual framework

To understand how encouraging hospitals to increase technical efficiency might affect hospital quality, we briefly review some key results from the theoretical literature on hospital behaviour. Policymakers contract hospital activity with the aim to keep costs of care down while also inducing hospitals to engage in quality improvement efforts (Chalkley and Malcomson, 1998a). Prospective payment systems where the fixed case-based tariffs are set at average costs of providers (yardstick competition), unlike cost reimbursement schemes, provide an incentive for hospitals to contain costs as hospitals can retain the cost savings. Such schemes may also induce hospitals to invest in quality if patient demand responds to quality, as more patients will mean higher hospital revenue for the hospital (Ma, 1994; Chalkley and Malcomson, 1998b). Only when demand is inelastic, fixed price systems may induce hospitals to lower costs at the expense of quality if the fixed tariff is low, though the effect of fixed price on quality in this context depends ultimately on hospitals' degree of altruism (Chalkley and Malcomson, 1998a).

A related strand of the theoretical literature has been concerned with the effect of payment schemes where providers are faced with two possible treatment options, typically one more intensive than the other, for a given diagnosis. Under prospective payment system, implementing different tariffs for each treatment type, rather than having a unique tariff based on patient diagnosis, gives hospitals an incentive to overprovide the most intensive (and profitable) treatment (Malcomson, 2005; Siciliani, 2006; Hafsteinsdottir and Siciliani, 2010). Keeping a unique tariff per diagnosis induces hospitals to under-provide the most intensive treatment option, only if hospitals' degree of altruism is sufficiently low (Hafsteinsdottir and Siciliani, 2010). Hafsteinsdottir and Siciliani (2010) further show that under prospective payment systems where fixed tariffs are based on the average cost across providers, refining tariffs incentivizes providers to over-provide the more intensive and profitable treatment, especially if low-severity patients benefit more from the most intensive treatment than the less intensive option.

In our setting, policymakers set a higher tariff for discharging low risk chest pain patients on the same day as admission (i.e., the less intensive treatment option), than for keeping patients overnight (i.e., most intensive treatment). However, in this context, the less intensive treatment option is the most profitable, given a higher tariff and the lower costs incurred and low-severity patients likely derive higher clinical benefits more from the less intensive treatment, i.e. same-day admission rather than overnight stays, given the risk of hospital-acquired infections and the discomfort of not being home. The literature suggests that hospitals would admit and discharge patients on the same day as long as the marginal savings do not exceed the marginal reductions in patient benefits. Hospitals may therefore have a financial incentive to overly treat patients as same-day admission when same-day discharge care becomes more profitable after the change in tariff, depending on how hospitals weigh profits considerations against clinical considerations.

4 Data

4.1 Sample

We use detailed patient-level administrative hospital data from English NHS hospitals, known as the Hospital Episodes Statistics data set. We extract hospital records for all emergency (i.e., unplanned) patients admitted to hospital with chest pain between April 2010 and March 2014, i.e., two years before and two years after the 2012 policy incentivising hospitals to discharge patients early, i.e. on the same day as admission, for patients with chest pain symptoms (N= 1,269,235 observations). The policy which introduced a bonus for discharging patients early is expected to encourage hospitals to reduce inpatient length of stay of low-risk patients mainly from one or two days to zero days where deemed clinically appropriate (Department of Health, 2012, p.60). Consequently, we exclude the most severely ill patients, which has two main advantages. First, more severely ill patients have little chance of being discharged on the same day as admission, as this is unlikely to be clinically appropriate. Including these patients would therefore dilute the impact of this policy, which we use for our IV strategy. Second, excluding the severely ill patients for which early discharge is not clinically appropriate increases the homogeneity of our sample. The choice to only remove relatively small proportions of the sample strikes a balance between homogeneity and generalisability of results.

We proxy severity of condition in several ways: aged over 95 years (top percentile); diagnosed with more than seven comorbidities as defined by Elixhauser (Elixhauser et al., 1998) (around two percent of the original sample); length of stay over seven days (around four percent of the initial sample); arriving from a care home or psychiatric institution (less than one percent of the original sample); patients with a previous admission for chest pain within the past 365 days (around 15% of the original sample)[56]. We also exclude patients who live more than 100 km from the admitting hospital (one percent of the original sample). While not a proxy of severity per se, distance may be a barrier to rapid discharge (e.g., due to concerns over excessive stress of travel; or availability of transport at certain hours of the day) or it might indicate data error.

Our IV strategy (explained in detail in the Methods section) exploits the variation in the effect of the bonus policy induced by the time at which patients arrive at the Emergency Department (hereafter ED). We therefore merge patient hospital records with the ED data set to obtain the exact hour at which each patient arrived at the Emergency Department. Around 73% of the sample successfully merges. Unmerged records may be due to patients who were admitted to a hospital bed without first attending the Emergency Department. For instance, patients may be given an urgent referral to the hospital from their family physician (General Practitioner, GP) in which case there is no ED record for these patients. Our final sample consists of 735,693 patients.

To understand the impact of using the merged sample, we present in Table C.1 in Appendix key summary statistics for patient characteristics for i) the sample before merging with the ED

[56]A history of chest pain admissions may indicate a more severe underlying condition. It might also suggest that the previous treatment was not intensive enough to provide sufficient relief.

data set and ii) our final merged sample. Patient characteristics are similar in both samples in terms of the number of comorbidities, patient age, gender, ethnicity, income deprivation, percentage of the sample suffering from cardiac risk factors or having had an emergency admission in the past year. This suggests that restricting our sample to the merged sample does not impact the representativeness of our sample. The only notable difference is that the rate of same-day discharge is higher in the original unmerged sample (48% versus 43% in our final sample). This could be due to patients who were referred and admitted to the hospital directly after seeing their GP. These patients are therefore more likely to be admitted during daytime (ie. GP office hours) rather than late at night and do not have to wait at the Emergency Department. This can speed up their admission and subsequent discharge and explain why they were on average more often discharged on the same day as admission. Importantly, because both samples are very similar in terms of the many (observed) patient characteristics, differences in same-day discharge rates are unlikely to be due to unobserved patient severity.

4.2 Dependent variable

Our dependent variable is an indicator variable equal to one if the patient has an emergency (unplanned) readmission within 28 days of being discharged from the hospital, and zero otherwise. Readmission rates are a commonly used measure of health outcomes alongside mortality rates (Friebel et al., 2018; O'Dowd, 2018). In our sample for chest pain patients, the average 30-day mortality rate is infrequent, at about 0.5%, making it less responsive to finer variations in hospital quality.

4.3 Independent variables

Our key independent variable is an indicator variable equal to one if a patient is admitted to hospital, following an Emergency Department attendance, and discharged on the same calendar day. Control variables include patient-level clinical and socio-economic characteristics such as patient age, sex, ethnicity (coded as white or non-white), number of Elixhauser comorbidities (Elixhauser et al., 1998) reported in the current hospital admission (current Elixhauser comorbidities) and in past hospital admissions during the year preceding the index chest pain admission (past Elixhauser comorbidities). We also control for whether the patient had any past (all cause) emergency hospital admission in the year preceding the index chest pain admission or whether the patient arrived at the Emergency Department by ambulance; both of which may indicate more severe symptoms or underlying medical conditions. Information on month and year of attendance are extracted from hospital admission data.

In some cases, chest pain may be the symptom of serious and potentially life-threatening cardiac or respiratory conditions[57]. We take into account the patients' cardiac or pulmonary comorbidities, which may affect both a doctor's decision to discharge patients quickly and patients' subsequent health outcomes (Prina et al., 2004). In particular, we include a separate indicator variable for whether the patient has been diagnosed with cardiac arrhythmia, congestive heart failure, valvular heart disease or pulmonary circulatory disorders during the index chest pain admission.

We use relative income deprivation in the area of the patient's residence as a proxy for their socioeconomic status. Income deprivation is measured as the proportion of people in a small geographic area (i.e., lower-layer super output areas (LSOAs)) who claim means-tested benefits (e.g., income-based employment and support allowance). We assign each LSOA to quintile groups of the national distribution of income deprivation, and map these to patients based on their geographic identifiers recorded in HES.

To implement our instrumental variable approach, we construct an indicator variable equal to one if the patient arrived at the Emergency Department between 6am and 5pm (daytime). Our instrument relies on variation in the effect of the policy incentivizing early patient discharge from the hospital. We observe that patients who attended the ED during daytime were more impacted by the policy: the rates of patients who were discharged on the same day as admission increased faster after the policy than for patients who attended the Emergency Department at night (see Section 5.2 on our instrumental strategy and Figure 3.1). Our in-

[57]See patient health information for chest pain from the Mayo Clinic: https://www.mayoclinic.org/diseases-conditions/chest-pain/symptoms-causes/syc-20370838 [Last accessed: 27/10/2020].

strument is the interaction of an indicator variable equal to one if the patient arrived at the ED during daytime and an indicator variable for the post-policy period 2012 onwards.

4.4 Summary statistics

Around nine percent of patients in our sample have an emergency readmission within 28 days of discharge from their index admission, and 43% were admitted and discharged on the same day, which we refer to as same-day discharge (Table 3.1).

Table 3.1: Summary statistics

	Mean	Std. Dev.	Min.	Max.
28-day emergency readmission	0.09	0.28	0	1
Same-day discharge (SDD)	0.43	0.49	0	1
Patient age	58.74	17.44	19	95
Male patient	0.53	0.50	0	1
White ethnicity	0.80	0.40	0	1
Elixhauser: current admission	1.19	1.19	0	6
Elixhauser: past admissions (1 year)	0.65	1.27	0	6
Congestive heart failure	0.04	0.19	0	1
Cardiac arrhytmia	0.10	0.29	0	1
Valvular disease	0.04	0.18	0	1
Pulmonary circulatory disorder	0.01	0.10	0	1
Past emergency admission	0.25	0.43	0	1
Arrived to the ED by ambulance	0.54	0.50	0	1
Distance to admitting hospital (km)	9.98	11.08	0	101
1st quintile - Least income deprived	0.15	0.36	0	1
2nd quintile	0.21	0.40	0	1
3rd quintile	0.17	0.38	0	1
4th quintile	0.23	0.42	0	1
5th quintile - Most deprived	0.24	0.43	0	1
Daytime ED arrival (6am to 5pm)	0.72	0.45	0	1
Observations	735693			

Notes: Same-day discharge patients are admitted to a hospital bed and discharged from the hospital on the same calendar day. Quintiles of income deprivation are based on the national distribution of income deprivation in England and measured for patients' small area of residence.

The average patient is 58 years old. 53% of patients are male, and 80% are of white ethnicity. The average patient has 1.19 comorbidities recorded in the index chest pain admission, and an additional 0.65 comorbidities recorded during hospital admissions in the previous year.

Four percent had congestive heart failure, ten percent had cardiac arrhythmia, four percent had valvular disease, and one percent suffered from pulmonary circulatory disorder. 25% had a past emergency admission. 54% attended by ambulance. The average patient resides about 10km from the admitting hospital. Almost 50% of the sample fall in the two most deprived quintiles of income deprivation. Nearly three-quarters (72%) of the patients arrived at the Emergency Department during daytime, defined as between 6am and 5pm.

5 Methods

5.1 Empirical strategy

Our aim is to measure the causal effect of being discharged on the same calendar day of the index admission on the probability of the patient being readmitted as an emergency within 28 days of hospital discharge. We employ the following regression model:

$$y_{iht} = \beta_0 + \alpha_h + \phi_t + \beta_1 SDD_{iht} + X'_{iht}\beta_2 + \epsilon_{iht}, \tag{3.1}$$

where y_{iht} is the probability of having an emergency readmission within 28 day of hospital discharge for patient i admitted to hospital h in year t. SDD_{iht} is an indicator variable equal to one if patient i was admitted to a hospital bed and discharged on the same day (same-day discharge, SDD). α_h is a vector of hospital fixed effects that control for time-invariant hospital factors[58]. ϕ_t is a set of indicator variables for each financial year, which allow for aggregate changes in quality of care over time that arise due to e.g. improvement in medical knowledge and guidelines, or technology advancement[59].

X_{iht} is a vector of patient characteristics comprising patient age (in 10-year bands), sex, ethnicity and the number of past and current Elixhauser comorbidities in categories (zero, one, two to three or four to six comorbidities) to allow for non-linear effects on post-discharge

[58]Hospital fixed effects would also account for unobserved regional differences, such as potential difference in the number of emergency departments in an area, difference in doctors' propensity to admit and discharge patients or differences in patients' living arrangements. For instance, in areas with higher proportion of people living alone, hospitals may keep patients overnight more often.

[59]For simplicity, hospital and year fixed effects are denoted in the same way in all equations.

health outcomes. Patient controls also include a set of indicator variables for being diagnosed with cardiac or pulmonary risk factors, for a past emergency admission, ambulance arrival to the Emergency Department and the quintile of income deprivation at the small area level. Patient controls also include a set of dummies for month of the year and day of the week of admission, to control for seasonality effects, which may affect both the probability of being discharged the same day and health outcomes. Importantly, we include a set of dummies for each hour of the day of ED arrival to control for the fact that patients who arrived late at night might have more severe symptoms or receive different standards of care, for example, due to lower staffing levels. Our coefficient of interest, β_1, measures the effect of being discharged on the same day on the probability of having an emergency readmission within 28 days. We estimate Equation (3.1) by Ordinary Least Squares using a linear probability model. All standard errors are clustered at the hospital level.

5.2 Instrumental variable approach

Despite the large set of patient controls, a potential endogeneity concern with estimating Equation (3.1) by OLS is that patients discharged on the same day may be unobservably different from patients admitted for an overnight stay. Any remaining unobserved differences in patient severity would therefore introduce an omitted variable bias in our estimate of β_1. In particular, we would expect patients who are admitted and discharged on the same day to be (unobservably) healthier on average than overnight patients, thus introducing a downward bias in the OLS estimate of the effect of same-day discharge treatment on readmissions.

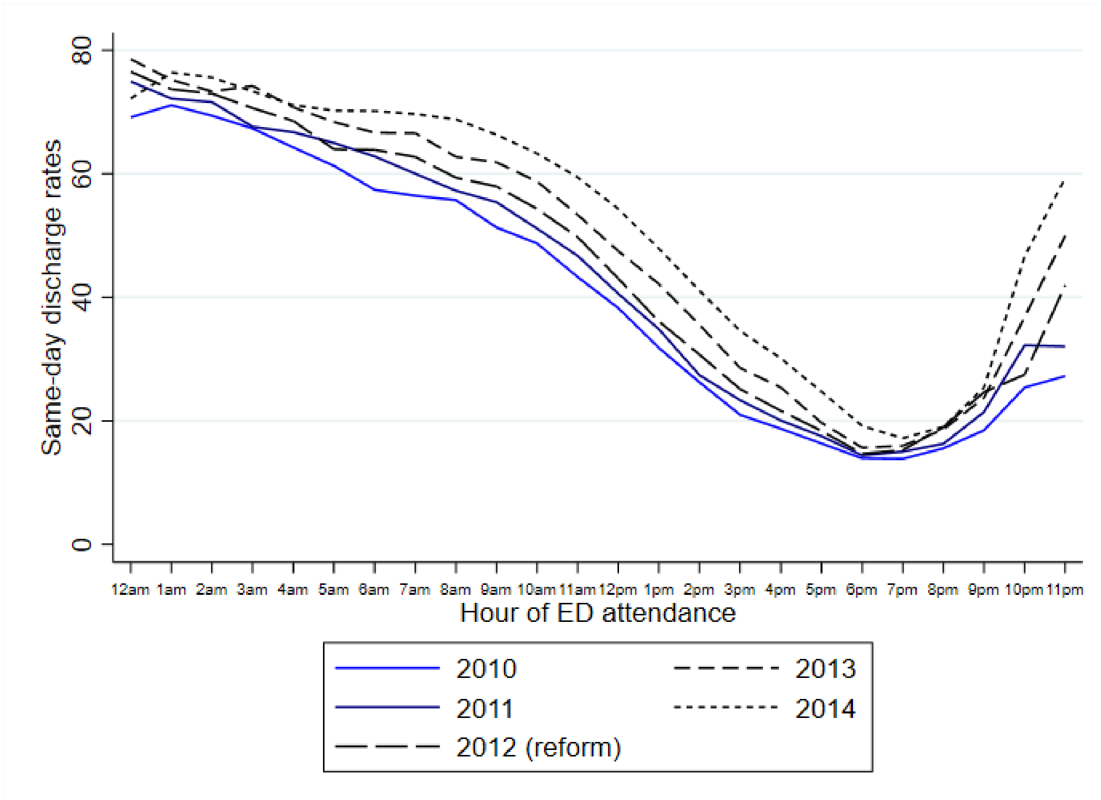
We use an instrumental variable approach to address this possible omitted-variable bias. We instrument the indicator variable for being discharged on the same day as admission by exploiting variation in patient exposure to a 2012 policy. The latter introduced a bonus payment (see Section 2) that has been shown to increase the rates of patients admitted and discharged on the same day nationally (Allen et al., 2016; Gaughan et al., 2019). Specifically, our IV strategy exploits the variation in the policy-induced increase in such rates for patients arriving to the ED at different times of the day. Our treatment of interest, referred to as same-day discharge, is defined according to the calendar day of admission and discharge. Patients who arrive to the ED in the early hours of the day have more time to be admitted and discharged on the same day than patients who arrive late at night. By construction, the rate of patients admitted and discharged on the same day is therefore highest for patients who attended the ED in the early hours of the day and decrease as the day passes because the number of hours left to be discharged before midnight of the same calendar day decreases. Figure 3.1 illustrates this pattern in our data. Note that we focus on the time of attendance at ED and not the time of admission to the hospital ward since the former is arguably exogenous to the hospital and not affected by clinical decisions that are susceptible to gaming[60]. Once patients arrive at the Emergency Department, ED personnel will conduct some initial screening and diagnostic work to determine patients' health status and the need for hospitalisation. As a result, patients who attended the ED in the evening will most likely be admitted sometime after midnight, i.e., in the next calendar day, and, as a consequence, we observe that the rate of patients discharged on the same day as admission starts rising again from 7pm onwards.

Figure 3.1 also shows that the rates of patients discharged early increased markedly after the start of the BPT policy, as the plotted lines (dashed and dotted) shift outwards after 2012. Importantly, the rates of patients discharged on the same day as admission increased faster after 2012 for patients who arrived at the ED during daytime, i.e. between 6am and 5pm. Conversely, such rates increased less for patients who arrived at the ED at night, indicating that they were less impacted by the payment policy. There are two possible reasons for this. First, patients who arrived in the evening may have been admitted just before midnight, thus leaving little room for hospitals to respond to the policy incentive by increasing the proportion of patients who are discharged without staying overnight. Alternatively, for patients who were admitted just after midnight hospitals already had ample time to discharge all appropriate patients during the same calendar day prior to the policy reform. As a result, the policy response was strongest for patients who arrived during daytime hours, when hospitals had most

[60]We are also limited by the depth of data recording in inpatient records. Whilst we have the exact time at which patients attend the ED, we only observe the date at which they are admitted to the hospital ward.

scope to actively seek to discharge patients early in response to the policy incentive.

Figure 3.1: Same-day discharge rates by arrival time at the Emergency Department, by year



Notes: Same-day discharge refers to patients who were admitted and discharged from the hospital on the same calendar day. Plot of the rates of same-day discharge patients per hour of patient arrival at the Emergency Department for each year of the sample (2010-2014).

Figure 3.1 provides the intuition for our instrumental variable strategy. We use these observed patterns of treatment rates over time to derive our instrument. The rest of this section lays out our IV strategy more formally. The first stage regression of our IV approach is the following:

$$SDD_{iht} = \delta_0 + \alpha_h + \phi_t + \delta_1(post_t * D_i) + X'_{iht}\delta_2 + \upsilon_{iht}, \tag{3.2}$$

where $post_t$ takes the value of one from 2012 onwards when the policy came into effect, and zero otherwise. D_i is an indicator variable for arriving at the Emergency Department during daytime, i.e. between 6am and 5pm. By defining daytime admissions to correspond to the hours where the reform had the most important effect on increasing the rates of same-day discharge treatment, we ensure that we have a strong first stage[61]. A patient's exposure to the policy is determined both by the year of admission relative to the timing of the reform and by the time of day at which they arrived at the ED. Our instrument is therefore the interaction of both indicator variables, $(post_t * D_i)$. After controlling for year fixed effects and hour of arrival fixed effects, the interaction term between being admitted after the reform and arriving at the ED during daytime[62] creates plausibly exogenous variation in same-day discharge treatment and serves as our instrumental variable[63].

In the second stage, the readmission indicator variable is regressed on the predicted probability of being discharged the same day, $\widehat{SDD}_{iht}$, obtained from the first-stage regression, as follows:

$$y_{iht} = \beta_0 + \alpha_h + \phi_t + \beta_1 \widehat{SDD}_{iht} + X'_{iht}\beta_2 + \epsilon_{iht}, \tag{3.3}$$

The IV results are estimated by standard two-stage least squares (2SLS), using the Stata command xtivreg. Robust standard errors are clustered on hospitals to account for possible correlation across patients within hospitals.

5.3 Instrument validity

For our instrumental variable strategy to be valid, our instrument must satisfy a set of assumptions. First, patient exposure to the policy, determined here by the time of arrival at the ED and financial year, must be significantly correlated with changes in the rates of patients

[61]We show results of alternative cut-offs for the definition of daytime arrival in robustness checks, section 6.

[62]Both indicator variables are perfectly collinear with the year dummies and the hour of arrival dummies. Therefore, only their interaction term is included in Equation (3.2).

[63]Note that our first stage is akin to a difference-in-differences estimate of the effect of the policy on the rates of same-day discharge treatment. Several studies (Duflo, 2001; Ma, 2019; Américo and Rocha, 2020)) have similarly exploited variation in the intensity of a national reform to infer the causal effect of a policy in a difference-in-differences type design by interacting a group dummy with a post-policy dummy. This uncovers the causal effect of the program, under the assumption that in the absence of the program, the increase in same-day discharge rates (or, years of schooling in Duflo (2001)) would not have been systematically different for patients who presented at the emergency department during day or night time (or, for regions with more or less schools constructed in the study by Duflo (2001)).

admitted and discharged on the same day (relevance condition). Figure 3.1 supports this notion and we provide additional evidence that this condition is satisfied in the next section where we discuss the results of the first stage.

Second, patient exposure to the policy should be exogenous, conditional on our set of controls (independence condition). Note that only the interaction of time of arrival at the emergency department and the post-policy dummy, ie. our instrument, is exogenous. This independence assumption, though untestable, is likely to hold: our instrument is the interaction of a major policy, whose timing was decided nationally and implemented for all hospitals at the same time, and the hour at which patients arrive at the Emergency Department with chest pain symptoms. We focus on an emergency condition: patients go to the ED after the onset of the chest pain symptoms, which precludes patient self-selection into hospital if they anticipate a certain treatment. The timing of the onset of symptoms and the subsequent arrival at the ED are also exogenously determined and not subject to hospital decision-making. We cannot rule out that patients who arrived late at night suffer from more severe or acute chest pain. However, potential differences in symptoms severity by time of day are accounted for via our set of hour fixed effects as long as their effect on healthcare seeking behaviour does not vary over financial years, which seems plausible in this context.

Third, our instrument should only impact patient health outcomes *indirectly* through the increase in the rate of same-day discharge (exclusion restriction). This assumes that patient exposure to the policy is unrelated to unobserved characteristics that may directly affect our outcome variables, namely emergency readmissions. We assess the likelihood of this assumption in the next section. Finally, the effect of our instrument should be monotonic (monotonicity assumption) (Imbens and Angrist, 1994). While our instrument might have had no effect on the probability of being discharged on the same day for certain groups of patients, there should be no patients who would have been discharged the same day before the reform but were admitted with an overnight stay after the reform (i.e. called 'defiers'). This scenario seems unlikely given that the policy provided a generous financial incentive for hospitals to avoid overnight stays for this population and the publication of clinical recommendations.

Further, the nationally recommended rate for chest pain patients was noticeably higher than the actual rate prior to the policy (60% recommended vs. 40% actual).

If these conditions hold, our IV estimand is the weighted average causal effect of being discharged on the same day for patients who were only discharged on the same day because of the policy but who would have otherwise stayed overnight (i.e., local average treatment effect, LATE) (Angrist and Imbens, 1995; Angrist and Pischke, 2008).

6 Results

6.1 First-stage results

Table 3.2 shows the effect of our instrument on the probability of being discharged on the same calendar day (first-stage results; equation (3.2)). Because our instrument utilises the increase in the rates of patients discharged on the same as admission induced by the bonus policy, the first stage results can be interpreted as the effect of the policy on same-day discharge. Our findings indicate that the bonus policy implemented in 2012 for chest pain increased the rates of same-day discharge by around five percentage points (11% effect relative to the average same-day discharge rate in the sample) for emergency chest pain patients. The effect is statistically significant at the 0.1% level. The policy effect is in line with the previous related literature (Allen et al., 2016; Gaughan et al., 2019) which find effects of similar magnitude for the same policy using different health conditions. The partial effective F-statistics for the instrument is 40, above the minimum recommended value of 23 (Montiel-Olea and Pflueger, 2013).

Table 3.2: Effect of the instrument on same-day discharge rates (first stage)

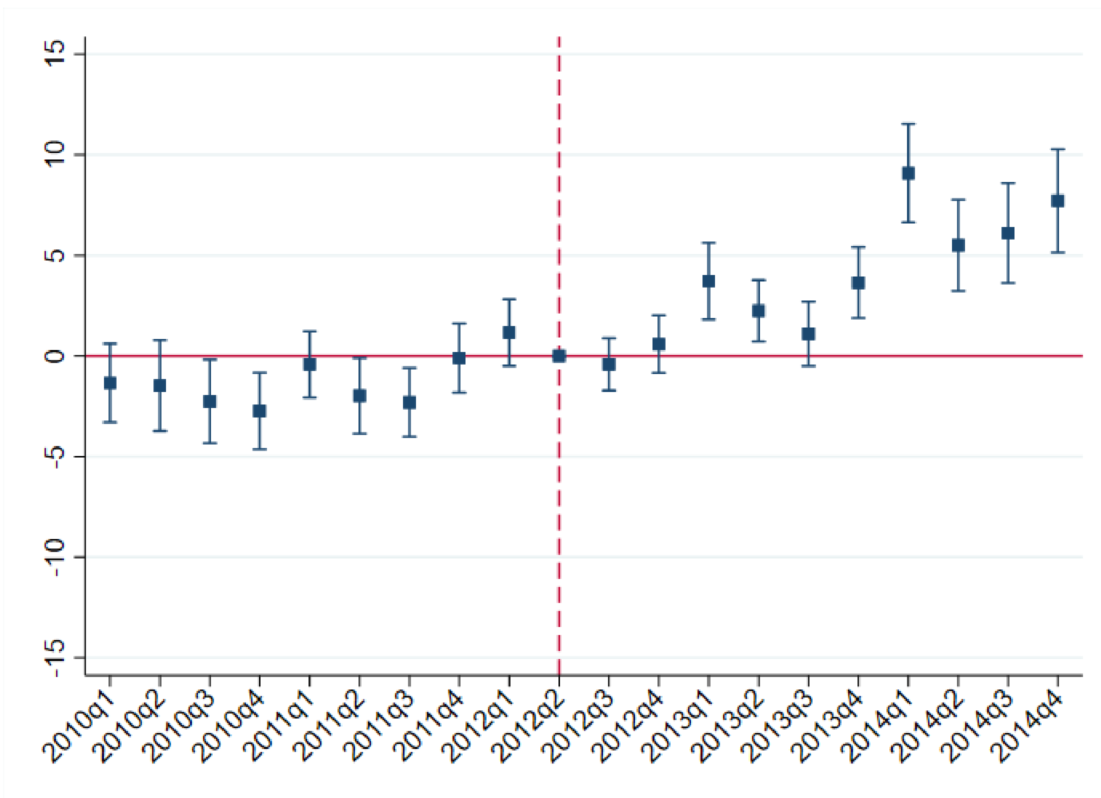
	Same-day discharge
Post-reform*daytime ED arrival (IV)	4.832***
	(6.325)
Hospital fixed effects	Yes
Year dummies	Yes
ED hour dummies	Yes
Patient controls	Yes
F-statistics	40
Observations	735693

Notes: Coefficients are expressed in percentage points. Same-day discharge refers to patients who are admitted and discharged from the hospital on the same calendar day. Daytime ED arrival is an indicator variable for patients who arrived at the Emergency Department between 6am and 5pm. Patient controls include patient age (in 10-year bins), sex, ethnicity, past and current Elixhauser comorbidities in categories, past emergency admission, ambulance arrival and quintile of income deprivation. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The full table of coefficients is shown in the Appendix, Table C.2. The results show that, as expected, patients with more Elixhauser comorbidities or who are diagnosed with cardiac or pulmonary risk factors are significantly less likely to be discharged on the day as admission. The effect is important, ranging from around six percentage points difference for patients with valvular disease to 19 percentage points differences in the same-day discharge probability for patients with four to six comorbidities. Older patients have also a significantly lower probability of being discharged on the same day. The effect of age on same-day discharge monotonically increases across five-year age categories, from two percentage point difference for patients aged 26 to 35 years old, up to 19 percentage points difference for patients over 85 years old, relative to patients between 19 and 25 years old. Results also show that, holding everything else constant, male patients have a lower probability of having a same-day discharge than female patients. The medical literature suggests that chest pain is often misdiagnosed in women (Martinez-Nadal et al., 2021), which could explain this result together with biological differences in risk factors.

To test whether the exclusion restriction assumption is likely to hold, we run a similar first-stage regression but substitute the post-policy dummy ($Post_t$) with dummies for calendar quarter for the whole period, similar to an event study design. Figure 3.2 plots the coefficients on the interaction between the indicator variable for daytime ED arrival and the calendar quarter fixed effects. Most of the coefficients before the first quarter of 2012 are not statistically different from zero, indicating that there were no differential trends in the rates of patients discharged early (same-day discharge) among daytime vs night-time ED attendance prior to the policy. After 2012, rates begin to diverge as evidenced by the positive and statistically significant coefficient estimates. Note that the effect of the policy increased over time, with noticeably larger impacts observed two years after the reform (2014).

Figure 3.2: Effect of the instrument on same-day discharge rates, by quarter of year



Notes: Plot of the coefficients (percentage points) and the 95% confidence intervals for the interaction of the dummy for daytime ED arrival (6am to 5pm) and quarter dummies on same-day discharge rates. Quarter 1 stands for January-March, Quarter 2: April-June, Quarter 3: July-September, Quarter 4: October-December. The red dashed line is the quarter when the bonus payment policy started. The reference quarter is the second quarter of 2012 (2012Q2), corresponding to the start of the scheme for chest pain condition.

These results confirm that the differential increase in same-day discharge rates between day and night-time patients is due to the reform, rather than to some unobserved differences between groups of patients. Unobserved differences could affect patient health outcomes directly, which would violate the exclusion restriction assumption. Figure 3.2 suggests that there is variation in same-day discharge rates across quarters of years; the first quarter (Q1), i.e. the winter months, is almost always associated with higher rates of same-day discharge care. While we account for some of the seasonality variations via our month dummies, same-day discharge rates still fluctuate across quarters. A possible interpretation is that same-day discharge practice is influenced by hospitals' bed occupancy rate (Harrison et al., 2013; Friebel et al., 2018), which may be higher during winter months given a higher incidence of flu or hip fractures occurring over winter.

6.2 Main results

Table 3.3 reports the OLS and IV results. The model with OLS coefficients indicates that same-day discharge care is associated with a lower risk of emergency readmissions, by 0.875 percentage points, equivalent to slightly under a 10% effect relative to the sample average readmission rate of nine percent. The coefficient is highly statistically significant at the 0.1% level. Once we instrument for the chance of being admitted and discharged on the same calendar day, the effect becomes statistically insignificant. The point estimate is positive but small in magnitude and statistically insignificant. The standard errors are larger than under OLS, as is expected with IV. The results are consistent with the hypothesis that OLS results suffer from omitted variable bias, in the form of a downward bias in the estimate of being discharged on the same day as admission on emergency readmissions. Overall, these results indicate that discharging patients on the same day as admission doesn't lead to a statistically significant higher risk of emergency readmission, even though the standard errors are large. The reduced-form results (i.e. the direct effect of the instrument on readmission rates) presented in the next section also point to a null but precisely estimated effect (see Table 3.4).

This suggests that, though imprecise, the IV estimate does point to a null effect.

Table 3.3: Effect of same-day discharge care on emergency readmissions, OLS and IV results

	Emergency readmission	
	OLS	IV
Same-day discharge (SDD)	-0.875***	1.901
	(0.115)	(3.040)
R^2	0.039	0.036
Observations	735693	735693

Notes: Coefficients are expressed in percentage points. Our instrument is the interaction of an indicator variable for daytime arrival at the Emergency Department (6am to 5pm) and a post-reform indicator. Patient controls include patient age (in 10-year bins), sex, ethnicity, past and current Elixhauser comorbidities in categories, past emergency admission, ambulance arrival and quintile of income deprivation. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The full model with all covariates is provided in the Appendix Table C.3. Patient characteristics have the generally expected effects on readmission rates in both models. For instance, patients with more comorbidities, with past emergency admission or patients who suffer from risk factors have a higher risk of readmission. Age has a non-linear effect on the probability of being readmitted. Young patients, between 19 to 25 years old, are at a higher risk of readmissions than patients up until 65, probably indicating unusually severe symptoms for the very young patients, but at a lower risk of patients older than 75 years old. We further observe that more deprived patients have slightly higher risk of readmission. Table C.3 also shows that day of admission is significantly associated with differences in the rates of readmission.

6.3 Reduced-form results

We report here the reduced-form results, equivalent to the direct effect of our instrument on readmission rates. Given the definition of our instrument, these correspond to the causal effect of the reform on patient readmission rates. Table 3.4 shows that we find no effect of the bonus payment policy on patient readmission rates. The coefficient is close to zero and is precisely estimated.

Table 3.4: Reduced-form results

	Emergency readmission
Post-reform*daytime ED arrival (IV)	0.092
	(0.148)
R^2	0.039
Observations	735693

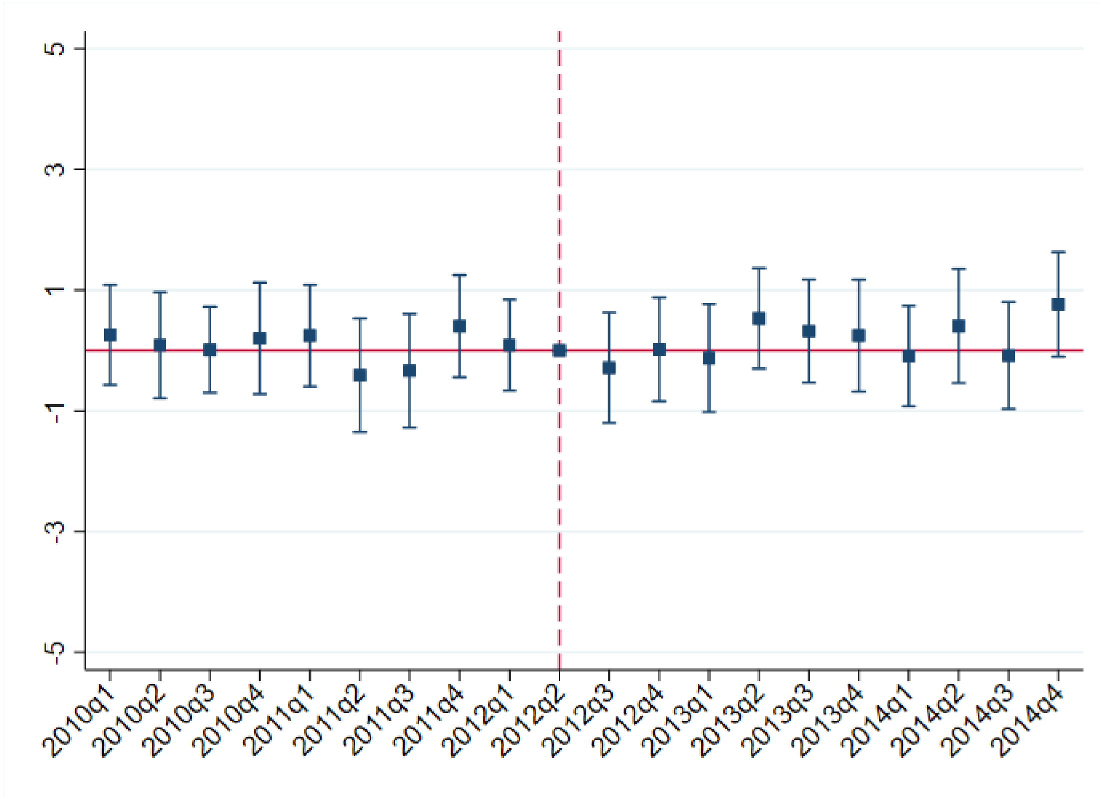
Notes: Coefficients are expressed in percentage points. Daytime ED arrival is an indicator variable for patients who arrived at the Emergency Department between 6am and 5pm. Patient controls include patient age (in 10-year bins), sex, ethnicity, past and current Elixhauser comorbidities in categories, past emergency admission, ambulance arrival and quintile of income deprivation. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We further run an event study design where the post-reform dummy is replaced by quarter-of-year dummies for the whole period, as previously done for the rates of same-day discharge patients in Figure 3.2. This exercise enables us to graph the trends in readmission rates over the whole period. Figure 3.3 shows that readmission rates are not affected by the reform. These findings are in line with results from Allen et al. (2016) which evaluate the effect of the same policy reform on readmission or mortality rates for cholecystectomy patients.

The absence of pre-trends in readmission rates, evidenced by the insignificant pre-reform coefficients, suggests that the reform most likely did not have a direct effect on health outcomes, other than by increasing the proportion of patients discharged on the same day as admission. This is further suggestive evidence that the exclusion restriction assumption should hold.

However, two features of our instrument, the policy, are noteworthy. First, under the bonus payment policy, hospitals received a higher payment for patients admitted and discharged on the same calendar day. Second, the policy might have raised awareness in the medical community about the existing clinical evidence and best practices in the management of chest pain patients. Financial help to hospitals and better dissemination of the medical evidence might have contributed to directly increase the quality of care provided to chest pain patients (Phelps, 2000; Celhay et al., 2019), which would violate the exclusion restriction. The design of our instrument mitigates such concerns: the effect of the policy is estimated over time and within hospitals, based on patients' time of arrival at the ED. Because any changes in the

Figure 3.3: Effect of the instrument on emergency readmission rates, by quarter of year



Notes: Plot of the coefficients (percentage points) and the 95% confidence intervals for the interaction of the dummy for daytime Emergency Department arrival and quarter dummies on 28-day emergency readmission rates. Quarter 1 stands for January-March, Quarter 2: April-June, Quarter 3: July-September, Quarter 4: October-December. The red dashed line is the quarter when the bonus payment policy started. The reference quarter is the second quarter of 2012 (2012Q2), corresponding to the start of the scheme for chest pain condition.

standards of care for chest pain patients would have likely affected all chest pain patients at the hospital (regardless of their time of arrival, ie. daytime or night time arrivals), the possible direct effect of the policy on quality of care is controlled for.

Overall, our baseline results point to a null causal effect, though the effect is not precisely estimated. This is a well-identified feature of instrumental variables: while IV estimates are unbiased, they are also less efficient than OLS estimates. In this context, our reduced-form results estimated by OLS indicate that our instrument, the policy, did not have a direct causal effect on patient health outcomes (Table 3.4). The point estimate is close to zero (0.92) and precisely estimated (SE= 0.148). Similarly, Figure 3.3 shows a consistent and null effect of the instrument on health outcomes over the whole period. This serves to reassure that our main IV results, though imprecise, do point to a null effect.

6.4 Robustness checks

We run a couple of robustness checks. First, we proxy hospital quality by patients' risk of having an (all-cause) emergency readmission within 28 days of hospital discharge. Though commonly used as an indicator of quality of care, emergency readmissions to hospital may also be affected by factors unrelated to the quality of care provided during the index hospital admission[64]. Emergency readmissions may include for instance admissions for hip fracture or stroke, which would introduce some measurement error in our estimates. In a robustness check, we re-define readmissions to include only readmissions with a chest-pain related diagnosis. Chest pain can be caused by issues around the heart or lungs. We use the ICD10 codes chapters to identify all conditions linked to the circulatory and pulmonary system (ICD codes I00-I99 and J00-J99, listed in Appendix, Table C.4). The average readmission rates for circulatory and with respiratory related causes is lower than all-cause readmissions, at around four to five percent of the sample. Results in Table 3.5 show a similarly null effect, with a point estimate close to zero.

Table 3.5: Results with (1) circulatory and (2) including respiratory-related readmissions

	(1)		(2)	
	OLS	IV	OLS	IV
Same-day discharge (SDD)	-0.849***	0.299	-0.869***	-0.433
	(0.062)	(2.333)	(0.064)	(2.487)
R^2	0.011	0.101	0.017	0.016
Observations	735693	735693	735693	735693

Notes: Coefficients are expressed in percentage points. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[64] In our sample, mortality rates are too low (around 0.5%) to introduce bias (e.g. survival bias) in our outcome measures.

Second, we define daytime admissions to be between 6am and 5pm, based on the observation that same-day treatment rates increased most for these patients, resulting in a strong first stage. In a sensitivity analysis, we also show the results from i) defining daytime arrivals to be between 7am and 7pm, which corresponds to the normal working hours of consultants, outside of call duties (British Medical Association, 2009), and ii) not imposing any definition of daytime arrival by interacting each hour of arrival at the emergency department with the post-reform dummy, forming a set of instruments. Results in Appendix, Table C.5 are essentially unchanged, except that the first stage is weaker and the IV results tend to be less precisely estimated.

Third, we run sensitivity analyses around our sample restrictions. We excluded most critically ill patients, ie., patients with a high number of comorbidities, long hospital stays or high number of previous hospital admissions. In Appendix Table C.6, we present OLS and IV results from our main specifications without these sample restrictions. The results hold, despite a weaker first stage (F-stat = 35). In addition, we run the reduced form analysis for the subsample of excluded patients only, to understand the effect of the reform on readmission rates for these patients. A concern might be that health outcomes could have deteriorated for these patients because of the reform's focus on less severely ill patients. Results, in Appendix Figure C.1, indicate no effect of the reform on readmission rates over time, suggesting that the reform did not have any adverse effect on health outcomes for these patients either.

Finally, the theoretical and empirical literature on hospital behaviour suggests that hospitals commonly respond to incentives by increasing the volume of patients treated for incentivised conditions (Chandra et al., 2011). In our context, hospitals could therefore obtain the bonus payment by admitting a higher number of (unobservably) less severely ill patients only for a short stay, which would introduce a bias in our dependent variable (i.e. lower readmission rates). We provide some evidence that hospitals did not engage in such strategies by showing that the evolution of hospitals' volume of admissions and average patient severity was not impacted by the policy. Appendix Table C.7 reports the results from an interrupted time series analysis at the hospital-quarter level which indicates that there was no change in trend in chest pain admissions or average patient severity after the reform.

7 Conclusions

This study investigates whether admitting and discharging patients within one day leads to worse health outcomes for chest pain patients. We find that discharging chest pain patients on the same day as admission is associated with lower readmissions. However, using an instrumental variable to account for unobserved patient severity shows that same-day discharge that has no causal effect on emergency readmissions. Our results provide causal evidence in an observational context that it is safe to discharge low risk chest pain patients on the same day as admission.

More broadly, this study sheds light on the effect of cost-reduction strategies on the quality of care in the hospital sector. Our findings suggest that cost reductions measures that target unnecessary inpatient lengths of stay have no harmful effects on quality of care. This corroborates results from Allen et al. (2016) who find that the English NHS policy that incentivised daycase surgery for cholecystectomy patients had no adverse effects in terms of patients' readmission or mortality rates. Policymakers may wish to encourage reduction in inpatient lengths of stay for certain clinical conditions, as a way of cutting down ineffective spending. Further, freeing up hospital beds may further allow to increase hospital admissions for other conditions, thus contributing to better technical efficiency of hospital resources. Alternatively, discharging patients earlier when possible would help reduce bed occupancy rates which may in turn have positive effects on quality of care.

However, a recent strand of literature has identified potential adverse effects of the austerity measures that followed the 2008 financial crisis on population health. The unintended effects of large budget cuts on patient health seem primarily driven by their impact on hospital staffing levels and resources (Vallejo-Torres et al., 2018; Borra et al., 2019; Arcà et al., 2020; Bordignon et al., 2020). This suggests that cost-containment strategies may target inpatient length of stay, up to a certain point after which measures may be detrimental to the quality of care provided if for e.g. health care staff is overstretched.

C Appendix

Table C.1: Patient characteristics for the sample after sample restriction (1), and final sample merged to Emergency Department records (2)

	(1)		(2)	
	Mean	SD	Mean	SD
28-day emergency readmission	0.09	0.28	0.09	0.28
Same-day discharge (SDD)	0.48	0.50	0.43	0.49
Patient age	58.16	17.57	58.74	17.44
Male patient	0.51	0.50	0.53	0.50
White ethnicity	0.80	0.40	0.80	0.40
Elixhauser: current admission	1.18	1.19	1.19	1.19
Elixhauser: past admissions (1year)	0.65	1.26	0.65	1.27
Congestive heart failure	0.04	0.19	0.04	0.19
Cardiac arrhytmia	0.09	0.29	0.10	0.29
Valvular disease	0.03	0.18	0.04	0.18
Pulmonary circulatory disorder	0.01	0.11	0.01	0.10
Past emergency admission	0.25	0.43	0.25	0.43
Income deprivation (LSOA)	0.17	0.12	0.17	0.12
Observations	1006803		735693	

Notes: Same-day discharge patients are admitted to a hospital bed and discharged from the hospital on the same calendar day. Income deprivation is expressed as the proportion of people experiencing income deprivation in patients' small area of residence called Lower-Super Output Areas (LSOAs).

Table C.2: First stage estimates (full table)

	Same-day discharge	
	Coef.	SE
Post-reform*Daytime ED arrival (IV)	4.832***	(0.764)
Hour of ED arrival (Ref. Midnight)		
1am	-0.499	(1.036)
2am	-1.346	(1.041)
3am	-2.894**	(1.053)
4am	-5.048***	(1.122)
5am	-7.921***	(1.153)
6am	-12.955***	(1.233)
7am	-15.443***	(1.255)
8am	-18.386***	(1.280)
9am	-21.109***	(1.282)
10am	-24.220***	(1.304)

11am	-29.189***	(1.346)
12pm	-34.958***	(1.414)
1pm	-40.837***	(1.470)
2pm	-46.950***	(1.511)
3pm	-52.492***	(1.546)
4pm	-56.364***	(1.555)
5pm	-60.735***	(1.576)
6pm	-62.031***	(1.456)
7pm	-62.734***	(1.520)
8pm	-60.112***	(1.573)
9pm	-55.431***	(1.909)
10pm	-47.580***	(2.858)
11pm	-42.922***	(5.831)
Current Elixhauser comorbidities (Ref. 0)		
1	-6.639***	(0.338)
2 to 3	-12.234***	(0.514)
4 to 6	-19.331***	(0.735)
Past Elixhauser comorbidities (Ref. 0)		
1	-1.093***	(0.198)
2 to 3	0.097	(0.269)
4 to 6	1.001*	(0.403)
Congestive heart failure	-4.761***	(0.367)
Cardiac arrhythmia	-2.320***	(0.236)
Valvular disease	-6.462***	(0.426)
Pulmonary circulatory disorder	-8.937***	(0.750)
Past emergency admission	-4.033***	(0.182)
Arrived to the ED by ambulance	-9.963***	(0.369)
Patient age groups (Ref. 19-25)		
26 to 35	-2.014***	(0.416)
36 to 45	-6.118***	(0.620)
46 to 55	-9.319***	(0.702)
56 to 65	-12.154***	(0.730)
66 to 75	-14.145***	(0.720)
76 to 85	-16.913***	(0.735)
>85	-19.065***	(0.727)
Male patient	-1.075***	(0.166)
White ethnicity	-0.540	(0.274)
1st quintile - Least income deprived		
2nd quintile	0.298	(0.168)
3rd quintile	0.114	(0.195)
4th quintile	0.019	(0.214)
5th quintile - Most deprived	-0.079	(0.276)
Distance to admitting hospital	-0.030	(0.048)
Distance to admitting hospital^2	-0.000	(0.001)
Day of week dummies (Ref. Sunday)		

Monday	1.690***	(0.298)
Tuesday	2.410***	(0.319)
Wednesday	2.417***	(0.287)
Thursday	2.221***	(0.288)
Friday	3.402***	(0.268)
Saturday	1.129***	(0.199)
Month dummies (Ref. January)		
February	-0.744**	(0.243)
March	0.719*	(0.299)
April	-2.700***	(0.324)
May	-2.603***	(0.286)
June	-2.765***	(0.308)
July	-2.383***	(0.286)
August	-3.152***	(0.276)
September	-2.771***	(0.287)
October	-2.362***	(0.295)
November	-1.768***	(0.265)
December	-0.825**	(0.288)
Year dummies (Ref. 2011)		
2010	-2.079***	(0.443)
2012	-1.390**	(0.482)
2013	1.969***	(0.532)
2014	6.110***	(0.769)
Constant	101.314***	(1.464)
Observations	735693	

Notes: Our instrument is the interaction of an indicator variable for daytime Emergency Department (ED) arrival (6am to 5pm) and a post-reform indicator. Coefficients are expressed in percentage points. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.3: Effect on patient 28-day emergency readmission, using OLS or IV model

	OLS		IV	
	Coef.	SE	Coef.	SE
Same-day discharge	-0.875***	(0.115)	1.901	(3.040)
Hour of ED arrival (Ref. Midnight)				
1am	-0.011	(0.261)	0.002	(0.259)
2am	0.181	(0.287)	0.219	(0.285)
3am	-0.078	(0.275)	0.003	(0.299)
4am	-0.090	(0.301)	0.052	(0.343)
5am	0.335	(0.280)	0.557	(0.369)
6am	0.143	(0.292)	0.426	(0.420)

7am	0.012	(0.304)	0.364	(0.505)
8am	-0.197	(0.286)	0.236	(0.579)
9am	-0.785***	(0.228)	-0.276	(0.608)
10am	-1.246***	(0.241)	-0.650	(0.723)
11am	-1.626***	(0.240)	-0.893	(0.851)
12pm	-1.582***	(0.223)	-0.689	(1.024)
1pm	-1.621***	(0.244)	-0.565	(1.184)
2pm	-1.494***	(0.237)	-0.268	(1.381)
3pm	-1.675***	(0.259)	-0.295	(1.534)
4pm	-1.476***	(0.264)	0.012	(1.680)
5pm	-1.583***	(0.257)	0.026	(1.785)
6pm	-1.602***	(0.253)	0.120	(1.913)
7pm	-1.433***	(0.284)	0.309	(1.935)
8pm	-0.716*	(0.304)	0.952	(1.867)
9pm	-0.760	(0.423)	0.775	(1.715)
10pm	-0.435	(0.643)	0.883	(1.647)
11pm	1.189	(1.894)	2.377	(2.289)
Current Elixhauser comorbidities (Ref. 0)				
1	0.801***	(0.092)	0.986***	(0.211)
2 to 3	1.944***	(0.107)	2.284***	(0.393)
4 to 6	4.514***	(0.287)	5.051***	(0.652)
Past Elixhauser comorbidities (Ref. 0)				
1	0.467***	(0.127)	0.497***	(0.129)
2 to 3	2.106***	(0.156)	2.103***	(0.155)
4 to 6	6.893***	(0.270)	6.866***	(0.274)
Congestive heart failure	0.733**	(0.237)	0.866**	(0.264)
Cardiac arrhytmia	0.617***	(0.141)	0.681***	(0.152)
Valvular disease	-0.550*	(0.220)	-0.371	(0.298)
Pulmonary circulatory disorder	1.296**	(0.389)	1.545***	(0.449)
Past emergency admission	6.558***	(0.112)	6.670***	(0.171)
Arrived to the ED by ambulance	1.350***	(0.081)	1.627***	(0.315)
Patient age groups (Ref. 19-25)				
26 to 35	-0.640**	(0.220)	-0.584**	(0.221)
36 to 45	-1.274***	(0.213)	-1.105***	(0.260)
46 to 55	-1.569***	(0.220)	-1.311***	(0.346)
56 to 65	-1.283***	(0.214)	-0.946*	(0.410)
66 to 75	-0.258	(0.236)	0.134	(0.480)
76 to 85	0.996***	(0.234)	1.465*	(0.572)
>85	3.501***	(0.315)	4.030***	(0.630)
Male patient	0.701***	(0.080)	0.731***	(0.084)
White ethnicity	1.465***	(0.090)	1.479***	(0.086)
1st quintile - Least income deprived				
2nd quintile	0.207	(0.111)	0.199	(0.113)
3rd quintile	0.406***	(0.115)	0.403***	(0.117)
4th quintile	0.600***	(0.128)	0.599***	(0.130)

5th quintile - Most deprived	0.935***	(0.124)	0.938***	(0.124)
Distance to admitting hospital	-0.010	(0.009)	-0.010	(0.009)
Distance to admitting hospital^2	0.000	(0.000)	0.000	(0.000)
Day of week dummies (Ref. Sunday)				
Monday	-0.829***	(0.113)	-0.876***	(0.122)
Tuesday	-1.004***	(0.122)	-1.071***	(0.146)
Wednesday	-0.828***	(0.133)	-0.896***	(0.155)
Thursday	-0.821***	(0.125)	-0.882***	(0.139)
Friday	-0.534***	(0.131)	-0.628***	(0.160)
Saturday	0.013	(0.146)	-0.019	(0.147)
Month dummies (Ref. January)				
February	0.052	(0.167)	0.073	(0.168)
March	-3.153***	(0.153)	-3.173***	(0.154)
April	0.441**	(0.144)	0.516**	(0.170)
May	0.320*	(0.155)	0.392*	(0.177)
June	0.226	(0.156)	0.303	(0.171)
July	0.377*	(0.158)	0.443*	(0.172)
August	0.120	(0.156)	0.207	(0.171)
September	0.350*	(0.163)	0.427*	(0.182)
October	-0.133	(0.169)	-0.067	(0.183)
November	-0.115	(0.148)	-0.066	(0.156)
December	0.169	(0.177)	0.192	(0.178)
Year dummies (Ref. 2011)				
2010	0.313**	(0.108)	0.370**	(0.124)
2012	-0.040	(0.135)	-0.098	(0.147)
2013	0.072	(0.142)	-0.079	(0.207)
2014	0.159	(0.156)	-0.107	(0.312)
Constant	5.363***	(0.372)	2.606	(3.009)
R^2	0.039		0.036	
Observations	735693		735693	

Notes: Our instrument is the interaction of an indicator variable for daytime Emergency Department (ED) arrival (6am to 5pm) and a post-reform indicator. Coefficients are expressed in percentage points. Standard errors (in parentheses) are clustered on hospitals. * p < 0.05, ** p < 0.01, *** p < 0.001.

Table C.4: List of ICD-10 codes for readmissions

ICD codes	Description
I00-I02	Acute rheumatic fever
I05-I09	Chronic rheumatic heart diseases
I10-I15	Hypertensive diseases
I20-I25	Ischemic heart diseases
I26-I28	Pulmonary heart disease and diseases of pulmonary circulation
I30-I52	Other forms of heart disease
I60-I69	Cerebrovascular diseases
I70-I79	Diseases of arteries, arterioles and capillaries
I80-I89	Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified
I95-I99	Other and unspecified disorders of the circulatory system
J00-J06	Acute upper respiratory infections
J09-J18	Influenza and pneumonia
J20-J22	Other acute lower respiratory infections
J30-J39	Other diseases of upper respiratory tract
J40-J47	Chronic lower respiratory diseasess
J60-J70	Lung diseases due to external agents
J80-J84	Other respiratory diseases principally affecting the interstitium
J85-J86	Suppurative and necrotic conditions of the lower respiratory tract
J90-J94	Other diseases of the pleura
J96-J99	Other diseases of the respiratory system

Table C.5: IV results from alternative instrument definitions

	Emergency readmission	
	(1)	(2)
Same-day discharge (SDD)	4.349	2.685
	(5.826)	(2.625)
Hospital fixed effects	Yes	Yes
Year dummies	Yes	Yes
ED hour dummies	Yes	Yes
Patient controls	Yes	Yes
First stage F-statistics	19.40	8.9
Observations	735693	735693

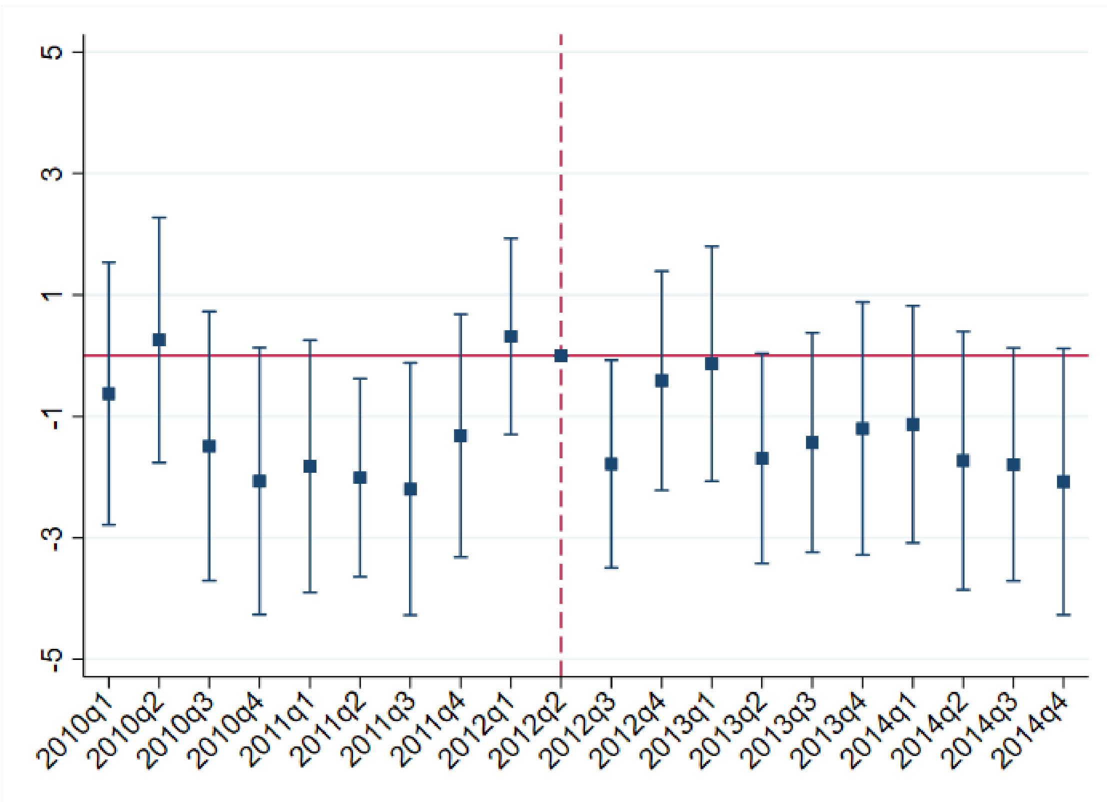
Notes: (1) - the instrument used is the interaction of daytime admission, defined as arrival at the ED between 7am and 7pm, with a post-policy dummy. (2) - the instruments are the interactions between a dummy for each hour of arrival at the ED and the post-policy dummy. Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table C.6: OLS and IV results, without exclusion of most severely ill patients

	Emergency readmission	
	OLS	IV
Same-day discharge (SDD)	-0.587***	3.407
	(0.128)	(4.152)
R^2	0.074	0.069
Observations	930377	930377

Notes: Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure C.1: Trends in emergency readmission rates by quarter of year, for the sample of severely ill patients



Notes: Plot of the coefficients (percentage points) and the 95% confidence intervals for the interaction of the dummy for daytime Emergency Department arrival and quarter dummies on 28-day emergency readmission rates. Quarter 1 stands for January-March, Quarter 2: April-June, Quarter 3: July-September, Quarter 4: October-December. The red dashed line is the quarter when the bonus payment policy started. The reference quarter is the second quarter of 2012 (2012Q2), corresponding to the start of the scheme for chest pain condition.

Table C.7: Evolution of hospital chest pain admissions and average comorbidities

	Hospital admissions	Average comorbidities
Year quarters	-2.661**	0.006*
	(0.876)	(0.002)
Post-reform (2012)	-7.904	0.024
	(15.011)	(0.027)
Post-reform (2012) × Year quarters	1.742	0.002
	(1.434)	(0.003)
Constant	438.344***	1.288***
	(6.893)	(0.013)
Hospital FE	Yes	Yes
Hospital-year quarters	3015	3015

Notes: Standard errors (in parentheses) are clustered on hospitals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Conclusions

This thesis considers several determinants of hospital quality of care, using the English National Health Service as a case study. While there has been a large literature on the effect of financial incentives and market structure on quality of care, this thesis focused on less often explored quality determinants related to the organisation of hospital care provision. The stress on the health care system from the surge in hospital demand experienced during the Covid-19 pandemic has highlighted the importance of hospital capacity planning. This may involve re-thinking hospital wards' organisation (e.g. allowing for Covid and non-Covid patients spots to limit nosocomial infections), ensuring sufficient workforce on the wards or encouraging same-day hospital discharge when clinically possible to allow for spare bed capacity. This thesis's findings shed light on these important aspects of the organisation of hospital care. In addition, this thesis highlighted the importance of accounting for endogeneity. The results in Chapters 1 and 3 present different conclusions, before and after endogeneity is accounted for. This indicates that addressing potential endogeneity is warranted, even when using rich administrative data sets. The evidence provided throughout this thesis has several policy implications, which are reviewed below.

Chapter 1 assesses the often-cited potential for hospital economies of scale in quality for a common planned orthopaedic surgery. Results show that hospital volume does not have a causal effect on patient health outcomes once the endogeneity of volume linked to hospital demand being responsive to quality is accounted for. This finding has two main implications. First, it suggests that economies of scale in quality cannot be an argument to further concentrate planned hip replacement care in England. Increasing volume in low-volume hospitals with poorer quality would not result in health outcome improvements in these hospitals. Our results are derived from the English hospital market but ought to extend to countries with similar health systems. A possible limitation to the generalisability of our findings however is that the English hospital market is quite concentrated and hospitals may already be operating at the flat end of the volume-outcome curve, even though we do also observe small hospitals (i.e. from 20 annual hip replacement cases).

Second, more broadly, findings show that the evidence around economies of scale in quality should account more systematically for the endogeneity of hospital volume, to avoid misleading policy implications. While under economies of scale, concentrating hospital activity into larger healthcare providers may be desirable to achieve quality gains, if quality determines volume through e.g., hospital reputation, concentrating care would not improve health outcomes; instead, mechanisms to enhance patient choice of healthcare providers might be preferred to generate improvements in quality.

There may be other reasons to concentrate hospital care, including for instance efficiency savings in technical equipment, which are not investigated here. However, in terms of quality considerations, even if higher volume increases quality or does not affect quality, the optimal policy may not necessarily be to further concentrate care as this could raise issues of access. Policies that consolidate care may indeed risk shifting some of the NHS costs onto patients and carers, by increasing travel times or transportation costs, particularly affecting patients from disadvantaged socio-economic backgrounds (Ferguson et al., 1997; Watkinson et al., 2021). Further, the implementation of centralisation policies should be carefully monitored to mitigate the potential adverse effects of discontinued care in certain hospitals (Friebel et al., 2018).

Chapter 2 focuses on the role of surgeons' organisation of activity on quality of care. Results indicate that surgeons who treated a patient for a hip fracture surgery after a few days off perform better. Breaks of four to six days reduce 30-day mortality rates after hip fracture by six percent relative to surgeons who were in the operating room the day before. The positive effect of short breaks is driven by surgeons with relatively lower volume of hip fracture practice. Findings also report that short breaks affect the type of surgical treatment chosen, holding fixed patient characteristics and the type of fracture.

Chapter 2 highlights the role of non-financial considerations in surgeons' performance and decision-making by showing that short breaks in surgical activity can improve patient survival after emergency hip fracture. These findings stress the importance of team organisation and work schedules on the quality of care provided. Alongside longer-term policies to increase recruitment of healthcare staff, likely rendered more difficult by the Covid-19 crisis and post-Brexit restrictions on mobility (Propper et al., 2020), arrangements of work schedules to ensure regular breaks could improve outcomes, especially for more junior surgeons who may be more prone to fatigue effects. This could also help staff retention if it improves surgeons' quality of work life.

Chapter 3 reviews the effect of shortening inpatient stays on health outcomes for patients who present at the Emergency Department with low-severity chest pain symptoms. Patients who are not kept overnight have better health outcomes as measured by 28-day readmission rates. However, when inpatient length of stay is instrumented for, being discharged early rather than staying overnight does not lead to significant differences in readmission rates. In common with Chapter 1, accounting for endogeneity - in this context linked to potential unobserved patient severity - affects the results.

The results from Chapter 3 have policy implications. They indicate that discharging chest pain patients on the same day as admission does not lead to worse health outcomes. More broadly, the findings suggest that cost reductions in inpatient length of stay, i.e., amounting to more systematically discharging low-severity patients early, can be achieved without harming quality of care. Our results, though limited to chest pain, focus on the largest emergency condition amongst the list of medical conditions for which clinical guidelines recommend a higher rate of same-day discharge. The results may well extend to a larger set of clinical conditions. More generally, policy initiatives could aim at shortening medically unnecessary inpatient stays. In England, a national programme for reducing length of stay (RLoS) for long inpatient stays of 21 days or more, encourages faster patient discharge through the dissemination of key principles (e.g., patient and family involvement in the decision to discharge, multidisciplinary teams). Increasing the supply of long-term care could also reduce the risk of delayed patient discharge from the hospital (a phenomenon called bed blocking) for frail patients and reduce hospital costs (Forder, 2009; Gaughan et al., 2015; Moura, 2021).

However, the related literature on the effect of austerity measures on health care systems has identified adverse effects of budgetary cuts on health outcomes, primarily driven by reductions in staffing and hospital capacity (Borra et al., 2019; Arcà et al., 2020). Taken together, these sets of results suggest that policymakers should be careful about the source of cost reductions, but that reducing inpatient length of stay for certain conditions and up to some point, could potentially cut down costs with no adverse effect on quality. Further, freeing hospital beds by discharging patients when medically possible could increase capacity for other inpatient surgeries, thus potentially improving the technical efficiency of hospital resources. Alternatively, keeping fewer patients overnight could result in lower bed occupancy rates, which may have positive effects on quality.

Overall, recognizing the complex nature of hospitals comprising both medical and administrative staff (Harris, 1977), policymakers may consider initiatives that focus on improving the organisation of hospital activity and workforce with potential positive repercussions on quality (Ali et al., 2018; Lagarde et al., 2019).

The work presented in this thesis has several limitations. Despite the richness of the data used, Chapters 1-2 are ultimately limited by the difficulty of separating the effect of individual surgeons from the joint effect of other health care staff and medical teams. While the data attribute patients to a single responsible surgeon, patient care involves more health care personnel, such as anaesthetists, trainee doctors and nursing staff who are not identified in the hospital records used. Chapter 1 investigates whether hospitals benefit from economies of scale in quality. In various robustness checks, it allows for the role of other factors, such as individual surgeon volume and characteristics or hospital staff composition, to impact quality of care. However, the composition and joint experience of (unobserved) medical teams can also affect quality (Chan, 2021). Similarly, in Chapter 2, controlling for the effect of joint experience or familiarity with the rest of the team may refine the understanding of the effect of breaks on patient health outcomes.

Data permitting, future work could seek to establish the relative causal contribution of surgeons' individual learning-by-doing, within-team coordination and hospital-based economies of scale in quality improvements. With data on individual surgeons' activity, future studies could investigate individual learning-by-doing by using surgeons' exit from practice as an exogenous shock on remaining surgeons' volume as done in Ramanarayanan (2008). Previous work has relied on workers' mobility across firms in various settings to disentangle worker effects from firm effects, assuming exogenous mobility (Abowd et al., 1999; Card et al., 2013; Molitor, 2018). Future work may also seek to uncover the effect of team work by exploiting settings where surgeons started working at more than one hospital, for example following the opening of independent sector treatment centres in England. The role of team effect and hospital-wide scale economies may be disentangled with precise data on hospital staff turnover. Hospital scale economies should be little affected by staff turnover, whereas if the effect of volume on quality is mainly driven by team effects, hospitals with high and low staff turnover would have widely different results.

Another data limitation of Chapter 2 is that the reason for surgeons' break is not known, though robustness checks are implemented to rule out several confounding mechanisms. Understanding on-the-job determinants of surgical practice would have important implications for workforce policies in health care. Future studies could study how work schedules that allow for regular breaks impact healthcare staff retention. Shortages of healthcare staff are an increasing issue across OECD countries and are likely to be exacerbated in England by the Covid-19 crisis and Brexit (Lee et al., 2019; Propper et al., 2020). In relation to the organisation of work activity, the economics literature has been concerned with potential skill depreciation after long interruptions in practice (Hockenberry and Helmchen, 2014). Given more detailed data on surgeons' career path, future studies could use the plausibly exogenous long breaks in surgical practice induced by parental leave.

The interpretation of the findings from Chapter 3 is to some extent limited by the imprecision of the causal results. The design of the instrumental variable has the considerable advantage that it accounts for a large range of possible confounders by exploiting patient exposure to a reform over time in a panel data framework. While it finds a null causal effect of same-day discharge treatment on emergency readmissions, the IV coefficient is not precisely estimated; the 95% confidence interval comprises the OLS point estimate. The reduced-form results and additional robustness checks also indicate a null causal effect with a point estimate close to zero, which mitigates concerns of a type II error.

Future work could extend analyses to investigate whether the effect of cost reduction strategies on patient health outcomes depends on hospitals' organisational characteristics. For instance, specialist hospitals or teaching units may be more apt to limit the potential adverse effects of austerity measures or ensure patient safety, via possibly better awareness of clinical evidence, state-of-the-art practice, or ability to detect complications.

Finally, the thesis implicitly assumes that hospital quality is determined in isolation from other sectors of care, such as primary, social or long-term care. A more accurate measure of health outcomes would need to account for other sectors of care and patients' care pathways. Hospital quality is likely impacted by patient access to and quality of care provided by General Practitioners (GPs) and the availability of long-term care (Forder, 2009; Gaughan et al., 2015; Pinchbeck, 2019). Chapter 1 controls for accessibility to primary care in hospitals' catchment area (proxied by distance to the closest GP). All three chapters also include hospital fixed effects, which control for geographic differences in other aspects of care as long as they are time-invariant over the study period.

Provided availability of linked data across sectors of care, future research could control for the degree of integration of care in patient care pathway, using evidence-based good practices. An example is the communication of post-hospitalisation clinical information to the patients' General Practitioner for follow-up care. Taking into account the coordination of health care actors across sectors of care could help improve our understanding of the relative contribution of the determinants of quality of care.

Bibliography

Abaluck, J., Agha, L., Kabrhel, C., Raja, A., and Venkatesh, A. (2016). The determinants of productivity in medical testing: Intensity and allocation of care. *American Economic Review*, 106(12):3730–3764.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Acemoglu, D. and Angrist, J. (2000). How large are human-capital externalities? Evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, 15:9–59.

Ali, M., , Salehnejad, R., and Mansur, M. (2018). Hospital heterogeneity: what drives the quality of health care. *The European Journal of Health Economics*, 19(3):385–408.

Allan, J. L., Johnston, D. W., Powell, D. J. H., Farquharson, B., Jones, M. C., Leckie, G., and Johnston, M. (2019). Clinical decisions and time since rest break: An analysis of decision fatigue in nurses. *Health Psychology*, 38(4):318–324.

Allen, T., Fichera, E., and Sutton, M. (2016). Can payers use prices to improve quality? Evidence from English hospitals. *Health Economics*, 25(1):56–70.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *The Journal of human resources*, XL(4):791–821.

Américo, P. and Rocha, R. (2020). Subsidizing access to prescription drugs and health outcomes: The case of diabetes. *Journal of health economics*, 72:102347–102347.

Anandaciva, S. (2020). *Financial debts and loans in the NHS*. Report. The King's Fund.

Angrist, J. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, New Jersey.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442.

Aouad, M., Brown, T. T., and Whaley, C. M. (2019). Reference pricing: The case of screening colonoscopies. *Journal of Health Economics*, 65:246–259.

Arcà, E., Principe, F., and Doorslaer, E. (2020). Death by austerity? The impact of cost containment on avoidable mortality in Italy. *Health Economics*, 29(12):1500–1516.

Argote, L. and Epple, D. (1990). Learning curves in manufacturing. *Science*, 247(4945):920–924.

Augenblick, N. and Nicholson, S. (2016). Ballot position, choice fatigue, and voter behaviour. *The Review of economic studies*, 83(2 (295)):460–480.

Avdic, D., Lundborg, P., and Vikström, J. (2019). Estimating returns to hospital volume: evidence from advanced cancer surgery. *Journal of Health Economics*, 63:81–99.

Ayoubi, C., Pezzoni, M., and Visentin, F. (2017). At the origins of learning: Absorbing knowledge flows from within the team. *Journal of Economic Behavior & Organization*, 134:374–387.

Baicker, K. and Chandra, A. (2004). Medicare spending, the physician workforce, and beneficiaries' quality of care. *Health Affairs*, 23(Suppl1):W4–184–W4–197.

Barrenho, E., Miraldo, M., Propper, C., and Walsh, B. (2021). The importance of surgeons and their peers in adoption and diffusion of innovation: An observational study of laparoscopic colectomy adoption and diffusion in England. *Social Science & Medicine*, 272:113715–113715.

Bartel, A. P., Beaulieu, N. D., Phibbs, C. S., and Stone, P. W. (2014). Human capital and productivity in a team environment: Evidence from the healthcare sector. *American Economic Journal: Applied Economics*, 6(2):231–259.

Bauer, H. and Honselmann, K. (2017). Minimum volume standards in surgery - Are we there yet? *Visceral Medicine*, 33(2):106–116.

Benkard, C. L. (2000). Learning and forgetting: The dynamics of aircraft production. *American Economic Review*, 90(4):1034–1054.

Besanko, D., Doraszelski, U., Kryukov, Y., and Satterthwaite, M. (2010). Learning-by-doing, organizational forgetting, and industry dynamics. *Econometrica*, 78(2):453–508.

Birkmeyer, J. and Dimick, J. (2004). Potential benefits of the new leapfrog standards: effect of process and outcomes measures. *Surgery*, 135(6):569–575.

Blanco, B., Kothari, A., Blackwell, R., Brownlee, S., Yau, R., Attisha, J., Ezure, Y., Pappas, S., Kuo, P., and Abood, G. (2017). "Take the volume pledge" may result in disparity in access to care. *Surgery*, 161(3):837–845.

Bloom, N., Lemos, R., Sadun, R., and Van Reenen, J. (2020). Healthy business? Managerial education and management in health care. *The Review of Economics and Statistics*, 102(3):506–517.

Bloom, N., Propper, C., Seiler, S., and Van Reenen, J. (2015). The impact of competition on management quality: Evidence from public hospitals. *The Review of Economic Studies*, 82(2 (291)):457–489.

Bloom, N. and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4):1351–1408.

Bordignon, M., Coretti, S., Piacenza, M., and Turati, G. (2020). Hardening subnational budget constraints via administrative subordination: The Italian experience of recovery plans in regional health services. *Health Economics*, 29(11):1378–1399.

Borra, C., Pons-Pons, J., and Vilar-Rodríguez, M. (2019). Austerity, healthcare provision, and health outcomes in Spain. *The European Journal of Health Economics*, 21(3):409–423.

Brachet, T., David, G., and Drechsler, A. M. (2012). The effect of shift structure on performance. *American Economic Journal: Applied Economics*, 4(2):219–46.

Brealey, S., Andronis, L., Dale, V., Gibbon, A. J., Gilbert, F. J., Hendry, M., Hood, K., King, D., and Wilkinson, C. (2012). The effect of waiting times from general practitioner referral to MRI or orthopaedic consultation for the knee on patient-based outcomes. *The British Journal of Radiology*, 85(1019):e1134.

Brekke, K., Gravelle, H., Siciliani, L., and Straume, O. (2014). Patient choice, mobility and competition among health care providers. In Levaggi, R. and Montefiori, M., editors, *Health Care Provision and Patient Mobility: Health Integration in the European Union*, pages 1–26. Springer, Milan.

Brekke, K., Siciliani, L., and Straume, O. (2017). Hospital mergers with regulated prices. *The Scandinavian Journal of Economics*, 119(3):597–627.

British Association for Ambulatory Emergency Care (2014). *Directory of Ambulatory Emergency Care for Adults*. Report.

British Association of Day Surgery (2006). *BADS Directory of Procedures*. Report. 3rd Edition.

British Medical Association (2009). The consultant handbook. Online. `https://www.bma.org.uk/media/1368/bma-2009-consultants-contract-handbook-march-2020.pdf`. [Accessed 11/04/2021].

British Medical Association (2020). Medical training pathway. Online. `https://www.bma.org.uk/advice-and-support/studying-medicine/becoming-a-doctor/medical-training-pathway`. [Accessed 05/11/2020].

British Medical Association (2021). Doctors and the European working time directive. Online. `https://www.bma.org.uk/pay-and-contracts/working-hours/european-working-time-directive-ewtd/doctors-and-the-european-working-time-directive`. [Accessed 11/01/2021].

British Orthopaedic Association (2007). *The care of patients with fragility fracture*. Report.

Cameron, A. C. (2009). *Microeconometrics using Stata*. Stata Press, College Station, Tex.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly Journal of Economics*, 128(3):967–1015.

Caruso, C. C. (2014). Negative impacts of shiftwork and long work hours. *Rehabilitation Nursing : the Official Journal of the Association of Rehabilitation Nurses*, 39(1):16–25.

Caruso, C. C., Bushnell, T., Eggerth, D., Heitmann, A., Kojola, B., Newman, K., Rosa, R. R., Sauter, S. L., and Vila, B. (2006). Long working hours, safety, and health: Toward a national research agenda. *American Journal of Industrial Medicine*, 49(11):930–942.

Castelli, A., Street, A., Verzulli, R., and Ward, P. (2015). Examining variations in hospital productivity in the English NHS. *The European Journal of Health Economics*, 16(3):243–254.

Celhay, P. A., Gertler, P. J., Giovagnoli, P., and Vermeersch, C. (2019). Long-run effects of temporary incentives on medical care productivity. *American Economic Journal: Applied Economics*, 11(3):92–127.

Chalkley, M. and Malcomson, J. M. (1998a). Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics*, 17(1):1–19.

Chalkley, M. and Malcomson, J. M. (1998b). Contracting for health services with unmonitored quality. *The Economic Journal*, 108(449):1093–1110.

Chan, D. (2021). Influence and information in team decisions: Evidence from medical residency. *American Economic Journal: Economic Policy*, 13(1):106–137.

Chan, D. C. (2016). Teamwork and moral hazard: Evidence from the emergency department. *Journal of Political Economy*, 124(3):734–770.

Chandra, A., Cutler, D., and Song, Z. (2011). Who ordered that? The economics of treatment choices in medical care. In Pauly, M. V., McGuire, T. G., and Barros, P. P., editors, *Handbook of Health Economics*, volume 2, pages 397–432. Elsevier.

Chandra, A., Finkelstein, A., Sacarny, A., and Syverson, C. (2016a). Health care exceptionalism? Performance and allocation in the US health care sector. *American Economic Review*, 106(8):2110–2144.

Chandra, A., Finkelstein, A., Sacarny, A., and Syverson, C. (2016b). Productivity dispersion in medicine and manufacturing. *American Economic Review*, 106(5):99–103.

Chandra, A. and Staiger, D. (2007). Productivity spillovers in health care: Evidence from the treatment of heart attacks. *Journal of Political Economy*, 115(1):103–140.

Chen, L. M., Jha, A. K., Guterman, S., Ridgway, A. B., Orav, E. J., and Epstein, A. M. (2010). Hospital cost of care, quality of care, and readmission rates: Penny wise and pound foolish? *Archives of Internal Medicine (1960)*, 170(4):340–346.

Chowdhury, M., Dagash, H., and Pierro, A. (2007). A systematic review of the impact of volume of surgery and specialization on patient outcome. *British Journal of Surgery Society*, 94(2):145–161.

Chung, I. H. (2015). School choice, housing prices, and residential sorting: Empirical evidence from inter-and intra-district choice. *Regional Science and Urban Economics*, 52:39–49.

Collewet, M. and Sauermann, J. (2017). Working hours and productivity. *Labour Economics*, 47:96–106.

Com-Ruelle, L., Or, Z., and Renaud, T. (2008). The volume-outcome relationship in hospitals: lessons from the literature. Report 135, Institut de Recherche et Documentation en Economie de la Santé.

Cook, A., Gaynor, M., Stephens Jr, M., and Taylor, L. (2012). The effect of a hospital nurse staffing mandate on patient health outcomes: Evidence from California's minimum staffing regulation. *Journal of Health Economics*, 31(2):340–348.

Cooper, Z., Gibbons, S., Jones, S., and McGuire, A. (2011). Does hospital competition save lives? Evidence from the English NHS patient choice reforms. *The Economic Journal (London)*, 121(554):F228–F260.

Cooper, Z., Gibbons, S., and Skellern, M. (2018). Does competition from private surgical centres improve public hospitals' performance? Evidence from the English National Health Service. *Journal of Public Economics*, 166:63–80.

Correia, S. (2016). Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Working Paper.

Costa-Ramón, A., Kortelainen, M., Rodríguez-González, A., and Sääksvuori, L. (2020). The long-run effects of cesarean sections. *Journal of Human Resources*.

Costa-Ramón, A. M., Rodríguez-González, A., Serra-Burriel, M., and Campillo-Artero, C. (2018). It's about time: Cesarean sections and neonatal health. *Journal of Health Economics*, 59:46–59.

Currie, J. and MacLeod, W. B. (2017). Diagnosing expertise: Human capital, decision making, and performance among physicians. *Journal of Labor Economics*, 35(1):1–43.

Cutler, D., Skinner, J. S., Stern, A. D., and Wennberg, D. (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.

Dafny, L. S. (2005). How do hospitals respond to price changes? *American Economic Review*, 95(5):1525–1547.

Danziger, S., Levav, J., and Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17):6889–6892.

Darr, E. D., Argote, L., and Epple, D. (1995). The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management Science*, 41(11):1750–1762.

David, G. and Brachet, T. (2011). On the determinants of organizational forgetting. *American Economic Journal: Microeconomics*, 3(3):100–123.

Dawson, J., Fitzpatrick, R., Carr, A., and Murray, D. (1996). Questionnaire on the perceptions of patients about total hip replacement. *Journal of Bone and Joint Surgery*, 78(2):185.

De Luca, G., Lisi, D., Martorana, M., and Siciliani, L. (2021). Does higher institutional quality improve the appropriateness of healthcare provision? *Journal of Public Economics*, 194(104356).

De Vries, L., Sturkenboom, M., Verhaar, J., Kingma, J., and Stricker, B. (2011). Complications after hip arthroplasty and the association with hospital procedure volume: A nationwide retrospective cohort study on 50,080 total hip replacements with a follow-up of 3 months after surgery. *Acta Orthopaedica*, 82(5):545–552.

Department of Health (2012). *A methodology for applying casemix adjustment - Annex C: coefficients for hip replacement models. Patient Reported Outcome Measures (PROMs) in England*. Report.

Department of Health (2013). *Managing NHS hospital consultants. Report by the Comptroller and Auditor General*. Report.

Dimick, J. B., Nicholas, L. H., Ryan, A. M., Thumma, J. R., and Birkmeyer, J. D. (2013). Bariatric surgery complications before vs after implementation of a national policy restricting coverage to centers of excellence. *Journal of the American Medical Association*, 309(8):792–799.

Doyle, J. J., Graves, J. A., and Gruber, J. (2017). Uncovering waste in US healthcare: Evidence from ambulance referral patterns. *Journal of Health Economics*, 54:25–39.

Doyle, J. J., Graves, J. A., Gruber, J., and Kleiner, S. A. (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy*, 123(1):170–214.

Dranove, D., Kessler, D., McClellan, M., and Satterthwaite, M. (2003). Is more information better? The effects of report cards on health care providers. *Journal of Political Economy*, 111(3):555–588.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4):795–813.

Dunn, P., McKenna, H., and Murray, R. (2016). *Deficits in the NHS 2016*. Report. The King's Fund.

Elixhauser, R., Steiner, M., Harris, M., and Coffey, M. (1998). Comorbidity measures for use with administrative data. *Medical Care*, 36(1):8–27.

Elwell-Sutton, T., Tinson, A., Greszczuk, C., Finch, D., Holt-White, E., Everest, G., Mihaylova, N., Wood, S., and Bibby, J. (2019). Creating healthy lives. A whole-government approach to long-term investment in the nation's health. Report. The Health Foundation.

Epstein, A. J. and Nicholson, S. (2009). The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics*, 28(6):1126–1140.

Eurostat (2020). Dataset - Health care expenditure by provider. Online. `https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlth_sha11_hp&lang=en`. [Accessed 22/01/2021].

Farrar, S., Yi, D., Sutton, M., Chalkley, M., Sussex, J., and Scott, A. (2009). Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis. *BMJ: British Medical Journal*, 339(7720):554–556.

Feng, Y., Pistollato, M., Charlesworth, A., Devlin, N., Propper, C., and Sussex, J. (2015). Association between market concentration of hospitals and patient health gain following hip replacement surgery. *Journal of Health Services Research Policy*, 20(1):11–17.

Ferguson, B., Posnett, J., and Sheldon, T. (1997). *Concentration and choice in the provision of hospital services. Summary Report.* Report. York, NHS Centre for Reviews and Dissemination.

Finkelstein, A., Gentzkow, M., and Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *The Quarterly Journal of Economics*, 131(4):1681–1726.

Flinn, F. and Armstrong, C. (2011). Junior doctors' extended work hours and the effects on their performance: the irish case. *International journal for quality in health care*, 23(2):210–217.

Forberg, J. L., Henriksen, L. S., Edenbrandt, L., and Ekelund, U. (2006). Direct hospital costs of chest pain patients attending the emergency department: a retrospective study. *BMC Emergency Medicine*, 6(1):6–6.

Forder, J. (2009). Long-term care and hospital utilisation by older people: an analysis of substitution rates. *Health Economics*, 18(11):1322–1338.

Francese, M., Piacenza, M., Romanelli, M., and Turati, G. (2014). Understanding inappropriateness in health spending: The role of regional policies and institutions in caesarean deliveries. *Regional Science and Urban Economics*, 49:262–277.

Friebel, R., Hauck, K., Aylin, P., and Steventon, A. (2018). National trends in emergency readmission rates: a longitudinal analysis of administrative data for England between 2006 and 2016. *BMJ Open*, 8(3):e020325.

Friedrich, B. and Hackmann, M. B. (2017). The returns to nursing: Evidence from a parental leave program. *NBER Working Paper Series*, (23174).

Frimmel, W. and Pruckner, G. J. (2020). Retirement and healthcare utilization. *Journal of Public Economics*, 184:104146.

Gaughan, J., Gravelle, H., Santos, R., and Siciliani, L. (2017). Long-term care provision, hospital bed blocking, and discharge destination for hip fracture and stroke patients. *International Journal of Health Economics and Management*, 17(3):311–331.

Gaughan, J., Gravelle, H., and Siciliani, L. (2015). Testing the bed-blocking hypothesis: Does nursing and care home supply reduce delayed hospital discharges? *Health Economics*, 24(S1):32–44.

Gaughan, J., Gutacker, N., Grašič, K., Kreif, N., Siciliani, L., and Street, A. (2019). Paying for efficiency: Incentivising same-day discharges in the English NHS. *Journal of Health Economics*, 68:102226.

Gaynor, M., Ho, K., and Town, R. J. (2015). The industrial organization of health-care markets. *Journal of Economic Literature*, 53(2):235–284.

Gaynor, M., Moreno Serra, R., and Propper, C. (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., Seider, H., and Vogt, W. (2005). The volume-outcome effect, scale economies, and learning-by-doing. *American Economic Review*, 95(2):243–247.

Gaynor, M. and Town, R. (2011). *Competition in Health Care Markets*, volume 2, pages 499–637. Elsevier.

Geweke, J., Gowrisankaran, G., and Town, R. (2003). Bayesian inference for hospital quality in a selection model. *Econometrica*, 71(4):1215–1238.

Gobillon, L. and Milcent, C. (2016). Evaluating the effect of ownership status on hospital quality: The key role of innovative procedures. *Annals of Economics and Statistics*, (121/122):161–186.

Gowrisankaran, G. and Town, R. (2003). Competition, payers, and hospital quality. *Health Services Research*, 38(6p1):1403–1422.

Gravelle, H., Santos, R., and Siciliani, L. (2014). Does a hospital's quality depend on the quality of other hospitals? A spatial econometrics approach. *Regional Science and Urban Economics*, 49:203–216.

Grašič, K., Mason, A. R., and Street, A. (2015). Paying for the quantity and quality of hospital care: The foundations and evolution of payment policy in England. *Health Economics Review*, 5(1):50–50.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, 7(1):98–119.

Gutacker, N., Bloor, K., Cookson, R., Gale, C. P., Maynard, A., Pagano, D., Pomar, J., and Bernal-Delgado, E. (2017). Hospital surgical volumes and mortality after coronary artery bypass grafting: using international comparisons to determine a safe threshold. *Health Services Research*, 52(2):863–878.

Gutacker, N., Bojke, C., Daidone, S., Devlin, N. J., Parkin, D., and Street, A. (2013). Truly inefficient or providing better quality of care? Analysing the relationship between risk-adjusted hospital costs and patients' health outcomes. *Health economics*, 22(8):931–947.

Gutacker, N., Siciliani, L., Moscelli, G., and Gravelle, H. (2016). Choice of hospital: which type of quality matters? *Journal of Health Economics*, 50:230–246.

Hafsteinsdottir, E. J. G. and Siciliani, L. (2010). DRG prospective payment systems: refine or not refine? *Health Economics*, 19(10):1226–1239.

Halla, M., Mayr, H., Pruckner, G. J., and Garcia-Gomez, P. (2016). Cutting fertility? The effect of cesarean deliveries on subsequent fertility and maternal labor supply.

Halm, E., Lee, C., and Chassin, M. (2002). Is volume related to outcome in health care? a systematic review and methodologic critique of the literature. *Annals of Internal Medicine*, 137(6):511–520.

Hamilton, B. and Hamilton, V. (1997). Estimating surgical volume—outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Health Economics*, 6(4):383–395.

Harris, J. E. (1977). The internal organization of hospitals: Some economic implications. *The Bell Journal of Economics*, 8(2):467–482.

Harrison, G., Zeitz, K., Adams, R., and Mackay, M. (2013). Does hospital occupancy impact discharge rates? *Australian Health Review*, 37(4):458–466.

Harrison, M., Graff, L., Roos, N., and Brownell, M. (1995). Discharging patients earlier from Winnipeg hospitals: Does it adversely affect quality of care? *Canadian Medical Association journal (CMAJ)*, 153(6):745–751.

Hauck, K. and Zhao, X. (2011). How dangerous is a day in hospital? A model of adverse events and length of stay for medical inpatients. *Medical Care*, 49(12):1068–1075.

Healthcare Quality Improvement Partnership (2018). National joint registry - 15th annual report. Report.

Hentschker, C. and Mennicken, R. (2018). The volume–outcome relationship revisited: practice indeed makes perfect. *Health Services Research*, 53(1):15–34.

Herbert, A., Wijlaars, L., Zylbersztejn, A., Cromwell, D., and Hardelid, P. (2017). Data resource profile: Hospital episode statistics admitted patient care (HES APC). *International Journal of EAidemiology*, 46(4):1093–1093i.

Ho, V. (2002). Learning and the evolution of medical technologies: the diffusion of coronary angioplasty. *Journal of Health Economics*, 21(5):873–885.

Ho, V. (2014). Learning by doing. In Culyer, A. J., editor, *Encyclopedia of Health Economics*, pages 141–145. Elsevier, San Diego.

Hockenberry, J. and Helmchen, L. (2014). The nature of surgeon human capital depreciation. *Journal of Health Economics*, 37:70–80.

Hockenberry, J., Lien, H., and Chou, S. (2008). The impacts of task repetition and temporal breaks in production on human capital and productivity. *Journal of Human Capital*, 2(3):303–335.

Hodgkin, D. and McGuire, T. G. (1994). Payment levels and hospital response to prospective payment. *Journal of Health Economics*, 13(1):1–29.

Horowitz, J., Keil, S., and Spector, L. (2009). Do charter schools affect property values? *The Review of Regional Studies*, 39(3):297–316.

Hudson, S., Hull, P., and Liebersohn, J. (2017). Interpreting instrumented difference-in-differences. Working paper. Available: `http://www.mit.edu/~liebers/DDIV.pdf` [Accessed: 05/04/2019.

Huesch, M. (2014). The impact of short breaks from cardiac surgery on mortality and stay length in California. *Journal for Healthcare Quality*, 36(5):42–49.

Hunt, T. C., Ambrose, J. P., Haaland, B., Kawamoto, K., Dechet, C. B., Lowrance, W. T., Hanson, H. A., and O'Neil, B. B. (2021). Decision fatigue in low-value prostate cancer screening. *Cancer*, 127(18):3343–3353.

Hussey, P. S., Wertheimer, S., and Mehrotra, A. (2013). The association between health care quality and cost a systematic review. *Annals of Internal Medicine*, 158(1):27–33.

Häkkinen, U., Rosenqvist, G., Peltola, M., Kapiainen, S., Rättö, H., Cots, F., Geissler, A., Or, Z., Serdén, L., and Sund, R. (2014). Quality, cost, and their trade-off in treating AMI and stroke patients in European hospitals. *Health Policy (Amsterdam)*, 117(1):15–27.

Ikenwilo, D. and Scott, A. (2007). The effects of pay and job satisfaction on the labour supply of hospital consultants. *Health Economics*, 16(12):1303–1318.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Jamalabadi, S., Winter, V., and Schreyögg, J. (2020). A systematic review of the association between hospital cost/price and the quality of care. *Applied Health Economics and Health Policy*, 18(5):625–639.

Januleviciute, J., Askildsen, J. E., Kaarboe, O., Siciliani, L., and Sutton, M. (2016). How do hospitals respond to price changes? Evidence from Norway. *Health Economics*, 25(5):620–636.

Judge, A., Chard, J., Learmonth, I., and Dieppe, P. (2006). The effects of surgical volumes and training centre status on outcomes following total joint replacement: Analysis of the hospital episode statistics for England. *Journal of Public Health*, 28(2):116–124.

Kahn, J. M., Ten Have, T. R., and Iwashyna, T. J. (2009). The relationship between hospital volume and mortality in mechanical ventilation: An instrumental variable analysis. *Health Services Research*, 44(3):862–879.

Kelly, E. and Stoye, G. (2016). New joints: private providers and rising demand in the English National Health Service. Report. Institute for Fiscal Studies.

Kessler, D. and McClellan, M. (2000). Is hospital competition socially wasteful? *Quarterly Journal of Economics*, 115(2):577–615.

Kizer, K. (2003). The volume–outcome conundrum. *The New England Journal of Medicine*, 349(22):2159–2161.

Kolstad, J. T. (2013). Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, 103(7):2875–2910.

Kruse, F. M., Stadhouders, N. W., Adang, E. M. M., Groenewoud, S., and Jeurissen, P. P. T. (2018). Do private hospitals outperform public hospitals regarding efficiency, accessibility, and quality of care in the European Union? A literature review. *The International Journal of Health Planning and management*, 33(2):e434–e453.

Lafond, S., Arora, S., Charlesworth, A., and McKeon, A. (2014). Into the red? The state of the NHS' finances. Report. The Nuffield Trust.

Lagarde, M., Huicho, L., and Papanicolas, I. (2019). Motivating provision of high quality care: it is not all about the money. *BMJ*, 366:l5210.

Lazear, E. P. and Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *The Journal of Economic Perspectives*, 21(4):91–114.

Lee, T., Propper, C., and Stoye, G. (2019). Medical labour supply and the production of healthcare. *Fiscal Studies*, 40(4):621–661.

Lindenauer, P. K., Remus, D., Roman, S., Rothberg, M. B., Benjamin, E. M., Ma, A., and Bratzler, D. W. (2007). Public reporting and pay for performance in hospital quality improvement. *The New England Journal of Medicine*, 356(5):486–496.

Linder, J. A., Doctor, J. N., Friedberg, M. W., Reyes Nieva, H., Birks, C., Meeker, D., and Fox, C. R. (2014). Time of day and the decision to prescribe antibiotics. *JAMA internal medicine*, 174(12):2029–2031.

Liu, Y., Wang, Z., and Xiao, W. (2017). Risk factors for mortality in elderly patients with hip fractures: a meta-analysis of 18 studies. *Aging clinical and experimental research*, 30(4):323–330.

Livingston, E. H. (2009). Bariatric surgery outcomes at designated centers of excellence vs nondesignated programs. *Archives of Surgery*, 144(4):319–325.

Longo, F., Siciliani, L., Moscelli, G., and Gravelle, H. (2019). Does hospital competition improve efficiency? The effect of the patient choice reform in England. *Health Economics*, 28(5):618–640.

Losina, E., Barrett, J., Mahomed, N. N., Baron, J. A., and Katz, J. N. (2004). Early failures of total hip replacement: effect of surgeon volume. *Arthritis & Rheumatism*, 50(4):1338–1343.

Luft, H., Hunt, S., and Maerki, S. (1987). The volume-outcome relationship: Practice-makes-perfect or selective-referral patterns? *Health Services Research*, 22(2):157.

Lundborg, P. and Majlesi, K. (2018). Intergenerational transmission of human capital: Is it a one-way street? *Journal of Health Economics*, 57:206–220.

Ma, C. A. (1994). Health care payment systems: Cost and quality incentives. *Journal of Economics & Management Strategy*, 3(1):93–112.

Ma, M. (2019). Does children's education matter for parents' health and cognition? Evidence from China. *Journal of Health Economics*, 66:222–240.

Malcomson, James, M. (2005). Supplier discretion over provision: Theory and an application to medical care. *The Rand Journal of Economics*, 36(2):412–432.

Martinez-Nadal, G., Miro, O., Matas, A., Cepas, P., Aldea, A., Izquierdo, M., Coll-Vinent, B., Garcia, A., Carbo, M., Manuel, O., Aguilo, S., Esteban, E., and Lopez-Barbeito, B. (2021). An analysis based on sex gender in the chest pain unit of an emergency department during the last 12 years. *European Heart Journal. Acute Cardiovascular Care*, 10(Supplement$_1$).

McConnell, K. J., Lindrooth, R. C., Wholey, D. R., Maddox, T. M., and Bloom, N. (2013). Management practices and the quality of care in cardiac units. *JAMA Internal Medicine*, 173(8):684–692.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Economics*, volume 4, pages 105–142. Academic Press, New York.

McKay, N. L. and Deily, M. E. (2008). Cost inefficiency and hospital health outcomes. *Health Economics*, 17(7):833–848.

Meacock, R., Anselmi, L., Kristensen, S. R., Doran, T., and Sutton, M. (2019). Do variations in hospital admission rates bias comparisons of standardized hospital mortality rates? A population-based cohort study. *Social Science & Medicine (1982)*, 235:112409–112409.

Meacock, R., Kristensen, S., and Sutton, M. (2014). Paying for improvements in quality: Recent experience in the NHS in England. *Nordic Journal of Health Economics*, 2(1).

Mesman, R., Faber, M. J., Berden, B. J. J. M., and Westert, G. P. (2017). Evaluation of minimum volume standards for surgery in the Netherlands (2003–2017): A successful policy? *Health Policy*, 121(12):1263–1273.

Miani, C., Ball, S., Pitchforth, E., Exley, J., King, S., Roland, M., Fuld, J., and Nolte, E. (2014). Organisational interventions to reduce length of stay in hospital: A rapid evidence assessment. *Health Services and Delivery Research*, 2(52):1–178.

Molitor, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy*, 10(1):326–356.

Monitor and NHS England (2014). *2014/15 National Tariff Payment System*. Report.

Montiel-Olea, J. L. and Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, 31(3):358–369.

Moscelli, G., Gravelle, H., Siciliani, L., and Gutacker, N. (2018a). The effect of hospital ownership on quality of care: Evidence from England. *Journal of Economic Behavior & Organization*, 153:322–344.

Moscelli, G., Gravelle, H., Siciliani, L., and Santos, R. (2018b). Heterogeneous effects of patient choice and hospital competition on mortality. *Social Science & Medicine*, 216:50–58.

Moura, A. (2021). Do subsidized nursing homes and home care teams reduce hospital bed-blocking? Evidence from Portugal. Working paper.

Mukoyama, T. (2006). Rosenberg's "learning by using" and technology diffusion. *Journal of Economic Behavior & Organization*, 61(1):123–144.

National Clinical Guidelines Centre (2011). *The management of hip fracture in adults*. Report. National Clinical Guidelines Centre.

Nembhard, D. A. (2000). The effects of task complexity and experience on learning and forgetting: A field study. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 42(2):272–286.

Nguyen, N. T., Hohmann, S., Slone, J., Varela, E., Smith, B. R., and Hoyt, D. (2010). Improved bariatric surgery outcomes for medicare beneficiaries after implementation of the medicare national coverage determination. *Archives of Surgery (Chicago, Ill. : 1960)*, 145(1):72–78.

NHS (2009). *NHS' chief executive's annual report 2008/09*. Report.

NHS (2020). Hip replacement. Online. `https://www.nhs.uk/conditions/hip-replacement/`. [Accessed 21/04/2020].

NHS Commissioning Board (2012). *Everyone counts: planning for patients 2013-14*. Report.

NHS Digital (2016). Estates returns information collection (ERIC - England. Online. https://digital.nhs.uk/data-and-information/publications/statistical/estates-returns-information-collection. [Accessed 14/10/2019].

NHS Digital (2017). *A guide to PROMs methodology*. Report.

O'Dowd, A. (2018). Emergency admissions rise 24% in past decade, costing almost £14bn a year. *BMJ (Clinical research ed.)*, 360:k1022.

OECD (2015). *Fiscal sustainability of health systems: Bridging health and finance perspectives*. OECD Publishing, Paris.

OECD (2017a). *Caring for quality in health - Lessons learnt from 15 reviews of health care quality*. Report. OECD Publishing.

OECD (2017b). *Tackling wasteful spending on health*. Report. OECD Publishing.

OECD (2019). *Health at a Glance 2019: OECD Indicators*. Report. OECD Publishing.

Office for National Statistics (2017). UK Health Accounts. Online. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2017. [Accessed 22/12/2020].

Ostendorf, M., van Stel, H. F., Buskens, E., Schrijvers, A. J. P., Marting, L. N., Verbout, A. J., and Dhert, W. J. A. (2004). Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. *The Journal of Bone and Joint Surgery. British volume*, 86(6):801–808.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of business & economic statistics*, 37(2):187–204.

Paterson, J. M., Williams, J. I., Kreder, H. J., Mahomed, N. N., Gunraj, N., Wang, X., and Laupacis, A. (2010). Provider volumes and early outcomes of primary total joint replacement in Ontario. *Canadian Journal of Surgery*, 53(3):175–183.

Pearce, S. M., Pariser, J. J., Patel, S. G., Anderson, B. B., Eggener, S. E., and Zagaja, G. P. (2015). The impact of days off between cases on perioperative outcomes for robotic-assisted laparoscopic prostatectomy. *World Journal of Urology*, 34(2):269–274.

Pencavel, J. (2015). The productivity of working hours. *The Economic Journal*, 125(589):2052–2076.

Persson, E., Barrafrem, K., Meunier, A., and Tinghög, G. (2019). The effect of decision fatigue on surgeons' clinical decision making. *Health Economics*, 28(10):1194–1203.

Phelps, C. E. (2000). Information diffusion and best practice adoption. In Culyer, A. J. and Newhouse, J. P., editors, *Handbook of Health Economics*, volume 1, book section 5, pages 223–264. Elsevier.

Picone, G. A., Sloan, F. A., Chou, S.-Y., and Taylor, D. H. (2003). Does higher hospital cost imply higher quality of care? *The Review of Economics and Statistics*, 85(1):51–62.

Pinchbeck, E. W. (2019). Convenient primary care and emergency hospital utilisation. *Journal of Health Economics*, 68:102242.

Prina, L. D., Decker, W. W., Weaver, A. L., High, W. A., Smars, P. A., Locke, G. R., and Reeder, G. S. (2004). Outcome of patients with a final diagnosis of chest pain of undetermined origin admitted under the suspicion of acute coronary syndrome: A report from the Rochester epidemiology project. *Annals of Emergency Medicine*, 43(1):59–67.

Propper, C., Stoye, G., and Zaranko, B. (2020). The wider impacts of the coronavirus pandemic on the NHS*. *Fiscal Studies*, 41(2):345–356.

Rachet-Jacquet, L., Gutacker, N., and Siciliani, L. (2021). Scale economies in the health sector: The effect of hospital volume on health gains from hip replacement surgery. *Journal of economic behavior & organization*, 190:704–729.

Ramanarayanan, S. (2008). Does practice make perfect: An empirical analysis of learning-by-doing in cardiac surgery. Working paper.

Ravi, B., Jenkinson, R., Austin, P., Croxford, R., Wasserstein, D., Escott, B., Paterson, J., Kreder, H., and Hawker, G. (2014). Relation between surgeon volume and risk of complications after total hip arhroplasty: Propensity score matched cohort study. *British Medical Journal*, 348(7961):14.

Reagans, R., Argote, L., and Brooks, D. (2005). Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science*, 51(6):869–881.

Schilling, M. A., Vidal, P., Ployhart, R. E., and Marangoni, A. (2003). Learning by doing something else: Variation, relatedness, and the learning curve. *Management Science*, 49(1):39–56.

Schreyögg, J. and Stargardt, T. (2010). The trade-off between costs and outcomes: The case of acute myocardial infarction. *Health Services Research*, 45(6p1):1585–1601.

Sfekas, A. (2009). Learning, forgetting, and hospital quality: an empirical analysis of cardiac procedures in Maryland and Arizona. *Health Economics*, 18(6):697–711.

Shafer, S. M., Nembhard, D. A., and Uzumeri, M. V. (2001). The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science*, 47(12):1639–1653.

Sheldon, T. A. (2004). The volume-quality relationship: Insufficient evidence for use as a quality indicator. *Quality and Safety in Health Care*, 13(5):325.

Shojania, K. and Forster, A. (2008). Hospital mortality: when failure is not a good measure of success. *Canadian Medical Association Journal*, 179(2):153–157.

Siciliani, L. (2006). Selection of treatment under prospective payment systems in the hospital sector. *Journal of Health Economics*, 25(3):479–499.

Singh, J., Kwoh, C., Boudreau, R., Lee, G., and Ibrahim, S. (2011). Hospital volume and surgical outcomes after elective hip/knee arthroplasty: a risk-adjusted analysis of a large regional database. *Arthritis & Rheumatism*, 63(8):2531–2539.

Skellern, M. (2017). The hospital as a multi-product firm: The effect of hospital competition on value-added indicators of clinical quality. *CEP Discussion Papers*.

Skinner, J. (2011). Causes and consequences of regional variations in health care. In Pauly, M. V., McGuire, T. G., and Barros, P. P., editors, *Handbook of Health Economics*, volume 2, pages 45–93. Elsevier.

Skinner, J. and Staiger, D. (2015). Technology diffusion and productivity growth in health care. *The Review of Economics and Statistics*, 97(5):951–964.

Spurgeon, A. (2003). *Working time: Its impact on safety and health*. International Labour Office [Imprint].

Stargardt, T., Schreyögg, J., and Kondofersky, I. (2014). Measuring the relationship between costs and outcomes: The example of acute myocardial infarction in German hospitals. *Health Economics*, 23(6):653–669.

Steven, T. F. (1991). The relationship between quality and cost: Pure and simple? *Inquiry (Chicago)*, 28(1):29–38.

Street, A., Gutacker, N., Bojke, C., Devlin, N., and Daidone, S. (2014). Variations in outcome and costs among nhs providers for common surgical procedures: Econometric analyses of routinely collected data. *Health Services and Delivery Research*, 2(1):1–90.

The King's Fund (2018). *The healthcare workforce in England: Make or Break?* Report.

The King's Fund (2020). *The NHS budget and how it has changed*. Report.

Thompson, P. (2007). How much did the liberty shipbuilders forget? *Management Science*, 53(6):908–918.

Thompson, P. (2010). Chapter 10 - Learning by Doing. In Hall, B. H. and Rosenberg, N., editors, *Handbook of the Economics of Innovation*, volume 1, pages 429–476. North-Holland.

Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, 2nd edition.

Tsai, A. C., Votruba, M., Bridges, J. F. P., and Cebul, R. D. (2006). Overcoming bias in estimating the volume-outcome relationship. *Health Services Research*, 41(1):252–264.

Tuominen, U., Sintonen, H., Hirvonen, J., Seitsalo, S., Paavolainen, P., Lehto, M., Hietaniemi, K., and Blom, M. (2010). Is longer waiting time for total knee replacement associated with health outcomes and medication costs? Randomized clinical trial. *Value in Health*, 13(8):998–1004.

Vallejo-Torres, L., García-Lorenzo, B., and Serrano-Aguilar, P. (2018). Estimating a cost-effectiveness threshold for the Spanish NHS. *Health Economics*, 27(4):746–761.

Van Gestel, R., Müller, T., and Bosmans, J. (2017). Does my high blood pressure improve your survival? Overall and subgroup learning curves in health. *Health Economics*, 26(9):1094–1109.

Varagunam, M., Hutchings, A., and Black, N. (2015a). Do patient-reported outcomes offer a more sensitive method for comparing the outcomes of consultants than mortality? A multilevel analysis of routine data. *BMJ Quality & Safety*, 24(3):195.

Varagunam, M., Hutchings, A., and Black, N. (2015b). Relationship between patient-reported outcomes of elective surgery and hospital and consultant volume. *Medical Care*, 53(4):310–316.

Verzulli, R., Fiorentini, G., Lippi Bruni, M., and Ugolini, C. (2017). Price changes in regulated healthcare markets: Do public hospitals respond and how? *Health Economics*, 26(11):1429–1446.

Watkinson, R. E., Sutton, M., and Turner, A. J. (2021). Ethnic inequalities in health-related quality of life among older adults in England: Secondary analysis of a national cross-sectional survey. *The Lancet Public Health*.

West, M. (2020). *What does the 2019 NHS Staff Survey truly tell us about how staff needs are being met?* Report. The King's Fund.

Wilkinson, E. (2015). UK NHS staff: stressed, exhausted, burnt out. *The Lancet (British edition)*, 385(9971):841–842.

Williams, S. and Buchan, J. (2006). *Assessing the new NHS consultant contract. A something for something deal?* Report. The King's Fund.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55(1):235–269.

Yoon, T. (2019). Quality information disclosure and patient reallocation in the healthcare industry: Evidence from cardiac surgery report cards. *Marketing Science*, 9(3):459–665.