

Essays in the Economics of Healthcare

Elena Ashtari Tafti

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Economics
University College London

November 13, 2023

Declaration

I, Elena Ashtari Tafti confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.”

Signature

Date

Abstract

This dissertation examines three phenomena, that are prevalent in healthcare, and their consequences for patient health outcomes. The focus of **Chapter 2** is the adoption and diffusion of robots in England for the surgical treatment of prostate cancer patients in the National Health Service (NHS). Exploiting quasi-random variation in the geographic allocation of robots, Chapter 2 shows that robots shorten patients' length of stay and decrease the incidence of adverse events from surgery, but their effects are heterogeneous and significantly depend on surgeons' skills. High-skilled surgeons benefit the least from using the technology, while lower-skilled surgeons appear to gain the most from it. **Chapter 3** studies the impact of hospital mergers on the quality of clinical care. Using the universe of hospital medical records in England, it examines all public hospital mergers after the introduction of hospital choice in 2006. There were 159 hospital sites involved in mergers over our sample period, comprising 13 transactions. Using an event study framework, this Chapter finds that mergers have immediate and persistent negative impacts on clinical quality. **Chapter 4** uses a unique source of information, Real-Time Location System (RTLS) Data, to study the effect of contact time on patient health outcomes. RTLS allows to perfectly observe the amount of time nurses spend with patients in the hospital. This Chapter exploits the granularity of this data to estimate a causal impact of contact time on patient outcomes by decomposing contact time into an endogenous and plausibly exogenous component and shows that direct contact between nurses and patients significantly reduces in-hospital mortality and accidents.

Impact Statement

The healthcare sector is of great significance across multiple dimensions within society. It plays a pivotal role in tackling global challenges like demographic shifts towards aging populations and addressing disparities in quality of life. By serving as the bedrock for developing healthier and more resilient communities, it emerges as a fundamental building block for fostering a prosperous and sustainable future.

By exploring the transformative potential of new technologies, the findings of **Chapter 2** hold promise for addressing long-standing inequalities and improving patient outcomes. This Chapter reveals that robots can significantly reduce hospital stays and decrease adverse events from surgery. However, it also highlights the role of surgical expertise in realizing these benefits. The advent of robots in healthcare, as illuminated by this research, represents a pivotal step towards enhancing patient experiences and safety, but also an important tool to reduce disparities in the quality of care.

Chapter 3 focuses on the intricate dynamics of hospital mergers. It unveils the complex relationship between hospital consolidation and patient outcomes. Recent trends in hospital consolidation have been notable within the healthcare industry, with a growing number of mergers and acquisitions taking place among healthcare providers. These trends have sparked considerable discussion and concern regarding their effects on patient care and outcomes. The results presented in this Chapter provide crucial insights into the ramifications of these mergers, shedding light on their immediate and lasting effects on clinical quality. In this sense, they provide essential evidence for policymakers, healthcare providers, and patient advocates to

closely monitor these trends and work collaboratively to balance the potential benefits of consolidation and the protection of patient access, affordability, and quality of care.

Chapter 4 examines the fundamental role of direct nurse-patient interactions within hospital settings. Employing innovative data, it uncovers the substantial impact of contact time on in-hospital mortality and accident rates. These findings underscore the invaluable contribution of personal care in the recovery process. Understanding the impact of nurse availability on patient care is vital for policymakers as they make decisions about healthcare staffing standards, funding allocation, and workforce planning. It informs decisions related to nurse-to-patient ratios, nurse training, and strategies for retaining nursing talent. This Chapter provides substantial evidence of the detrimental effect that nurse shortages may have on patients.

Collectively, this dissertation illuminates critical facets of the healthcare landscape, offering tangible insights and evidence-based recommendations. Its research not only enriches the academic discourse but also has the potential to inform policy decisions and practical strategies aimed at enhancing patient health outcomes in healthcare systems worldwide.

Acknowledgements

My time at UCL and the IFS is unique in my heart. I had the most fulfilling personal and intellectually challenging experience, and I would not change a single thing. Thanks to my exceptional supervisors, Marcos Vera Hernandez and Aureo de Paula, I genuinely felt heard and supported. I also want to share a special appreciation to Thomas P. Hoe, an incredible mentor, role model, and friend.

I am truly appreciative of my twin brother Oliver, whose dedication went above and beyond. He spent countless hours discussing, sharing insightful perspectives, and motivating me to strive for more. He is an exceptional human and economist.

I want to express gratitude to my younger brother Antonio for always believing in me, and giving me strength on my weakest days. From the beginning, he has been a cherished gift in my life.

Thanks to my wonderful friends Mikkel, Alex, Hugo, Tom, Ludovica, Alexia, Giulia, Marta, Sharu and Rocco. I am truly blessed to have so many exceptional individuals in my life.

I want to express my heartfelt appreciation to my family, including my grandma, aunts, uncles, and all my cousins, who have consistently celebrated my accomplishments with unwavering enthusiasm. Nonno, I miss you. Mom, Dad, I love you dearly. Marzia, Adil, Benedetta, Elio; I would not be here without you. Kita, thank you for being my editor always and forever.

Lastly, I want to thank my partner and best friend Lodovico. His attentive listening, constant support, and boundless love have been truly remarkable. Thank you for forcing me to go on holiday.

Go Gunners!

Contents

1	Introduction	18
2	Technology, Skills, and Performance: The Case of Robots in Surgery	21
2.1	Introduction	21
2.2	Robotic surgery for prostate cancer	26
2.3	Data and Institutional Context	29
2.4	Measuring Skills	34
2.5	Econometric model	39
2.6	Exogenous variation in treatment probability	47
2.7	Results	57
2.8	Conclusive remarks	70
3	Killer Deals? The Impact of Hospital Mergers on Clinical Quality	72
3.1	Introduction	72
3.2	Data	76
3.3	Empirical Methodology	80
3.4	Results	86
3.5	Extension	90
3.6	Conclusions	94
4	Does Contact Time Matter For Patient Outcomes?	98
4.1	Introduction	98
4.2	Data	102

Contents 13

4.3	Empirical Strategy	108
4.4	Results	111
4.5	Conclusions	112
5	General Conclusions	114
	Appendices	115
	Bibliography	128

List of Figures

2.1	Picture of Da Vinci surgical system	27
2.2	Comparison of incisions	28
2.3	Hospital level diffusion of robotic surgery	30
2.4	Diffusion of robotic surgery in the NHS	32
2.5	Volume of robotic and traditional radical prostatectomies	33
2.6	Surgical interventions as a share of prostate cancer diagnosis	33
2.7	Distribution of surgical skills	38
2.8	Key empirical facts	39
2.9	Variation of instrumental variables in sample data	49
2.10	Average relative distance to robotic hospital and to closest hospital	53
2.11	Probability of robotic approach at Z_{days} values	55
2.12	Estimated probability of robotic approach from selection equation - at Z_{dist} values	55
2.13	Test for monotonicity of the instruments	56
2.14	Common support	57
2.15	Length of stay – high vs low skilled surgeons	60
2.16	Adverse event – high vs low skilled	60
2.17	Estimated probability of robotic approach by skill	64
2.18	Estimated probability of robotic approach by age	65
2.19	MTE curve – Normal	66
2.20	MTE curve – Polynomial	66
2.21	MTE curve – Semiparametric	67

2.22	MTE curve – Normal with area fixed effects	67
2.23	Policy simulations - MTE and PRTE	71
3.1	Hospital sites involved in mergers, 2006-2015	77
3.2	Number of NHS mergers 2006-2015	77
3.3	Number of NHS sites involved in mergers 2006-2015	78
3.4	Event study estimates of the impact of hospital mergers on un- planned readmissions	86
3.5	Event study estimates of the impact of hospital mergers on in hos- pital mortality	87
3.6	Event study estimates of the impact of hospital mergers on un- planned readmissions- Placebo test	88
3.7	Event study estimates of the impact of hospital mergers on in hos- pital mortality - Placebo test	88
3.8	Impact of hospital mergers on unplanned readmissions	89
3.9	Impact of hospital mergers on mortality	90
3.10	Hospital-Specialty HHI for HWM categories	93
3.11	Hospital-Specialty HHI for HWR categories	94
3.12	Relationship between effect of mergers and pre-merger HHI	95
3.13	Relationship between the effect of mergers and pre-merger HHI	95
4.1	RTLS components	103
4.2	Tracking movements in a day	103
4.3	Bed-side and Room-level Contact Time	105
4.4	Distribution of Contact Time	106
4.5	Minutes of Bed-Side Contact Time by the Hour (Patients)	106
1	Distribution of surgical skills	123
2	MTE curve – Normal with surgeon level SRR	123
3	MTE curve – Normal with surgeon level SRR	125
4	MTE curve – Normal with sample restriction by experience	126
5	MTE curve – Normal with experience dummies	127

List of Tables

2.1	Radical Prostatectomy Patients	31
2.2	Correlation of surgical outcomes and Z_{days} (Pre-Robots) - Linear regression coefficients	49
2.3	AMI patients mortality and Z_{dist}	52
2.4	Relevance of instruments	54
2.5	Heterogeneity in causal effects - Normal model	61
2.6	Heterogeneity in causal effects – Normal model with local area fixed effects	62
2.7	Heterogeneity in causal effects - Normal model with sample restriction	63
2.8	Length of stay - Conventional estimates	69
2.9	Adverse Event – Conventional estimates	70
3.1	Balancing of characteristics treatment and control group	79
3.2	Summary statistics for quality measures	84
3.3	Summary statistics for competition measure	92
4.1	Patients Summary Statistics	107
4.2	Nurse Summary Statistics	108
4.3	OLS regression - Dependent variable binary indicator of patient death	112
4.4	OLS regression - Dependent variable binary indicator of patient ICU transfer	112

4.5	OLS regression - Dependent variable binary indicator of patient accident	113
1	Patient's Characteristics	115
2	Hospital Trusts - Sample Summary Statistics - 2004/2017	115
3	Association of robotic approach and adverse event - OLS Regression	116
4	Association of robotic approach and log length of stay - OLS Regression	117
5	Estimated impacts of robotic approach on log length of stay	118
6	Estimated impacts of robotic approach on probability of adverse event	118
7	Estimated coefficients - Surgeon level SRR	121
8	Estimated coefficients - Surgeon level SRR	121
9	Estimated coefficients controlling for hospital's year of adoption	121
10	Estimated coefficients - surgeon experience	122
11	Estimated effects - surgeon experience	122
12	Estimated coefficients - Skills measured with fixed effect model	124
13	Estimated effects - Skills measured with fixed effect model	124

Chapter 1

Introduction

This dissertation studies the healthcare sector from an economist's perspective. Underpinning these studies is the idea that economic theories and models can help us understand how the healthcare market operates and how policies can be designed to make healthcare more efficient and accessible for everyone.

The first economic phenomenon I explore is technology adoption. Economists study technology adoption to understand its impact on productivity, economic growth, and market dynamics. Analyzing how businesses and individuals adopt and utilize technology provides insights into efficiency gains, income distribution, and overall societal progress. Notably, understanding the dynamics of technology adoption in healthcare is crucial, given the sector's reliance on innovation. **Chapter 2** investigates the potential of robots to reduce variation in patient outcomes. Across and within occupations, individuals differ substantially in their level of skills, and healthcare providers, such as surgeons and doctors, are no different (Chan et al., 2022; Currie and MacLeod, 2017; Kolstad, 2013a). Differences in providers' skills generate inequality and can exacerbate systematic disparities in access (Finkelstein et al., 2016; Chandra and Skinner, 2003; Deaton, 2003). I show that, in England, the diffusion of robots coincided with an improvement in average surgical performance and convergence in outcomes between high and lower-skilled surgeons. I exploit quasi-random variation in the geographic allocation of robots to study whether this is attributable to the adoption of robotic surgery. Using administrative data

on prostate cancer patients, the most common type of cancer in men in the United Kingdom (UK), I show that robots played a fundamental role. I find that robotic surgery improves surgeons' performance. The robot reduces post-operative length of stay and morbidity across patients. However, my analysis shows that these effects are highly heterogeneous, and technological gains significantly depend on the skills of the surgeon. High-skilled surgeons benefit the least from using the technology, while lower-skilled surgeons appear to gain the most from it. This result suggests that the robot exhibits *decreasing* returns in skills, which means that it complements lower-skilled surgeons more strongly than higher-skilled ones. With traditional surgery, the patients of high skilled surgeons are four percentage points less likely to experience an adverse event than those of lower-skilled surgeons. However, with the robot, they are around one percentage point less likely to experience these events. A similar pattern emerges for the length of stay. As differences in patient outcomes between high and lower-skilled surgeons shrink, my analysis thus suggests that the robot may have the potential to reduce variation in patient outcomes. This effect appears to ensue from lower skilled surgeons performing significantly more poorly without any technological aid, and the technology equalizing them to high skill surgeons.

The second economic phenomenon I study is related to market consolidation. Economists investigate how mergers influence consumer welfare, particularly in terms of pricing, product quality, variety, and availability. Mergers can affect consumers positively through improved offerings or negatively through reduced competition and increased prices. Recent trends in hospital consolidation have been notable within the healthcare industry, with a growing number of mergers and acquisitions taking place among healthcare providers. **Chapter 3** studies the impact of hospital mergers on the quality of clinical care. Using the universe of hospital medical records in England it examines all public hospital mergers after the introduction of hospital choice in 2006. There were 159 hospital sites involved in mergers over our sample period, comprising 13 transactions. Using an event study framework to evaluate the population of hospital mergers in the English NHS between 2006

and 2015, it finds that mergers have immediate and persistent negative impacts on clinical quality.

Lastly, this dissertation focuses on how interactions between service providers and consumers affect the provision of services in a very special type of firm: hospitals. **Chapter 4** uses a unique source of information, Real-Time Location System (RTLS) Data, to study the effect of contact time on patient health outcomes. The setting of this paper is the New Cross Hospital in Wolverhampton, England. This is a large district general hospital part of the Royal Wolverhampton NHS Trust (RWT). The Trust is one of the main healthcare providers in the West Midlands, covering acute, community, and primary care services. In 2013, RWT partnered with a United States technology company to develop a real-time patient flow and tracking solution. This application was intended to support staff in delivering care and to enhance efficiency through the process of providing real-time operational information across clinical areas. RTLS allows to perfectly observe the amount of time nurses spend with patient in hospital. This Chapter exploits the level of granularity of this data to estimate a causal impact of contact time on patient outcomes. By decomposing contact time into an endogenous and plausibly exogenous component, this Chapter shows that direct contact between nurses and patients significantly reduces in-hospital mortality and accidents.

Chapter 2

Technology, Skills, and Performance: The Case of Robots in Surgery

2.1 Introduction

Disparities in access and quality of services concern regulators and policy makers. This is particularly true in healthcare, where substantial effort has been devoted to study why differences in patient outcomes across areas and providers persist, even after controlling for patient risk Skinner (2011). Providers' use of alternative treatments may explain part of this phenomenon (Tsugawa et al., 2017; Birkmeyer et al., 2013b). Health outcomes appear nonetheless to be only marginally affected by it (Molitor, 2018). In fact, heterogeneity in healthcare providers' skills may be at the root of this variation (Chandra and Staiger, 2020; Hull, 2018; Chandra and Staiger, 2007).

In this paper, I investigate the potential of robots to reduce variation in patient outcomes. Across and within occupations, individuals differ substantially in their level of skills, and healthcare providers, such as surgeons and doctors, are no different (Chan et al., 2022; Currie and MacLeod, 2017; Kolstad, 2013a). Differences in providers' skills generate inequality and can exacerbate systematic disparities in access (Finkelstein et al., 2016; Chandra and Skinner, 2003; Deaton, 2003). I show that, in England, the diffusion of robots coincided with an improvement in average

surgical performance and convergence in outcomes between high and lower-skilled surgeons. I exploit quasi-random variation in the geographic allocation of robots to study whether this is attributable to the adoption of robotic surgery. Using administrative data on prostate cancer patients, the most common type of cancer in men in the United Kingdom (UK), I show that robots played a fundamental role.

The literature in economics has mostly thought of robots as competing against human labor in the production of different tasks (Acemoglu and Restrepo, 2020; Humlum, 2019). However, in many applications, the robot is meant to aid rather than substitute workers. Surgical robots are fully operated by surgeons and are an extension of their users. I anticipate that, in this case, any potential return from using the technology will depend on the interaction between the human and robotic capabilities.

Robotic technology may exacerbate variation in surgical performance, or may be a solution to this problem if its returns are decreasing in surgeons' skills. I show that robotic surgery reduces variation in patient outcomes, and this reduction is caused by what I estimate to be more significant improvements among lower skilled surgeons.

Part of my contribution is to identify the impact of this technology in the presence of both heterogeneous treatment effects and a selection problem. To this day, medical evidence that robotic surgery improves patient outcomes, relative to the more invasive alternative, has been at best inconclusive (Coughlin et al., 2018; Yaxley et al., 2016; Robertson et al., 2013; Bolla et al., 2012). Existing studies are based on small and selected samples (Neuner et al., 2012) and are not designed to identify causal effects (Ho et al., 2013). If the potential of robotic surgery to improve performance depends on surgical skills, small sample studies will reflect only part of the picture. Moreover, suppose the uptake of this technology is also heterogeneous across the skills' distribution. In that case, any naive correlation will speak more to the characteristics of the adopters rather than the technology itself. Importantly, when treatment effects are heterogeneous, surgeons and patients may choose

the robot based on their specific technological gains (Björklund and Moffitt, 1987). Regression-adjusted comparisons between robotic and traditional surgery would, in this case, provide misleading estimates if adoption is informed by unobserved factors that influence selection.

To identify causal effects, I use an approach introduced by Björklund and Moffitt (1987) and generalized by Heckman and Vytlacil (2005) that concentrates on the marginal treatment effect (MTE). In this context, the MTE is the average effect of robots on the outcome of individuals at a particular margin of indifference between robotic and traditional surgery. With this approach, I identify the causal effects of robots on patient outcomes and how these depend on surgical skills. I focus on two patient outcomes: the speed of recovery (i.e. post-operative length of stay) and the occurrence of adverse events from surgery (i.e. post-operative morbidity). These are two dimensions of surgical performance that matter to physicians, patients, and policymakers (Lotan, 2012), and robotic surgery should have a measurable effect on them because it increases precision and requires smaller incisions (Higgins et al., 2017; Coelho et al., 2010; Lowrance et al., 2010; Nelson et al., 2007). I use a single risk-adjusted indicator of surgeons' patient outcomes to measure skills. Because I expect the robot to impact surgeons' performance, I estimate this indicator using data from the years preceding the introduction of this technology nationally. In fact, the indicator is measured when all operations were carried out without technological aid and is not affected by the surgeons' adoption behavior.

Identification of causal effects in the MTE framework requires, in most cases, no stronger assumptions than standard instrumental variable methods, but poses a more substantial burden on the instrument (Cornelissen et al., 2016). Indeed, this method requires at least one instrumental variable to be continuous. I exploit the staggered adoption of robots over time to construct two instruments that arguably satisfy the conditions for identification.

In England, the acquisition of surgical robots has been managed by individual hospitals (Lam et al., 2021). This process resulted in an uneven distribution of robots

geographically and created differences in the availability of the technology over time. I argue that the timing of the patient cancer diagnosis, relative to his closest hospital adopting the robot, induces a variation in the probability of robotic surgery that is uncorrelated to patient outcomes. Further, as in McClellan et al. (1994); McClellan and Newhouse (1997) and Gowrisankaran and Town (1999), I argue that the patient relative distance to a hospital with the robot affects the probability of robotic surgery but is plausibly uncorrelated to outcomes.

I find that robotic surgery improves surgeons' performance. The robot reduces post-operative length of stay and morbidity across patients. However, my analysis shows that these effects are highly heterogeneous, and technological gains significantly depend on the skills of the surgeon. High skilled surgeons benefit the least from using the technology, while lower skilled surgeons appear to gain the most from it. This result suggests that the robot exhibits *decreasing* returns in skills, which means that it complements lower skilled surgeons more strongly than higher skilled ones. With traditional surgery, the patients of high skilled surgeons are four percentage points less likely to experience an adverse event than those of lower skilled surgeons. However, with the robot, they are around one percentage point less likely to experience these events. A similar pattern emerges for length of stay. As differences in patient outcomes between high and lower skilled surgeons shrink, my analysis thus suggests that the robot may have the potential to reduce variation in patient outcomes. This effect appears to ensue from lower skilled surgeons performing significantly more poorly without any technological aid, and the technology equalizing them to high skill surgeons.

That said, I uncover a strong pattern of negative selection. High skilled surgeons use the technology more intensively, while lower skilled ones use it less despite their higher returns. Surgeons generally appear to use the robot on younger and less complex patients, but on all patients highly skilled surgeons are more likely to use the robot. Similarly, the MTE curve is downward sloping, with higher resistance to treatment associated with larger improvements in patient outcomes. Heterogeneous actual or perceived costs to adopt the technology may explain this result

(Suri, 2011).

This paper builds on several literatures. An influential body of work has documented heterogeneity in skills and treatment rates across healthcare providers. Abaluck et al. (2016), Currie and MacLeod (2017), and Chan et al. (2022) show that doctors differ in their ability to diagnose patients. Part of this literature focuses on the role of comparative advantage to explain providers' treatment decisions. In Chandra and Staiger (2007) productivity spillovers generate heterogeneity in returns which may induce some hospitals to use a certain treatment more intensively. In a recent paper, Breg (2022) shows that tradeoffs between multiple dimensions of health may explain differences in treatment rates. Chandra and Staiger (2020) conclude that most hospitals overuse treatments in part because of incorrect beliefs about their comparative advantage. I add to this literature by showing that the adoption of new technologies may limit the extent to which skills heterogeneity affect patient outcomes, but that some providers may under use the innovation, therefore limiting its potential.

More broadly, this paper contributes to the literature studying the effects of technology on the labor market. This literature focuses, for the most part, on the way technology affects workers across education levels (Acemoglu and Autor, 2011). I concentrate instead on within occupation and task effects. A recent focus of this literature have been robots. Unlike Acemoglu and Restrepo (2020) and Humlum (2019), I study the effects of robots on workers' performance rather than wages or employment, and I study robots in abstraction from automation. Hence, I bring a novel perspective to the study of the relationship between skills and technologies.

Lastly, I contribute to a new literature studying the effects of robots in healthcare. Using data from the United States (US), Horn et al. (2022) show that adopting a robot drives prostate cancer patients to the hospital. Maynou et al. (2021) describe a similar pattern for the UK and shows that the adoption of robots correlates with reduced readmissions and length of stay. Maynou et al. (2022) discusses how the use of robots for prostate cancer patients affected their diffusion in other specialties

in the UK.

The paper proceeds as follows. Section 2.2 describes surgical robots and their use for prostate cancer surgery. Section 4.2 presents the data. Section 2.4 discusses how I measure surgeons' skills and provides the empirical facts that have motivated this work. Section 2.5 presents the econometric model and the conditions required for identification and estimation of the parameters. Section 2.6 introduces the instrumental variables I will use to identify the model parameters and discusses their validity. Section 2.7 summarizes the results. Finally, Section 2.8 concludes.

2.2 Robotic surgery for prostate cancer

The uses of robotics in surgery were hypothesized as far back as 1967, but it took nearly 30 years and the National Aeronautics and Space Administration (NASA) to complete the first functional surgical robot (George et al., 2018).

The only type of robot currently available in the US and the UK is the da Vinci surgical system. This is manufactured by the California-based company and market leader Intuitive. The robot has three components which I show in Figure 2.1:

1. a viewing and control console that the surgeon uses,
2. a vision cart that holds the endoscopes and provides visual feedback, and
3. a manipulator arm unit that includes three or more arms.

The instruments, including a video camera, are attached to the robotic arms and controlled directly by the surgeon. The robotic arms not only allow to work through incisions much smaller than what would be required for human hands but also to work at scales, where hand tremors would pose fundamental limitations (Tonutti et al., 2017). The console consists of multiple components, including finger loops, joysticks, and foot pedals, that allow movements to go through the robotic arms. The robotic joysticks require less force to manipulate than standard tools (Jayant Ketkar et al., 2022), and an adjustable seat and arm support allow surgeons to adapt the machine to their bodies. By providing articulation, implementing filtering of tremors,

and simulating tactile sensations, the surgeon's dexterity and eye-hand coordination are enhanced, thereby subjectively improving surgical performance (Tonutti et al., 2017).

Figure 2.1: Picture of Da Vinci surgical system



Although robots have found several applications in surgery, this paper focuses on robotic surgery for prostate cancer (or radical prostatectomy (RP)). Prostate cancer is the most common cancer in men in the UK; that's 129 men are diagnosed with prostate cancer every day, and more than 11,500 die yearly from it.¹ I restrict my attention to this operation because the robot has played a notable role in transforming how surgeons perform it (Hussain et al., 2014).

In the US, the diffusion of robots for prostate cancer surgery has been incredibly rapid. In 2003, less than 1 percent of surgeons in the US performed this procedure robotically. Seven years later, already 86 percent of the 85,000 men who had prostate cancer surgery had a robot-assisted operation. Eventually, by 2014, robotic surgery accounted for up to 90 percent of radical prostatectomies across the US.² This trend has been similar in England where, by 2014, the majority of cases (62.7 percent) were performed robotically (Marcus et al., 2017).

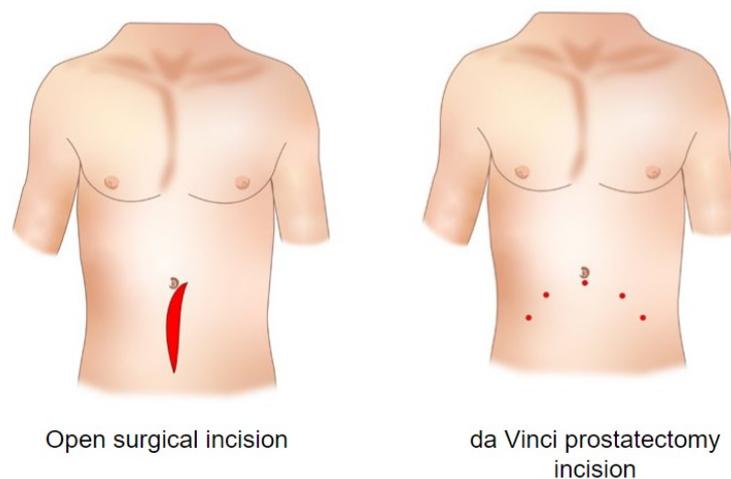
Before robots, prostate cancer surgery was usually performed with an 'open'

¹<https://prostatecanceruk.org>

²<https://www.nature.com/articles/d41586-020-01037-w>

method because the prostate is hard to access with conventional tools. In the ‘open’ method, the surgeon makes a single large incision that allows seeing the area of interest and operate.³ From an oncological perspective, robotic surgery is equivalent to traditional surgery; they are both practical to remove cancer when this is confined to the prostate. However, robotic surgery promised to reduce blood loss, pain, scarring, infections, and average length of stay (among others) by replacing the practice of cutting patients open with a technique that involved only a few small incisions (see Figure 2.2) and complex manual tools.

Figure 2.2: Comparison of incisions



Note: Comparison of incisions required for traditional and robotic radical prostatectomy

Generally, medical technology is considered to be valuable if the benefits of medical advances exceed the costs (Cutler and McClellan, 2001). Robotic surgery is now the standard for the removal of prostate cancer, but doubts remain on whether the supposed benefits outweigh the costs of this technology (Davies, 2022). Indeed, among the most significant barriers to adopting robotic surgery are the high costs associated with the purchase and maintenance of robots (Marcus et al., 2017). Lam et al. (2021) suggests that the median cost of acquisition of the da Vinci robot in England is £1,350,000, with a median yearly maintenance cost of £492,000. Moreover,

³Other minimally invasive approaches, such as laparoscopy, had also been available before robotic surgery but had limited popularity because of the problematic position of the prostate. Throughout this paper, I will refer to all approaches that do not involve using robots as traditional surgery.

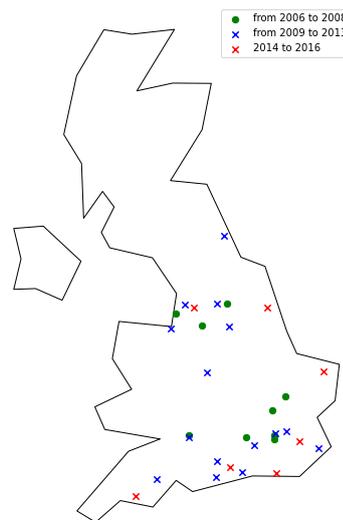
robotic technology requires the surgeon and the hospital to change their practices significantly. Robots usually necessitate a dedicated operating room, which is built for this purpose in many cases. Both surgeons and nurses also need specialized training. Operating using the console requires significant coordination between the head surgeon and the assistant working at the bedside. Any technical drawback during the operation is risky for the patient, but also prolongs operation time and generates inefficiencies for the hospital (Compagni et al., 2015).

2.3 Data and Institutional Context

The data I use comes from the Hospital Episodes Statistics (HES). HES is an administrative data set covering the universe of inpatient discharges from the English National Health Service (NHS). HES provides detailed demographic and clinical information about the patient, including age, sex, ethnicity, admission date, discharge date, and up to 20 recorded diagnoses. Geographical information, such as where patients receive treatment and their area of residence, is also available.

In England, health care is publicly funded and free for all UK residents. Hospitals in the NHS provide care to patients and are reimbursed by the government under nationally agreed tariffs. Planned or elective care is rationed through waiting times and requires an initial referral from a primary care physician (known as a General Practitioner or GP). Patients are entitled to choose a hospital for treatment when the treatment is planned. The choice of which hospital to attend is made with the support of the patient's GP. Hospitals cannot refuse patients, but will schedule admissions and cancel treatments if there is a lack of capacity.

Although equitable accessibility of resources is part of the NHS constitution, the acquisition of surgical robots in England has been managed by individual hospitals (Lam et al., 2021). The adoption of surgical robots has occurred in the absence of guidelines, leaving to the individual provider the decision to adopt the technology and the development of best practices. A recent study suggests that at least 25 percent of hospitals own a robot in England (Lam et al., 2021), but to this day there is no account of the location and utilization of robots in the NHS.

Figure 2.3: Hospital level diffusion of robotic surgery

Note: Produced using HES. In green the hospitals using for the first time a robotic operation code for RP in 2006. In blue those using for the first time a robotic operation code for RP between 2007 and 2009. In red those observed using for the first time a robotic operation code for RP after 2009.

Using HES, I am able to identify and collect data on all operations that involve a surgical robot. In fact, HES provides a record of all procedures performed by NHS hospitals in England and the method used to perform them (e.g. traditional or robotic). Moreover, for each admission, HES identifies the consultant in charge of the operation. HES allows me then to determine the date of the first robotic RP within each hospital, which I will consider as the date of adoption of the technology. In Figure 2.3, I present the location of the hospitals adopting the robot. I do this over three windows of time; from 2006 to 2008, from 2009 to 2013, and from 2014 to 2015. In my identification strategy, I will exploit differences in adoption timing.

Eventually, my sample comprises all radical prostatectomies occurring throughout NHS England from 2004 to 2017 for a total of 62,258 admissions, 25,208 of which are performed with a robot. Table 2.1 summarizes the characteristics of patients for both the traditional and the robotic approach.

HES shows that prostate cancer surgery is England's most commonly performed robotic operation. In Figure 2.4, I plot the number of robotic operations in the NHS vis a vis the number of robotic operations in urology (of which RP is the most

Table 2.1: Radical Prostatectomy Patients

	Full sample		Traditional		Robotic	
	mean	sd	mean	sd	mean	sd
Age	63.061	6.570	63.237	6.550	62.835	6.589
White	0.725	0.446	0.764	0.424	0.675	0.468
Black	0.035	0.184	0.031	0.172	0.041	0.199
Asian	0.014	0.119	0.015	0.121	0.014	0.117
Other	0.225	0.418	0.190	0.393	0.270	0.444
Diabetes	0.076	0.265	0.071	0.257	0.082	0.275
Heart disease	0.035	0.183	0.033	0.178	0.037	0.189
Metastatic cancer	0.015	0.122	0.013	0.113	0.018	0.133
Liver disease	0.007	0.086	0.005	0.074	0.010	0.099
Rural-Urban Indicator	5.411	0.986	5.392	0.935	5.436	1.047
Rank of income deprivation	15531	8471	15421	8432	15646	8511
Rank of health deprivation	16382	9060	16150	9007.111	16626	9110
Rank of education deprivation	17186	9123	16691	9230	17706	8979
Elective admission	0.996	0.066	0.995	0.067	0.996	0.065
Waiting time	39.574	32.518	42.103	33.675	36.679	30.889
Length of stay	3.274	3.024	4.305	3.380	1.944	1.750
Length of stay (pre-operative)	0.330	1.089	0.475	1.190	0.144	0.910
Length of stay (post-operative)	2.944	2.892	3.830	3.212	1.800	1.877
Adverse event	0.144	0.351	0.186	0.389	0.090	0.286
Observations	61839		34829		27010	

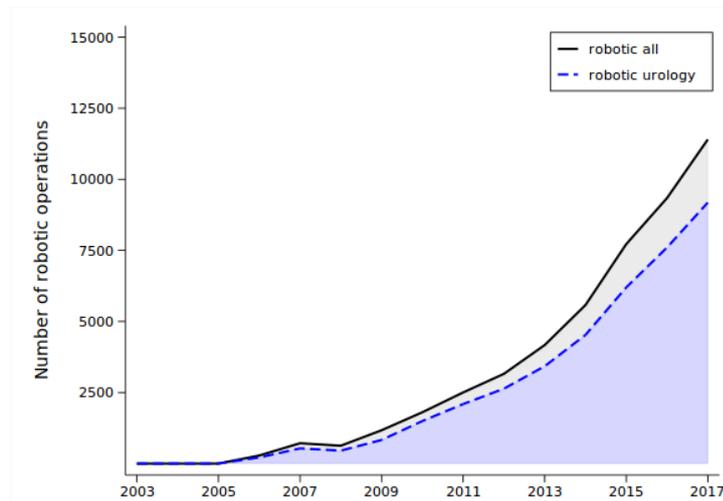
* Note: Source HES. Sample of patients undergoing RP from 2004 to 2007. Patients identified using OPCS code for operations and procedures.

common operation). The figure shows that urology dominates the field of robotic surgery. The first notable use of robots for urology is in 2007. Only five years after 2013 robots start to diffuse in other specialities, but uptake is significantly slower.

The data shows that in England, the use of robotic surgery for RP grew from 5 percent in 2007 to 80 percent in 2017. In Figure 2.5, I plot the total number of RP by surgical approach from 2003 to 2017. The steady increase in the number of robotic operations coincided with a decrease in the number of traditional surgeries. Hence, a clear pattern of substitution toward this technology (Maynou et al., 2021). Moreover, the figure shows a remarkable increase in the number of RPs over time, with the number of patients undergoing this operation almost doubling from 2009 to

2017. In fact, this period is characterized by a considerable increase in prostate cancer diagnoses. Figure 2.6 displays the number of prostate cancer diagnoses and the share of patients opting for RP over time. However, the share of patients undergoing surgery remains relatively stable.

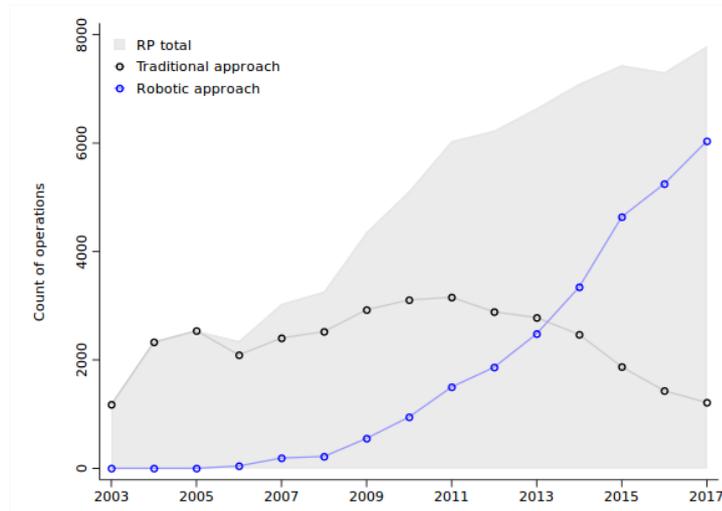
Figure 2.4: Diffusion of robotic surgery in the NHS



Note: The picture shows the number of robotic operations by year for urology compared to all other specialties in which robots are used. The data is from the Hospital Episodes Statistics.

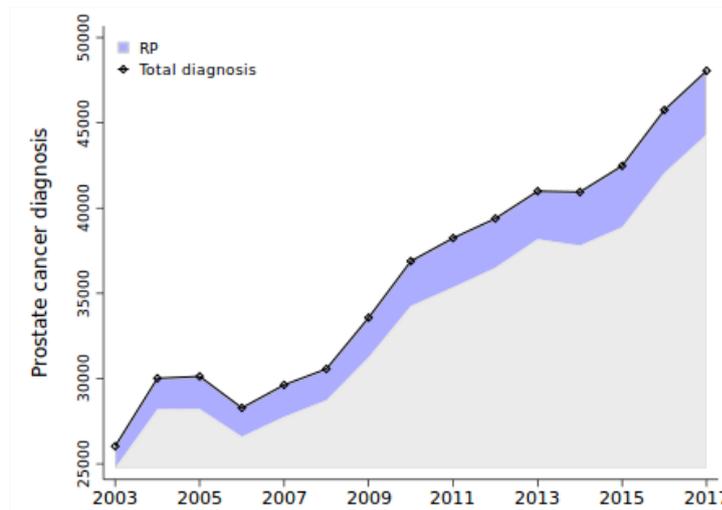
From HES, I identify two patient outcomes for which I evaluate the effect of robots. Namely, patients' length of stay and the occurrence of adverse events from surgery. I focus on these patient outcomes for several reasons. First, they are important margins of performance for patients. Undoubtedly, patients desire to spend fewer days in the hospital and to minimize the number of complications from surgery. If robotic surgery would improve these outcomes, patients would clearly benefit from it. Second, these are important cost drivers to the system and often considered when evaluating whether a technology is worth adopting (Lotan, 2012). Third, the medical literature considers that — if any — robotic technology should have measurable benefits on these two margins (Higgins et al., 2017; Coelho et al., 2010; Lowrance et al., 2010; Nelson et al., 2007). Robotic surgery allows operating using small and compact tools that can fit into narrow incisions. For this reason, the procedure is less invasive and should therefore increase the speed of recovery (or reduce length of stay). Further, because these tools allow for higher precision, the

Figure 2.5: Volume of robotic and traditional radical prostatectomies



Note: Graph produced using HES. The shaded gray area represents the total number of RP performed by NHS Hospitals in England. The black dots represent the number of radical prostatectomies performed using the traditional approach. The blue dots represent the number of radical prostatectomies performed using the robotic approach.

Figure 2.6: Surgical interventions as a share of prostate cancer diagnosis



Note: Graph produced using HES. The shaded blue area represents the number of RP performed by NHS hospitals in England. The shaded gray area represents the number of patients with prostate cancer that have undergone radio therapy treatment. The black line represents the total number of patients diagnosed with prostate cancer.

incidence of complications should diminish. Lastly, these are outcomes that I can reliably measure from the data I have.

The length of stay in hospital of a patient undergoing surgery can be decomposed into two parts; pre- and post-operative. Pre-operative length of stay refers to the number of days between the date of admission and the date of operation. This is believed to be primarily determined by hospital management and should therefore reflect efficiency rather than performance (Cooper et al., 2010a). Post-operative length of stay refers to the number of days a patient spends in the hospital after surgery. A shorter post-operative length of stay suggests that the patient recovered quickly, while a prolonged one may indicate the occurrence of complications in the operating theatre (Strother et al., 2020). Consequently, I concentrate on the effect of robots on post-operative length of stay, which I measure for each patient as the number of days between the operation date and the date of discharge.

I identify adverse health events, likely to be the result from the operation being suboptimally performed, by exploiting the panel dimension of my data. I focus on three adverse events: in-hospital deaths, 30 days emergency readmissions, and complications arising within two years of operation that require surgical interventions. The latter class of events includes urinary complications and erectile dysfunctions. These are common side effects of prostate cancer surgery and are often employed to measure surgical performance.⁴

Table 2.1 summarizes both margins of surgical performance. The average post-operative length of stay in the sample is 2.9 days, and more than 14 percent of individuals appear to have experienced an adverse event.

2.4 Measuring Skills

Skills are not directly observable and notoriously difficult to measure. The measurement most commonly called upon in economics is some indicator of educational attainment (Borghans et al., 2001), but when all those performing a job must have the

⁴I will not be able to detect erectile dysfunctions that are treated with medical interventions with the data I have.

same level of education, this approach is infeasible. In some occupations, however, the product of one's work is observable, and its quality can be attributed to the skills of the individual. For example, Birkmeyer et al. (2013a) shows a clear relationship between surgical skills and patient outcomes.

In line with the medical literature, I use patients' post-operative outcomes as a proxy measure of surgeons' skills. I focus on two adverse events, namely within 30 days in hospital deaths and readmissions. Using patient outcomes to compare surgeons requires however some way of risk-adjustment. The purpose of the risk adjustment is to remove differences in health and other risk factors that impact observed outcomes, thereby enabling a more accurate comparison across surgeons that treat individuals of varying clinical complexity. In fact, surgeons work on patients that vary in their level of health and deal with cases of various complexity.

My objective is to produce a single risk-adjusted indicator of skills. To compare outcome rates from different populations of patients, I adapt a risk-adjustment methodology developed in Horwitz et al. (2014) for the Centers for Medicare Medicaid Services (CMS). I compute the skills measure in two steps. In the first step, I estimate a random coefficient model with a surgeon random intercept.

Let Y_{ij} for patient i operated by surgeon j denote the binary outcome equal to one if the patient experiences post-operative morbidity. X_{ij} denotes a set of risk factors identified by the medical literature to influence the outcome of patient j . Let M denote the number of surgeons and M_j the number of prostatectomies performed by surgeon j . I assume that the outcome is related linearly to the covariates via a Logit function:

$$\text{logit}(\text{Prob}(Y_{ij} = 1)) = \alpha_j + \beta X_{ij} \quad (2.1)$$

$$\alpha_j = \mu + \omega_j$$

$$\omega_j \sim \mathcal{N}(0, \tau^2)$$

α_j represents the surgeon specific random intercept; μ is the adjusted average outcome over all surgeons; and τ^2 is the between surgeons variance component. The

component ω_j will be estimated it using the empirical Bayes posterior mean. The empirical Bayes estimate will capture variation in post-operative morbidity at the surgeon level for observationally similar patients. The conditional distribution of the binary indicator given the random effects is assumed to be Bernoulli, with the probability of an adverse event determined by the logistic cumulative distribution function. I present the X_{ij} set of k patient level covariates included in the model in Table 1.

In the second step, I use the regression estimates from Equation 2.1 to compute a surgeon's Standardized Risk Ratio (SRR) of post-operative morbidity, which I use to proxy the surgeon's skills. The SRR is the ratio between what Horwitz et al. (2014) calls the *predicted* and *expected* post-operative morbidity. The *predicted* number of adverse events for a surgeon j is calculated as the sum of the predicted probability for each patient $\in M_j$, including surgeon j random effect α_j . The *expected* number of adverse events for a surgeon j is calculated as the sum of the predicted probability of readmission for each patient $\in M_j$, ignoring the surgeon specific random effect. This is the probability of an adverse event given the estimated parameters, but where τ is zero, equivalently, this is the probability of an adverse event when the dispersion in α_j is set to zero.

In practice, I compute these terms as follows:

$$\text{predicted}_j = \sum_{i \in j} \text{logit}^{-1}(\widehat{\alpha}_j + \widehat{\beta}X_{ij}) \quad (2.2)$$

$$\text{expected}_j = \sum_{i \in j} \text{logit}^{-1}(\widehat{\mu} + \widehat{\beta}X_{ij}) \quad (2.3)$$

My indicator of skills is then is:

$$\text{Skills}_j = \frac{\text{predicted}_j}{\text{expected}_j}. \quad (2.4)$$

A value of 1 indicates that the level of post-operative morbidity for surgeon j is as

expected given her pool of patients. When the ratio is above (below) 1 it indicates that the surgeon is under- (over-) performing relative to the national average. I estimate the model parameters using data from 2005 to 2007, a period prior to the diffusion of robots in the NHS. Skills are then measured when all operations were performed with the traditional method. In this way, the skill level is not endogenous to the use of the technology.

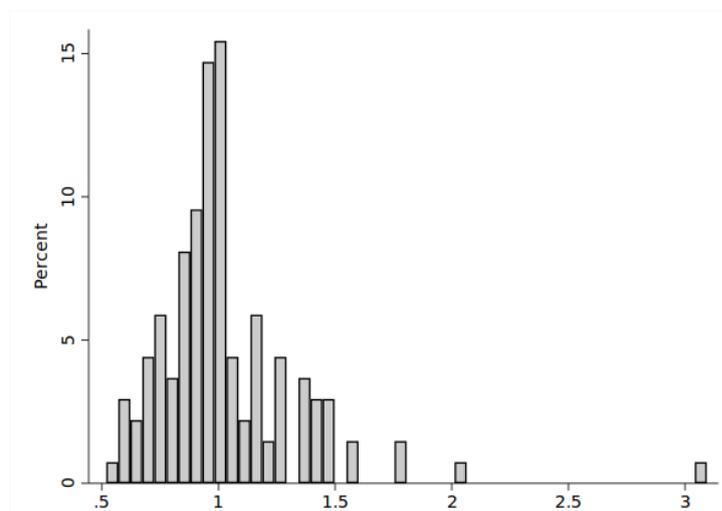
In practice, I perform this estimation at the hospital level. I do this because the majority of hospitals have 1 to 2 surgeons operating prostate cancer patients, and in most cases, one surgeon is significantly lower volume. What I am going to measure will reflect then more closely the average performance of the hospital's team of surgeons. As the median number of surgeons per hospital in my sample is two, I believe this simplification is unlikely to be significant. In turn, I am able to estimate this measure for 144 hospitals. To show that this simplification is unlikely to be relevant, I also test my baseline specification when I compute the measure at the surgeon level. I show the distribution of the measure of skills with surgeons' random intercept in Figure 1 of the Appendix.

In Figure 2.7, I show how surgeons' skills are distributed according to this measure. There is substantial variation in the skills of surgeons pre-robot. The standard deviation is 0.4, and the distribution is characterized by long tails to the right, suggesting that some surgeons perform particularly poorly.

Key facts on robotic surgery, skills, and performance

I start my analysis by showing in Figure 2.8 some correlations between robotic surgery, skills, and performance. I group surgeons into two categories; top and bottom surgeons. Top surgeons are identified as those above the 20th percentile of the distribution of skills (low post-operative morbidity), bottom surgeons are identified as those below the 80th percentile (high post-operative morbidity).

The first fact that emerges is that surgeons at the top of the distribution of skills appear to use the technology more intensively. These surgeons start using the robot before anyone else, and by their second year of use, they operate on more than

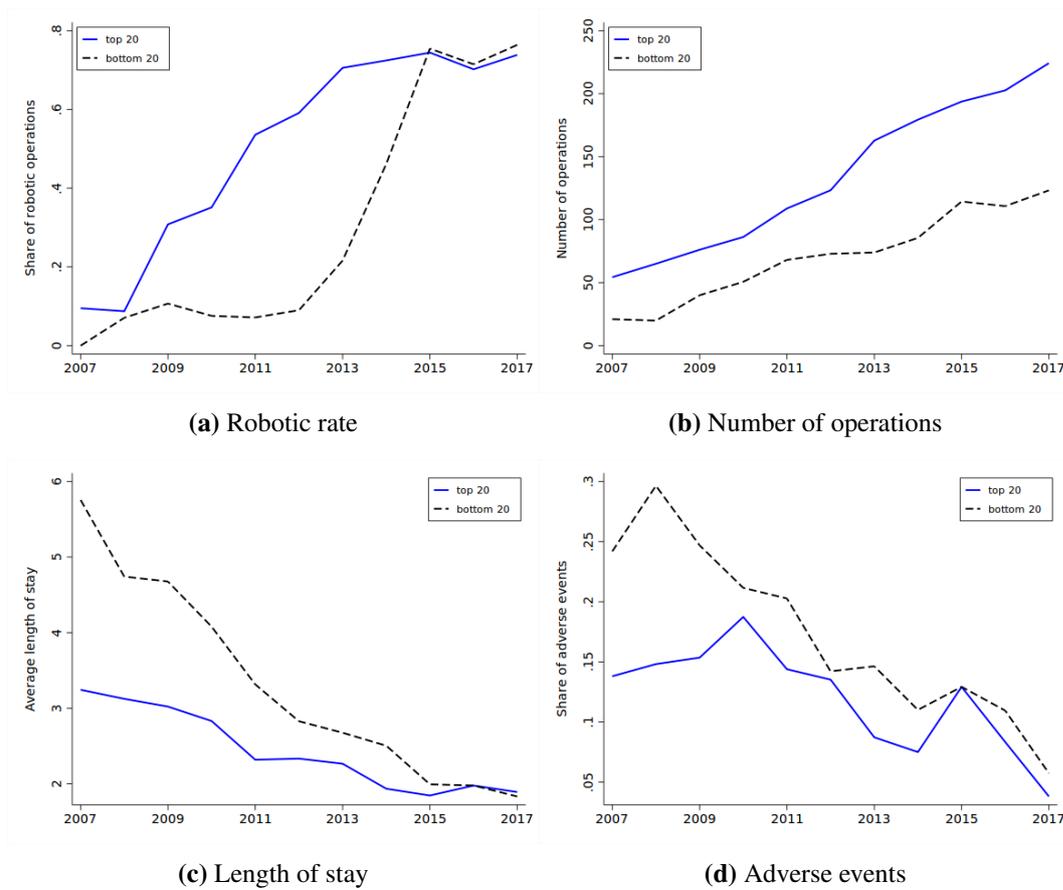
Figure 2.7: Distribution of surgical skills

Note: Distribution of skills measure (i.e. post-operative morbidity standardised risk ratio). Measure computed as the ratio between predicted and expected morbidity (deaths and readmissions). Predicted and expected post-operative morbidity are obtained by estimating the logistic model described in Section 4. Hospital random intercept for predicted post-operative morbidity. Estimates using all prostatectomy patients from 2005 to 2007.

20 percent of their patients using the technology. It takes five more years for the surgeons at the bottom of the distribution to use the technology at a similar rate. By the end of the sample period, however, both groups use the robot at a similar rate and almost 80 percent of patients are operated on with the robot in 2017.

The second fact is that over this period there has been a substantial improvement in surgical performance. Post-operative length of stay and morbidity have decreased respectively by 57 and 73 percent from 2007 to 2017. But, there has also been a convergence in surgical performance between surgeons at the top and the bottom of the skill distribution. In 2007, patients operated on by high-skilled surgeons experienced 3.5 days of post-operative length of stay, while lower skilled surgeons had an average of 6 days. By 2017, this was down to around 2 days for both groups. A similar trend can be observed when inspecting the share of patients experiencing an adverse event from surgery. For both outcomes indeed, by the end of the sample the raw outcomes of high and low skilled surgeons are about the same.

Generally, regardless of skills, there has been an increase in the number of patients under the care of these surgeons. This is consistent with the increase in the number

Figure 2.8: Key empirical facts

Note: Top quality if skills pre-robot period above the 20th percentile, bottom quality if skills pre-robot below 80th percentile. Mean share of robotic operations is computed as the number of operations per year using the robot over the total number of operations at the hospital level. The rate of adverse events is computed as the number of patients experiencing an adverse event from surgery over the total number of operations.

of prostate cancer diagnosis we observe in this period. But, it also appears consistent with the findings of Neuner et al. (2012) and more recently by Horn et al. (2022) and Maynou et al. (2021). This increase is nonetheless more significant for high skilled surgeons.

2.5 Econometric model

My empirical strategy is tailored to the presence of heterogeneous treatment effects and the possibility of selection into treatment. My hypothesis is that surgeon's skills will induce substantial heterogeneity in treatment effects, but this could also arise because patients differ in their observed and unobserved characteristics. For

example, the returns from using the robot may depend on the age of the patient, or on whether the patient suffers from diabetes and other comorbidities. Selection occurs because neither patients nor surgeons are randomly allocated to the robotic approach, and the choice of treatment may be endogenous to their observed and unobserved characteristics, and surgery could be selected on the basis of their anticipated effects from treatment (Zhou and Xie, 2019). Surgeons may choose to use the robot only on patients for which they expect a substantial improvement in their outcomes, and opt for traditional surgery otherwise. Regardless of how the allocation of treatment occurs, a selection bias will arise if this process is non-random.

The most commonly used approach to deal with selection on unobservables is the instrumental variable (IV) method. In the IV approach, an external variable (i.e. the instrument) is used to distil out an exogenous variation in the probability of treatment (Banerjee and Basu, 2021). In this paper, I use a different method and employ an approach first pioneered by Björklund and Moffitt (1987) and subsequently developed in Heckman and Vytlačil (2005). This approach focuses on the identification and estimation of the marginal treatment effects (MTE).

The MTE is the average treatment effects for people with a particular unobserved variable value that influences selection. Identification of MTE is intuitively similar to the IV, but is more informative in the presence of heterogeneous effects in some cases (Cornelissen et al., 2018). Heckman and Vytlačil (2005) shows that the MTE is the foundation of all population level treatment effects. For example, the average treatment effect (ATE) is the unweighted average of the MTEs, and it is point identified for $0, 1 \in \text{supp } P(Z)$ (Heckman and Vytlačil, 2001). The average treatment effect on the treated (ATT) is a weighted average of the MTEs where individuals with low values of the unobserved variable value that influences selection are given heavier weights. The average treatment effect on the untreated (ATU) is a weighted average of the MTEs where heavier weights are given to individuals whose unobserved variable value that influences selection is high.

In this section, I first describe the MTE in its theoretical set-up, introduce terminol-

ogy and notation, as well as the foundational assumptions needed for identification. I then present how I apply this framework to my specific context, and introduce the additional assumptions I impose for identification and estimation.

General MTE framework

The building block of the MTE approach is the generalized Roy model of binary treatment choice (Roy, 1951). In this model, the individual can have one of two potential outcomes, Y_1 and Y_0 , depending on the choice of treatment $D \in [0, 1]$. For each individual, depending on the choice of treatment, only one outcome is actually observable. Both outcomes depend on some observed characteristics X , that are not determined by D , and an unobserved component which is additively separable:

$$Y_0 = h_0(X) + \varepsilon_0 \quad (2.5)$$

$$Y_1 = h_1(X) + \varepsilon_1 \quad (2.6)$$

$h_D(X) \equiv E[Y_D|X]$ for $D \in [0, 1]$ and ε_0 and ε_1 are error terms of mean zero conditional on X .

The treatment choice is represented by an index threshold crossing model

$$D = 1[D^* \geq 0] \quad (2.7)$$

where a person chooses $D = 1$ whenever the latent variable $D^* \geq 0$. The latent choice is a function of observable Z characteristics and an additively separable component V :

$$D^* = g(Z) - V \quad (2.8)$$

From the point of view of the econometrician Z is observed while V is not (Carneiro et al., 2011). The Z vector may include some or all of the variables in X , but crucially includes a continuous variable that affects outcomes only via the treatment status (i.e. a continuous instrument for D). As V enters the expression with a negative sign, this is called resistance to treatment. This a continuously distributed ran-

dom variable representing all unobserved factors that make an individual less likely to choose $D = 1$. Importantly, no restriction is imposed on the relationship between (Y_1, Y_0) and V , so that individuals may select on the basis of their anticipated return from treatment, or treatment effect.

Two assumptions are maintained to specify and identify the MTE using the method of local instrumental variables (Heckman and Vytlacil, 1999):

Assumption 1. $(\varepsilon_0, \varepsilon_1, V)$ are statistically independent of Z conditional on X (*Independence*).

Assumption 2. $g(\cdot)$ is a non-trivial function of Z conditional on X (*Rank condition*).

To specify the MTE, the decision rule is conventionally expressed in terms of the propensity score $P(Z)$, i.e., the probability of treatment given the observed covariates:

$$\begin{aligned} P(Z) &\equiv P(D = 1|Z) \\ &= P(D^* \geq 0|Z) = P(g(Z) - V \geq 0|Z) \\ &= F_{V|Z}(g(Z)) \\ &= F_{V|X}(g(Z)) \end{aligned}$$

where $F_{V|X}(\cdot)$ is the cumulative distribution function of V given X .

The decision rule in terms of the propensity score is

$$\begin{aligned} D &= 1[D^* \geq 0] \\ &= 1[g(Z) - V \geq 0] \\ &= 1[F_{V|X}(g(Z)) - F_{V|X}(V) \geq 0] \\ &= 1[P(Z) - U \geq 0] \end{aligned}$$

where the variable $U \equiv F_{V|X}(V)$ represents the quantiles of the distribution of the

unobserved resistance to treatment V , which by definition follows a standard uniform distribution.

The MTE, is defined by the following conditional expectation:

$$\begin{aligned} & E[Y_1 - Y_0 | X = x, U = u] \\ &= h_1(X) - h_0(X) + E[\varepsilon_1 - \varepsilon_0 | X = x, U = u] \\ &\equiv MTE(x, u) \end{aligned}$$

It is the average gain from treatment for individuals with characteristics $X = x$, and indifferent between treatments at the propensity score $P(Z) = u$. Variation in the $MTE(x, u)$ over values of u reflects how treatment effect varies with different quantiles of the unobserved resistance to treatment.

The MTE is closely related to the LATE. The model, as presented, combined with **Assumption 1.** and **Assumption 2.**, is equivalent to the Imbens and Angrist (1994) conditions of independence and monotonicity for the interpretation of the IV estimands as a local average treatment effects (LATE) (Vytlacil, 2002). The LATE is the average treatment effects on the compliers, individuals in a given range of U , while the MTE is this effect at a specific value of U .

Application of the MTE to robotic surgery

In practice, I have two margins over which to evaluate treatment effects. Namely, the logarithm of the patient length of stay in hospital and a binary indicator of adverse event from surgery. I will assume that both outcomes and the choice of treatment depend linearly on the patient's characteristics X_i , and the surgeon's skills $Skills_j$.

Assumption 3. Y_{1ij}, Y_{0ij} and D_{ij}^* are a linear function of X_i and $Skills_j$

$$Y_{1ij} = \beta_1 X_i + \delta_1 Skills_j + \varepsilon_{1ij} \quad (2.9)$$

$$Y_{0ij} = \beta_0 X_i + \delta_0 Skills_j + \varepsilon_{0ij} \quad (2.10)$$

Note that the return to using the robot (i.e., $Y_{1ij} - Y_{0ij}$) varies across individuals with different observed (X_i and $Skills_j$) and unobserved characteristics (ε_{0ij} and ε_{1ij}). This is an important feature of this framework, which emphasizes heterogeneity in returns (and the distinction between the returns for average and marginal individuals) (Carneiro et al., 2011).

As both patients and surgeons jointly determine the course of treatment, the decision to use the robot will also be a linear function of surgeons' skills and patients' characteristics. Importantly, the decision to take treatment depends on a continuous variable Z_i that does not enter the outcome equation (i.e., the instrument).

$$D_{ij} = 1[D^* \geq 0] \quad (2.11)$$

$$D_{ij}^* = \beta_d X_i + \delta_d Skills_j + \gamma_d Z_i - V_{ij} \quad (2.12)$$

Patients that are observationally similar will be allowed to differ in their treatment because of V . For example, if either the surgeon or the patient dislikes the robot, this will be captured by V . Variation in Z will allow me to identify the parameters of the model. I will present the variables included in Z in Section 2.6.

The equivalent representation in terms of the propensity score is:

$$D = 1 \text{ if } P(X_i, Skills_j, Z_i) \geq U, \text{ and } D = 0 \text{ otherwise.} \quad (2.13)$$

Individuals are treated with the robot if the propensity score exceeds the quantile of the distribution of V_{ij} at which the individual is located (Cornelissen et al., 2016).

The observed outcome can then be expressed as:

$$Y_{ij} = Y_{0ij} + D_{ij} \left[\underbrace{(\beta_1 - \beta_0) X_i}_{\Delta_1} + \underbrace{(\delta_1 - \delta_0) Skills_j}_{\Delta_2} + \underbrace{\varepsilon_{1ij} - \varepsilon_{0ij}}_{\Delta_3} \right] \quad (2.14)$$

Where Y_{ij} is either the length of stay or an indicator for whether the patient i has experienced an adverse event from surgery.

The effect of robotic surgery is the sum of Δ_1 , Δ_2 , and Δ_3 . Δ_1 reflects what arises from the characteristics of the patient. For example, Δ_1 will be negative if the technology makes an older patient less likely to experience an adverse event from surgery. Δ_2 reflects gains that arise from the way technology combines with skills and is my quantity of interest. My interpretation is the following:

- A negative Δ_2 implies that the technology complements more strongly individuals with lower skills (decreasing returns in skills);
- A positive Δ_2 implies that higher skilled surgeons experience larger improvements in patient outcomes relative to lower skilled surgeons (increasing returns in skills).

Lastly, Δ_3 is the individual specific idiosyncratic effect from treatment. An important feature of this framework is then that the return from using the robot depends on both observed and unobserved characteristics.

The marginal treatment effect of robotic surgery at $Skills = s, X = x$ and $U = u$ is:

$$MTE(s, x, u) = E(Y_{1ij} - Y_{0ij} | X_i = x, Skills_j = s, U_{ij} = u) \quad (2.15)$$

I will assume that the MTE is additively separable in its components (Brinch et al., 2017):

Assumption 4. $E[\varepsilon_{1ij} - \varepsilon_{0ij} | X_i = x, Skills_j = s, U_{ij} = u]$ does not depend on x and s (Additive Separability)

Under Assumption 1 to 4, the MTE can be represented as:

$$MTE(s, x, u) = x(\beta_1 - \beta_0) + s(\delta_1 - \delta_0) + E(\varepsilon_{1ij} - \varepsilon_{0ij} | U_{ij} = u) \quad (2.16)$$

and the expected outcome of individual i operated by surgeon j is:

$$E[Y_{ij}|X_i = x, Skills_j = s, P(Z) = p] = X_i\beta_0 + Skills_j\delta_0 + pX_i(\beta_1 - \beta_0) + pSkills_j(\delta_1 - \delta_0) + K(p)$$

where $K(p) \equiv \int_0^p E(\varepsilon_{1ij} - \varepsilon_{0ij}|U = u)du$ is a function of the propensity score p and captures all the ‘essential heterogeneity’ in the outcomes. $K(p)$ can be estimated either nonparametrically or with some functional form restrictions.

As shown in Carneiro et al. (2011), the derivative of the outcome Y with respect to p identifies the MTE for individuals with $X = x$, $S = s$, and $U = p$.

$$\begin{aligned} \frac{\partial E[Y|X = x, Skills = s, P(Z) = p]}{\partial p} &= x(\beta_1 - \beta_0) + s(\delta_1 - \delta_0) + \frac{\partial K(p)}{\partial p} \\ &= MTE[X = x, Skills = s, U = p] \end{aligned}$$

The intuition is simple. Increasing the propensity score by a small amount shifts previously indifferent individuals into treatment and changes the observed outcome. By taking the derivative with respect to the propensity score, we obtain the change in Y (i.e., the treatment effect) at a given margin of indifference. As the $K(p)$ component only depends on p , patient and surgeon’s characteristics do not affect the shape of the MTE curve, which implies that I can identify the MTE over the unconditional support of $P(Z)$, jointly generated by the instruments and the covariates, as opposed to the support of $P(Z)$ conditional on covariates.

Estimation of the MTE allows me therefore to identify complementarities between robots and skills, but also to determine whether there is selection on gains from observed or unobserved characteristics. The parameter $\delta_1 - \delta_0$ could be positive or negative depending on whether surgeons of higher skills have higher or lower returns from using the robot. The derivative of $K(p)$ will similarly tell us whether returns are increasing or decreasing in the unobserved component V . In the education literature, the component V is usually thought as the negative of unobserved ability (Carneiro et al., 2011). Under this interpretation, if an individual with higher

unobserved ability had higher returns, the $K(p)$ function should be declining in V .

2.6 Exogenous variation in treatment probability

The MTE framework requires at least one continuous instrumental variable to be included in the selection equation (Heckman and Vytlacil, 2005). The instrument must satisfy the same conditions required by Imbens and Rubin (1997) for identification of the LATE (Vytlacil, 2002). First, it should affect treatment but be plausibly independent of potential outcomes (Y_1, Y_0) . Second, it should affect selection into treatment monotonically. Moreover, ideally, the instrument should have enough variation to generate a propensity score with full support (Cornelissen et al., 2016). I use the fact that robots have been acquired under no centralized strategy, leading to a staggered adoption, to build two instrumental variables that exploit the fact that an individual's access to robotic surgery will vary according to where they live and to the timing of their cancer diagnosis.

Diagnosis timing instrument definition and validity

I propose a novel instrument that exploits diagnosis timing to detect an exogenous variation in the probability of robotic surgery. I will refer to this instrument with the name Z_{days} , and compute it for each patient as:

$$Z_{days} = t - T_R \quad (2.17)$$

where t is the date on which the patient received his diagnosis of prostate cancer⁵, and T_R is the date on which his closest hospital performed its first robotic assisted prostatectomy. I expect that a patient diagnosed after T_R will be more likely to get treated than one diagnosed earlier. The intuition is simple, individuals tend to visit their closest hospital for most issues, hence adoption by the closest hospital raises the probability of robotic surgery.

To satisfy the exclusion restriction, I require the timing of adoption to be random

⁵As diagnosis of prostate cancer requires a biopsy which is performed in hospital, the diagnosis date is identifiable using the HES data.

relative to the individual health status, and hence unrelated to his potential outcomes. Consequentially, Z_{days} should affect the outcomes only through its effect on the patient's likelihood to receive robotic surgery. To provide evidence that this is actually the case, I test whether the instrument has an effect on the surgical outcomes of patients undergoing a radical prostatectomy prior to the introduction of robots to the NHS. For these patients, Z_{days} cannot affect selection into treatment because treatment is not available to them, which means that the first stage effect is by definition null. Hence, any effect of the instrument on the outcomes of these patients would suggest the presence of another channel of impact, and a violation of the exclusion restriction.

Table 2.2 presents the result of this exercise. Column 1 to 3 show the coefficients estimated from a OLS regression of log length of stay on Z_{days} for increasingly richer specifications. The sample comprises all prostatectomy patients operated in the NHS in 2003. The coefficient on Z_{days} is not statistically significant. Column 4 to 6 show the coefficients estimated from a OLS regression of a binary indicator of adverse events on Z_{days} for increasingly richer specifications. The coefficient in column 4 is negative and statistically significant, but after controlling for patient characteristics, this correlation disappears. Overall, this is suggestive that the exclusion restriction is likely to be satisfied conditional on the covariates included in the model.

I show how Z_{days} is distributed in Figure 2.9. The average patient is diagnosed almost a year before his closest hospital has adopted the robot. Consistently, the distribution exhibits a longer tail to the left, i.e., more patients being diagnosed prior their closest hospital has started performing robotic prostate cancer surgery.

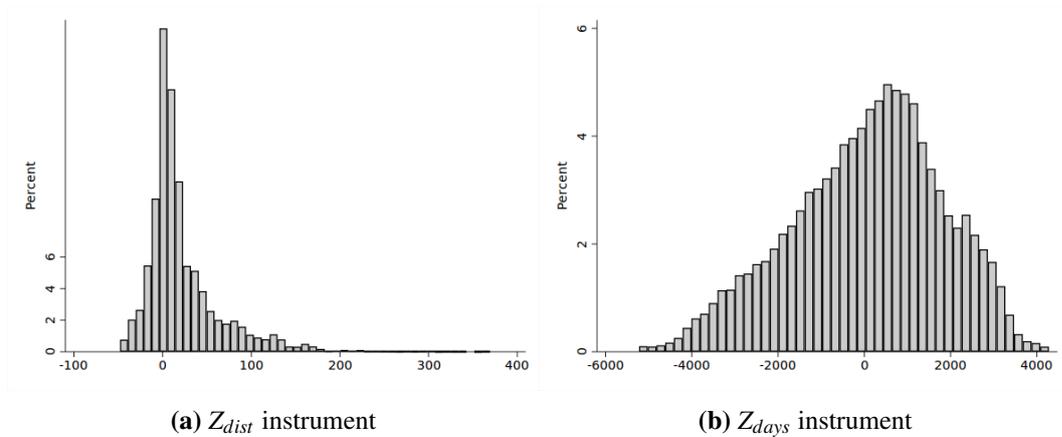
Relative distance instrument definition and validity

In their seminal contribution, McClellan et al. (1994) use differential distances to alternative types of hospitals as independent predictors of how heart attack patients will be treated. More recently, Card et al. (2019) employ a similar instrument in the context of delivery choices of mothers in the US. Card et al. (2019) use the relative

Table 2.2: Correlation of surgical outcomes and Z_{days} (Pre-Robots) - Linear regression coefficients

	Length of stay			Adverse event		
	(1)	(2)	(3)	(4)	(5)	(6)
Z_{days}	-0.049 (0.047)	-0.028 (0.047)	0.072 (0.049)	-0.117* (0.051)	-0.035 (0.05)	-0.033 (0.053)
Patient control	No	Yes	Yes	No	Yes	Yes
Year-month	No	No	Yes	No	No	Yes
Day of the week	No	No	Yes	No	No	Yes
Z_{days}	-2707	-2709	-2709	-2707	-2709	-2709
N	5566	5549	5549	5574	5557	5557

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Linear regression model estimated using OLS. Coefficients and standard errors multiplied by 100. Three significant figures displayed. Model in (2)-(3)-(5)-(6) control for age, age squared, 10 comorbidity dummies, ethnicity, rural urban indicator. Sample of radical prostatectomy patients in 2003.

**Figure 2.9:** Variation of instrumental variables in sample data

Note: Panel (a) plots the instrument Z_{dist} defined as the relative distance between the patients nearest hospital capable of offering robotic assisted radical prostatectomy and the closest hospital offering traditional radical prostatectomy. The distance is expressed in kilometers. Panel (b) plots the instrument Z_{days} defined as the number of days from the patient diagnosis of prostate cancer and the closest hospital to the patient adopting the robot. The date of adoption is the earliest date in which the hospital performs a robotic assisted radical prostatectomy.

distance from a mother's home zip code to the nearest high c-section hospital versus the nearest low c-section hospital as an instrumental variable for delivery at a high c-section hospital.

Inspired by this body of work, I use as an additional instrument the differential distance from the patient's residence to a hospital capable of providing robotic surgery. The idea is that relative distances approximately randomize patients to different likelihoods of receiving treatment. In other words, a patient closer to a hospital offering robotic surgery will be more likely to be operated on with the robot for reasons unrelated to his health. I refer to this variable as Z_{dist} , and I compute it for each patient as;

$$Z_{dist} = D_R - D_T, \quad (2.18)$$

where D_R is the geographic distance between the patient and the nearest hospital with a robot in the year the patient is operated, and D_T is the geographic distance between the patient and the nearest hospital without the robot.

Data on where a patient lives in HES is limited to the postal area, but HES includes information on the patient GP. Hence, I use the postcode of the patient's GP to proxy for his location. In England, individuals have to register to a GP to obtain a referral, which is necessary to access non-emergency services from hospitals. As patients can only register to GP practices in proximity to their home address, I believe the GP's postcode is a good proxy for the location of the patient.

A criticism of this type of instruments is that patients who live nearer to a hospital offering a given treatment — or for this matter to any hospital — may differ in terms of their underlying health because they have better access to care, or access to higher quality care (Hadley and Cunningham, 2004). If this was the case, the instrument would be invalid. To limit this concern, I control directly for the distance between the individual and his closest hospital, and for whether this is a teaching hospital. In this way, relative distance comparisons occur only within groups of individuals that have similar quality and access to care.

Nevertheless, it may still be that relative distance is correlated to health outcomes in a way not accounted for by the model. To investigate the plausibility of such a story, I test whether relative distance to a robotic hospital can predict the health outcomes of individuals who had a heart attack (clinically referred to as an Acute Myocardial Infarction, or AMI).

Under the exclusion restriction, relative distance should only affect patients' outcomes through its effect on the probability of receiving robotic surgery. The treatment of AMI does not involve robotic surgery, and for this reason, relative distance should have no relationship with the health outcomes of patients with this condition. But, if there was non-random sorting of individuals across locations in such a way that relative distance was correlated with better (or worse) health, this would surely emerge in this relationship. I focus on AMI patients for two reasons. First, cardiovascular diseases, of which AMI is the primary manifestation, have a high mortality rate and therefore a well-defined health outcome to test for. Second, mortality from AMI is often associated with poverty or low access to social support (Mookadam and Arthur, 2004). This means that AMI mortality can serve as a proxy for both individuals' health and physical well-being, and of economic and social risk factors. I estimate the relationship between relative distance and AMI mortality only for patients admitted to the hospital from the emergency department, which account for 68 percent of the total admissions for AMI from 2006 to 2010. Table 2.3 presents the estimates from a logistic regression where the dependent variable is hospital death and the independent variable of interest is the instrument Z_{dist} computed for my sample of AMI patients. When I control for patient characteristics and the time period of the operation, I find no statistically significant relationship between AMI mortality and the instrument.

Lastly, I test my baseline model under the inclusion of area fixed effects.⁶ As hospitals adopt the robot at different dates, the relative distance will change for patients living in the same area. I exploit this variation and estimate the model within small

⁶These are postal area fixed effects (i.e., the first four digits of the patient postcode of residence, which is available in HES)

Table 2.3: AMI patients mortality and Z_{dist}

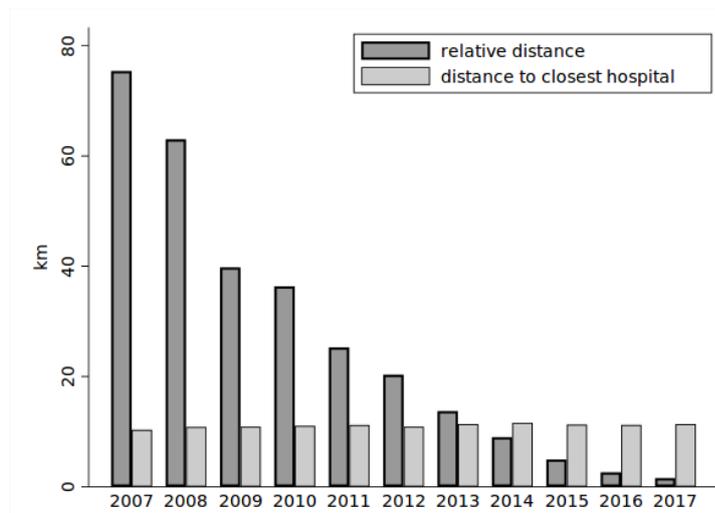
	(1)	(2)	(3)
Z_{dist}	0.045** (0.016)	-0.016 (0.018)	0.031 (0.019)
Distance closest hospital		0.269* (0.107)	0.111 (0.130)
Year-month	No	Yes	Yes
Day of the week	No	Yes	Yes
Patient control	No	No	Yes
Deaths (%)	19	19	19
Z_{dist}	68.64	68.64	68.75
N	68467	68467	67882

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Linear regression model estimated with OLS. Demographic controls are age, age squared, ethnicity and a rural urban indicator. Clinical controls include age, age squared, 10 comorbidity dummies, ethnicity, rural urban indicator. Sample of AMI patients from 2005 to 2009. Coefficients and standard errors multiplied by 100.

geographic cells, which allows for tighter handling of non-random selection than most studies using this type of instrument. A notable exception is Cornelissen et al. (2018), which estimates marginal treatment effects of child care. In this paper, the staggered rollout of a policy granting universal child-care in Germany creates variation in the availability of childcare slots across both geography and cohorts, thus allowing the authors to include in the model municipality fixed effects. As in Cornelissen et al. (2018), I restrict the area dummies to having the same effect in the treated and untreated outcome equations, so they have no influence on the treatment effect. I show how Z_{dist} is distributed in Figure 2.9. The average relative distance is 19 km. This varies substantially over time. The value of the instrument in 2007 was 80 km for the average patients. By 2012 this was down to 20 km, while in 2017 the closest hospital to the average patient offers robotic surgery.

Relevance, monotonicity, and common support assumptions

To show that the instruments are relevant, I estimate a Probit regression where the dependent variable is a binary indicator of the robotic approach regressed on Z_{dist} , Z_{time} , and a large set of individual clinical and demographic controls. Coefficients and marginal effects are presented in Table 2.4, where the columns denote increas-

Figure 2.10: Average relative distance to robotic hospital and to closest hospital

Note: Relative distance computed as the difference between the patient's distance to the closest hospital offering robotic technology and the distance to the closest hospital offering only traditional surgery. The patient location is proxied with the location of his GP. Hospitals date of adoption is identified from HES as the earliest data when a robotic RP is performed.

ingly richer specifications. Column 7 represents the selection equation, which I will discuss in more details in Section 2.7.

Table 2.4 shows that both instruments are statistically significant in predicting whether the patient will be operated with the robot. Z_{dist} has a positive coefficient in all specifications. This indicates that the longer it passes, after the closest hospital has adopted the robot, the more likely the patient is of getting robotic surgery.

In Figure 2.11, I show the average predicted probability evaluated as different values of this instrument. The figure shows how the probability of receiving robotic surgery changes at different values of the instrument. An individual diagnosed two years before his closest hospital has adopted the robot has a 0.4 probability of being treated, while for an individual diagnosed two years after the probability is 25 percent higher. Z_{days} has, instead, a negative coefficient. This indicates that the higher the relative distance, the less likely is the patient to receive robotic surgery. In Figure 2.12, I show the average predicted probability evaluated as different values of this instrument. An individual whose value of Z_{dist} is 30 km has a probability of being treated of 0.4, doubling this distance reduces this probability by almost fifty

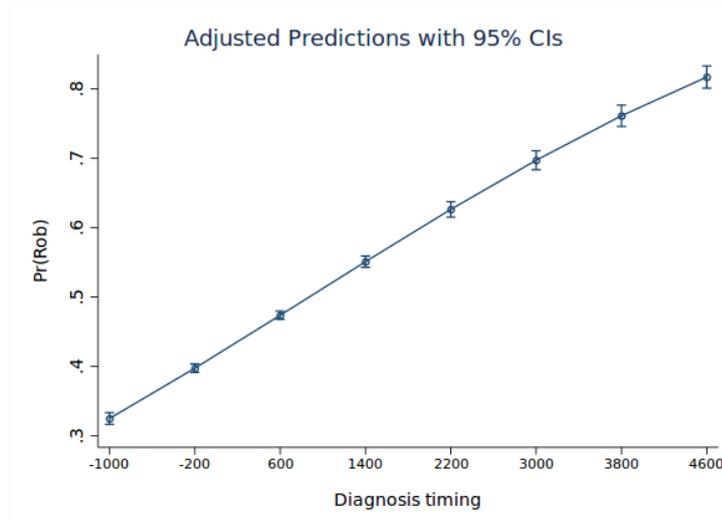
Table 2.4: Relevance of instruments

	(1)	(2)	(3)	(4)	(5)	(6)
Coefficients						
Z_{dist}	-1.95*** (0.038)		-1.04*** (0.032)	-1.07*** (0.032)	-1.1*** (0.033)	-0.99*** (0.034)
Z_{days}		0.057*** (0.001)	0.041*** (0.001)	0.041*** (0.001)	0.042*** (0.001)	0.024*** (0.001)
Marginal effects						
Z_{dist}	-0.651*** (0.008)		-0.307*** (0.009)	-0.316*** (0.009)	-0.319*** (0.009)	-0.272*** (0.009)
Z_{days}		0.0159*** (0.000)	0.0122*** (0.001)	0.0119*** (0.001)	0.0122*** (0.001)	0.007*** (0.001)
Demographic	No	No	No	Yes	Yes	Yes
Clinical	No	No	No	Yes	Yes	Yes
Year-month	No	No	No	Yes	Yes	Yes
Day of the week	No	No	No	Yes	Yes	Yes
Area	No	No	No	No	Yes	Yes
Robot (%)	48	44	49	49	49	49
Z_{dist}	21.83		21.83	21.86	21.86	21.86
Z_{days}		68	389	387	387	387
N	53937	58906	52671	52572	52572	52572

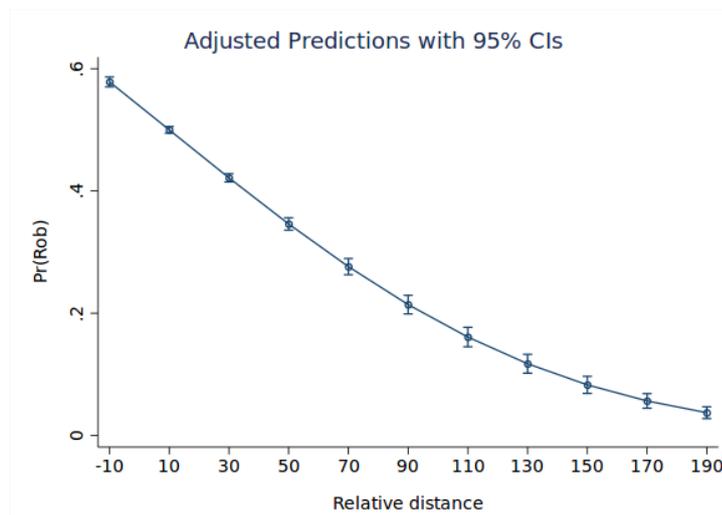
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parentheses. Coefficients, standard errors, and margins multiplied by 100. Probit regression with dependent variable indicator of robotic approach. Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. Area controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. SRR is the standardized risk ratio for post-operative morbidity (interpreted as the inverse of skills).

percent.

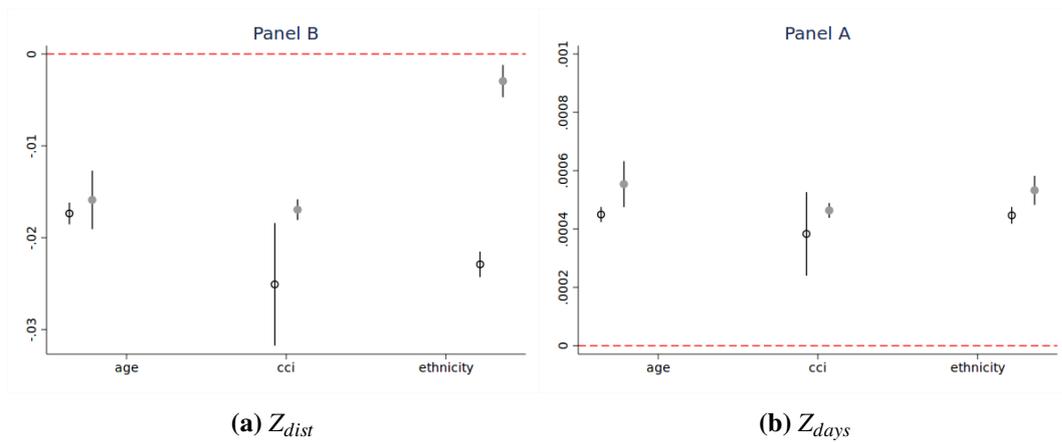
The instruments should affect the probability of treatment in a monotone way. In other words, there should be no defiers (Imbens and Rubin, 1997). I believe that this arguably satisfied by both instruments. It is indeed unlikely that an individual would opt for traditional surgery for a reduction in the distance to a robotic hospital. Similarly, there is no reason to believe that as time passes, from the adoption of the closest hospital, a patient would opt for traditional surgery. To corroborate that this is actually the case, I estimate the selection equation for different subgroups of the population. Specifically, I estimate the first stage separately for individuals above and below the age of 55, residing in areas above and below the mean level of urban development, with different case complexity as measured by the Charlson Comorbidity Index (CCI), and finally for white individuals and for those of other ethnic backgrounds. I present the coefficients on the instruments, estimated using a

Figure 2.11: Probability of robotic approach at Z_{days} values

Note: Probit regression estimates, dependent variable robotic approach. Marginal probability estimated at different value of relative diagnosis timing. Covariates in the model at means, include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest hospital is a teaching hospital, and instrument Z_{dist} . Model controls for month-year and day of the week. Model includes continuous measure of surgical skills. Standard errors computed with delta method.

Figure 2.12: Estimated probability of robotic approach from selection equation - at Z_{dist} values

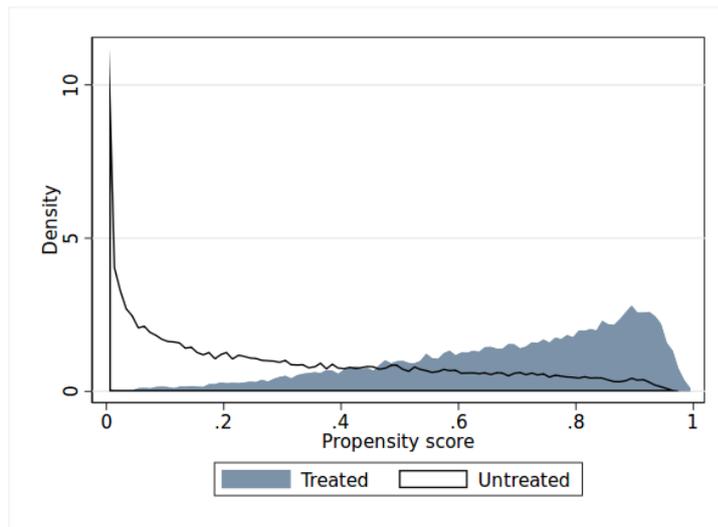
Note: Probit regression estimates, dependent variable robotic approach. Marginal probability estimated at different value of relative distance to hospital offering robotic approach. Covariates in the model at means, include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest hospital is a teaching hospital, and instrument Z_{days} . Model controls for month-year and day of the week. Model includes continuous measure of surgical skills. Standard errors computed with delta method.

Figure 2.13: Test for monotonicity of the instruments

Note: OLS regression for subsets of the population. Age above and below 55. CCI above and below 2. Ethnicity white and all other ethnicity. Coefficients estimated using logistic regression. Dependent variable is a binary indicator of whether the individual has been operated using the robot. Demographic controls are age, age squared, ethnicity and a rural urban indicator. Clinical controls are a set of ten comorbidity dummies. All models are estimated using year, month, and day of the week fixed effects.

logistic regression, for the subgroups of interest in Figure 2.13. Z_{dist} has always a negative coefficient indicating that increasing the relative distance to a robotic hospital weakly decreases patient's propensity to undergo robotic surgery regardless of the cell of patients demographics I focus on. Similarly, Z_{days} has always a positive coefficient when statistically significant. In all cases, the estimated effect of diagnosis timing on the choice of robotic surgery is the same, affecting positively the choice, suggesting that there are no defiers.

Finally, under Assumption 4, the instruments should generate sufficient variation across the observable characteristics to generate a propensity score $P(Z)$ with full common support. In Figure 2.14, I present the unconditional support jointly generated by the instruments and covariates. The instruments create a common support in the estimated propensity score that spans virtually the full unit interval. This is crucial to compute the treatment effect of the treated (ATT) and the treatment effect on the untreated (ATU).

Figure 2.14: Common support

Note: Unconditional support jointly generated by instruments and covariates. Covariates in the model include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest hospital is a teaching hospital, instrument Z_{dist} , and Z_{days} . Model controls for year-month and day of the week fixed effects. Model includes continuous measure of surgical skills.

2.7 Results

I will estimate the MTE using the local instrumental variable method introduced by Heckman and Vytlacil (1999). I estimate the selection equation (i.e., Equation 2.11) using a Probit regression model, from which I derive the propensity score \hat{p} . The model includes the two instruments, controls for distance to the closest hospital, and an indicator for whether the closest hospital to the patient is a teaching hospital. I present the variables included in X_i in Table 1. $Skills_j$ are alternatively added as a continuous variable or as a high skilled indicator (i.e., above the median of the distribution of skills). In all specifications, I include day of the week, month, and year fixed effects. I will model the outcomes both parametrically and non-parametrically (partially-linear) in terms of the unobserved term $K(p)$.⁷ Heckman et al. (2006) provide a detailed discussion of different estimation methods.

⁷I want to acknowledge that this can be easily done using Stata thanks to a command from Andresen (2018).

Skills and technological gains

In Table 2.5 to 2.7, I test three different specifications under the assumption of joint normality of the error terms. Table 2.5 presents the baseline model where the outcomes depend on distance to the closest hospital, an indicator for whether the closest hospital to the patient is a teaching hospital, and the patient characteristics presented in Table 1. The model includes year, month, and day of the week fixed effects all interacted with the propensity score. In Table 2.6, I add postal area fixed effects to control for time invariant differences across neighborhoods. In Table 2.7, I test the baseline specification on a restricted sample of surgeons for which I can observe at least 50 operations in the period pre-robots (2005-2007).

Column 1 provides the coefficient on skills for the selection equation (Equation 2.11). Column 2 and 3 present, respectively, the coefficients δ_0 and $\delta_1 - \delta_0$ estimated from Equation 2.14. Column 2 provides the estimates for log length of stay, and Column 3 for the adverse event indicator. The coefficient δ_0 speaks to the way skills affect patient outcomes when traditional surgery is used. The coefficient on skills interacted with the propensity score speaks to the level of heterogeneity in treatment effects that depends on the skills of the surgeon (i.e., $\delta_1 - \delta_0$). I test the model using either a continuous measure of skills or a binary variable that takes value 1 if the surgeon's skills are above the median of the distribution.

Under all model specifications, the coefficient on skills δ_0 is negative and statistically significant for both patient outcomes. With traditional surgery, high skilled surgeons' patients have better outcomes than the patients of lower skilled surgeons. This is not unexpected, as finding otherwise would have questioned the validity of my measure of skills. The coefficient interacted with the propensity score $\delta_1 - \delta_0$ is instead positive for both outcomes. Treatment effects from using the robot depend on the skills of the surgeon. For length of stay, the coefficient interacted with the propensity score is positive and statistically significant, suggesting that the treatment effect is stronger the lower the skills of the surgeon. Length of stay decreases from using the robot, but more significantly for lower skilled surgeons. The same

is true for the adverse event indicator, although the coefficient is not statistically significant in the baseline specification. In turn, these results suggest limited complementarities of the robot with high skilled surgeons.

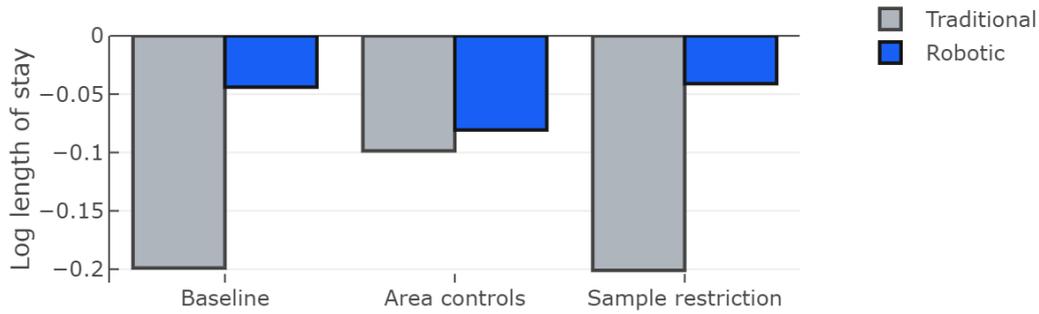
In Figure 2.15 and 2.16, I provide a graphical representation of the difference in performance between high and low skilled surgeons under traditional (the gray bars) and robotic surgery (the blue bars). For both outcomes, the difference between high and low skilled surgeons shrinks when using the robot. For example, in the model with area fixed effects, the patients of high skilled surgeons are 4 percentage points less likely to experience an adverse event from surgery. However, the treatment effect is almost five percentage points more negative for lower skilled surgeons. Actually, in some cases it appears that, with the robot, patients of low skilled surgeons are less likely to experience an adverse event from surgery relative to the patients of high skilled surgeons. This result points to an equalizing effect of the technology.

In the Appendix, I test the robustness of this result to a number of different model specifications. In Table 10, I show the coefficients on skills estimated under the inclusion of a measure of surgeons' experience. In Table 2.6, I use an alternative measure of surgeons' skills which I derive from a fixed effect model rather than the model presented in Section 2.4. Further, in Table 9 I test the baseline specification under the inclusion of dummies that indicate the year the hospital has adopted the robot. Lastly, I test a specification where I employ the instrument Z_{dist} as a binary indicator that takes value 1 if Z_{dist} is positive and takes value 0 otherwise.

Selection into robotic surgery

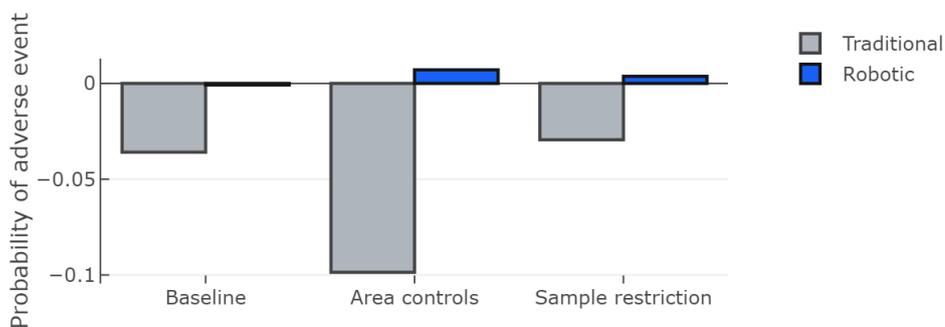
Comparing the coefficients from the selection equation (Column 1) to the estimates from the outcome equations (Column 2 and Column 3) allows identifying whether surgeons of different quality select based on their gains. This is not the case. Lower skilled surgeons have the largest gains from using the robot, but are also less likely to use it on any given patient. Hence, the estimates uncover a pattern of negative selection on gains.

Figure 2.15: Length of stay – high vs low skilled surgeons



Note: High skilled surgeons above the median of skills. Displays the value of δ_0 the coefficient on High Skilled indicator for the estimated outcome equation with dependent variable log length of stay (in gray). Displays in blue the value of δ_1 obtained by adding to the coefficient on High Skilled indicator * Propensity score $\delta_1 - \delta_0$ the estimated δ_0 . The baseline model controls for age, age squared, indicator for white ethnic profile, ten comorbidity dummies, distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator, year-month, and day of the week fixed effects all interacted with the propensity score. Instruments used to estimate the propensity score are Z_{dist} and Z_{days} . The model Area control includes postal area fixed effects not interacted with the propensity score. The model Sample restriction estimates the baseline specification using data from surgeons that are observed operating on at least fifty patients in the period 2005-2007.

Figure 2.16: Adverse event – high vs low skilled



Note: High skilled surgeons above the median of skills. Displays in gray the value of δ_0 the coefficient on High Skilled indicator for the estimated outcome equation with dependent variable indicator of adverse event. Displays in blue the value of δ_1 obtained by adding to the coefficient on High Skilled indicator * Propensity score $\delta_1 - \delta_0$ the estimated δ_0 . The baseline model controls for age, age squared, indicator for white ethnic profile, ten comorbidity dummies, distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator, year-month, and day of the week fixed effects all interacted with the propensity score. Instruments used to estimate the propensity score are Z_{dist} and Z_{days} . The model Area control includes postal area fixed effects not interacted with the propensity score. The model Sample restriction estimates the baseline specification using data from surgeons that are observed operating on at least fifty patients in the period 2005-2007.

Table 2.5: Heterogeneity in causal effects - Normal model

	(1) Selection equation	(2) Length of stay	(3) Adverse event
Continuous Skills			
Skills	0.368*** (0.017)	-0.319*** (0.011)	-0.0319*** (0.007)
Skills * Propensity score		0.272*** (0.026)	0.0265 (0.014)
Binary Skills			
High skilled	0.261*** (0.014)	-0.199*** (0.010)	-0.036*** (0.007)
High skilled * Propensity score		0.155*** (0.018)	0.035** (0.011)
Year-Month FE	Yes	Yes	Yes
Day of the week FE	Yes	Yes	Yes
<i>N</i>	50203	49215	50203

* Standard errors bootstrapped with 100 repetitions $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Column(1) dependent variable binary indicator of robotic surgery. Estimated using Probit regression model. Column (2) and (3), coefficients of regressors not interacted with the propensity score measure effects on the outcome in the untreated state (δ_0). Coefficients of regressors interacted with the propensity score measure effects the difference of the effects between the treated and the untreated state ($\delta_1 - \delta_0$). Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. The controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. Skills, a continuous variable, is measured using the standardized risk ratio for post-operative morbidity (deaths and readmissions) computed using 2005-2007 data. High skilled indicator takes value 1 if skills above the median of the distribution of skills. Instruments used to estimate the propensity score are Z_{dist} and Z_{days} . Estimation of coefficients under the assumption of normality of unobserved components.

In Column 1 of each table, I show the coefficients on skills from the estimated selection equation (i.e., Equation 2.11). The dependent variable is a binary indicator for whether the patient has been operated with robotic surgery. The results show that surgical skills are an important determinant of whether the patient is operated with the robot. The coefficient on skills is positive and statistically significant, and this is true using both skills as a continuous measure or the high-skilled indicator.

To illustrate the magnitude of this relationship, in Figure 2.17, I show graphically how the probability of using the robot depends on skills. These are the marginal effects at different levels of my measure of skills, which I have normalized to be

Table 2.6: Heterogeneity in causal effects – Normal model with local area fixed effects

	(1)	(2)	(3)
	Selection equation	Length of stay	Adverse event
Continuous Skills			
Skills	0.567*** (0.036)	-0.264*** (0.016)	-0.022* (0.009)
Skills * Propensity score		0.263*** (0.025)	0.049*** (0.012)
Binary Skills			
High skilled	0.121*** (0.027)	-0.099*** (0.015)	-0.040*** (0.007)
High skilled * Propensity score		0.018 (0.016)	0.047*** (0.010)
Year-Month FE	Yes	Yes	Yes
Day of the week FE	Yes	Yes	Yes
Area FE	Yes	Yes	Yes
<i>N</i>	48083	47139	48083

Standard errors bootstrapped with 100 repetitions $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Column(1) dependent variable binary indicator of robotic surgery. Estimated using Probit regression model. Column (2) and (3), coefficients of regressors not interacted with the propensity score measure effects on the outcome in the untreated state (δ_0). Coefficients of regressors interacted with the propensity score measure effects the difference of the effects between the treated and the untreated state ($\delta_1 - \delta_0$). Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. The controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. Skills, a continuous variable, is measured using the standardized risk ratio for post-operative morbidity (deaths and readmissions) computed using 2005-2007 data. High skilled indicator takes value 1 if skills above the median of the distribution of skills. Instruments used to estimate the propensity score are Z_{dist} and Z_{days} . Estimation of coefficients under the assumption of normality of unobserved components. Model estimated using postal area fixed effects, not interacted with the propensity score.

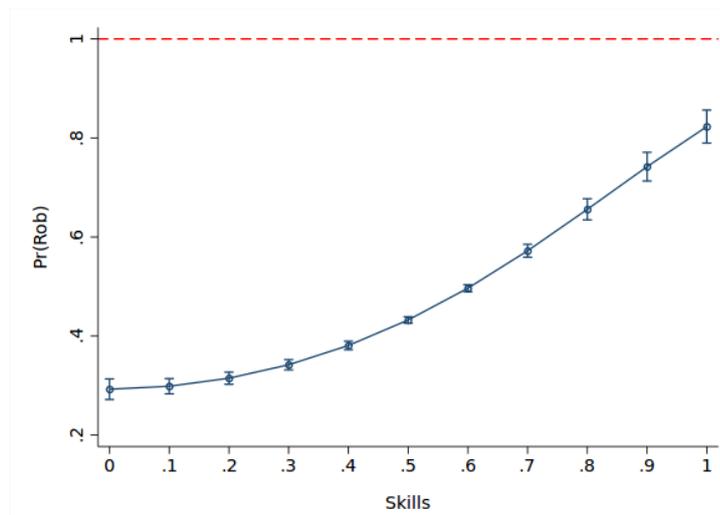
between 0 and 1. The rest of the covariates are held at their mean value. The figure shows that a patient whose surgeon is at the top of the distribution of skills will almost certainly be operated with the robot. On the other hand, a patient whose surgeon is at the bottom of the distribution will have 1 in 10 chances to be operated with it. For the high-skilled indicator, the value of the margin is the difference in the probability of using the robot between high and lower skilled surgeons. High-skilled surgeons' average predicted probability of using the robot is 0.58 while for the rest is 0.38, they are 30 percent more likely to use the robot on an average patient.

Table 2.7: Heterogeneity in causal effects - Normal model with sample restriction

	(1) Selection equation	(2) Length of stay	(3) Adverse event
Continuous Skills			
Skills	0.180*** (0.017)	-0.324*** (0.012)	-0.033*** (0.009)
Skills * Propensity score		0.352*** (0.025)	0.057*** (0.016)
Binary Skills			
High skilled	0.271*** (0.015)	-0.201*** (0.016)	-0.029** (0.010)
High skilled * Propensity score		0.160*** (0.027)	0.026 (0.015)
Year-Month FE	Yes	Yes	Yes
Day of the week FE	Yes	Yes	Yes
Area FE	Yes	Yes	Yes
<i>N</i>	48083	47139	48083

* Standard errors bootstrapped with 100 repetitions $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Column(1) dependent variable binary indicator of robotic surgery. Estimated using Probit regression model. Column (2) and (3), coefficients of regressors not interacted with the propensity score measure effects on the outcome in the untreated state (δ_0). Coefficients of regressors interacted with the propensity score measure effects the difference of the effects between the treated and the untreated state ($\delta_1 - \delta_0$). Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. The controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. Skills, a continuous variable, is measured using the standardized risk ratio for post-operative morbidity (deaths and readmissions) computed using 2005-2007 data. High skilled indicator takes value 1 if skills above the median of the distribution of skills. Instruments used to estimate the propensity score are Z_{dist} and Z_{days} . Estimation of coefficients under the assumption of normality of unobserved components. Sample is restricted to surgeons for which I observe at least 50 operations in the period pre-robots.

Generally, more complex patients appear to be less likely to be operated with robotic surgery. Patients that have a comorbidity, or are older, have a lower probability of getting the robotic approach, regardless of whether they are operated by a high or a lower skilled surgeon. However, high skilled surgeons use the robot more intensively for all patients. In Figure 2.18, I show how the predicted probability varies by age for surgeons above and below the median of skills. For both types of surgeons, the likelihood of using the robot diminishes with the age of the patient. But, at all age levels, high skilled surgeons are more likely to operate with the robot. The fact that lower skilled surgeons use the robot less intensively, conditional on patient characteristics, suggests they face a higher cost (actual or perceived) to use

Figure 2.17: Estimated probability of robotic approach by skill

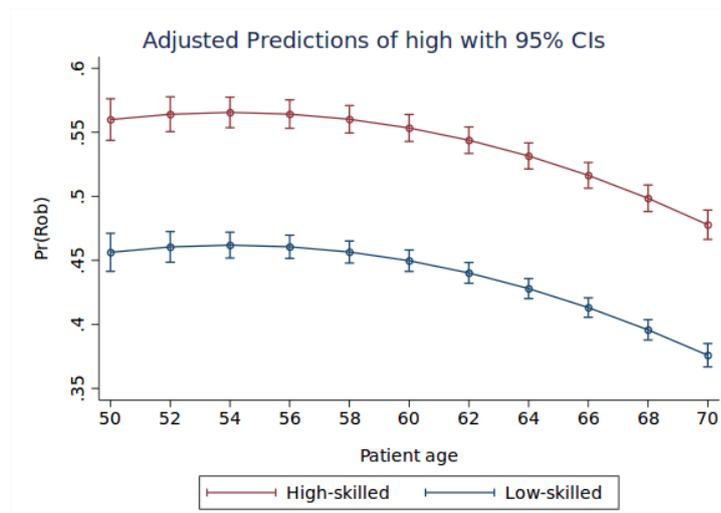
Note: Probit regression estimates, dependent variable robotic approach. Marginal probability estimated at values of skills measure. Covariates in the model at means, include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest hospital is a teaching hospital, instrument Z_{dist} , and Z_{days} . Includes a squared term for skills. Model controls for month-year and day of the week. Delta method for standard errors.

the technology (Chandra and Staiger, 2020; Suri, 2011).

Returns to treatment based on unobserved characteristics

Using the model parameters, I can estimate the MTE curve that relates the returns from using the robot to the unobserved resistance to treatment. As a first step, I estimate the $K(p)$ component parametrically under joint normality of the error terms. Under this assumption, the outcome and choice equation can be jointly estimated using the method of maximum likelihood (Carneiro et al., 2011). The estimated MTE under this assumption is shown in Figure 2.19.

The MTE curve mimics the pattern of negative selection found on observables. The relationship between the unobserved resistance to treatment V and the gains from treatment is consistently negative for the length of stay, and homogeneity can be rejected at all conventional levels of statistical significance. This implies that the patients most likely to undergo robotic surgery, based on their unobserved characteristics (which may include some characteristic of the surgeon), have the lowest returns from the treatment. The shape of MTE curve for the adverse event indicator

Figure 2.18: Estimated probability of robotic approach by age

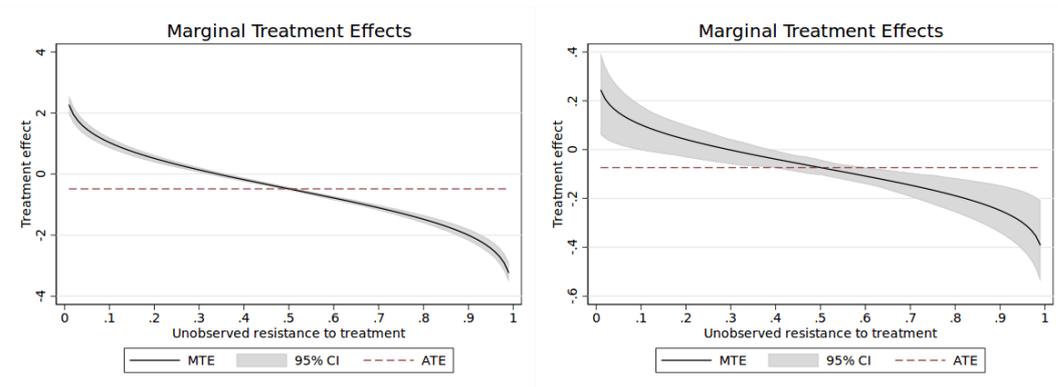
Note: Adjusted predictions with 95 per cent confidence interval. Probit regression estimates, dependent variable robotic approach. Marginal probability estimated at different value of patient age. Covariates in the model at means, include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest hospital is a teaching hospital, instrument Z_{dist} , and Z_{days} . Model controls for month-year and day of the week. High-skilled indicator takes value 1 if SRR above median of the distribution. Standard errors computed with delta method.

suggests a similar story, but we can't reject homogeneity on unobservable characteristics.

In Figure 2.20, I relax the assumption of joint normality and let the function $K(p)$ be approximated by a polynomial in p . Estimation in this case is achieved by a two-step procedure discussed in Heckman et al. (2006). For length of stay, the results are almost unchanged and the shape is remarkably similar to what described earlier. For the probability of adverse event, however, we are able to get more precise estimates under which we can exclude homogeneous effects.

Lastly, I estimate $E(Y|P(Z) = p)$ semi-parametrically and compute its derivative with respect to p . The parameters in this case are estimated from a partial linear regression of Y on X and $P(Z)$, and the estimation of $K(p)$ is achieved by a local polynomial regression. Still, the MTE curve suggests negative selection for length of stay and the adverse event indicator.

Figure 2.19: MTE curve – Normal

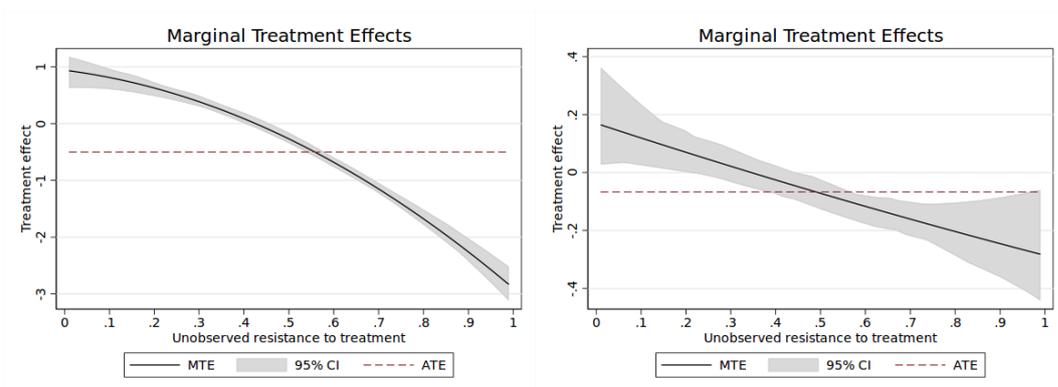


(a) Length of stay

(b) Adverse event

Note: Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on log length of stay (a) and probability of adverse event (b). The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p , parametrically under the assumption of $K(p)$ is normal. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon’s skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions.

Figure 2.20: MTE curve – Polynomial

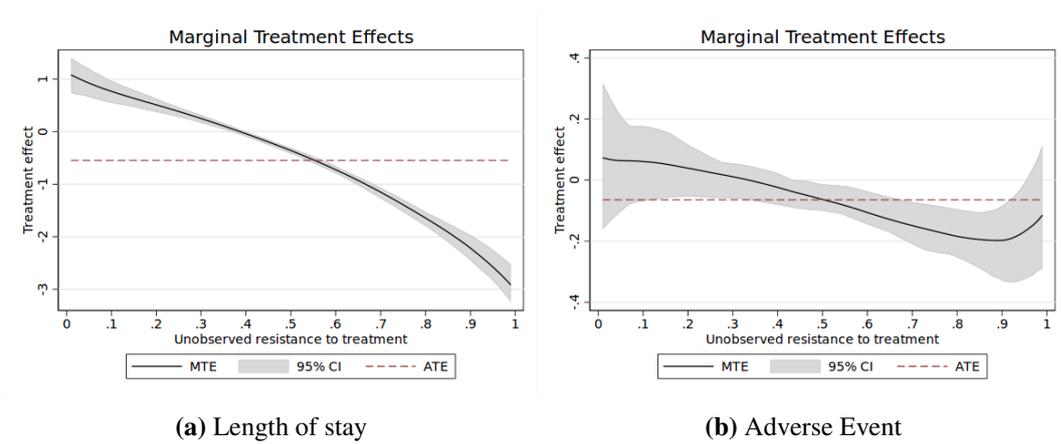


(a) Length of stay

(b) Adverse Event

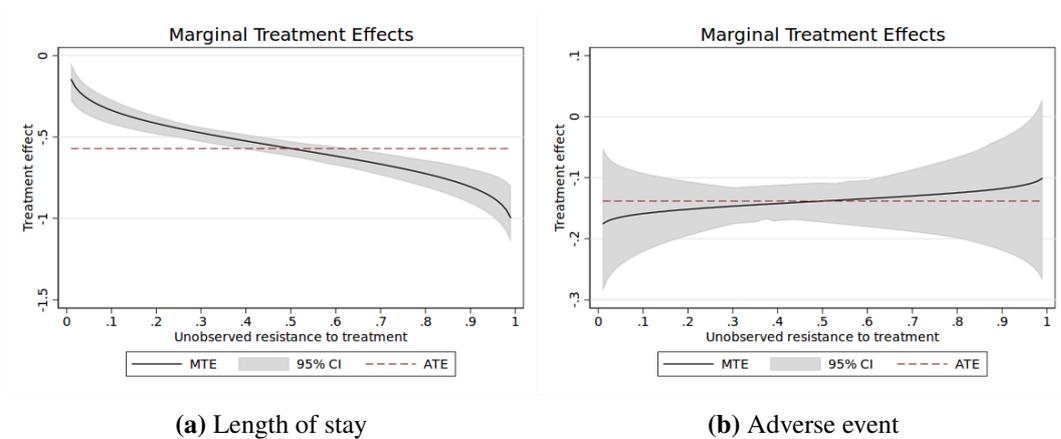
Note: Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on log length of stay (a) and probability of adverse event (b). The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p , parametrically under the assumption of $K(p)$ is a polynomial of degree 2. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon’s skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions

Figure 2.21: MTE curve – Semiparametric



Note: Includes area fixed effects not interacted with the propensity score. Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on log length of stay (a) and probability of adverse event (b). The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p , semi-parametrically. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon's skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions.

Figure 2.22: MTE curve – Normal with area fixed effects



Note: Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on log length of stay (a) and probability of adverse event (b). The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p , parametrically under the assumption of $K(p)$ is normal. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon's skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions. Include area fixed effects (not interacted with propensity score).

Conventional treatment effects and policy simulation

In Table 3.1 and Table 2.9, I show the treatment effects parameters, which I compute by appropriately integrating over the MTE curve. The robot improves significantly the performance of surgeons. The effect of using the robot is negative and statistically significant. The ATE is always negative regardless of the specification. This means that the robot on average improves surgical performance. The robot reduces length of stay and the probability that the patient experiences an adverse event from surgery. Consistent with the pattern of selection I have uncovered, the average treatment effect on the untreated (ATU) is more negative than the effect on the treated (ATT). In some cases actually the ATT is positive, indicating that these patients would have been better off with traditional surgery. The patients that would benefit the most from being operated with the robot are the untreated group. Notice that Mogstad et al. (2021) show that with more than one instrument, the monotonicity condition required for identification of the LATE can only be satisfied if choice behavior is effectively homogeneous.

As a conclusive exercise, I exploit the structure of the model to conduct a policy simulation. Following Heckman and Vytlacil (2005); Carneiro et al. (2011), I consider a class of policies that change $P(Z)$, the probability that the patient is operated with the robot, but that do not affect the potential outcomes or the unobservable characteristics in the model. Heckman and Vytlacil (2005) show how to compute the Policy Relevant Treatment Effect (PRTE) which is the mean effect from going to the baseline policy to an alternative policy per net person shifted in to treatment.

I compute this parameter for a counterfactual scenario in which I assign to lower skilled surgeons the same probability of using the robot as high skilled surgeons. Basically, I evaluate effects if lower skilled surgeons were mandated to use the robot with the same intensity as high skilled ones. This policy simulation speaks to a hypothetical counterfactual scenario in which the costs or barriers that limit the use of the robot by lower skilled surgeons were lifted. For example, suppose that lower skilled surgeons use the robot less because they have *fewer* of them. Then,

Table 2.8: Length of stay - Conventional estimates

	(1) Normal	(2) Normal FE	(3) Polynomial	(4) Semiparametric
ATE	-0.483*** (0.026)	-0.571*** (0.026)	-0.501*** (0.033)	-0.546*** (0.033)
ATT	0.649*** (0.067)	-0.408*** (0.026)	0.619*** (0.057)	0.525*** (0.062)
ATUT	-1.570*** (0.067)	-0.727*** (0.044)	-1.578*** (0.069)	-1.576*** (0.075)
LATE	-0.343*** (0.024)	-0.596*** (0.024)	-0.366*** (0.023)	-0.414*** (0.027)
Year-Month	Yes	Yes	Yes	Yes
Day of the week	Yes	Yea	Yes	Yes
Area FE	No	Yes	No	No
<i>N</i>	49215	47139	49215	49215

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors bootstrapped with 100 repetitions. The dependent variable is the logarithm of post-operative length of stay. Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. Area controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. All specifications are estimated using the instruments Z_{dist} and Z_{days} and include the continuous measure of surgeon's skills.

this policy counterfactual shows what would happen to the average treatment effect if lower skilled surgeons had the same number of robots as high skilled surgeons. In a different vein, suppose that lower skilled surgeons dislike the robot and that's why they use it less intensively than high skilled surgeons. In this case, the policy counterfactual speaks to a situation in which the lower skilled surgeons liked the robot as much as the high skilled surgeons. The results of this exercise are shown in Figure 2.23 for both margins of performance. The PRTE is always more negative than the ATE indicating that inducing lower skilled surgeons to use the robot more intensively would generate larger gain from the adoption of robots.

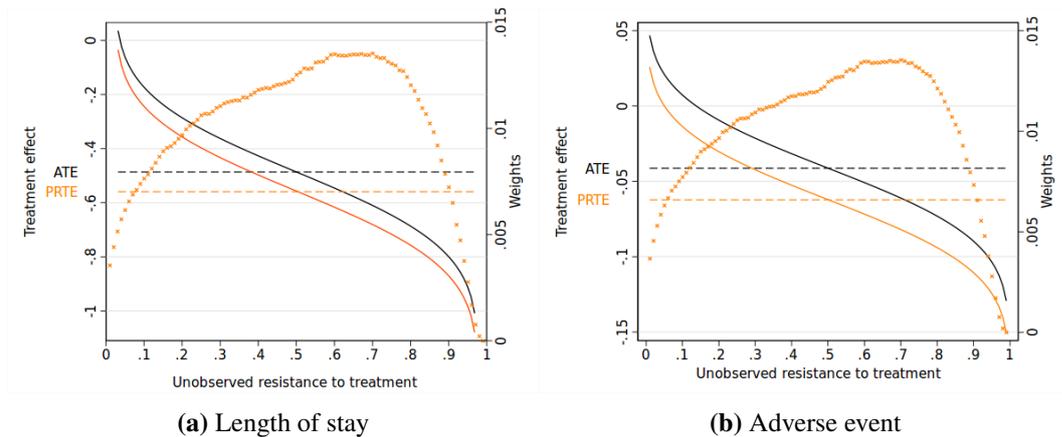
Table 2.9: Adverse Event – Conventional estimates

	(1)	(2)	(3)	(4)
	Normal	Normal FE	Polynomial	Semiparametric
ATE	-0.073*** (0.015)	-0.138*** (0.019)	-0.067*** (0.017)	-0.065** (0.0211)
ATT	0.0623 (0.032)	-0.144*** (0.019)	0.0718* (0.033)	0.043 (0.039)
ATUT	-0.205*** (0.038)	-0.133*** (0.039)	-0.202*** (0.043)	-0.169*** (0.049)
LATE	-0.067*** (0.012)	-0.136*** (0.016)	-0.065*** (0.012)	-0.0713*** (0.015)
Year-Month	Yes	Yes	Yes	Yes
Day of the week	Yes	Yes	Yes	Yes
Area FE	No	Yes	No	No
<i>N</i>	49215	47139	49215	49215

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors bootstrapped with 100 repetitions. The dependent variable is the logarithm of post-operative length of stay. Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. Area controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. All specifications are estimated using the instruments Z_{dist} and Z_{days} and include a continuous measure of surgeon's skills pre-robot. Skills are measured using the SRR.

2.8 Conclusive remarks

This paper shows that thinking of innovations in abstraction from the characteristics of their users limits our view of what technologies can achieve. Using the case of robots in surgery, I showed that new technologies might help reduce variation in workers' performance. This is a significant finding in healthcare, where disparities in access and quality are a central concern of regulators and policymakers. Nevertheless, it can be applied to any context where service delivery should be of consistent quality regardless of the individual in charge. The adoption of robots in surgery has been criticized because the literature, so far, has not reached a conclusive agreement on whether robots improve the outcomes of patients relative to traditional surgery. I show that outcomes improve by using the robot, but also that robots have the potential to reduce variation in patient outcomes arising from heterogeneity in surgeons' skills. I have shown that the robot helps lower skilled

Figure 2.23: Policy simulations - MTE and PRTE

Note: Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on log length of stay (a) and probability of adverse event (b). The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Unobserved heterogeneity, modeled as a function of the propensity score, p , parametrically under the assumption of $K(p)$ is normal. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control for age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon's skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions. In orange the estimated effects from policy simulation. Crosses indicate the weights.

surgeons perform almost as well as high skilled surgeons. However, my analysis suggests that lower skilled surgeons may face a higher cost of using the robot. Although, they have the highest gains, they are less likely to use the robot on any given patient. More research is needed to identify the reason lower skilled surgeons use the robot less than their high skilled colleagues. Policies that encourage the adoption of these technologies may be welfare enhancing.

Chapter 3

Killer Deals? The Impact of Hospital Mergers on Clinical Quality

Joint work with Thomas P. Hoe

3.1 Introduction

Industry consolidation has often raised concern among economists. These issues have come to the fore again lately, with recent research showing the extent of consolidation globally (Loecker and Eeckhout, 2018) and other work highlighting the limited benefits (Grullon et al., 2019) or negative impacts (Gutierrez Gallardo and Philippon, 2017) associated with mergers. These concerns have been especially acute in the hospital market where there has been significant consolidation (Cutler and Morton, 2013). Mergers in this setting have been shown to increase hospital prices, which feed through to higher insurance premiums (Dafny and Lee, 2019; Schmitt, 2018), and to slow wage growth in local labor markets (Prager and Schmitt, 2019). While hospital mergers have also been found to lead to reduced costs (Schmitt, 2017), other work has shown that these benefits are sometimes limited and can be outweighed by other transaction costs (Craig et al., 2019).

This paper provides new evidence on the impact of hospital mergers on clinical quality of care. These outcomes are notoriously difficult to measure, but there are clear antitrust concerns. Increasing market power may be exercised through price

increases or a deterioration in product quality, and the latter can be a matter of life or death in hospital markets. These concerns have been raised on multiple occasions, and antitrust authorities have blocked proposed mergers on this basis.¹

We use the universe of public hospital medical records in England to examine all public hospital mergers after the introduction of hospital choice in 2006. There were 159 hospital sites involved in mergers over our sample period, comprising 13 transactions. An advantage of our setting is that hospital reimbursement is unaffected by mergers—hospital services are free at the point of care and hospital payments are made and set centrally by the government—which allows us to study the impact of mergers on quality of care in isolation from other changes.

We use an event-study framework and focus on two measures of clinical quality: 30-day unplanned readmissions and 30-day in-hospital mortality. To account for the fact that hospitals may treat different patient populations before and after mergers, we risk adjust our quality measures following the latest methodologies developed for the Center for Medicare and Medicaid Services (Leora, 2012). These methods use a series of hierarchical logistic regression models, which control for a variety of observable patient characteristics, to compute standardized risk ratios (SRRs). These ratios report the number of ‘excess deaths’ (or excess readmissions) at a particular hospital relative to the expected number of deaths nationally for a comparable patient pool.

Our analysis tracks SRRs for the two years prior to and following a hospital merger. We perform this analysis for all acute hospital sites involved in a merger. The event study estimates indicate how clinical quality at these hospitals, as measured

¹A debated case in the UK was for example the merger proposal between The Royal Bournemouth and Christchurch Hospitals NHS Foundation Trust and Poole Hospital Foundation Trust. The Competition Commission prohibited the merger in a decision in 2013 on the basis of competition concerns and the absence of demonstrated benefits (Schiraldi, 2019). Similarly, in 2014, the U.S. Court of Appeals for the Sixth Circuit supported the FTC decision to block a hospital merger in Toledo, Ohio, between the area’s most extensive healthcare system, ProMedica, and one of its rivals, St. Luke’s Hospital. The FTC showed that health plans could obtain more competitive prices from ProMedica when St. Luke’s existed as a suitable alternative. The loss of that alternative would have increased ProMedica’s negotiating clout and left health plans vulnerable to ProMedica’s price demands (Ramirez, 2014).

by excess deaths and readmissions, evolved relative to national trends.

We find evidence that hospital mergers have immediate and persistent negative impacts on the quality of care. Risk-adjusted mortality is flat in the pre-merger period, suggesting that mergers are not a case of poorly performing hospitals being acquired. In the post-merger period, mortality increases in the month after a merger and remains above baseline levels for the next two years. There is a similar pattern with readmissions, with some evidence of an increase two months prior to a merger.

As a placebo test, we compute equivalent estimates for hospital sites not involved in mergers. The event study estimates for these hospitals are flat in pre- and post-merger periods.

The magnitude of the estimated merger effects is substantive. On average, the mergers we study increased the likelihood of a patient dying by 0.4 percentage points, or 27 percent, relative to the baseline mortality rate of 1.4 percent. Similarly, the likelihood of readmission increased by 0.9 percentage points (11 percent relative to a baseline of 8.3 percent).

Scaling these effects up to the patient population suggests that mergers led to approximately 98 additional deaths and 218 additional readmissions per year. Under conservative assumptions on the value of a life-year, the additional deaths are valued at around £11 million, which is approximately 4 percent of average annual hospital costs. This is the same order of magnitude as previous estimates of cost synergies from mergers in other settings (Schmitt, 2017; Craig et al., 2019).

These results illustrate that mergers can have severe consequences for the quality of care. These effects could plausibly outweigh any cost savings from mergers. As we develop this work further, we intend to explore the incentives and mechanisms leading to the quality deterioration. In particular, we will test whether the impacts are associated with changes in market power (e.g., HHI variations) or specific decisions taken by hospitals in the post-merger period (e.g., internal department reorganizations). We also plan to address several methodological issues, such as how we

compute standard errors and the choice of covariates in the risk adjustment methodology.

Our work contributes to the literature on the effects of mergers in the healthcare setting. The large majority of these studies analyze mergers between hospitals in the US market and focus on the effect of consolidation on prices (Lewis and Pflum, 2017), wages (Prager and Schmitt, 2019), and/or on the impact on hospital-insurer negotiations (Dafny et al., 2019). Fewer studies investigate whether hospital mergers and consolidation have an impact on quality. Ho and Hamilton (2000) compares the quality of hospital care before and after mergers in California between 1992 and 1995. Quality is measured using mortality for heart attack and stroke patients, 90-day readmission for heart attack patients, and discharge within 48 hours for normal newborn babies. Propper et al. (2012) examines the impact of public hospital mergers in England over the period 1997 to 2006 on a number of outcomes including emergency heart attack mortality. Both of these studies find little to no impact on the quality of care. The most robust evidence linking market structure to quality of care is Cooper et al. (2011) and Gaynor et al. (2013). These studies evaluate the impact of introducing patient choice in hospital markets in England in 2006, finding that mortality rates for heart attack patients were reduced by more in areas where hospitals were exposed to more competition in comparison to areas with less competition.

This paper complements the existing literature by directly analyzing mergers – i.e. changes in market structure – and focusing on hospital-wide measures of clinical quality. The rich administrative data that we have available makes this possible since we observe the entire population of patients over a long time period and have detailed information on patient characteristics to construct state-of-the-art risk adjusted quality measures. While this result is in line with standard predictions from industrial organization, to the best of our knowledge, this is the first paper providing empirical evidence that mergers negatively impact clinical quality. Although more research is needed, our conclusion has potential implications for antitrust policy in healthcare markets.

The remainder of this paper is organized as follows. Section 2 describes the data. Section 3 outlines our empirical specifications. Section 4 presents our results. Section 5 sets out a framework for analyzing how our results may relate to changes in market competition. Section 6 concludes.

3.2 Data

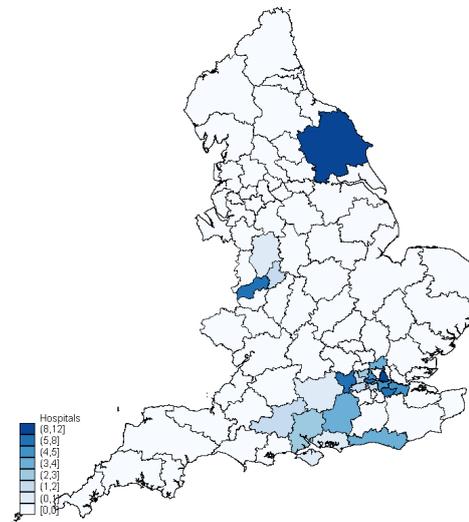
Our primary source of data is the Hospital Episodes Statistics (HES), a comprehensive administrative database containing all visits to public hospitals in England. We combine HES data with information from NHS Digital about the opening and closing dates of all public hospitals. Together these datasets allow us to identify hospital mergers and observe the treatment and health outcomes of all patients in England.

Identifying Hospital Mergers

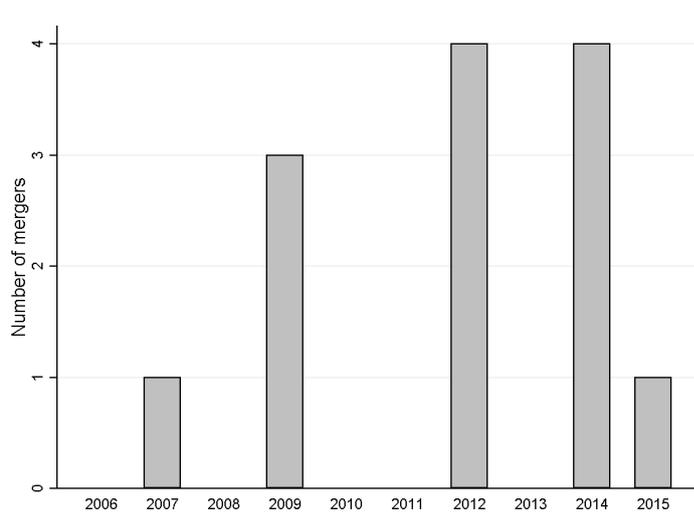
We use NHS Digital data on openings and closures of hospital sites to identify hospital mergers and validate our final list using a variety of other sources of information. We restrict our focus to mergers occurring after 2006, the year marking the introduction of patient choice in the NHS and the prospective payment system. Under the prospective payment system hospitals are reimbursed through fixed prices set and paid centrally by the government (Gaynor et al., 2013).

We identify 25 hospital mergers occurring during the period 2006-2015. We limit our analysis to acute hospitals, the leading providers of hospital-based services in England (Gaynor et al., 2013), resulting in a total of 16 mergers. From those, we exclude two mergers involving hospitals already involved in a previous merger.

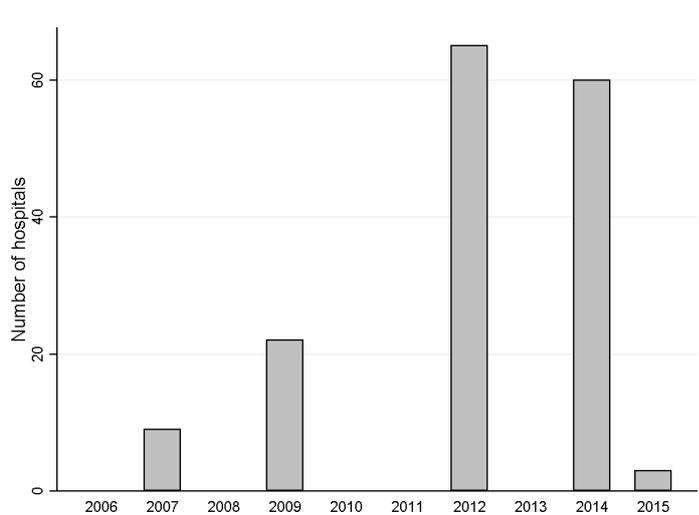
Mergers are somewhat concentrated across time and space. Geographically, mergers occurred in three main areas: London and the south, the northwest near Liverpool and Manchester, and the northeast near Newcastle, see Figure 3.1. The bulk of mergers occurred between 2012 and 2014. Figure 3.2 shows the number of mergers occurring between 2006 and 2015. 3.3 shows the number of hospitals involved in a merger in the period 2006-2015.

Figure 3.1: Hospital sites involved in mergers, 2006-2015

Notes: The figure shows the location of acute hospital sites involved in mergers in England between 2006 and 2015. We display the distribution of hospitals by postal area. Postcode areas are used by Royal Mail for the purposes of directing mail within the United Kingdom. The postcode area is the largest geographical unit used and forms the initial characters of the alphanumeric UK postcode.

Figure 3.2: Number of NHS mergers 2006-2015

Notes: The Figure shows the number of mergers between Acute NHS Trusts in England between 2006-2015.

Figure 3.3: Number of NHS sites involved in mergers 2006-2015

Notes: The Figure shows the number of hospital sites involved in mergers in England between 2006-2015.

Hospital Episodes Statistics

The HES data covers the universe of inpatient discharges from NHS hospitals in England, comprising over 9 million admissions per year. The data includes detailed patient clinical information (e.g., diagnoses, operations), patient demographic information, administrative information (e.g., methods of admission and discharge), and geographical information such as where patients are treated and where they live.

We apply three selection rules to the data before conducting the analysis. First, for each year between 2006 and 2015 we select a 50% random sample of patients admitted to the hospital. Second, we select only hospitals that belong to an acute trust.² Third, we drop from our sample any hospital for which the data required for the analysis is not available.

We divide hospitals into a treatment and a control group. The treatment group is composed by hospitals that have been involved in a merger between 2006 and 2015, with the remaining hospitals making up the control group. We have 1,199 and 139 hospitals in the control and treatment groups, respectively. Table 3.1 shows the characteristics of the treatment and control group at the trust and hospital (i.e.,

²We define as *acute trust* any NHS trust that in the Estates Returns Information Collection data has had one site recorded as "General and Acute" from 1999 onwards.

site) level. On average, control-group trusts have a more significant number of hospital sites compared to the treatment group. However, once we restrict to acute trusts, the gap disappears. Control trusts also have, on average, a more significant number of occupied and available beds and many employees. The average site in the control group has relatively fewer patients than the average site in the treatment group, although this difference is not statistically significant.

Table 3.1: Balancing of characteristics treatment and control group

	Treatment	Control	P-value	N-difference
<i>Panel A: Trust characteristics</i>				
Total number of sites	4.5 (5.051)	8.311 (11.76)	[0.000]	0.298
Acute sites	2.088 (0.794)	1.959 (1.236)	[0.210]	0.087
Available beds	696.7 (265.5)	937.7 (435.7)	[0.000]	0.472
Occupied beds	617.6 (257.9)	746.9 (399.4)	[0.000]	0.272
Total staff	4539.7 (2472.2)	5105.4 (2552.5)	[0.057]	0.159
<i>Panel B: Hospital characteristics</i>				
Admissions per month	277.4 (328.8)	275.4 (366.3)	[0.954]	0.004
Raw mortality rate	0.018 (0.070)	0.028 (0.109)	[0.003]	0.079
Raw readmission rate	0.137 (0.250)	0.081 (0.146)	[0.020]	0.194

Notes: Column 3 reports p-values from a test of equality of means carried out by OLS regression of each characteristic on a dummy for assignment to treatment. Standard errors are clustered at the site level. Column 4 reports normalized differences are computed following Imbens and Wooldridge (2009). The sample comprises acute sites in England in the period 2006-2015. The control group includes all acute sites that have not been involved in a merger in the period of interest; the control group comprises all acute sites that have been involved in a merger in the period of interest. The number of sites, acute sites, available beds, occupied beds and total staff is at the year level and retrieved from the ERIC data. The number of patients, readmission rates and mortality rates are computed at the monthly level from the HES data.

3.3 Empirical Methodology

We use an event-study framework to study the impact of hospital mergers on risk-adjusted measures of clinical quality. In this section, we describe our methodology for computing quality and then set out the event-study specification.

Measures of Clinical Quality

To measure clinical quality we use standard risk-adjustment methodologies employed by the Center for Medicare and Medicaid Services (CMS). Specifically, we compute two measures: the Hospital Wide Risk Standardized Mortality measure (henceforth HWM) and Hospital Wide Risk Standardized Readmission measure (henceforth HWR) developed for CMS by the Yale New Haven Health System / Center for Outcomes Research and Evaluation (YNHHS/CORE).

The methods that underpin the HWM and HWR are routinely used by the US federal government to evaluate and disseminate information on quality of care at US hospitals. Most notably, the Hospital Readmission Reduction Program one of the largest pay-for-performance policies in health care globally uses similar measures to financially penalise hospitals with excess readmissions (Gupta, 2017). Similarly, risk-adjusted mortality measures are publicized through hospital report cards that allow patients compare the quality of participating hospitals (Kolstad, 2013b).

The HWM measure is a single hospital level summary score that reports the risk-standardized rate of deaths within 30 days of hospital discharge for any condition. Mortality is an unwanted outcome for the majority of patients and when assessed among appropriate individuals, it provides a concrete signal of quality by capturing the result of care processes as well as the impact of both optimal care and adverse events. The measure is computed using a sample of patients for which survival was most likely the primary goal when entering the hospital, and for which improved care quality could have been reasonably expected to impact the chance of survival. The measure is meant to cover all deaths of patients admitted to hospital that died either while in hospital or within 30 days of discharge. From our data, we cannot

observe deaths that occur outside the hospital. For this reason, we limit the measure to cover only deaths of patients admitted to acute sites that occur in hospital. Because overall mortality rate for patients admitted more than once is higher than for those with only one admission, the measure randomly selects one admission for each patient in the measurement period. Random selection is meant to reflect that the outcome of an admission can be either survival or death.

The HWR measure is a single hospital level summary score that reports the risk-standardized rate of unplanned readmission within 30 days of hospital discharge for any condition. A readmission is an admission to an acute care hospital within 30 days of discharge from an acute care hospital. Readmissions may be planned or unplanned. A planned readmission is intentional and scheduled as part of the patient's plan of care. Unplanned readmissions are acute clinical events experienced by a patient that require urgent hospital admission. Hospital readmissions, of this kind, are disruptive to patients and costly to the healthcare system. Higher than expected unplanned readmission rates suggest lower quality of care and are the focus of quality measurement. Because planned readmissions are not a signal of quality of care, these patient outcomes are excluded from the measure.

Following the CMS methodologies, we compute HWM and HWR in two steps. In the first step we estimate a regression model that risk adjusts the relevant outcome, i.e. unplanned readmission or mortality. In the second step we use the first-stage regression estimates to compute specialty-level Standardized Risk Ratios (SRR) which are then aggregated to produce the HWM or HWR.

Risk-adjustment Model

CMS estimates the following hierarchical logistic regression separately for each specialty cohort c and year t

$$\Pr(y_{icht} = 1 | x_i, c, h, t) = F(\alpha_{ct} + \beta_{cht} + \gamma_{ct}x_i + \varepsilon_{icht}) \quad (3.1)$$

$$\beta_{cht} \sim \mathcal{N}(0, \theta_{ct}^2) \quad (3.2)$$

$$\varepsilon_{icht} \sim \mathcal{N}(0, \sigma_{ct}^2), \quad (3.3)$$

where $F(\cdot)$ is the logistic function and y_{icht} is a binary variable indicating the readmission or death of patient i in specialty cohort c at hospital h in year t . The remaining terms are x_i a vector of (assumed exogenous) patient-level characteristics, α_{ct} the intercept, β_{cht} a hospital-specific fixed effect, and ε_{icht} an error term capturing any over- or under-dispersion.

The model is estimated separately for each specialty cohort. A cohort is a group of discharge conditions or procedure categories typically cared for by the same team of specialists. In line with CMS, we estimate the model for five specialty cohorts for the HWR measure and for fifteen specialty cohorts for the HWM measure.³ The vector x_i is common across specialty cohorts and includes age, a set of comorbidity indicators, and a set of diagnosis fixed effects.⁴

For each estimation, we use the baseline sample described in section 3.2 and then follow CMS by making a number of nuanced data exclusions that are specific to the HWM or HWR measures.⁵

³For the HWM measure, we group non-surgical patients into cancer, cardiac, gastrointestinal, infectious disease, orthopedics, pulmonary, renal, and other conditions categories, while we group surgical patients into cancer, cardiothoracic, general surgery, neurosurgery, orthopedics, and other surgical procedures categories. For the HWR measure, we group non-surgical patients into medical, neurology, cardiovascular, and cardiorespiratory categories, while we group surgical patients in the single category surgery.

⁴The comorbidity indicators are as follows: for the HWR measure, the model controls for 31 risk variables that encompass 74 comorbidity conditions categories.; and for the HWM measure, the model controls for 19 risk variables that encompass 38 comorbidity categories. We use the 12 months prior to the admission to identify the patients' comorbid conditions. The diagnosis fixed effects are included only for those categories with more than 1,000 admissions.

⁵The HWM sample excludes the following patients: those who have died in hospital, those that have been transferred between hospitals during the admission, those who have been admitted

Constructing Standardized Risk Ratios

Standardized Risk Ratios (SRRs) are a function of the ‘predicted’ and ‘expected’ outcomes (i.e., mortality and readmission) for each specialty-hospital combination. The expected outcome is the number of deaths, or readmissions, that would occur if a particular set of patients were treated by an average hospital (i.e., the national average expected performance). The predicted outcome is the equivalent number for a *specific* hospital. We compute these terms at the specialty-hospital-month level as follows

$$\text{predicted}_{chm} = \sum_{i \in c,h,m} F(\hat{\alpha}_{ct} + \hat{\beta}_{cht} + \hat{\gamma}_{ct}x_i) \quad (3.4)$$

$$\text{expected}_{chm} = \sum_{i \in c,h,m} F(\hat{\alpha}_{ct} + \hat{\beta}_{ct}x_i), \quad (3.5)$$

where m is a month within year t . So while Equation (3.1) is estimated annually, we use these estimates to construct SSRs which are measured at a monthly frequency.⁶

We then compute the specialty-level SRR as follows

$$\text{SRR}_{chm} = \frac{\text{predicted}_{chm}}{\text{expected}_{chm}}. \quad (3.6)$$

An SRR of 1 indicates that the number of deaths (or readmissions) in specialty c at hospital h in month m are in line with the number of deaths expected nationally at hospitals treating similar patients during that year. An SRR above (below) 1 indicates that the hospital is under- (over-) performing relative to the national average.

Finally, a hospital-wide SRR is computed as the geometric mean of specialty-

for psychiatric and metastatic cancer diagnosis, rehabilitation treatments, and those that have been enrolled in a hospice in the 12 months prior. The HWR sample excludes the following patients: those who have died in hospital, that have been transferred between hospitals during the admission; that have been discharged against medical advice, who have been admitted for psychiatric and cancer diagnosis or rehabilitation treatments.

⁶Ideally we would run the risk adjustment regression monthly, however the sample sizes do not allow us to do this with sufficient precision.

specific SRRs

$$\text{SRR}_{hm} = \exp \left\{ \frac{\sum_c w_{chm} \log(\text{SRR}_{chm})}{\sum_c w_{chm}} \right\} \quad (3.7)$$

where the weights, w_{chm} , correspond to the monthly volume of admissions in each specialty cohort.

We run the risk-adjustment estimation and SRR computations for mortality and readmission separately, giving us two measures of SRR_{hm} for each hospital-month in our data. We refer to these aggregated variables as HWM and HWR for simplicity.

Table 3.2 shows summary statistics for the two measures over the period of interest. The average HWR is larger than one, indicating a higher-than-expected rate of readmissions. The average HWM is less than one, indicating a lower-than-expected rate of in-hospital deaths. Some hospitals appear to be outliers with the value of the measure greatly exceeding the rest of the observations.

Table 3.2: Summary statistics for quality measures

	Mean	SD	Min	Max
Hospital Wide Readmissions	1.050	0.505	0.242	12.764
Hospital Wide Mortality	0.831	1.004	0.019	34.861

Notes: The Table displays summary statistics for the risk adjusted hospital mortality and readmissions measure. The sample over which they are calculated includes all hospitals involved in a merger that belong to a trust providing acute-care services in England. The level of observation is the hospital in a given month-year.

Event Study Framework

We use the following regression specification to estimate the impact of mergers on hospital quality over time

$$y_{hs} = \mu_s + \sum_{k=-24}^{-11} \kappa_k + \sum_{k=-13}^{24} \kappa_k + u_{th} \quad (3.8)$$

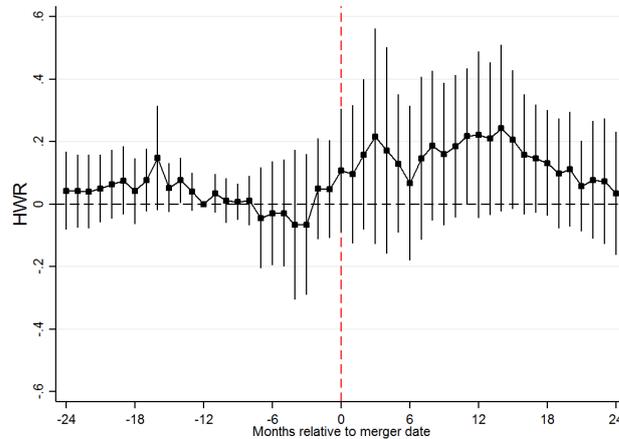
where y_{hs} is one of the two measures of hospital quality, HWM or HWR, in period s for hospital h . μ_s are calendar time fixed effects, κ_k are event time fixed effects, and u_{hs} is an error term.

We estimate Equation (3.8) using hospitals involved in mergers as described in section 3.2. The parameters of interest are the κ_k , which measure the outcome at event time k relative to the base category which we set at 12 months prior to the merger. The κ_k parameters for periods prior to the merger ($k < 0$) allow us to test for pre-trends or impacts on quality that occur before the merger is formally completed. We are primarily interested in κ_k parameters for periods after the merger ($k > 0$). We anticipate that $\kappa_k = 0$ for $k < 0$ (i.e. no pre-trends) and we test the null hypothesis that $\kappa_k = 0$ for $k > 0$, meaning mergers have no impact on hospital quality. Since our dependent variables HWM and HWR are measured relative to the expected national average, if we find that $\kappa_k > 0$ in the post-merger period this implies that clinical quality at merged hospitals has deteriorated relative to hospitals that did not merge.

As a placebo test, we also estimate Equation (3.8) for hospitals not involved in mergers and impose artificial mergers at the dates of the actual mergers in our data. In this case, we anticipate that $\kappa_k = 0$ for all k .

We estimate Equation (3.8) using OLS and cluster standard errors by hospital site. In this specification, we treat the hospital quality measures as fixed, which will understate the standard errors because it omits any estimation error from the risk-adjustment exercise. In future work, we plan to bootstrap the standard errors to account for the two-stage nature of the estimation.

Figure 3.4: Event study estimates of the impact of hospital mergers on unplanned readmissions



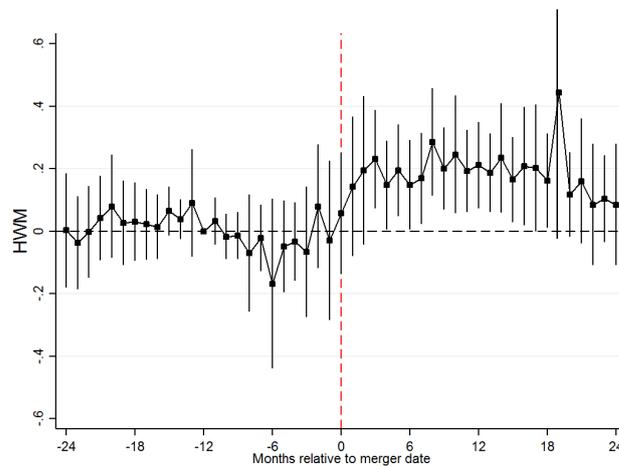
Notes: The Figure plots event study estimates and corresponding 95 percent confidence bands. The dependent variable is the Hospital Wide Risk Standardized Readmission Measure - HWR. The unit of observation is a hospital site. Estimation from OLS regression with standard errors clustered at the hospital level. The regression includes calendar time fixed effects. Sample comprises all hospitals involved in a merger that belong to a trust providing acute-care services in England.

3.4 Results

Baseline Results

We start our analysis by plotting the event study estimates from equation (3.8) for our two measures of quality in Figure 3.4 and 3.5. The coefficients measure the difference in care quality relative to the baseline period for the hospitals involved in a merger. We include calendar time fixed effects that control for changes in quality common to all hospitals. Standard errors are clustered at the hospital level. We find a negative and significant effect of mergers on the quality of care, with both measures displaying an increase in the period after the merger (i.e., a reduction in the quality of hospital care).

Our risk-adjusted mortality measure increases in the month following a merger and remains above the baseline level for the following two years. The picture is broadly similar when using risk-adjusted readmissions, with some evidence of an increase in quality two months prior the merger. Reassuringly, the coefficients are close to zero

Figure 3.5: Event study estimates of the impact of hospital mergers on in hospital mortality

Notes: The Figure plots event study estimates and corresponding 95 percent confidence bands. Dependent variable is the Hospital Wide Risk Standardised Mortality Measure - HWM. The unit of observation is a hospital site. Estimation from OLS regression with standard errors clustered at the hospital level. The regression includes calendar time fixed effects. Sample comprises all hospitals involved in a merger that belong to a trust providing acute-care services in England.

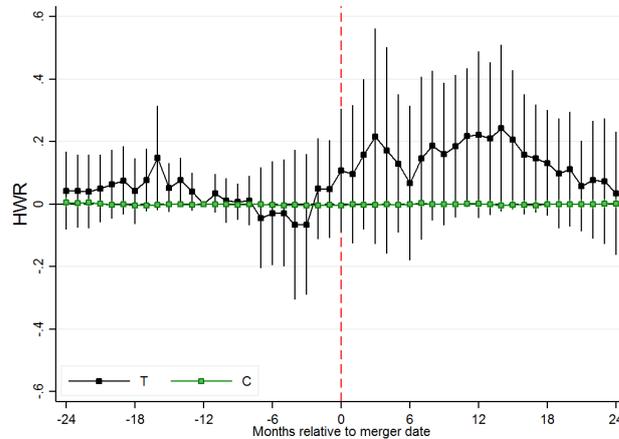
in the time preceding the event indicating that we are not capturing the acquisition of poorly performing hospitals.

The mergers we study increase the likelihood of a patient dying by 0.4 percentage points, or 27 per cent relative to the baseline mortality rate of 1.4 per cent. Similarly, the likelihood of readmission increases by 0.9 percentage points (11 per cent relative to a baseline of 8.3 per cent). This effect corresponds to around 60,000 excess deaths, and 140,000 excess readmissions due to the mergers in the period we study. We estimate the total value of life lost from mergers to exceed 13 millions, 4% of hospital costs.⁷

As a placebo test, we compute equivalent estimates for hospital sites not involved in mergers. The event study estimates for these hospitals are entirely flat in the pre- and post-merger periods, see Figure 3.6 and 3.7.

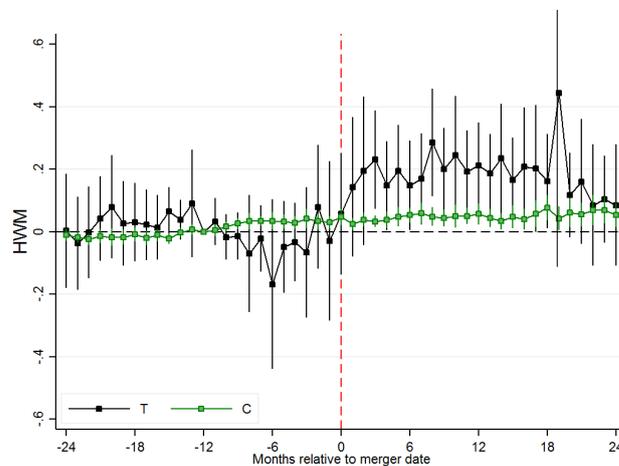
⁷This value is calculated under the assumption that the average patient that died in hospital had two years of life left to leave with value of life per year being 111,860.66 pounds.

Figure 3.6: Event study estimates of the impact of hospital mergers on unplanned readmissions- Placebo test

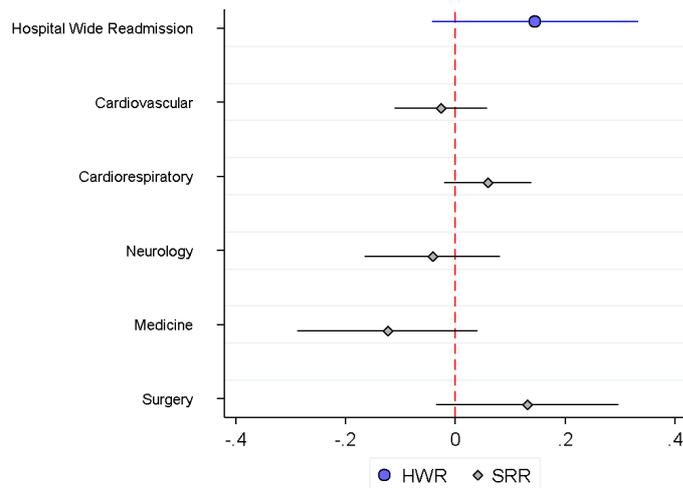


Notes: The Figure plots event study estimates and corresponding 95 percent confidence bands. Dependent variable is the Hospital Wide Risk Standardised Readmission Measure - HWR. The unit of observation is a hospital site. Estimation from OLS regression with standard errors clustered at hospital level. The regression includes calendar time fixed effects. Treatment sample comprises all hospitals involved in a merger that belong to a trust providing acute-care services in England. Control sample comprises all hospitals not involved in a merger that belong to a trust providing acute-care services in England.

Figure 3.7: Event study estimates of the impact of hospital mergers on in hospital mortality - Placebo test



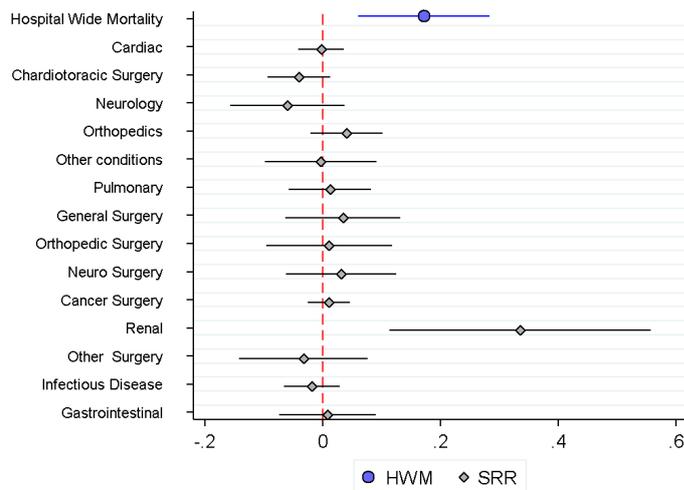
Notes: The Figure plots event study estimates and corresponding 95 percent confidence bands. Dependent variable is the Hospital Wide Risk Standardised Readmission Measure - HWR. The unit of observation is a hospital site. Estimation from OLS regression with standard errors clustered at hospital level. The regression includes calendar time fixed effects. Treatment sample comprises all hospitals involved in a merger that belong to a trust providing acute-care services in England. Control sample comprises all hospitals not involved in a merger that belong to a trust providing acute-care services in England.

Figure 3.8: Impact of hospital mergers on unplanned readmissions

Notes: The Figure plots the estimates and corresponding 95 percent confidence bands of the difference in quality before and after the merger. Quality is measured using the Hospital Risk Standardised Mortality (HWM) measure, and the mortality Standardised Risk Ratio (SRR) for the specialty cohort estimates. The unit of observation is a hospital site. Estimation from OLS regression. The regression includes calendar time fixed effects. Standard errors are clustered at site level. Sample comprises all hospitals involved in a merger that belong to a trust providing acute-care services in England.

Subgroup Analysis

We estimate the average effect of mergers on hospital quality at specialty cohort level with the intent of uncovering heterogeneous effects. The results are shown in Figure 3.8 and 3.9. Here, the coefficients represent the difference between the average quality of merging hospitals before and after the event. We include calendar time fixed effects that control for changes in quality common to all hospitals. Standard errors are clustered at the hospital level. In Figure 3.9 we break down the analysis into the 15 specialty cohort components of the HWM measure. The effect of mergers on quality appears to vary across different specialties, although in most cases we can't distinguish it from zero with a high level of accuracy. In Figure 3.9, we repeat the analysis using the 5 specialty cohort components of the HWR measure. Similarly to the HWM measure, the effect of mergers on quality is not homogeneous, with patients undergoing surgical operations being the most negatively affected from the merger.

Figure 3.9: Impact of hospital mergers on mortality

Notes: The Figure plots the estimates and corresponding 95 percent confidence bands of the difference in quality before and after the merger. Quality is measured using the Hospital Risk Standardised Mortality (HWM) measure, and the mortality Standardised Risk Ratio (SRR) for the specialty cohort estimates. The unit of observation is a hospital site. Estimation from OLS regression. The regression includes calendar time fixed effects. Standard errors are clustered at site level. Sample comprises all hospitals involved in a merger that belong to a trust providing acute-care services in England.

3.5 Extension

In this section we sketch out one possible mechanism through which mergers can negatively affect the quality of care: through their impact on market competition. Weaker market competition is believed to reduce hospitals incentives to provide better quality services in markets with fixed prices. As of now we can't conclude on whether this is the ultimate mechanism through which mergers impact hospital quality.

Mergers and Market Competition

Gaynor et al. (2013) finds compelling evidence that competition in the NHS reduced acute myocardial infarction mortality and patient length of stay. These findings are consistent with predictions from economic theory. In the absence of price competition, hospitals have to compete on quality to attract patients, hence the higher the number of competitors in a market the higher hospitals incentives to increase quality (Gaynor, 2006).

Mergers reduce the number of competitors in an industry and may therefore weaken competition and increase the merging parties market power leading to a reduction in quality. To produce some descriptive evidence on whether a *competition story* is behind the effects we found we compute a measure of market competition and relate this measure to changes in quality following the mergers.

We choose to measure competition using the Herfindahl Hirschman Index (HHI) which does not require price inputs and is therefore suitable to analyse the environment in which UK hospitals operate. The HHI is the sum of squared market shares of each firm competing in the market, ranging from 0 (perfect competition) to 10000 (monopoly). Following Gaynor et al. (2013), we calculate a hospital level HHI in two steps. First, we produce General Practice (GP) level HHI. Second, we aggregate HHIs at the hospital level. Equation 3.9 shows the first stage

$$HHI_{jc} = \sum_{i=1}^N S_{ic}^2 \quad (3.9)$$

i denotes a hospital that competes on the market of GP j for specialty c . S_{ic} is the GP j 's share of patients that have been inpatient in hospital i for specialty c expressed as a whole number, not a decimal. We compute S_{ic} using the HES data.

A critical input for this measure is determining the relevant market over which to compute market shares. We define a different relevant market for each GP to mirror the structure of the NHS where patients choose hospitals in conjunction with their doctors (Cooper et al., 2010b). We restrict the geographical extension of each GP market using the 80th percentile of the distance travelled by patients from the GP location to the hospital. This distance approximate patients willingness to travel. More specifically, the GP-relevant market is a circle centered around the GP location with radius equal to the 80th percentile of the distance. All acute hospitals that are located within this circle are potential competitors. In addition, we define a different relevant market for each specialty for two reasons. First, there is little substitutability between specialties. Second, patients needing different treatments may have different willingness to travel.

For each hospital, the hospital-specialty HHI is then computed as a weighted average of GP-specialty HHIs. Equation 3.10 shows the second stage, that is the HHI at hospital level.

$$HHI_{hc} = \sum_{j=1}^J s_{jc} \cdot HHI_{jc} \quad (3.10)$$

s_{hic} represents the share of patients that hospital h receives from GP j for specialty c . HHI_{jc} represents the GP level HHI for practice j and specialty c . Table 3.3 shows summary statistics for the site-specialty HHI. On average, site-specialty markets appear to be moderately concentrated in the period of interest.

Table 3.3: Summary statistics for competition measure

	Mean	SD	Min	Max
Herfindahl-Hirschman Index	1792.958	2278.348	0	10000

Notes: The Table displays summary statistics for HHI index for site-specialty relevant markets over the period 2006-2013. Sample comprises all hospitals operating in England.

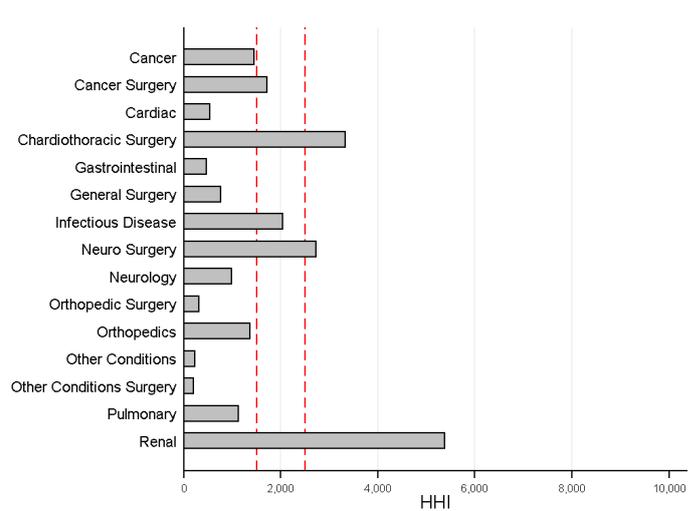
Inspecting the Mechanism

Antitrust authorities generally consider markets in which the HHI is between 1,500 and 2,500 points to be moderately concentrated and, beyond 2,500 points to be highly concentrated. Mergers occurring in concentrated markets are most likely to give cause for concern.⁸

In order to assess whether the effect found in the previous section can be the result of increased concentration and market power, we compute the pre-merger HHI for each specialty cohort. Figure 3.10 shows the average specialty HHI for the pre-merger period for the HWM groupings. The HHI of 4 of the 15 HWM specialty categories falls in the range of moderately concentrated; the HHI of 3 specialty categories falls in the highly concentrated range. Mergers occurring in these markets have ex-ante the potential to reduce quality.

Interestingly, some of the specialty cohorts for which we find a negative impact of mergers on hospital quality display a significantly high level of market concen-

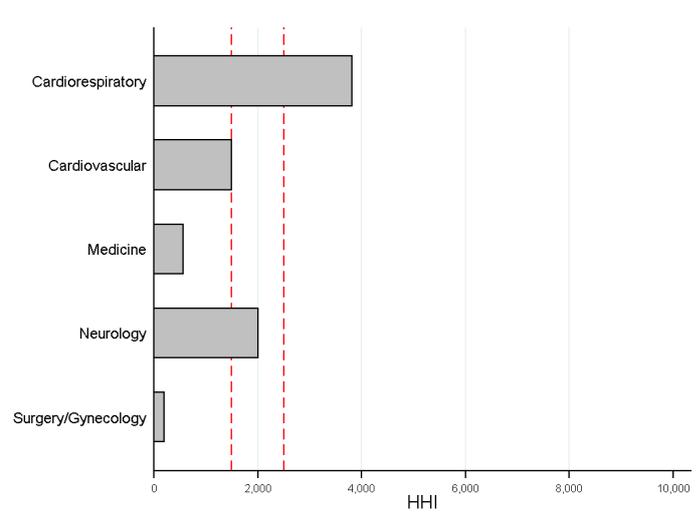
⁸In particular, mergers occurring in such markets that result in changes in HHI greater than 100 potentially raise competitive concerns; mergers in highly concentrated markets that result in a change in HHI greater than 200 are presumed to be very likely to enhance market power⁹.

Figure 3.10: Hospital-Specialty HHI for HWM categories

Notes: The Figure plots the average hospital-specialty HHI for the period -1 relative to the date of the merger. Sample comprises hospitals that have been involved in a merger in the period 2006-2015. Groupings of patients from HWM methodology. Red dashed lines denote antitrust authorities thresholds for moderately concentrated ($1500 \leq HHI < 2500$) and highly concentrated ($HHI \geq 2500$).

tration pre-merger. For example, *Renal* for which we find a statistically significant negative impact of mergers on quality has a pre-merger HHI of more than 5000. Figure 3.11 shows the average specialty HHI for the pre-merger period for the HWR groupings. *Cardiorespiratory* which we have found to experience a decrease in quality after the mergers shows high level of market concentration pre-merger. In a similar vein, *Medicine* for which we have found no negative impact of mergers on quality shows a low level of HHI pre-merger. On the other hand, *Surgery* shows a very low level of concentration pre-merger and a sizable negative impact of mergers on quality.

To further examine whether the competition mechanism could be a plausible explanation in our context we compute the average impact of mergers at hospital level and relate this to the level of pre-merger market competition faced by the hospital. Figure 3.12 shows the relationship between the average site-level effect of mergers on the HWM measure and the level of pre-merger market competition. Figure 3.13 shows the relationship between the average site-level effect of mergers on the HWM measure and the level of pre-merger market competition. There is a mild positive

Figure 3.11: Hospital-Specialty HHI for HWR categories

Notes: The Figure plots the average hospital-specialty HHI for the period -1 relative to the date of the merger. Sample comprises hospitals that have been involved in a merger in the period 2006-2015. Groupings of patients from the HWR methodology. Red dashed lines denote antitrust authorities thresholds for moderately concentrated ($1500 < HHI < 2500$) and highly concentrated ($HHI > 2500$).

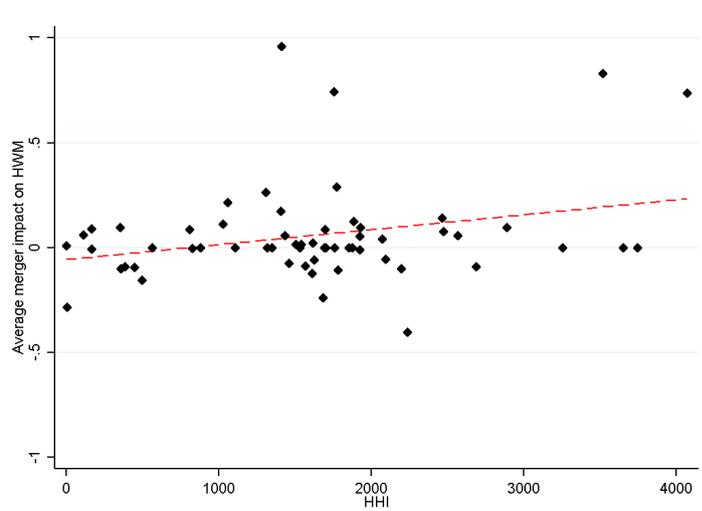
relationship between the level of pre-merger market competition and the impact of mergers on our measures of hospital quality. An increase in the HHI pre-merger is loosely correlated with an increase in the effect of mergers on the HWM and HWR measures (i.e. a reduction in quality following the merger). No direct conclusion can be extracted from these figures. In the future we plan to further develop this area of the analysis.

3.6 Conclusions

Mergers and acquisitions raise a number of antitrust concerns, chief among them that market consolidation may lead to higher prices or poorer product quality. Both of these effects are incredibly important in health care markets, which have experienced significant merger activity over the past two decades, and where prices and quality can literally be a matter of life or death.

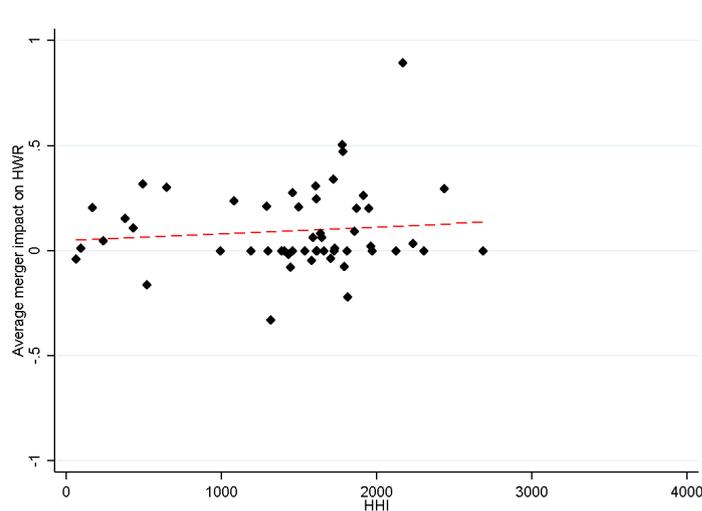
This paper provides novel evidence on the effect of mergers on quality of clinical care. We use an event study framework to evaluate 159 public hospitals involved in mergers in the English NHS during 2006 to 2015.

Figure 3.12: Relationship between effect of mergers and pre-merger HHI



Notes: The Figure displays the relationship between average effect of mergers at hospital level and hospital level HHI in the period pre-merger. Average effect of merger estimated with OLS regression controlling for calendar time fixed effects.

Figure 3.13: Relationship between the effect of mergers and pre-merger HHI



Notes: The Figure displays the relationship between average effect of mergers at hospital level and hospital level HHI in the period pre-merger. Average effect of merger estimated with OLS regression controlling for calendar time fixed effects.

We find evidence that hospital mergers have immediate, persistent and statistically significant negative impacts on quality of care. On average, the mergers we study increased the likelihood of a patient dying by 0.4 percentage points, or 27 per cent relative to the baseline mortality rate of 1.4 per cent. Similarly, the likelihood of readmission increased by 0.9 percentage points (11 per cent relative to a baseline of 8.3 per cent). Under very conservative assumptions about the valuation placed on life-years, we estimate that the deterioration in health outcomes is valued at approximately 4 per cent of average annual hospital costs. These effects are the same order of magnitude to the estimated effects of mergers on hospital costs.

There are a number of important caveats to this work. First, our work has focused on mergers between acute hospitals. In the future, it would be important to extend the analysis to specialists hospitals. This may uncover positive effects of mergers on hospital quality arising from specialisation economies. Second, the risk adjustment methodology could be further developed. In this work we risk adjust mortality and readmissions using the CMS methodology. This methodology has been developed for US hospitals and in particular Medicare patients. This is a limited share of population which could be very different from NHS patients. In the future we plan to develop our risk adjustment methodology further with the purpose of capturing health shifters relevant to the UK setting. Third, as of now we could improve on the way we estimate our coefficients standard errors. In the future we plan to refine the way we compute standard errors by using bootstrapping techniques. Lastly, we plan to investigate the channels through which mergers impact the quality of hospital care. In particular, we plan to investigate whether the impacts are associated with changes in market power or specific decisions taken by hospitals in the post-merger period (e.g. internal department reorganizations).

Based on the analysis to date, we conclude that hospital mergers may have first order impacts on clinical quality and this should be an important consideration for antitrust authorities. The steps outlined above will help us understand these effects in more detail. Our specific estimates of course apply to the English setting, where hospitals are publicly owned and prices are fixed. While this is an ideal setting to

evaluate the clinical quality impacts, understanding how these results generalize to more complex settings is an important direction for future

Chapter 4

Does Contact Time Matter For Patient Outcomes?

Joint work with Ann-Marie Cannaby, Vanda Carter, David C. Chan, Jonathan Gruber, and Stephenson Strobel

4.1 Introduction

In 2023, 71% of the National Health Service (NHS) staff who have direct contact with patients said they did not have the amount of time they would like to have to help them.¹

Although widely recognized as an essential component of good care and patient experience (Aiken et al., 2000), little is known about the relationship between contact time and patient hospital outcomes. This study aims to provide a first estimate of this relationship.

Contact time between medical staff and patients is generally considered a vital dimension of a patient stay in hospital (Barker et al., 2016; Griffiths et al., 2019). It is meant to allow the hospital staff to target, coordinate, and monitor treatment and contributes to the patient feeling cared for and supported (Aiken et al., 2000; Duffield et al., 2011; Westbrook et al., 2011). But does contact time matter for

¹<https://www.theguardian.com/society/2023/jul/24/most-nhs-staff-say-they-dont-have-enough-time-to-spend-with-patients>

patients' outcomes?

Empirical research in this area has been challenging. Data on contact time is not widely available and difficult to collect on a large scale. Existing studies have generally relied on self-reported logs or involved an external observer recording staff movements and activities (Dall'Ora et al., 2021; Edwards et al., 2009; Hendrich et al., 2008; Yen et al., 2018b). Both approaches suffer from significant limitations and are hard to scale. Time-motion analyses require a human observer to record and characterize activities over time. This is costly, time-consuming, and captures only limited time frames and locations. Using an observer to collect this type of data may also induce individuals to change their behavior, leading to biased results (Mahtani et al., 2018). Self-reported activity logs require staff to record, without any outside interference, how they spend their time during the day (Antinaho et al., 2015). This method is subject to recall bias, as staff members often overestimate their time with patients (Donaldson and Grant-Vallone, 2002), besides being an important ask to staff who are already significantly overworked.

However, data limitations have not been the only impediment to investigating the relationship between contact time and patient outcomes. Identifying a causal impact requires a plausible exogenous variation in contact time as this may correlate with observable and unobservable patient characteristics. Staff for example can be expected to spend more time with sicker patients, which may induce a spurious correlation between health outcomes and contact time. Policy experiments that have changed contact time have been challenging to come across, and to this day no existing study has attempted to address this challenge.

This paper uses a novel and unique data source, real-time location system (RTLS) data, to overcome these challengers. RTLS allows organizations to observe the movements of objects and individuals across space in real-time (Cannaby et al., 2022b). These systems have been deployed in healthcare, where they are increasingly leveraged to organize patient and staff flow (Overman, 2022). Data capture is a crucial capability of RTLS; every movement is recorded and stored in the system's

memory (Cannaby et al., 2022a). RTLS data allow us to have a privileged view of the interactions between patients and staff in the hospital and to compute contact time accurately across an entire hospital. Having quantified the amount of contact time patients receive from nurses; we analyze whether contact time has a detectable effect on patient outcomes. We focus on three patient outcomes: mortality, transfers to the Intensive Care Units (ICU), and accidents. These outcomes capture essential dimensions of the patient's hospital stay. Mortality is a crucial result for both the patient and the hospital. ICU transfers indicate deterioration in patients' health. Accidents are considered a marker of the quality-of-care patients receive during a hospital admission stay and the level of oversight provided.

We estimate the relationship between patient outcomes and contact time using a fixed effect Ordinary Least Squares regression to provide evidence of causal effects. Exploiting the richness of our data, we decompose the amount of contact time a patient receives into an endogenous and a likely exogenous component. We show that contact time that arises from interactions at the patient bed is endogenous as sicker patients require more care, so nurses spend more time with them at their bedside. However, we believe that a significant portion of the time nurses spend with patients is unrelated to their characteristics and can be considered unrelated to their health outcomes. In particular, we argue that the amount of time nurses spend in the patient room is uncorrelated to other determinants of patient health once we control for the patient bed-side level contact time. Under this assumption, our results suggest that contact time has a negative effect on both in-hospital mortality and accidents. An increase in contact time reduces mortality by 0.03% and accidents by 0.01%. This is evidence of the importance of direct care in hospitals for patient outcomes.

This paper contributes to two strands of literature. First, an extensive literature in health services that measures contact time as an end in itself. Among all, Butler et al. (2018) has the approach closest to us. They use a sensor-based measurement of contact time within an intensive care unit and show that nurses spend 32% of their time with patients. Other studies use less sophisticated contact measure-

ments; for example, direct observation of nurses taking vital signs suggests that this is a time-consuming process (Dall'Ora et al., 2021). This type of observation also suggests that hands-on tasks of nurses occur during the day and that a plurality of time is spent within the nursing station and charting (Yen et al., 2018a). This relates directly to concerns within medicine on the over-proliferation of non-clinical work. Surveys suggest that physicians find the amount of time devoted to paperwork and administrative tasks has increased to the detriment of clinical work (Government of Nova Scotia, 2020). The introduction of electronic health records has been associated with an increase in time spent on documentation by 8% among consultant physicians (Joukes et al., 2018) and nearly 45% increase in documentation tasks within the ICU (Carayon et al., 2015). Surgical residents spend nearly 24 hours on documentation during a week (Cook et al., 2010). Overall, the impact of introducing electronic health records seems to increase the overall documentation time (Baumann et al., 2018) which is in line with survey responses that suggest an increase in documentation in lieu of face-to-face contact time. This increase in documentation is associated with physician burnout (OMA, 2021). Our contribution to this literature is to show that direct contact with patients matters. Trends towards additional administrative tasks will likely result in poorer patient outcomes.

This links to a second piece of literature on nurse staffing. Much of the current theory on nurse staffing levels and patient outcomes relates to better oversight. More nurses mean more oversight due to more direct observation of patients (Shekelle, 2013). There has been mixed evidence that additional nurse staffing improves outcomes. Most associations between patient outcomes and staffing levels show improvements in patient outcomes with higher staffing levels (Aiken et al., 2000; Griffiths et al., 2019; Haegdorens et al., 2019; Musy et al., 2021; Needleman et al., 2011; Zaranko et al., 2022). Policy experiments have been more challenging to find. A notable exception has been the adoption of minimum nursing ratios in Australia, which was associated with 7% reductions in mortality and readmissions (McHugh et al., 2021). Similar mandates in California have demonstrated mixed evidence; while surveys of nurses report better quality of care (Aiken et al., 2010), adminis-

trative data suggests no improvement in patient outcomes like mortality despite the policy improving staffing ratios (Cook et al., 2010). We provide a plausible mechanism underpinning this literature: direct contact time improves patient outcomes.

This paper proceeds as follows. In Section 4.2 we describe the RTLS system, our measure of contact time, and the data on patient outcomes. In Section ?? we present our empirical strategy and results. Finally, Section 4.4 concludes.

4.2 Data

The setting of this paper is the New Cross Hospital in Wolverhampton, England. This is a large district general hospital part of the Royal Wolverhampton NHS Trust (RWT). The Trust is one of the main healthcare providers in the West Midlands, covering acute, community, and primary care services.

In 2013, RWT partnered with a United States technology company to develop a real-time patient flow and tracking solution (Nash, 2014). This application was intended to support staff in delivering care and to enhance efficiency through the process of providing real-time operational information across clinical areas. The RTLS requires four main components:

- A locating node (i.e., multiple boxes on the ceiling covering clinical areas),
- A location server (the computer that receives all the data),
- A user application (the software that interprets the data),
- The tag (typically a badge worn by members of staff, patients, and equipment)

RTLS tags can be worn by staff – on a lanyard or clip – while patients have it attached to a wrist or ankle via a bracelet. The technology transmits continuous location data that is unique to the tagged person or item in real or near-real-time (Kamel Boulos and Berry, 2012).

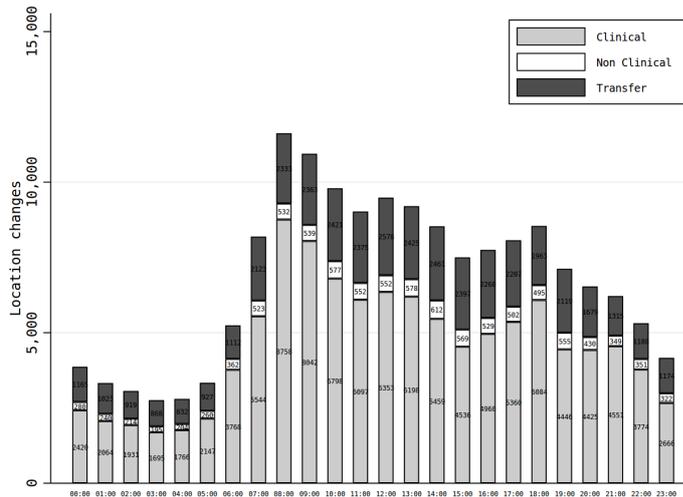
The RTLS at RWT provides real-time tracking of all staff and patients across 564,916 square feet of the New Cross hospital site. As of April 2022, RWT has

Figure 4.1: RTLS components



Note: The figure shows the locating node and the badges worn by staff and patients.

Figure 4.2: Tracking movements in a day



Note: Total location changes by hour and location type. Source RTLS data at RWT.

issued 7218 staff badges, and all patients are badged on admission for their stay within the hospital (Cannaby et al., 2022b). We show the privileged view that the RTLS buys us into the workings of the hospital in Figure 4.2. Each observation in our data is a location change; someone going in and out of a location. We classify hospital areas into clinical (e.g., bays), non-clinical (e.g., waiting rooms), and transfer (e.g., corridors, halls) and track how many location changes we observe in each of these over the course of a day. Clinical areas are the places where most movements occur. Between 8 and 9 AM on average, we can observe more than 8500 staff movements in and out of locations.

Patients’ contact time with nurses

We utilize RTLS data from April 2016 to April 2019 to measure how much contact time patients receive from nursing staff, including registered nurses and health care

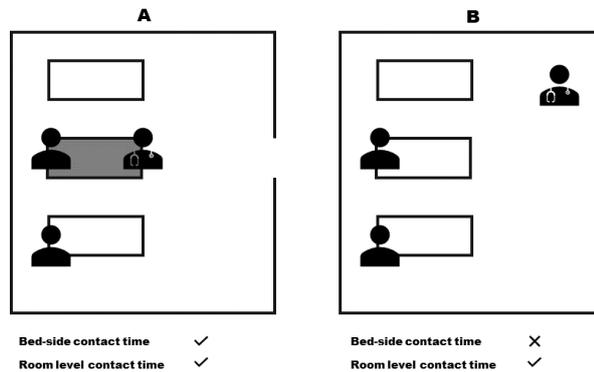
support workers (RNs and HCSWs).

The starting point is to identify time lapses when the patient and the nurse interact at a given location. We focus on two types of interactions: those at the patient's bed and those in the patient's room.² Our sample comprises more than 20 million interactions between patients and nursing staff.

We compute bed-side contact time as the total number of minutes a patient spends with staff at the location identified as the patient's own bed. We define room-level contact time as time spent by nurses and patients together in the room where the patient's bed is located, regardless of whether the interaction is detected at the patient's bed side. Figure 4.3 presents two scenarios illustrating the difference between bed-side and room-level contact time. In Scenario A, the nurse is at the bed-side of Patient 1 (in gray); this interaction will feed into Patient 1 bed-side contact time but will also account for both patients' room-level contact time. The interaction in Scenario B will account for both patients' room-level contact time but contribute to neither bed-side level contact time. Contact time is then computed as the total number of minutes nurses spend with patients at their bed (bed-side contact time) or in the patient's room (room-level contact time).

Figure 4.4 presents the bed-side and room-level contact time distribution. An observation in our data corresponds to a unique patient-day combination. The top and bottom sides of the box are the lower and upper quartiles. The box covers the interquartile interval, where 50% of the data is found. The horizontal dotted line in gray that splits the box in two is the median, and the horizontal dotted line in blue indicates the mean. The median bed-side contact time is 16 minutes, while the mean is 22. The median room-level contact time is 7 hours, while the mean is 8.

²We don't have data on the assignment of patients to beds or rooms, but we can reliably identify assignments using the RTLS data itself. We compute the number of hours we observe each patient in hospital on a given day. The system tracks the average patient in our sample for 15 hours daily. The patient's bed is identified as the bed location where the patient spends most of his/her day. For most of the patients in our data, the assigned bed accounts for 90 percent of the time they are observed in the hospital indicating that our imputation method is reasonable. To further check the accuracy of the imputation, we examine whether the patient diagnosis codes are consistent with the location of his/her bed. This check substantiates that we can reliably identify the location of the patient's own bed.

Figure 4.3: Bed-side and Room-level Contact Time

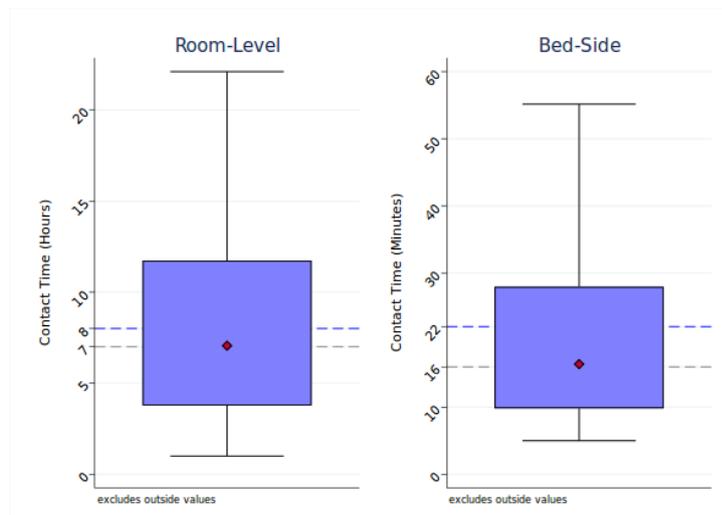
Note: Panel A shows what accounts for bed-side level contact time. Panel B shows what accounts for room-level contact time. All bed-side contact time accounts as room-level contact time

Both distributions display significant heterogeneity, with different patients receiving significantly different amounts of contact time. Bed-side contact time is also concentrated on particular moments of the day, as seen in Figure 4. The most significant amount of bed-side contact time is between 8 and 12 AM. This aligns with previous findings documenting that nurses' work is not distributed equally across a 12-hour shift and that hands-on tasks mainly occur between 7 and 11 AM (Yen et al., 2018b).

Complementary Data

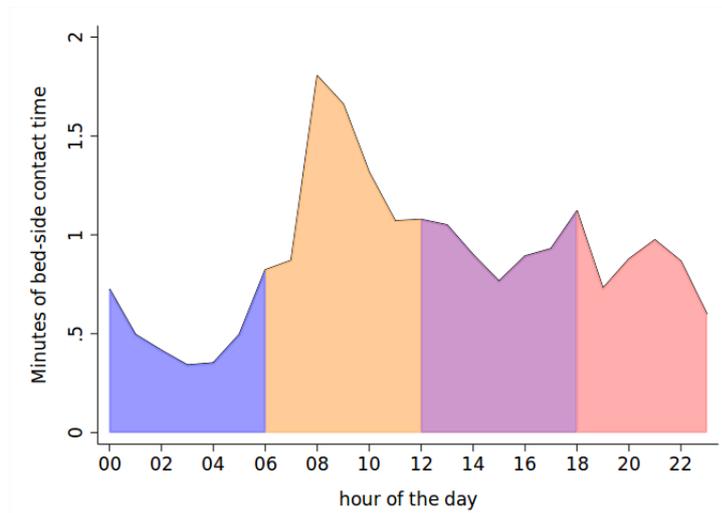
We complement the RTLS data with information on patient characteristics from the hospital's inpatient episode dataset. This includes patient clinical and demographic information. From this data, we identify the date of the patient's death. Further, we use data from the incident reporting system data at RWT, which collects all reported adverse events involving patients, staff, and visitors at the hospital. We observe the time and the day when these incidents occur, the location, whether the adverse event has caused any harm to the patient, and the type of event (e.g., pressure ulcer, wrong medication). Finally, we monitor deterioration in a condition necessitating higher levels of care with any transfer to the intensive care unit at RWT, which we detect

Figure 4.4: Distribution of Contact Time



Note: Scatter plot of bed-side and room-level contact time. Does not include outside values. Bed-side contact time is minutes per day and patient. Room-level contact time is in hours per day and patient.

Figure 4.5: Minutes of Bed-Side Contact Time by the Hour (**Patients**)



Note: Average number of minutes per hour patients and nurses are observed together at the patient's bed.

using RTLS data.

Our sample of patients includes all individuals admitted for inpatient care to the hospital during the observation period across 20 inpatient wards. 60 percent of patients are males, 78 percent are white, and the average patient is 64 years old and has a length of stay of 20 days. 0.3 percent of patients experience death in the hospital, 1.2 experience an accident, and 0.09 are observed being transferred to the ICU during their hospital stay.

Our sample of staff from the RTLS includes all RNs and HCSWs employed by the hospital between 2016 and 2019. RNs are qualified nurses who coordinate access and deliver prescribed care to patients. RNs have either a foundation or bachelor's degree in nursing and are registered practitioners with the Nursing/Midwifery Council (NMC). HCSWs work under the guidance of a healthcare professional, such as a nurse or a doctor. In a hospital setting, they may assist with the patient's hygiene needs and help mobilize and monitor patients' conditions. For part of the nurses in our sample, we can identify their gender, nationality, and experience in the NHS. For this subset, we can observe that 75 percent are female, 81 percent are British, and the average nurse has ten years of experience.

Table 4.1: Patients Summary Statistics

	mean	sd	min	max
Share Female	0.39	0	0	1
Share White	0.78	0	0	1
Age	64.5	16	15	90
Charlson Index	2.3	2	0	15
Length of Stay	19.5	25	0	583
Daily Bed-Side Contact Time (Minutes)	22.3	19.2	5	593.8
Daily Room-Level Contact Time (Hours)	8.2	5.3	1	22.1
Mortality	0.003	0	0	1
Accidents	0.012	0	0	1
ICU Transfers	0.009	0	0	1
Observations	66825			

* Note: Source RWT inpatient records, 2016/2019.

Table 4.2: Nurse Summary Statistics

	mean	sd	min	max
Panel A				
Share of Registered Nurses	0.64	0	0	1.0
Days Worked Monthly	13.63	5	1	31.0
Observations	203601			
Panel B				
Share Female	0.75	0	0	1.0
Share British	0.81	0	0	1.0
Years of Experience	10.18	9	0	52.0
Observations	63989			

* Note: Source RWT RTLS and HR records, 2016/2019.

4.3 Empirical Strategy

Our empirical strategy is tailored to identifying causal effects in the presence of endogeneity, which arises because the amount of contact time a patient receives is likely a function of their characteristics and, hence, their outcomes. Omitted health variables, in particular, are a concern in this context. For example, nurses may spend more time with patients whose condition is deteriorating rapidly. The nurses can observe this health measure, but the econometricians can't. This in turn could bias any estimate such that we might expect a much reduced or even negative relationship between contact time and outcomes.

An ideal experiment to address this concern would be a randomized control trial (RCT) where we could exogenously change the amount of contact time each patient receives. Within an RCT, averaging differences in outcomes between the control and treatment groups would be sufficient to identify a causal effect of contact time.

We do not observe any exogenous change in contact time which would be appropriate to approximate this setting. However, we believe the granularity of our data allows us to distinguish between the contact time patients receive because of their health status and the contact time patients receive because of other reasons unrelated to their health. We lay out our identification strategy below.

Identification Argument

Assume that the following relationships can express the true model of contact time and patient outcomes:

$$Y_{irt} = \beta_0 + \beta_1 CT_{irt} + \beta_2 X_{irt} + \beta_3 H_{irt} + \varepsilon_{irt} \quad (4.1)$$

$$CT_{irt} = f(X_{irt}, H_{irt}) \quad (4.2)$$

Where Y_{irt} is a binary indicator of whether a patient i residing in room r has experienced negative health outcomes on day t (i.e., death, ICU transfer, or accident) and CT_{irt} is the amount of contact time patient i receives on day t . X_{irt} and H_{irt} represent respectively all observed and unobserved factors affecting the outcome and the amount of contact time the patient receives. Lastly, ε_{irt} and u_{irt} represent two idiosyncratic components.

As we can at best only observe X_{irt} , the model we can seek to estimate would be:

$$Y_{irt} = \alpha + \beta_1 CT_{irt} + \beta_2 X_{irt} + \eta_{irt} \quad (4.3)$$

$$\eta_{irt} = \beta_3 H_{irt} + \varepsilon_{irt} \quad (4.4)$$

Estimating β_1 in this case would lead to a biased result because we cannot control for H_{irt} . To attenuate this problem and provide evidence on the causal relationship between contact time and patient outcomes, we make a simple distinction between what we call bed-side and room-level contact time.

We argue that bed-side level contact time is endogenous as sicker patients require more direct care and, for this reason, are given more contact time by nurses. However, we assume that the amount of time nurses spend in the patient room is likely unrelated to any unobserved determinants of health once we control for the amount bed-side contact time patient i receives on day t . In particular, we assume contact

time is the result of the following relationships:

$$CT_{irt} = CT_{irt}^B + CT_{irt}^R \quad (4.5)$$

$$CT_{irt}^B = h(X_{irt}, H_{irt}) \quad (4.6)$$

$$CT_{irt}^R = \sum_{j \neq i}^{N_r} g_j(CT_{jrt}^B) + g_i(CT_{irt}^B) \quad (4.7)$$

In this model, the amount of contact time patient i receives is given by the sum of CT_{irt}^B , the bed-side level contact time, and CT_{irt}^R the room-level contact time received by patient i on day t . CT_{irt}^B is a function of both observed and unobserved factors affecting the outcome and, in this sense, endogenous. CT_{irt}^R is a function of the bed-side contact time of the patients in room r and patient i contact time. The bed-side contact time of the patients in room r is unrelated to patient i characteristics and hence, once we control for CT_{irt}^B room-level contact time is unrelated to any unobserved determinant of health.

Estimation

In practice, we estimate the following model:

$$Y_{irt} = \alpha + \beta^B CT_{irt}^B + \beta^R CT_{irt}^R + \gamma N_{rt} + \delta X_{irt} + W_{rt} + \varepsilon_{irt} \quad (4.8)$$

Y_{irt} is a binary indicator of whether a patient i residing in room r has experienced negative health outcomes on day t (i.e., death, ICU transfer, or accident). CT_{irt}^B is the bed-side level contact time received by patient i on day t . CT_{irt}^R is the room-level contact time received by patient i on day t . N_{rt} is the number of patients in room r on day t . W_{rt} represents the interaction between the ward where room r is located and the day t . X_{irt} includes patient i clinical information (e.g., age, comorbidities) and the characteristics of the average patient in room r , as well as an additional proxy of health status which is the distance between the patient bed and the nursing station. We cluster standard errors at the ward level.

4.4 Results

In Table 4.3 we show the results from estimating the model in Equation 4.8 where the dependent variable is a binary indicator of the patient death on day t .

In column (1) we control for ten comorbidity indicators, primary diagnosis code, age, age squared, the average characteristics of the patients in the room, the distance between the patient bed and the nursing stations, and the interaction between ward and day. The coefficient on room-level contact time is negative and statistically significant, indicating that an increase in contact time is associated with a diminished probability of death. In column (2) we control for bed-side contact time. The coefficient on room-level contact time becomes more negative. Adding the number of patients in the room as an additional control in column (3) does not change the result.

The estimates suggest that one minute more of contact time is associated with a -0.000635 percentage point change in the probability of death, or a one-unit increase in contact time reduces in-hospital mortality by 0.3%. This reduction is not negligible, as it suggests that for ten more minutes of additional contact, patient mortality would drop by almost 3%. Assuming that for a 12-hour shift 40 percent of the time is spent in direct contact with patients, adding one more nurse to a ward with ten patients would result in a drop in mortality of 8.64 percent. This result is in line with Zaranko et al. (2023) where the authors show that on average, an extra 12-hour shift by an RN was associated with a reduction in the odds of a patient death of 9.6%.

In Table 4.4 we show the results from estimating the model in Equation 4.8 where the dependent variable is a binary indicator of the patient being transferred to ICU on day t . The estimates are negative but not statistically significant, and we cannot exclude a zero effect for this outcome.

In Table 4.5 we show the results from estimating the model in Equation 4.8 where the dependent variable is a binary indicator of the patient experiencing an accident

Table 4.3: OLS regression - Dependent variable binary indicator of patient death

	(1)	(2)	(3)
Room-level CT	-0.00000332*** (0.000000690)	-0.00000635*** (0.00000116)	-0.00000635*** (0.00000114)
Bed-side CT		0.000114*** (0.0000200)	0.000112*** (0.0000197)
Room patients (N)			0.000669** (0.000206)
Share of deaths	0.002	0.002	0.002
Room-level CT (mean)	439.154	439.154	439.154
Bed-side CT (mean)	16.572	16.572	16.572
N	227395	223996	227395

Standard errors in parentheses are clustered at the ward level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.4: OLS regression - Dependent variable binary indicator of patient ICU transfer

	(1)	(2)	(3)
Room-level CT	-0.00000340 (0.00000315)	-0.00000350 (0.00000260)	-0.00000356 (0.00000261)
Bed-side CT		0.00000124 (0.0000252)	0.00000278 (0.0000237)
Room patients (N)			0.000579 (0.000283)
Share of icu transfers	0.004	0.004	0.004
Room-level CT (mean)	439.154	439.154	439.154
Bed-side CT (mean)	16.572	16.572	16.572
N	227395	223996	227395

Standard errors in parentheses are clustered at the ward level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

on day t . The coefficient on room-level contact time becomes negative and statistically significant. Also, in this case, adding the number of patients controlling for bed-side level contact time makes the coefficient more negative. The estimates suggest that a one-unit increase in contact time reduces the probability of the patient experiencing an accident by 0.1%.

4.5 Conclusions

Using novel data from RTLS we showed that contact time between patients and nurses significantly affects the health outcomes of patients in the hospital. The estimates suggest that one minute more of contact time is associated with a -0.000635

Table 4.5: OLS regression - Dependent variable binary indicator of patient accident

	(1)	(2)	(3)
Room-level CT	-0.00000329** (0.00000108)	-0.0000120*** (0.00000111)	-0.0000119*** (0.00000109)
Bed-side CT		0.000334*** (0.0000207)	0.000329*** (0.0000201)
Room patients (N)			0.000460 (0.000261)
Share of Accidents	0.013	0.013	0.013
Room-level CT (mean)	439.154	439.154	439.154
Bed-side CT (mean)	16.572	16.572	16.572
N	227395	223996	227395

Standard errors in parentheses are clustered at the ward level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

percentage point change in the probability of death, or a one-unit increase in contact time reduces in-hospital mortality by 0.3% and accidents by 0.1%. These results are achieved under a critical assumption: once we control for bed-side contact time, room-level contact time is not correlated to unobservable patient health characteristics. Further work is needed to provide more evidence on the validity of this assumption. However, we believe this paper is an important step in delivering substantive evidence that contact time matter for patient outcomes.

Chapter 5

General Conclusions

This dissertation has provided three core results. First, it showed that robots shorten patients' length of stay and decrease the incidence of adverse events from surgery, but their effects are heterogeneous and significantly depend on surgeons' skills. High-skilled surgeons benefit the least from using the technology, while lower-skilled surgeons appear to gain the most from it. Second, it found that mergers have immediate and persistent negative impacts on clinical quality. Lastly, it provided evidence of the effect that direct contact between nurses and patients can have on in-hospital mortality and accidents. Subsequent research will further explore the underlying mechanisms driving these discoveries.

Appendix

Table 1: Patient's Characteristics

Variable	Type	Skills model	MTE model
Age	Continuous	X	X
Age squared	Continuous		X
Ethnicity	Categorical	X	
White	Binary		X
Myocardial infarction	Binary	X	X
Peripheral vascular disease	Binary	X	X
Cerebrovascular disease	Binary	X	X
Dementia	Binary	X	X
Chronic pulmonary disease	Binary	X	X
Rheumatic disease	Binary	X	X
Peptic ulcer disease	Binary	X	X
Mild liver disease	Binary	X	X
Moderate liver disease	Binary	X	X
HIV/AIDS	Binary	X	X
Diabetes	Binary	X	X
Any malignancy (e.g. lymphoma)	Binary	X	
Congestive heart failure	Binary	X	X
Admission method	Categorical	X	

Table 2: Hospital Trusts - Sample Summary Statistics - 2004/2017

	Early Adopter		Follower		Late Adopter	
	mean	sd	mean	sd	mean	sd
Sites	5.987	4.971	10.235	11.678	4.136	3.482
Acute sites	1.494	1.034	1.960	1.233	1.687	0.986
Specialist sites	0.420	0.695	0.460	1.081	0.127	0.370
Available beds	852	484	1089	372	763	316
Occupied beds	729	410	879	292	647	283
Capital investment	18254	12083	14586	9141	10963	8993
Patient costs	3270	2135.607	3355	1675	2429	1342
Teaching hospital	0.444	0.511	0.400	0.498	0.147	0.356
Observations	18		30		75	

Ordinary least squares regression

I estimate the relationship between robotic surgery and patients' outcomes using a linear regression model. In Table 3 the dependent variable is the logarithm of the patient length of stay in hospital. In Table 4 the dependent variable is an indicator for the occurrence of an adverse event following the surgery. In both Tables the columns represent sequentially richer models where control variables and fixed effects are added to the initial linear regression. The independent variable of interest for all models is a dummy variable that takes value one if the patient has been op-

erated with robotic surgery and zero otherwise. The sample includes all patients that have undergone a radical prostatectomy in a NHS hospital in England. Table 3 shows that patients operated with the robot experience lower length of stay in hospital. The coefficient in all specifications is negative and statistically significant. The negative relationship is robust to the inclusion of patient, hospital characteristics, and year and hospital fixed effects. Table 4 shows that patients that are operated with the robot are also less likely to experience an adverse event from surgery. The coefficient in all specifications is negative and statistically significant. The negative relationship is robust to the inclusion of patient characteristics, and year and hospital fixed effects. Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. The controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator.

Table 3: Association of robotic approach and adverse event - OLS Regression

	(1)	(2)	(3)	(4)
Robot	-0.768*** (0.057)	-0.648*** (0.056)	-0.407*** (0.065)	-0.323*** (0.046)
Patient characteristics	No	Yes	Yes	Yes
Year-Month	No	No	Yes	Yes
Day of the week	No	No	Yes	Yes
Hospital FE	No	No	No	Yes
<i>N</i>	61225	50884	50884	50884

Standard errors clustered at hospital level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Instrumental variable approach

I use an instrumental variable approach to estimate the local average treatment effect for log length of stay. I use either or both Z_{dist} and Z_{days} as an instrument for the probability of the robotic approach. I employ a two-step procedure suggested by Wooldridge (2015) whereby instead of using the instrument directly I use the predicted probabilities from a first stage Probit estimation in the second stage. I show the first and second stage estimates in Table 5. In Column 1 to 3, I show the first stage estimates from a Probit regression of an indicator of robotic approach on

Table 4: Association of robotic approach and log length of stay - OLS Regression

	(1)	(2)	(3)	(4)
Robot	-0.0958*** (0.0105)	-0.0809*** (0.00951)	-0.0438*** (0.00958)	-0.0639*** (0.0114)
Patient characteristics	No	Yes	Yes	Yes
Year-Month	No	No	Yes	Yes
Day of the week	No	No	Yes	Yes
Hospital FE	No	No	No	Yes
<i>N</i>	61839	51424	51424	51424

Standard errors clustered at hospital level

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the instrument(s) and controls. In Column 4 to 6, I show the second stage estimates where I instrument the indicator of robotic approach with the predicted values from the Probit first stage. Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. The controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. YM indicates year-month controls, DOW indicates day of the week controls. The LATE obtained in all specifications is negative and statistically significant. Notice that Mogstad et al. (2021) show that with more than one instrument, the monotonicity condition required for identification of the LATE can only be satisfied if choice behavior is effectively homogeneous.

Bivariate probit model

Chesher (2005) shows that the assumptions required for justification of two stage procedures are incompatible with a discrete outcome. For this reason, to provide a benchmark to the MTE estimates, I use a bivariate Probit model to test the impact of robotic surgery on the dependent variable adverse event. I maintain the assumption of joint normality of errors, exogeneity, and relevance conditions for the instruments. Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. The controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. YM indicates year-month controls, DOW indicates day of

Table 5: Estimated impacts of robotic approach on log length of stay

	(1)	(2)	(3)	(4)	(5)	(6)
	FS	FS	FS	IV	IV	IV
Z_{dist}	-0.127*** (0.003)		-0.093*** (0.003)			
Z_{days}		0.036*** (0.001)	0.027*** (0.001)			
Robot				-0.412*** (0.023)	-0.221*** (0.023)	-0.307*** (0.019)
Patient characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month	Yes	Yes	Yes	Yes	Yes	Yes
Day of the week	Yes	Yes	Yes	Yes	Yes	Yes
N	51423	50203	50203	50883	49795	49795

Robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the week controls. Table 6 shows the coefficients estimated for the simultaneous Probit model. In Panel A, the dependent variable is the indicator of adverse event (adverse event=1). In Panel B, the dependent variable is an indicator of robotic approach (robot=1).

Table 6: Estimated impacts of robotic approach on probability of adverse event

	(1)	(2)	(3)
	Bi-Probit	Bi-Probit	Bi-Probit
Panel A			
Robot	-0.430*** (0.059)	-0.448*** (0.072)	-0.459*** (0.059)
Panel B			
Z_{dist}	-0.014*** (0.001)		-0.014*** (0.001)
Z_{days}		0.0002*** (0.00002)	0.0001*** (0.00001)
N	49253	48081	48081

Robust standard errors.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Skills measured using surgeon level random intercept model

I test my baseline specification using an alternative measure of skills where I use surgeon level random intercepts instead of hospital level random intercepts to compute the ratio between expected and predicted adverse events describes in Section

2.4. I show how the resulting measure is distributed in Figure 1. In Table 7, I test my baseline specification when I use this measure as a continuous variable, in Table 8 I use a dummy variable that indicates that the surgeon is above the median distribution of skills. In Figure 2 I show the MTE curve with the continuous measure of skills. In Figure 3 I show the MTE curve with the indicator of high skilled. Standard errors bootstrapped with 100 repetitions. The dependent variable is the logarithm of post-operative length of stay. Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. Area controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. All specifications are estimated using the instruments Z_{dist} and Z_{days} and include a continuous measure of surgeon's skills pre-robot.

Controlling for hospital year of adoption

In Table 9, I test the robustness of my baseline results to the inclusion of a set of dummies that control for the year the hospital has adopted the robot. I test this specification using my continuous measure of skills. The adoption year is imputed from the data as the first year the hospital is observed operating a patient with robotic surgery. Standard errors bootstrapped with 100 repetitions. The dependent variable is the logarithm of post-operative length of stay. Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. Area controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. All specifications are estimated using the instruments Z_{dist} and Z_{days} and include a continuous measure of surgeon's skills pre-robot. Skills are measured using the SRR.

Surgeons' experience

In Column 1 and 3 of Table 10, I test the robustness of my results by estimating the baseline specification, restricting the sample to surgeons observed working since 2003. In Column 2 and 4, I test the robustness of my results by estimating the baseline specification under the inclusion of a set of dummies identifying the first year

the surgeon is observed in my data. I show the relevant MTE curves in Figure 4 and Table 5. Standard errors bootstrapped with 100 repetitions. The dependent variable is the logarithm of post-operative length of stay. Demographic controls include age, age squared, indicator for white ethnic profile. Clinical controls include ten comorbidity variables. Area controls include distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator. All specifications are estimated using the instruments Z_{dist} and Z_{days} and include a continuous measure of surgeon's skills pre-robot. Skills are measured using the SRR.

Fixed effect model of surgeons' skills

I test the baseline specification under two different measures of skills that I derive from a fixed effect model. In Table 13, I show the coefficients estimated from a model where I measure skills using the estimated fixed effect from a logistic regression of the binary indicator of adverse event of patient characteristics and a hospital fixed effect. I use the standardized hospital fixed effect as my measure of skills in Column 1 and 2. I use the standardized risk ratio computed using the fixed effects as my measure of skills in Column 2 and 3. The procedure to derive the standardized risk ratio is analogous to the one I use in Section 4.2. However, the predicted component includes the hospital fixed effects while the expected component sets the hospital fixed effect to zero. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to the closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon's skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions. Skills are measured for the period 2005-2007.

Table 7: Estimated coefficients - Surgeon level SRR

	(1)	(2)	(3)
	Selection	Length of stay	Adverse event
Skills (Surgeon SRR)	0.126*** (0.013)	-0.170*** (0.008)	-0.020*** (0.005)
Skills (Surgeon SRR) * Propensity score		0.121*** (0.023)	0.046*** (0.010)
Year-Month	Yes	Yes	Yes
Day of the week	Yes	Yes	Yes
<i>N</i>	50203	49215	50203

Standard errors in parentheses, bootstrapped with 100 repetitions

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Estimated coefficients - Surgeon level SRR

	(1)	(2)	(3)
	Selection	Length of stay	Adverse event
High Skilled (Surgeon SRR)	0.056*** (0.015)	-0.145*** (0.012)	-0.065*** (0.007)
High Skilled (Surgeon SRR) * Propensity score		0.029 (0.018)	0.092*** (0.010)
Year-Month	Yes	Yes	Yes
Day of the week	Yes	Yes	Yes
<i>N</i>	50203	49215	50203

Standard errors in parentheses, bootstrapped with 100 repetitions

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: Estimated coefficients controlling for hospital's year of adoption

	(1)	(2)	(3)
	Selection equation	Length of stay	Adverse event
Skills	0.122*** (0.022)	-0.150*** (0.013)	-0.031*** (0.009)
Skills * Propensity score		0.134*** (0.026)	0.044*** (0.013)
Year-Month	Yes	Yes	Yes
Day of the week	Yes	Yes	Yes
<i>N</i>	42378	41466	42378

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Estimated coefficients - surgeon experience

	(1)	(2)	(3)	(4)
	Length of Stay		Adverse Event	
Skills	-0.289*** (0.019)	-0.295*** (0.011)	-0.046** (0.016)	-0.0282*** (0.007)
Skills * Propensity score	0.279*** (0.053)	0.258*** (0.024)	0.0773** (0.029)	0.00910 (0.013)
Year-Month	Yes	Yes	Yes	Yes
Day of the week	Yes	Yes	Yes	Yes
Experience dummies	No	Yes	No	Yes
<i>N</i>	11824	49215	12028	50203

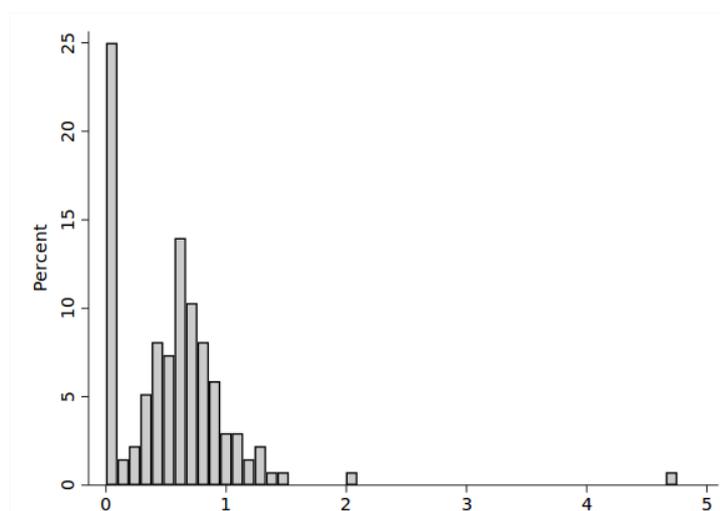
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ **Table 11:** Estimated effects - surgeon experience

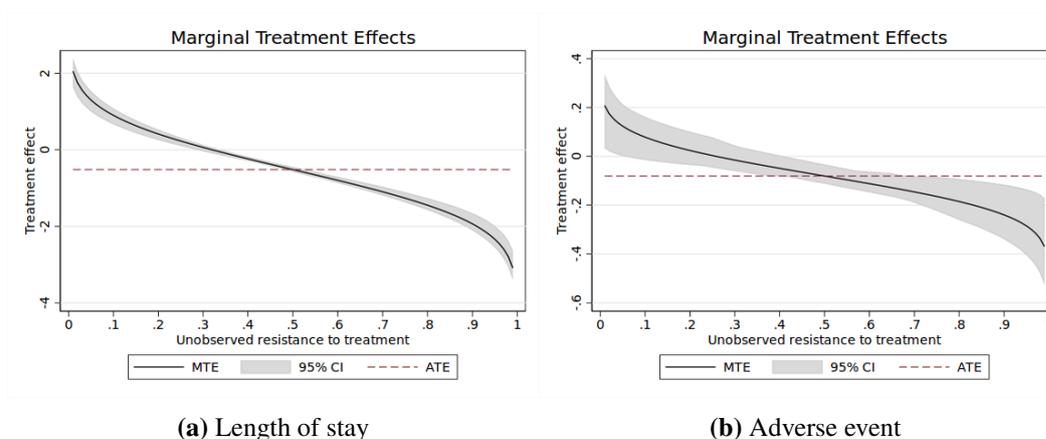
	(1)	(2)	(3)	(4)
	Length of stay		Adverse event	
ATE	-1.112*** (0.107)	-0.409*** (0.028)	0.139 (0.0768)	-0.0598*** (0.0159)
ATT	-0.0052 (0.089)	0.262*** (0.052)	-0.082 (0.052)	0.026 (0.034)
ATUT	-1.671*** (0.173)	-1.055*** (0.071)	0.250* (0.121)	-0.143*** (0.039)
LATE	-0.542*** (0.049)	-0.303*** (0.022)	-0.0459 (0.026)	-0.0602*** (0.018)
Year-Month	Yes	Yes	Yes	Yes
Day of the week	Yes	Yes	Yes	Yes
Experience dummies	No	Yes	No	Yes
<i>N</i>	11824	49215	12028	50203

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1: Distribution of surgical skills

Note: Distribution of skills measure (i.e. post-operative morbidity standardised risk ratio). Measure computed as the ratio between predicted and expected morbidity (deaths and readmissions). Predicted and expected post-operative morbidity are obtained by estimating a hierarchical logistic model accounting for patients' clinical and demographic characteristics. Surgeon random effects for predicted post-operative morbidity. Estimates using all prostatectomy patients from 2005 to 2007.

Figure 2: MTE curve – Normal with surgeon level SRR**(a)** Length of stay**(b)** Adverse event

Note: Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on log length of stay (a) and probability of adverse event (b). The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p , parametrically under the assumption of $K(p)$ is normal. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon's skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions. SRR computed using surgeon random intercept and included as a continuous measure.

Table 12: Estimated coefficients - Skills measured with fixed effect model

	(1)	(2)	(3)	(4)
	Length of stay	Adverse Event	Length of stay	Adverse Event
Skills (FE)	- 0.110*** (0.004)	-0.0196*** (0.003)		
Skills (SRR FE)			-0.061*** (0.003)	-0.009*** (0.002)
Skills (FE) * Propensity Score	0.132*** (0.008)	0.018*** (0.005)		
Skills (SRR FE) * Propensity Score			0.083*** (0.009)	0.004 (0.006)
<i>N</i>	49215	50203	49215	50203

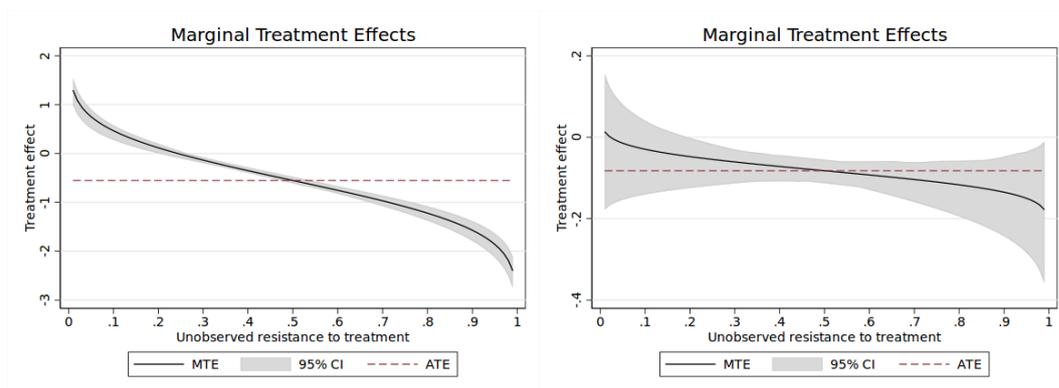
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ **Table 13:** Estimated effects - Skills measured with fixed effect model

	Skills (FE)		Skills (SRR FE)	
	(1)	(2)	(3)	(4)
	Length of stay	Adverse Event	Length of stay	Adverse Event
ATE	-0.535*** (0.030)	-0.077*** (0.017)	-0.554*** (0.024)	-0.078*** (0.015)
ATT	0.411*** (0.056)	0.014 (0.027)	0.287*** (0.054)	0.012 (0.028)
ATUT	-1.445*** (0.073)	-0.165*** (0.039)	-1.362*** (0.067)	-0.164*** (0.035)
LATE	-0.394*** (0.025)	-0.069*** (0.012)	-0.444*** (0.017)	-0.081*** (0.012)
<i>N</i>	49215	50203	49215	50203

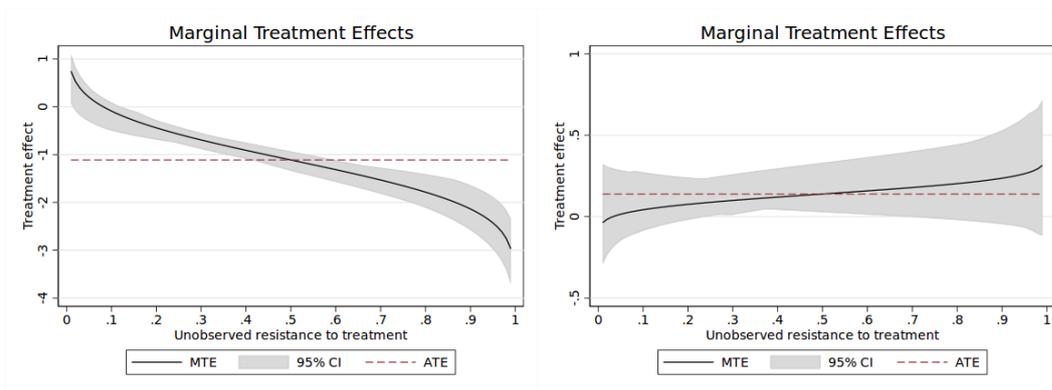
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 3: MTE curve – Normal with surgeon level SRR**(a)** Length of stay**(b)** Adverse event

Note: Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on log length of stay (a) and probability of adverse event (b). The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p , parametrically under the assumption of $K(p)$ is normal. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon's skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions. SRR computed using surgeon random intercept and included as dummy variable that takes value 1 if surgeons above median.

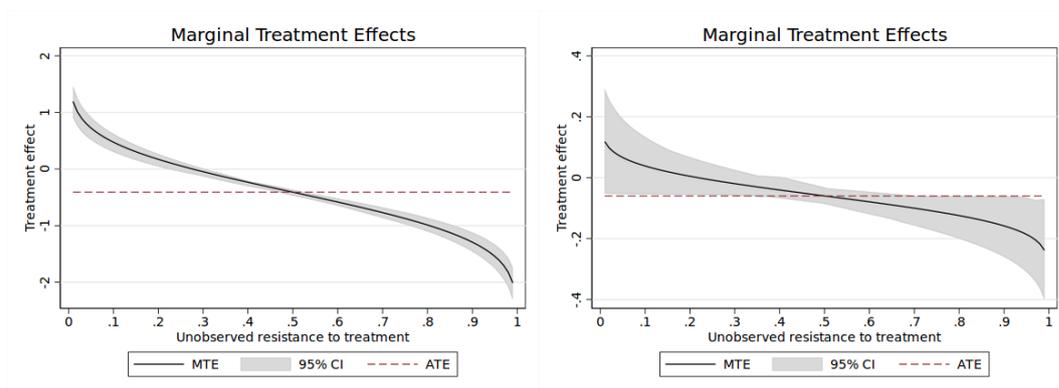
Figure 4: MTE curve – Normal with sample restriction by experience



(a) Length of stay

(b) Adverse event

Note: Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on log length of stay (a) and probability of adverse event (b). The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p , parametrically under the assumption of $K(p)$ is normal. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon's skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions. Include area fixed effects (not interacted with propensity score). Sample restricted to surgeons working in 2003.

Figure 5: MTE curve – Normal with experience dummies**(a)** Length of stay**(b)** Adverse event

Note: Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on log length of stay (a) and probability of adverse event (b). The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p , parametrically under the assumption of $K(p)$ is normal. All specifications use the instruments Z_{dist} Z_{days} as the excluded variables, and control age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, surgeon's skills (measured in the period pre-robot), and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions. Include area fixed effects (not interacted with propensity score). Model includes dummies for the first year the surgeon is observed working.

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):3730–64.

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier.

Acemoglu, D. and Restrepo, P. (2020). Robots and jobs: Evidence from us labor markets. *Journal of Political Economy*, 128(6):2188–2244.

Aiken, L. H., Clarke, S. P., and Sloane, D. M. (2000). Hospital staffing, organization, and quality of care: cross-national findings. page 9.

Aiken, L. H., Sloane, D. M., Cimiotti, J. P., Clarke, S. P., Flynn, L., Seago, J. A., Spetz, J., and Smith, H. L. (2010). Implications of the California Nurse Staffing Mandate for Other States: Implications of the California Nurse Staffing. *Health Services Research*, 45(4):904–921.

Andresen, M. E. (2018). Exploring marginal treatment effects: Flexible estimation using stata. *The Stata Journal*, 18(1):118–158.

Antinaho, T., Kivinen, T., Turunen, H., and Partanen, P. (2015). Nurses' working time use—how value adding it is? *Journal of nursing management*, 23(8):1094–1105.

- Banerjee, S. and Basu, A. (2021). Estimating endogenous treatment effects using latent factor models with and without instrumental variables. *Econometrics*, 9(1).
- Barker, H. R., Griffiths, P., Mesa-Eguiagaray, I., Pickering, R., Gould, L., and Bridges, J. (2016). Quantity and quality of interaction between staff and older patients in uk hospital wards: A descriptive study. *International Journal of Nursing Studies*, 62:100–107.
- Baumann, L. A., Baker, J., and Elshaug, A. G. (2018). The impact of electronic health record systems on clinical documentation times: A systematic review. *Health Policy*, 122(8):827–836.
- Birkmeyer, J. D., Finks, J. F., O'reilly, A., Oerline, M., Carlin, A. M., Nunn, A. R., Dimick, J., Banerjee, M., and Birkmeyer, N. J. (2013a). Surgical skill and complication rates after bariatric surgery. *New England Journal of Medicine*, 369(15):1434–1442.
- Birkmeyer, J. D., Reames, B. N., McCulloch, P., Carr, A. J., Campbell, W. B., and Wennberg, J. E. (2013b). Understanding of regional variation in the use of surgery. *The Lancet*, 382(9898):1121–1129.
- Björklund, A. and Moffitt, R. (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics*, 69(1).
- Bolla, M., van Poppel, H., Tombal, B., Vekemans, K., Da Pozzo, L., De Reijke, T. M., Verbaeys, A., Bosset, J.-F., Van Velthoven, R., Colombel, M., et al. (2012). Postoperative radiotherapy after radical prostatectomy for high-risk prostate cancer: long-term results of a randomised controlled trial (eortc trial 22911). *The Lancet*, 380(9858):2018–2027.
- Borghans, L., Green, F., and Mayhew, K. (2001). Skills measurement and economic analysis: an introduction. *Oxford Economic Papers*, pages 375–384.
- Breg, N. (2022). Medical technologies with comparative advantages on different dimensions: Evidence from hysterectomy.

- Brinch, C. N., Mogstad, M., and Wiswall, M. (2017). Beyond late with a discrete instrument. *Journal of Political Economy*, 125(4):985–1039.
- Butler, R., Monsalve, M., Thomas, G. W., Herman, T., Segre, A. M., Polgreen, P. M., and Suneja, M. (2018). Estimating Time Physicians and Other Health Care Workers Spend with Patients in an Intensive Care Unit Using a Sensor Network. *The American Journal of Medicine*, 131(8):972.e9–972.e15. Publisher: Elsevier.
- Cannaby, A.-M., Carter, V., Hoe, T., Strobel, S., Tafti, E. A., Baker, R., Nash, C., Whatley, V., and Gray, R. (2022a). The feasibility and validity of using a real time location system (RTLS) to measure bedside contact time. *Journal of Research in Nursing*, 27(5):421–433.
- Cannaby, A.-M., Carter, V., Strobel, S., and Tafti, E. A. (2022b). Measuring nursing movements using real time tracking data at the Royal Wolverhampton Hospital Trust. *Spatial and Spatio-temporal Epidemiology*, 43:100543.
- Carayon, P., Wetterneck, T. B., Alyousef, B., Brown, R. L., Cartmill, R. S., McGuire, K., Hoonakker, P. L., Slagle, J., Van Roy, K. S., Walker, J. M., Weinger, M. B., Xie, A., and Wood, K. E. (2015). Impact of electronic health record technology on the work and workflow of physicians in the intensive care unit. *International Journal of Medical Informatics*, 84(8):578–594.
- Card, D., Fenizia, A., and Silver, D. (2019). The health impacts of hospital delivery practices. Technical report, National Bureau of Economic Research.
- Carneiro, P., Heckman, J. J., and Vytlačil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, 101(6).
- Chan, D. C., Gentzkow, M., and Yu, C. (2022). Selection with variation in diagnostic skill: Evidence from radiologists. *The Quarterly Journal of Economics*, 137(2):729–783.
- Chandra, A. and Skinner, J. S. (2003). Geography and racial health disparities.

- Chandra, A. and Staiger, D. O. (2007). Productivity spillovers in health care: evidence from the treatment of heart attacks. *Journal of political Economy*, 115(1):103–140.
- Chandra, A. and Staiger, D. O. (2020). Identifying sources of inefficiency in health-care. *The quarterly journal of economics*, 135(2):785–843.
- Chesher, A. (2005). Nonparametric identification under discrete variation. *Econometrica*, 73(5):1525–1550.
- Coelho, R. F., Rocco, B., Patel, M. B., Orvieto, M. A., Chauhan, S., Ficarra, V., Melegari, S., Palmer, K. J., and Patel, V. R. (2010). Retropubic, laparoscopic, and robot-assisted radical prostatectomy: a critical review of outcomes reported by high-volume centers. *Journal of endourology*, 24(12):2003–2015.
- Compagni, A., Mele, V., and Ravasi, D. (2015). How early implementations influence later adoptions of innovation: Social positioning and skill reproduction in the diffusion of robotic surgery. *Academy of Management Journal*, 58(1):242–278.
- Cook, A., Gaynor, M., Stephens, M., and Taylor, L. (2010). The Effect of Hospital Nurse Staffing on Patient Health Outcomes: Evidence from California’s Minimum Staffing Regulation. Technical Report w16077, National Bureau of Economic Research, Cambridge, MA.
- Cooper, Z., Gibbons, S., Jones, S., and McGuire, A. (2010a). Does hospital competition improve efficiency? an analysis of the recent market-based reforms to the english nhs.
- Cooper, Z., Gibbons, S., Jones, S., and McGuire, A. (2010b). Does hospital competition save lives?: evidence from the english nhs patient choice reforms. Technical report, London School of Economics and Political Science, LSE Library.
- Cooper, Z., Gibbons, S., Jones, S., and McGuire, A. (2011). Does hospital competi-

- tion save lives? evidence from the english nhs patient choice reforms. *Economic Journal*, 121(554):228–260.
- Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2016). From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics*, 41:47–60.
- Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who benefits from universal child care? estimating marginal returns to early child care attendance. *Journal of Political Economy*, 126(6):2356–2409.
- Coughlin, G. D., Yaxley, J. W., Chambers, S. K., Occhipinti, S., Samaratunga, H., Zajdlewicz, L., Teloken, P., Dungleison, N., Williams, S., Lavin, M. F., et al. (2018). Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: 24-month outcomes from a randomised controlled study. *The Lancet Oncology*, 19(8):1051–1060.
- Craig, S., Grennan, M., and Swanson, A. (2019). Mergers and marginal costs: New evidence on hospital buyer power. Working paper.
- 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. M. and McClellan, M. (2001). Is technological change in medicine worth it? *Health affairs*, 20(5):11–29.
- Cutler, D. M. and Morton, F. S. (2013). Hospitals, market share, and consolidation. *JAMA*, 310(18):1964–70.
- Dafny, Leemore S., K. H. and Lee, R. S. (2019). The price effects of cross-market mergers: Theory and evidence from the hospital industry. *RAND Journal of Economics*, 50(2).

- Dafny, L., Ho, K., and Lee, R. S. (2019). The price effects of cross-market mergers: theory and evidence from the hospital industry. *The RAND Journal of Economics*, 50(2):286–325.
- Dall’Ora, C., Griffiths, P., Hope, J., Briggs, J., Jeremy, J., Gerry, S., and Redfern, O. C. (2021). How long do nursing staff take to measure and record patients’ vital signs observations in hospital? A time-and-motion study. *International Journal of Nursing Studies*, 118:103921.
- Davies, B. (2022). Waiting for godot robot.
- Deaton, A. (2003). Health, inequality, and economic development. *Journal of economic literature*, 41(1):113–158.
- Donaldson, S. I. and Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of business and Psychology*, 17:245–260.
- Duffield, C., Diers, D., O’Brien-Pallas, L., Aisbett, C., Roche, M., King, M., and Aisbett, K. (2011). Nursing staffing, nursing workload, the work environment and patient outcomes. *Applied nursing research*, 24(4):244–255.
- Edwards, A., Fitzpatrick, L.-A., Augustine, S., Trzebucki, A., Cheng, S. L., Presseau, C., Mersmann, C., Heckman, B., and Kachnowski, S. (2009). Synchronous communication facilitates interruptive workflow for attending physicians and nurses in clinical settings. *International journal of medical informatics*, 78(9):629–637.
- 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.
- Gaynor, M. (2006). What Do We Know About Competition and Quality in Health Care Markets? NBER Working Papers 12301, National Bureau of Economic Research, Inc.

- 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).
- George, E. I., Brand, C. T. C., Marescaux, J., et al. (2018). Origins of robotic surgery: from skepticism to standard of care. *JSLs: Journal of the Society of Laparoendoscopic Surgeons*, 22(4).
- Government of Nova Scotia (2020). Physician Administrative Burden Survey – Final Report. Technical report, Office of Regulatory Affairs and Service Effectiveness.
- Gowrisankaran, G. and Town, R. J. (1999). Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics*, 18(6).
- Griffiths, P., Maruotti, A., Recio Saucedo, A., Redfern, O. C., Ball, J. E., Briggs, J., Dall’Ora, C., Schmidt, P. E., and Smith, G. B. (2019). Nurse staffing, nursing assistants and hospital mortality: retrospective longitudinal cohort study. *BMJ Quality & Safety*, 28(8):609–617.
- Grullon, G., Larkin, Y., and Michaely, R. (2019). Are US Industries Becoming More Concentrated? *Review of Finance*, 23(4):697–743.
- Gupta, A. (2017). Impacts of performance pay for hospitals : The readmissions reduction program.
- Gutierrez Gallardo, G. and Philippon, T. (2017). Declining Competition and Investment in the U.S. *NBER Working Paper*, (w23583).
- Hadley, J. and Cunningham, P. (2004). Availability of safety net providers and access to care of uninsured persons. *Health services research*, 39(5):1527–1546.
- Haegdorens, F., Van Bogaert, P., De Meester, K., and Monsieurs, K. G. (2019). The impact of nurse staffing levels and nurse’s education on patient mortality in medical and surgical wards: an observational multicentre study. *BMC Health Services Research*, 19(1):864.

- Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The review of economics and statistics*, 88(3):389–432.
- Heckman, J. J. and Vytlacil, E. (2001). Policy-relevant treatment effects. *American Economic Review*, 91(2):107–111.
- Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3).
- Heckman, J. J. and Vytlacil, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences*, 96(8).
- Hendrich, A., Chow, M. P., Skierczynski, B. A., and Lu, Z. (2008). A 36-hospital time and motion study: how do medical-surgical nurses spend their time? *The Permanente Journal*, 12(3):25.
- Higgins, R. M., Frelich, M. J., Bosler, M. E., and Gould, J. C. (2017). Cost analysis of robotic versus laparoscopic general surgery procedures. *Surgical endoscopy*, 31(1):185–192.
- Ho, C., Tsakonas, E., Tran, K., Cimon, K., Severn, M., Mierzwinski-Urban, M., Corcos, J., and Pautler, S. (2013). Robot-assisted surgery compared with open surgery and laparoscopic surgery: clinical effectiveness and economic analyses.
- Ho, V. and Hamilton, B. H. (2000). Hospital mergers and acquisitions: does market consolidation harm patients? *Journal of Health Economics*, 19(5):767–791.
- Horn, D., Sacarny, A., and Zhou, A. (2022). Technology adoption and market allocation: The case of robotic surgery. *Journal of Health Economics*, 86:102672.
- Horwitz, L. I., Partovian, C., Lin, Z., Grady, J. N., Herrin, J., Conover, M., Montague, J., Dillaway, C., Bartczak, K., Suter, L. G., et al. (2014). Development and use of an administrative claims measure for profiling hospital-wide

- performance on 30-day unplanned readmission. *Annals of internal medicine*, 161(10_Supplement):S66–S75.
- Hull, P. (2018). Estimating hospital quality with quasi-experimental data. *Available at SSRN 3118358*.
- Humlum, A. (2019). Robot adoption and labor market dynamics. *Princeton University*.
- Hussain, A., Malik, A., Halim, M. U., and Ali, A. M. (2014). The use of robotics in surgery: a review. *International journal of clinical practice*, 68(11):1376–1382.
- Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2).
- Imbens, G. W. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, 64(4).
- Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.
- Jayant Ketkar, H., Carnahan, S., and Greenwood, B. N. (2022). Do new job tools improve women’s performance in male-dominated fields? evidence from robotic surgery. *Evidence from Robotic Surgery (March 28, 2022)*.
- Joukes, E., Abu-Hanna, A., Cornet, R., and de Keizer, N. (2018). Time Spent on Dedicated Patient Care and Documentation Tasks Before and After the Introduction of a Structured and Standardized Electronic Health Record. *Applied Clinical Informatics*, 09(01):046–053.
- Kamel Boulos, M. N. and Berry, G. (2012). Real-time locating systems (rtls) in healthcare: a condensed primer. *International journal of health geographics*, 11(1):1–8.
- Kolstad, J. T. (2013a). Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, 103(7):2875–2910.

- Kolstad, J. T. (2013b). Information and quality when motivation is intrinsic: Evidence from surgeon report cards. *American Economic Review*, 103(7).
- Lam, K., Clarke, J., Purkayastha, S., and Kinross, J. M. (2021). Uptake and accessibility of surgical robotics in england. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 17(1):1–7.
- Leora, Horwitz, C. P. Z. L. J. H. J. G. M. C. J. M. C. D. K. B. L. S. J. R. S. B. H. K. E. D. (2012). Hospital-wide all-cause unplanned readmission measure.
- Lewis, M. S. and Pflum, K. E. (2017). Hospital systems and bargaining power: evidence from out-of-market acquisitions. *The RAND Journal of Economics*, 48(3):579–610.
- Loecker, J. D. and Eeckhout, J. (2018). Global market power. NBER Working Paper No. 24768.
- Lotan, Y. (2012). Is robotic surgery cost-effective: no. *Current opinion in urology*, 22(1):66–69.
- Lowrance, W. T., Elkin, E. B., Jacks, L. M., Yee, D. S., Jang, T. L., Laudone, V. P., Guillonneau, B. D., Scardino, P. T., and Eastham, J. A. (2010). Comparative effectiveness of prostate cancer surgical treatments: a population based analysis of postoperative outcomes. *The Journal of urology*, 183(4):1366–1372.
- Mahtani, K., Spencer, E. A., Brassey, J., and Heneghan, C. (2018). Catalogue of bias: observer bias. *BMJ evidence-based medicine*, 23(1):23–24.
- Marcus, H. J., Hughes-Hallett, A., Payne, C. J., Cundy, T. P., Nandi, D., Yang, G.-Z., and Darzi, A. (2017). Trends in the diffusion of robotic surgery: A retrospective observational study. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 13(4):e1870.
- Maynou, L., Mehtsun, W. T., Serra-Sastre, V., and Papanicolas, I. (2021). Patterns of adoption of robotic radical prostatectomy in the united states and england. *Health Services Research*.

- Maynou, L., Pearson, G., McGuire, A., and Serra-Sastre, V. (2022). The diffusion of robotic surgery: Examining technology use in the english nhs. *Health Policy*, 126(4):325–336.
- McClellan, M., McNeil, B. J., and Newhouse, J. P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: analysis using instrumental variables. *Jama*, 272(11).
- McClellan, M. and Newhouse, J. P. (1997). The marginal cost-effectiveness of medical technology: a panel instrumental-variables approach. *Journal of Econometrics*, 77(1):39–64.
- McHugh, M. D., Aiken, L. H., Sloane, D. M., Windsor, C., Douglas, C., and Yates, P. (2021). Effects of nurse-to-patient ratio legislation on nurse staffing and patient mortality, readmissions, and length of stay: a prospective study in a panel of hospitals. *The Lancet*, 397(10288):1905–1913.
- Mogstad, M., Torgovitsky, A., and Walters, C. R. (2021). The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review*, 111(11):3663–98.
- Molitor, D. (2018). The evolution of physician practice styles: evidence from cardiologist migration. *American Economic Journal: Economic Policy*, 10(1):326–56.
- Mookadam, F. and Arthur, H. M. (2004). Social support and its relationship to morbidity and mortality after acute myocardial infarction: systematic overview. *Archives of internal medicine*, 164(14):1514–1518.
- Musy, S. N., Endrich, O., Leichtle, A. B., Griffiths, P., Nakas, C. T., and Simon, M. (2021). The association between nurse staffing and inpatient mortality: A shift-level retrospective longitudinal study. *International Journal of Nursing Studies*, 120:103950.
- Nash, C. (2014). Infection prevention in ‘safe hands’. *British Journal of Nursing*, 23(21):1123–1123.

- Needleman, J., Buerhaus, P., Pankratz, V. S., Leibson, C. L., Stevens, S. R., and Harris, M. (2011). Nurse Staffing and Inpatient Hospital Mortality. *New England Journal of Medicine*, 364(11):1037–1045. Publisher: Massachusetts Medical Society.
- Nelson, B., Kaufman, M., Broughton, G., Cookson, M. S., Chang, S. S., Herrell, S. D., Baumgartner, R. G., and Smith, J. A. (2007). Comparison of length of hospital stay between radical retropubic prostatectomy and robotic assisted laparoscopic prostatectomy. *The Journal of urology*, 177(3):929–931.
- Neuner, J. M., See, W. A., Pezzin, L. E., Tarima, S., and Nattinger, A. B. (2012). The association of robotic surgical technology and hospital prostatectomy volumes: increasing market share through the adoption of technology. *Cancer*, 118(2):371–377.
- OMA (2021). Healing the Healers: System-Level Solutions to Physician Burnout. Technical report, Ontario Medical Association.
- Overman, D. (2022). Ai platform increases efficiency at tampa general hospital. *AXIS Imaging News*.
- Prager, E. and Schmitt, M. (2019). Employer consolidation and wages: Evidence from hospitals. *Washington Center for Equitable Growth Working Paper*.
- Propper, C., Laudicella, M., and Gaynor, M. (2012). Can governments do it better? merger mania and hospital outcomes in the english nhs. Working papers, Imperial College, London, Imperial College Business School.
- Ramirez, E. (2014). Antitrust enforcement in health care — controlling costs, improving quality. *New England Journal of Medicine*, 371(24):2245–2247.
- Robertson, C., Close, A., Fraser, C., Gurung, T., Jia, X., Sharma, P., Vale, L., Ramsay, C., and Pickard, R. (2013). Relative effectiveness of robot-assisted and standard laparoscopic prostatectomy as alternatives to open radical prostatectomy for

- treatment of localised prostate cancer: a systematic review and mixed treatment comparison meta-analysis. *BJU international*, 112(6):798–812.
- Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2).
- Schiraldi, Pasquale, W. R. (2019). Does hospital competition reduce rates of patient harm in the english nhs? *Working Paper*.
- Schmitt, M. (2017). Do hospital mergers reduce costs? *Journal of health economics*, 52:74–94.
- Schmitt, M. (2018). Multimarket contact in the hospital industry. *American Economic Journal: Economic Policy*, 10(3).
- Shekelle, P. G. (2013). Nurse–Patient Ratios as a Patient Safety Strategy. *Annals of Internal Medicine*, 158(5_Part_2):404–409. Publisher: American College of Physicians.
- Skinner, J. (2011). Causes and consequences of regional variations in health care. In *Handbook of health economics*, volume 2, pages 45–93. Elsevier.
- Strother, M. C., Michel, K. F., Xia, L., McWilliams, K., Guzzo, T. J., Lee, D. J., and Lee, D. I. (2020). Prolonged length of stay after robotic prostatectomy: causes and risk factors. *Annals of Surgical Oncology*, 27(5):1560–1567.
- Suri, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica*, 79(1):159–209.
- Tonutti, M., Elson, D. S., Yang, G.-Z., Darzi, A. W., and Sodergren, M. H. (2017). The role of technology in minimally invasive surgery: state of the art, recent developments and future directions. *Postgraduate Medical Journal*, 93(1097).
- Tsugawa, Y., Jha, A. K., Newhouse, J. P., Zaslavsky, A. M., and Jena, A. B. (2017). Variation in physician spending and association with patient outcomes. *Jama internal medicine*, 177(5):675–682.

- Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1).
- Westbrook, J. I., Duffield, C., Li, L., and Creswick, N. J. (2011). How much time do nurses have for patients? a longitudinal study quantifying hospital nurses' patterns of task time distribution and interactions with health professionals. *BMC health services research*, 11:1–12.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.
- Yaxley, J. W., Coughlin, G. D., Chambers, S. K., Occhipinti, S., Samaratunga, H., Zajdlewicz, L., Duglison, N., Carter, R., Williams, S., Payton, D. J., et al. (2016). Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study. *The Lancet*, 388(10049):1057–1066.
- Yen, P.-Y., Kelley, M., Lopetegui, M., Loversidge, J., Chipps, E. M., Gallagher-Ford, L., and Buck, J. (2018a). Nurses' Time Allocation and Multitasking of Nursing Activities: A Time Motion Study.
- Yen, P.-Y., Kellye, M., Lopetegui, M., Saha, A., Loversidge, J., Chipps, E. M., Gallagher-Ford, L., and Buck, J. (2018b). Nurses' time allocation and multitasking of nursing activities: a time motion study. In *AMIA annual symposium proceedings*, volume 2018, page 1137. American Medical Informatics Association.
- Zaranko, B., Sanford, N. J., Kelly, E., Rafferty, A. M., Bird, J., Mercuri, L., Sigsworth, J., Wells, M., and Proper, C. (2022). Nurse staffing and inpatient mortality in the English National Health Service: a retrospective longitudinal study. *BMJ Quality & Safety*, pages bmjqs–2022–015291.
- Zaranko, B., Sanford, N. J., Kelly, E., Rafferty, A. M., Bird, J., Mercuri, L., Sigsworth, J., Wells, M., and Proper, C. (2023). Nurse staffing and inpatient

mortality in the english national health service: a retrospective longitudinal study. *BMJ quality & safety*, 32(5):254–263.

Zhou, X. and Xie, Y. (2019). Marginal treatment effects from a propensity score perspective. *Journal of Political Economy*, 127(6).