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

Elena Ashtari Tafti

University College London[1]

7 November 2022

Abstract. This paper investigates the potential of new technologies to reduce disparities in the provision of healthcare services. Differences in providers' skills may cause variation in patient outcomes. The adoption of innovations, like robots, can attenuate this problem if technological gains are decreasing in users' skills or may exacerbate existing variation in performance otherwise. I show that, in England, the diffusion of surgical robots coincided with an improvement in average surgical performance and a convergence in outcomes between high and lower-skilled surgeons for prostate cancer patients. I study whether this pattern can be attributed to the adoption of robots using the universe of inpatient admissions to the National Health Service (NHS). To identify the effects of robotic surgery on patient outcomes, I exploit quasi-random variation in the geographic allocation of robots, allowing for selection and heterogeneity in treatment effects. I find that robots shorten patients' length of stay in hospital and decrease the incidence of adverse events from surgery, but their effects significantly depend on surgeons' skills. The robot has little impact on the performance of highly skilled surgeons, while lower skilled surgeons gain the most from it. I also uncover a strong pattern of negative selection on both observable and unobservable characteristics. Although the attainable gains are higher for lower-skilled surgeons, they use the robot the least. My results suggest that the potential benefit of a new technology largely depends on how it combines with the skills of the individual users.

[1] elena.tafti.17@ucl.ac.uk. I wish to acknowledge the financial support provided by the Institute of Fiscal Studies for the completion of my doctoral thesis. I would also like to show my appreciation to my supervisors, Aureo de Paula and Marcos Vera Hernandez, their door was always open when I needed them. I am also grateful for the support I have received from David C. Chan. I am thankful to Antonio Ashtari Tafti, Oliver Ashtari Tafti, Nathaniel Breg, Alexander Clyde, Andrew Chesher, Pedro Carneiro, Pete Cook, Yasmin Khan, Alice Kuegler, Jonas Hjort, Thomas P. Hoe, Thomas Lazarowicz, Mikkel Mertz, Lars Nesheim, Nikita Roketskiy, and George Stoye for their comments and feedback on this project. Many thanks also to Martin Eckhoff Andresen for answering my questions on estimation the model using Stata. Lastly, I am grateful for Lodovico Felice de Vito.

1 Introduction

Disparities in access and quality of services concern regulators and policy markers. This is particularly true in healthcare, but also relevant in education, or the justice system. As a result, 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, 2013). 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 a structural 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 suggest 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 describes surgical robots and their use for prostate cancer surgery. Section 3 presents the data. Section 4 discusses how I measure surgeons' skills and provides the empirical facts that have motivated this work. Section 5 presents the econometric model and the conditions required for identification and estimation of the parameters. Section 6 introduces the instrumental variables I will use to identify the model parameters and discusses their validity. Section 7 summarizes the results. Finally, Section 8 concludes.

2 Robotic surgery and the treatment of 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 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).

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.[2] 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

[2] https://prostatecanceruk.org

Figure 1: Picture of a Da Vinci surgical system

Note: Picture shows the Da Vinci Robot surgical system from Intuitive Inc. On the left the surgeon sitting at the console. Above the operating bed the robotic arms. On the right the vision cart.

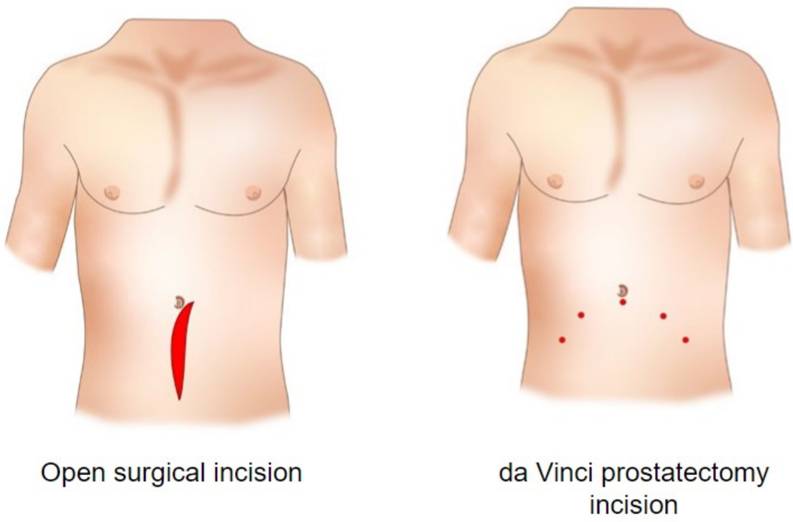
had a robot-assisted operation. Eventually, by 2014, robotic surgery accounted for up to 90 percent of radical prostatectomies across the US.[3] This trend has been similar in England where, by 2014, the majority of cases (62.7 percent) were performed robotically (Marcus et al., 2017).

Before robots, prostate cancer surgery was usually performed with an 'open' 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.[4] 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) and complex manual tools.

[3] https://www.nature.com/articles/d41586-020-01037-w

[4] Other 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.

Figure 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, 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).

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.

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 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.

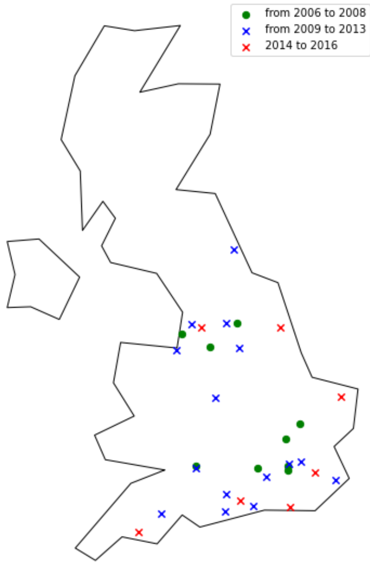
Eventually, my sample comprises all radical prostatectomies occurring throughout NHS England from 2004 to 2017 for a total of 62,258 admissions, 25208 of which are performed with a robot. Table 1 summarizes the characteristics of patients for both the traditional and the robotic approach.

Table 1: Radical Prostatectomy Patients— Sample Summary Statistics —2004/2017

	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.

Figure 3: Hospital level diffusion of robotic surgery—Timing of adoption

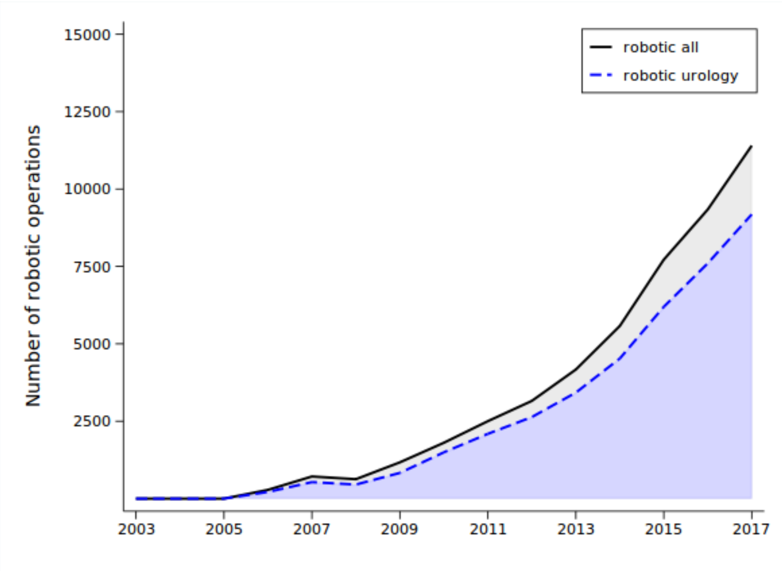


Note: Produced using HES. The green crosses represent hospitals that are observed using for the first time a robotic operation code for RP in 2006. The blue dots represent hospitals that are observed using for the first time a robotic operation code for RP between 2007 and 2009. The red dots represent hospitals that are observed using for the first time a robotic operation code for RP after 2009.

HES shows that prostate cancer surgery is England's most commonly performed robotic operation. In Figure 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 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 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 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.

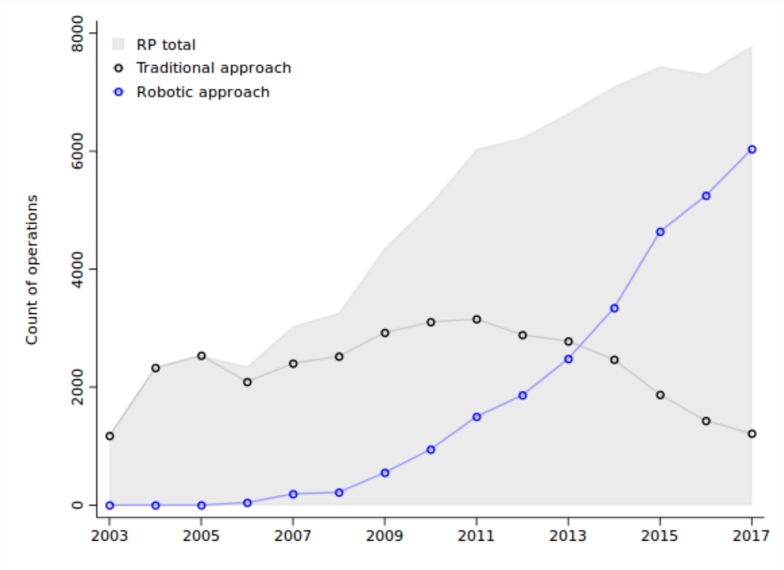
Figure 4: Diffusion of robotic surgery in the NHS



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.

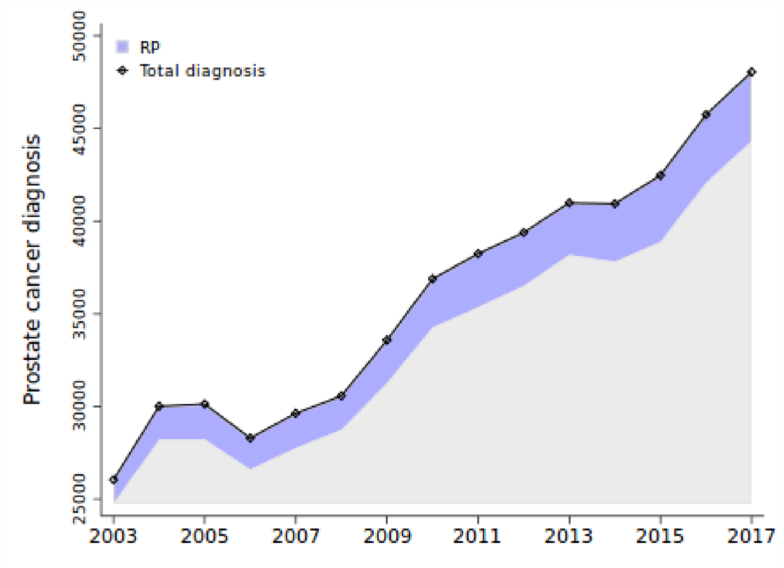
From HES, I identify two patient outcomes for which I evaluate the effect of robots. Namely, patients' length of stay and the occurrence of adverse events from surgery. I focus on these patient outcomes for several reasons. First, they are important margins of performance for patients. Undoubtedly, patients 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 incidence of complications should diminish. Lastly, these are outcomes that I can reliably measure from the data I have.

Figure 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 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.

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., 2010). 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.[5]

Table 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 from surgery.

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 same level of

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

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:

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

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

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

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 represents the empirical Bayes estimate also known as posterior mean estimate of the random effect. 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 10.

In the second step, I use the regression estimates from Equation 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:

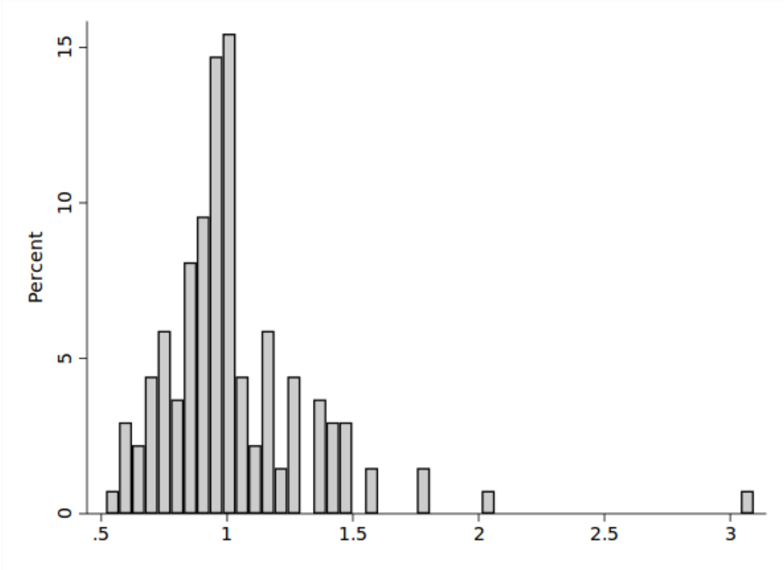
$$\text{predicted}_j = \sum_{i \in j} logit^{-1}\left(\widehat{\alpha}_j + \widehat{\beta} X_{ij}\right) \tag{2}$$

$$\text{expeted}_j = \sum_{i \in j} logit^{-1}\left(\widehat{\mu} + \widehat{\beta} X_{ij}\right) \tag{3}$$

My indicator of skills is then is:

$$\text{Skills}_j = \frac{\text{predicted}_j}{\text{expected}_j}. \tag{4}$$

Figure 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 a hierarchical logistic model accounting for patients' clinical and demographic characteristics. Hospital fixed effects for predicted post-operative morbidity. Estimates using all prostatectomy patients from 2005 to 2007.

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. In practice, I perform this estimation at the hospital level. The median number of surgeons per hospital in my sample is two, which means this simplification is unlikely to be significant. Moreover, the majority of surgeons are observed operating for the full period of observation. 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 Figure 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 tales to the right, suggesting that some surgeons perform particularly poorly.

Key facts on robotic surgery, skills, and performance

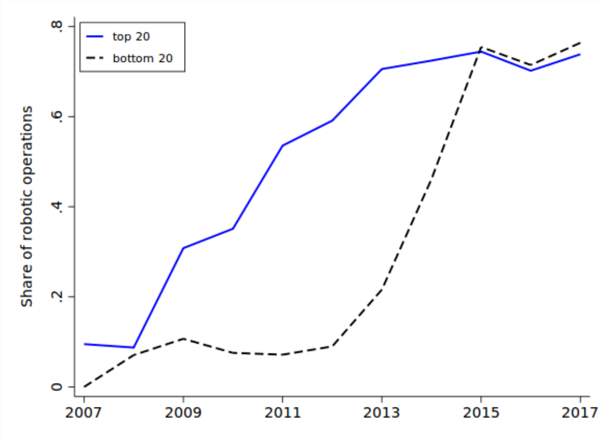
I start my analysis by showing in Figure 8 some correlations between robotic surgery, skills, and performance. I group surgeons into two categories; top and bottom surgeons. Top surgeons are identified as those above the 20th percentile of the distribution of skills (low post-operative morbidity), bottom 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 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.

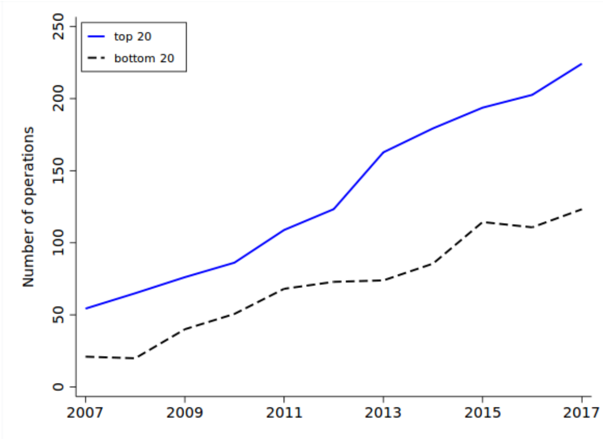
The second fact is that over this period there has been a substantial improvement in surgical performance. Post-operative length of stay and morbidity have decreased respectively by 57 and 73 percent from 2007 to 2017. But, there has also been a convergence in surgical performance between surgeons at 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 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.

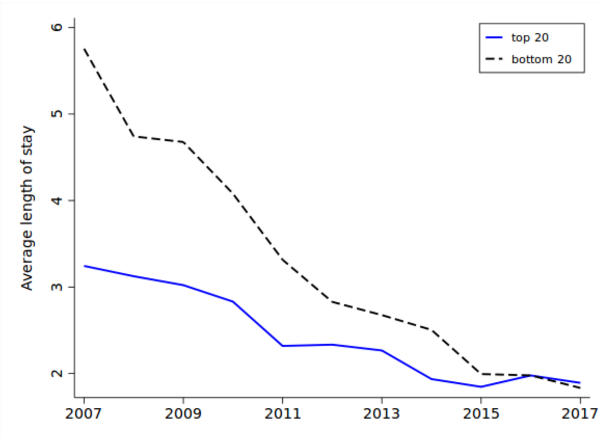
Figure 8: Key empirical facts



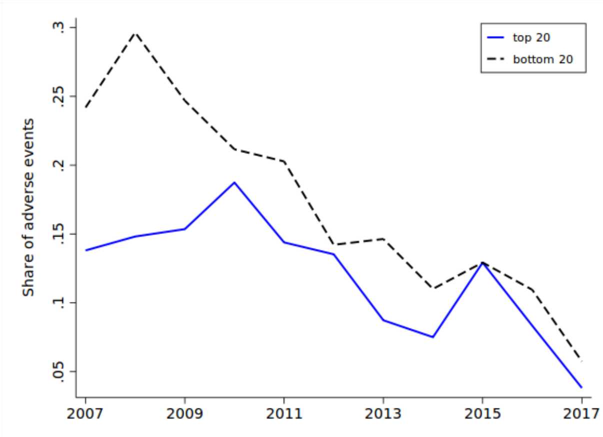
(a) Robotic rate



(b) Number of operations



(c) Length of stay



(d) Adverse events

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.

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 not 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 Vytlacil (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 Vytlacil (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 $0, 1 \in$ supp $P(Z)$ (Heckman and Vytlacil, 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 terminology 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:

$$Y_0 = h_0(X) + \epsilon_0 \tag{5}$$

$$Y_1 = h_1(X) + \epsilon_1 \tag{6}$$

$h_D(X) \equiv E[Y_D|X]$ 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

$$D = 1[D^* \geq 0] \tag{7}$$

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:

$$D^* = g(Z) - V \tag{8}$$

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 random 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.** $(\epsilon_0, \epsilon_1, V)$ 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:

$$\begin{aligned} & E[Y_1 - Y_0 | X = x, U = u] \\ &= h_1(X) - h_0(X) + E[\epsilon_1 - \epsilon_0 | X = x, U = u] \\ &\equiv MTE(x, u) \end{aligned}$$

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 $Skills_j$

$$Y_{1ij} = \beta_1 X_i + \delta_1 Skills_j + \epsilon_{1ij} \tag{9}$$

$$Y_{0ij} = \beta_0 X_i + \delta_0 Skills_j + \epsilon_{0ij} \tag{10}$$

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).

$$D_{ij} = 1[D^* \geq 0] \tag{11}$$

$$D^*_{ij} = \beta_d X_i + \delta_d Skills_j + \gamma_d Z_i - V_{ij} \tag{12}$$

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 6.

The equivalent representation in terms of the propensity score is:

$$D = 1 \text{ if } P(X_i, Skills_j, Z_i) \geq U, \text{ and } D = 0 \text{ otherwise.} \tag{13}$$

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:

$$Y_{ij} = Y_{0ij} + D_{ij}[\underbrace{(\beta_1 - \beta_0)X_i}_{\Delta_1} + \underbrace{(\delta_1 - \delta_0)Skills_j}_{\Delta_2} + \underbrace{\epsilon_{1ij} - \epsilon_{0ij}}_{\Delta_3}] \tag{14}$$

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:

$$MTE(s, x, u) = E(Y_{1ij} - Y_{0ij} | X_i = x, Skills_j = s, U_{ij} = u) \tag{15}$$

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

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

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

$$MTE(s, x, u) = x(\beta_1 - \beta_0) + s(\delta_1 - \delta_0) + E(\epsilon_{1ij} - \epsilon_{0ij} | U_{ij} = u) \tag{16}$$

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

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

where $K(p) \equiv \int_0^p E(\epsilon_{1ij} - \epsilon_{0ij} | U = u)du$ 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.

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 exogeneous 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:

$$Z_{days} = t - T_R \tag{17}$$

where t is the date on which the patient received his diagnosis of prostate cancer[6], 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 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 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

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

Table 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.

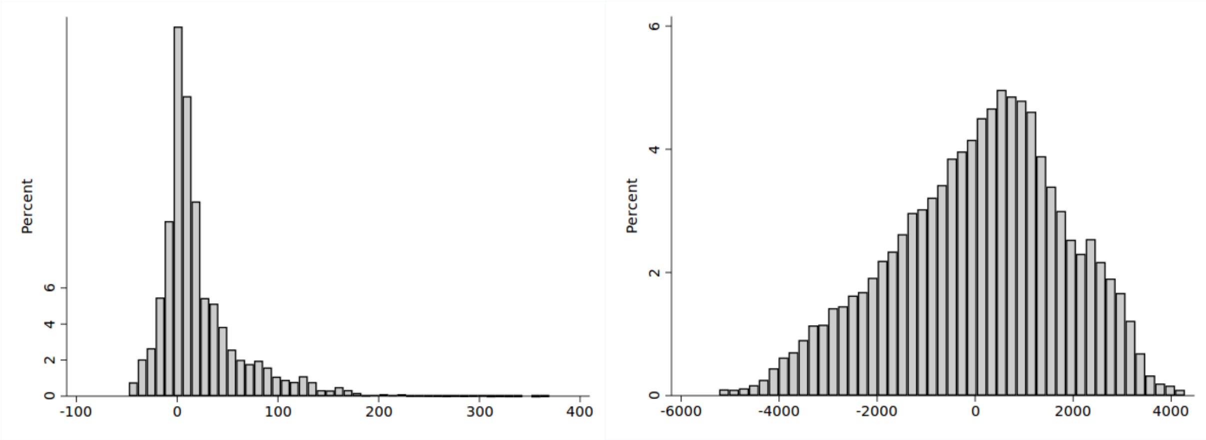
Z_{days} for increasingly richer specifications. The coefficient in column 4 is negative and statistically significant, but after controlling for patient characteristics, this correlation disappears. Overall, this is suggestive that the exclusion restriction is likely to be satisfied conditional on the covariates included in the model.

I show how Z_{days} is distributed in Figure 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 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.

Figure 9: Variation of instrumental variables in sample data



(a) Z_{dist} instrument (b) Z_{days} instrument

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.

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;

$$Z_{dist} = D_R - D_T, \tag{18}$$

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

Table 3: Correlation between AMI patients mortality and Z_{dist} - Linear regression coefficients

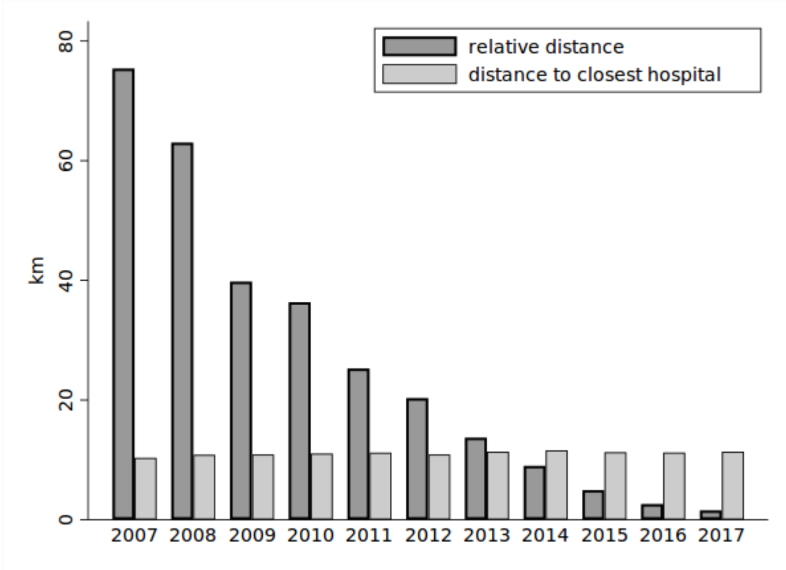
	(1)	(2)	(3)
Z_{dist}	0.045**	-0.016	0.031
	(0.016)	(0.018)	(0.019)
Distance closest hospital		0.269*	0.111
		(0.107)	(0.130)
Year-month	No	Yes	Yes
Day of the week	No	Yes	Yes
Patient control	No	No	Yes
Deaths (%)	19	19	19
Z_{dist}	68.64	68.64	68.75
N	68467	68467	67882

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.

from 2006 to 2010. Table 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. 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 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.

Figure 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.

I show how Z_{dist} is distributed in Figure 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 4, where the columns denote increasingly richer specifications. Column 7 represents the selection equation, which I will discuss in more details in Section 7.

Table 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.

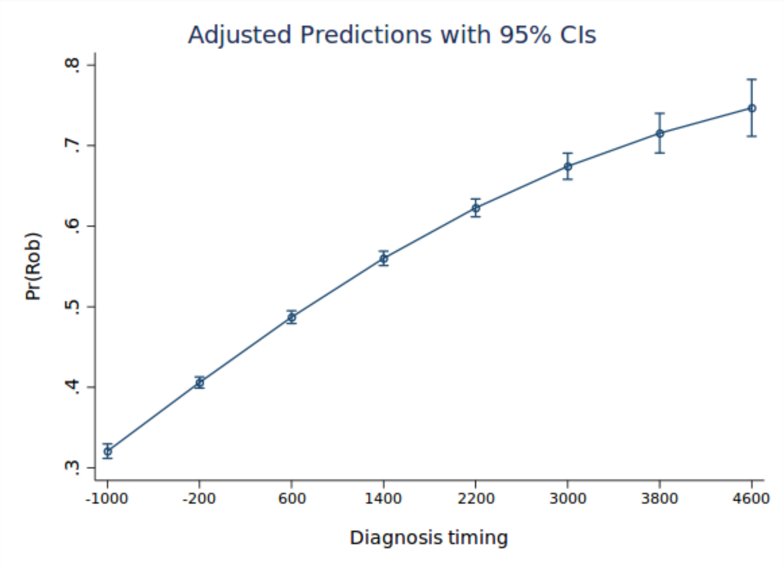
Table 4: Relevance of instruments - Probit regression dependent variable indicator of robotic surgery

	(1)	(2)	(3)	(4)	(5)	(6)
Coefficients						
Z_{dist}	-1.95***		-1.04***	-1.07***	-1.1***	-0.99***
	(0.038)		(0.032)	(0.032)	(0.033)	(0.034)
Z_{days}		0.057***	0.041***	0.041***	0.042***	0.024***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Marginal effects						
Z_{dist}	-0.651***		-0.307***	-0.316***	-0.319***	-0.272***
	(0.008)		(0.009)	(0.009)	(0.009)	(0.009)
Z_{days}		0.0159***	0.0122***	0.0119***	0.0122***	0.007***
		(0.000)	(0.001)	(0.001)	(0.001)	(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. YM indicates year-month controls, DOW indicates day of the week controls. SRR is the standardized risk ratio for post-operative morbidity (interpreted as the inverse of skills).

In Figure 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 12, I show the average

Figure 11: Estimated probability of robotic approach from selection equation - at Z_{days} values

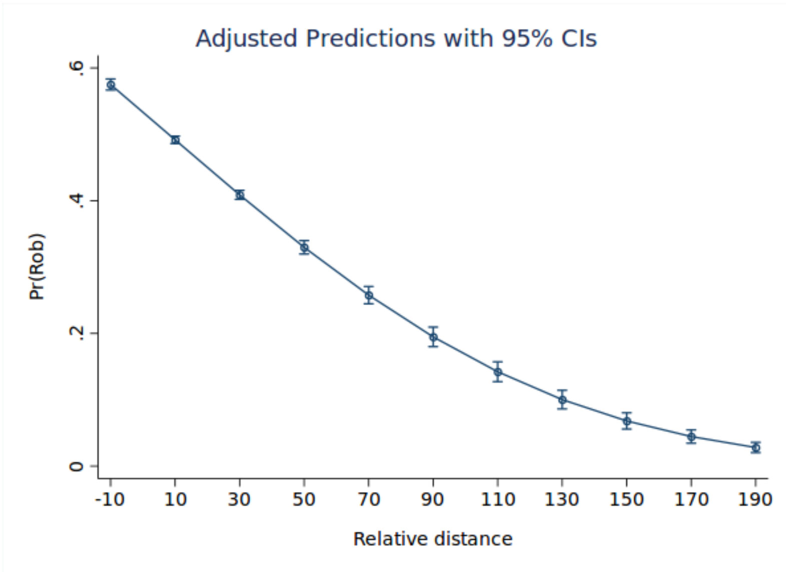


Note: Probit regression estimates, dependent variable robotic approach. Marginal probability estimated at different value of relative diagnosis timing. Covariates in the model at means, include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest 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.

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 percent.

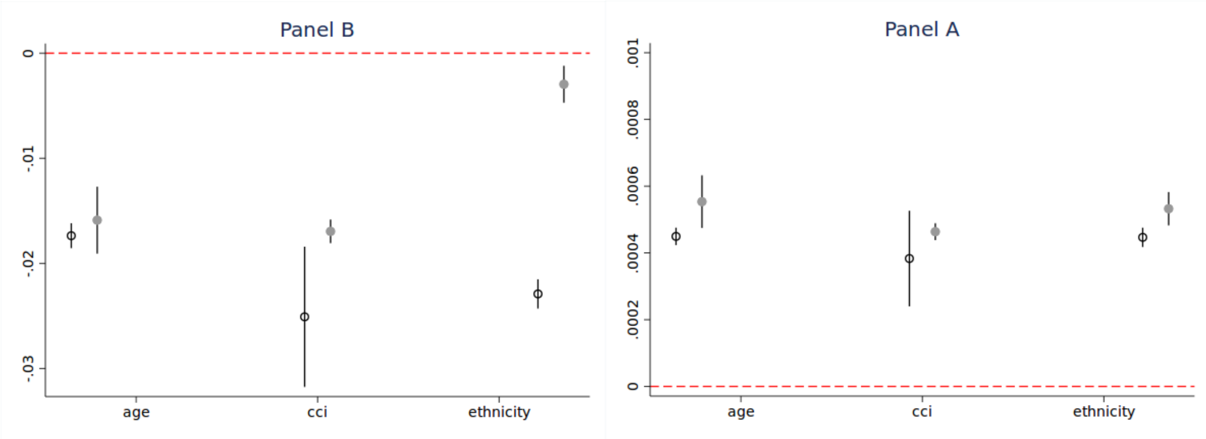
The instruments should affect the probability of treatment in a monotone way. In other words, there should be no defiers (Imbens and Rubin, 1997). I believe that this arguably satisfied by both instruments. It is indeed unlikely that an individual would opt for traditional surgery for a reduction in 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

Figure 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 13: Test for monotonicity of the instruments



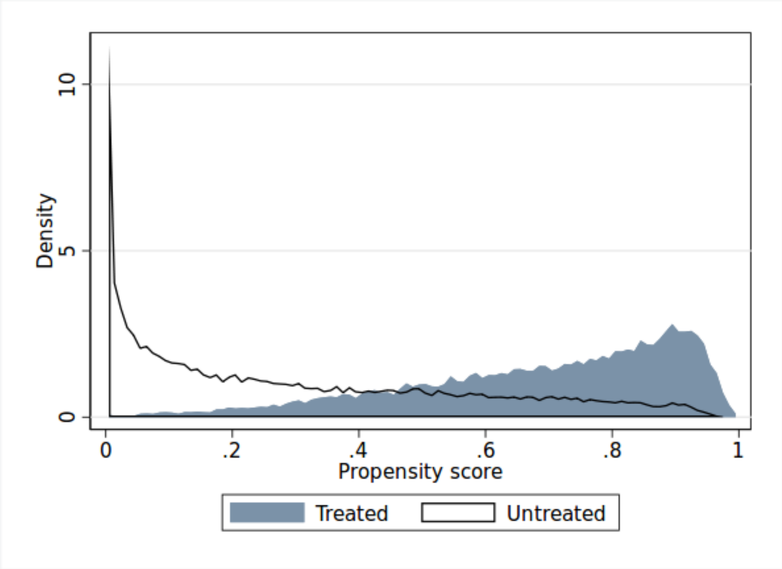
(a) Z_{dist} (b) Z_{days}

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.

instruments, estimated using a logistic regression, for the subgroups of interest in Figure 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 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 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 month-year and day of the week. Model includes continuous measure of surgical skills.

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 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 10. $Skills_j$ are alternatively added as a continuous variable or as a high skilled indicator. 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.

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

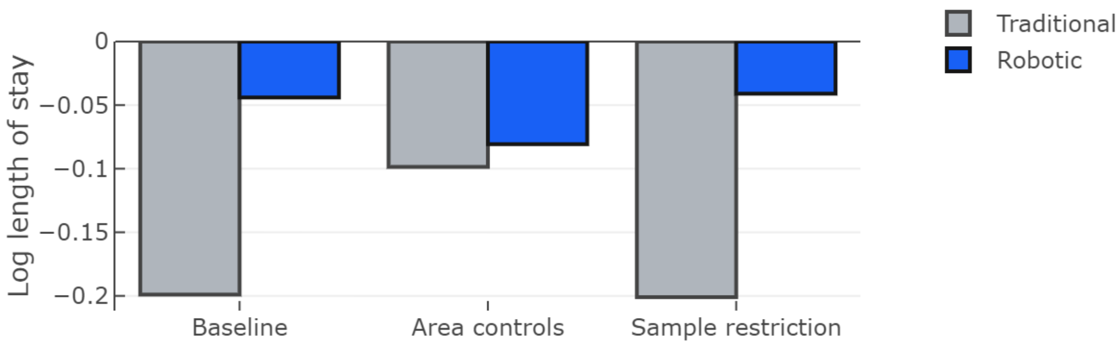
Skills and technological gains

In Table 5 to 7, I test three different specifications under the assumption of joint normality of the error terms. Table 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 10. The model includes year, month, and day of the week fixed effects all interacted with the propensity score. In Table 6, I add postal area fixed effects to control for time invariant differences across neighborhoods. In Table 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 11). Column 2 and 3 present, respectively, the coefficients δ_0 and $\delta_1 - \delta_0$ estimated from Equation 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,

Figure 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 lenght 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.

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 15 and 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.

Table 5: Heterogeneity in causal effects - Normal model

	(1) Selection equation	(2) Length of stay	(3) Adverse event
Continuous Skills			
Skills	0.368***	-0.319***	-0.0319***
	(0.017)	(0.011)	(0.007)
Skills * Propensity score		0.272***	0.0265
		(0.026)	(0.014)
Binary Skills			
High skilled	0.261***	-0.199***	-0.0359***
	(0.014)	(0.010)	(0.007)
High skilled * Propensity score		0.155***	0.0351**
		(0.018)	(0.011)
Year-Month FE	Yes	Yes	Yes
Day of the week FE	Yes	Yes	Yes
N	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. YM indicates year-month controls, DOW indicates day of the week controls. Skills, a continuous variable, is measured using the standardized risk ratio for post-operative morbidity (deaths and readmissions) computed using 2005-2007 data. Instruments used to estimate the propensity score are Z_{dist} and Z_{days}. Estimation of coefficients under the assumption of normality of unobserved components.

Table 6: Heterogeneity in causal effects – Normal model with local area fixed effects

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

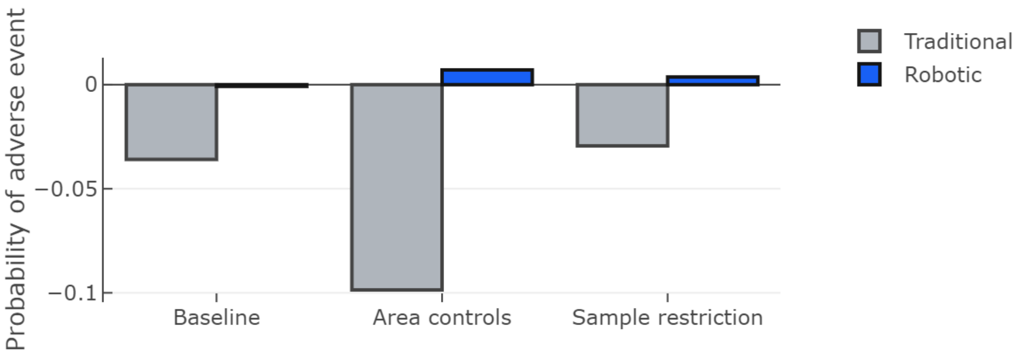
Standard errors bootstrapped with 100 repetitions $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Column(1) dependent variable binary indicator of robotic surgery. Estimated using Probit regression model. Column (2) and (3), coefficients of regressors not interacted with the propensity score measure effects 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. YM indicates year-month controls, DOW indicates day of the week controls. Skills, a continuous variable, is measured using the standardized risk ratio for post-operative morbidity (deaths and readmissions) computed using 2005-2007 data. 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.

Table 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***	-0.324***	-0.033***
	(0.017)	(0.012)	(0.009)
Skills * Propensity score		0.352***	0.057***
		(0.025)	(0.016)
Binary Skills			
High skilled	0.271***	-0.201***	-0.029**
	(0.015)	(0.016)	(0.010)
High skilled * Propensity score		0.160***	0.026
		(0.027)	(0.015)
Year-Month FE	Yes	Yes	Yes
Day of the week FE	Yes	Yes	Yes
Area FE	Yes	Yes	Yes
N	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. YM indicates year-month controls, DOW indicates day of the week controls. Skills, a continuous variable, is measured using the standardized risk ratio for post-operative morbidity (deaths and readmissions) computed using 2005-2007 data. 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.

Figure 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.

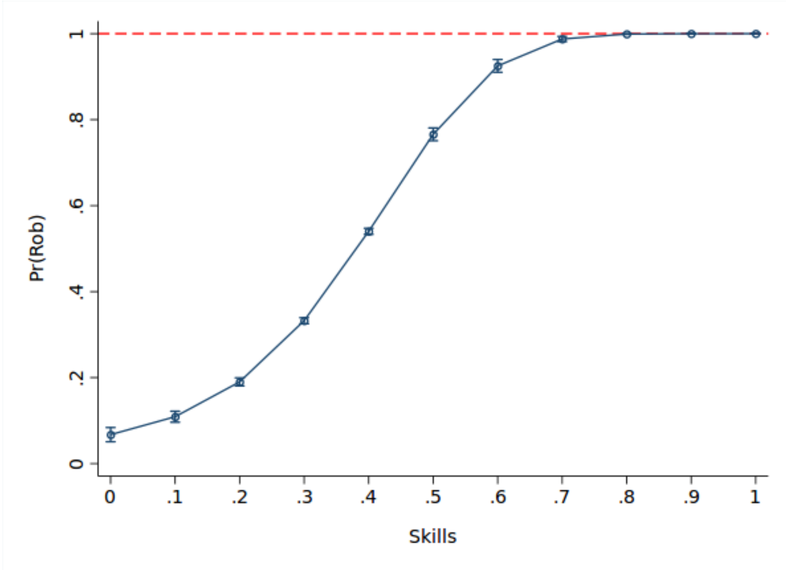
Selection into robotic surgery

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.

In Column 1 of each table, I show the coefficients on skills from the estimated selection equation (i.e., Equation 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 17, I show graphically how the probability of using the robot depends on skills. These are the marginal effects

Figure 17: Estimated probability of robotic approach by skill

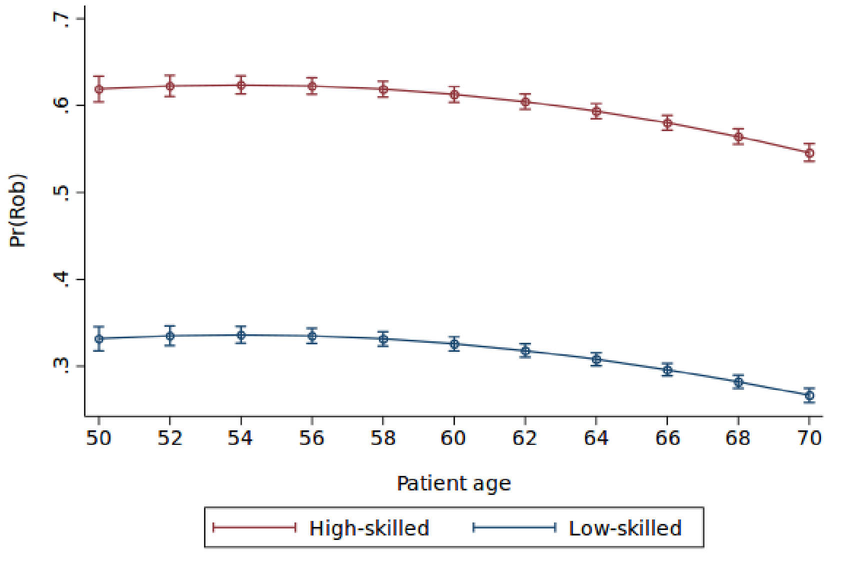


Note: Probit regression estimates, dependent variable robotic approach. Marginal probability estimated at values of skills measure. Covariates in the model at means, include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest hospital is a 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.

at different levels of my measure of skills, which I have normalized to be 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.

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 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

Figure 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.

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 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 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 characterstic of the surgeon), have the lowest returns from the treatment. The shape of MTE curve for the adverse event indicator suggests a similar story, but we can't reject homogeneity on unboservable characteristics.

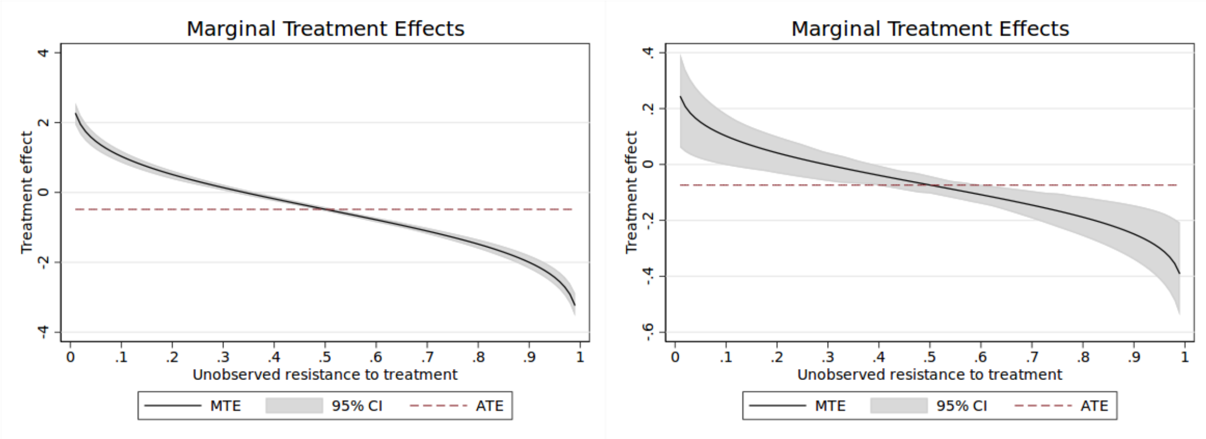
In Figure 20, I relax the assumption of joint normality and let the function $K(p)$ be approximated by a polynomial in p. Estimation in this case is achieved by a two-step procedure discussed in Heckman et al. (2006). For length of stay, the results are almost unchanged and the shape is remarkably similar 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.

Conventional treatment effects and policy simulation

In Table 8 and Table 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

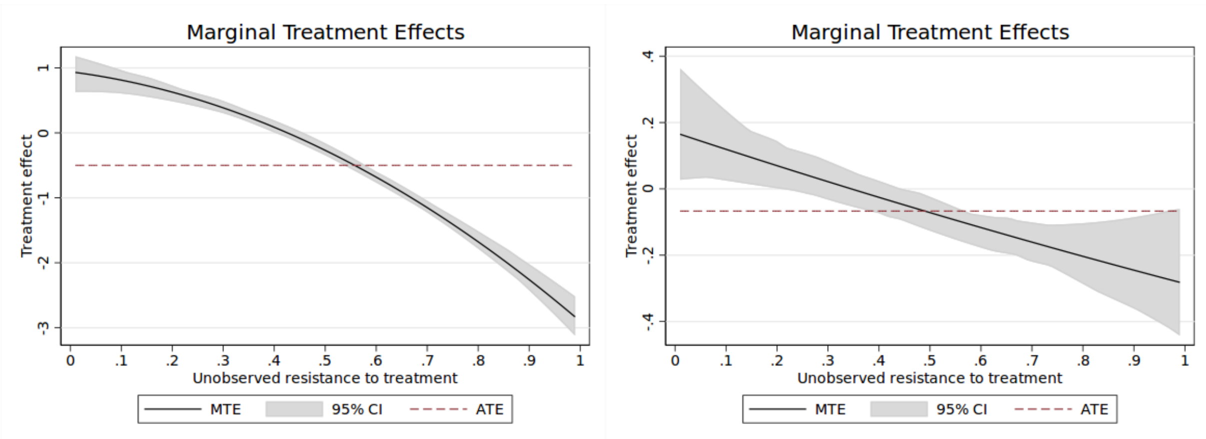
Figure 19: MTE curve - Normal



(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

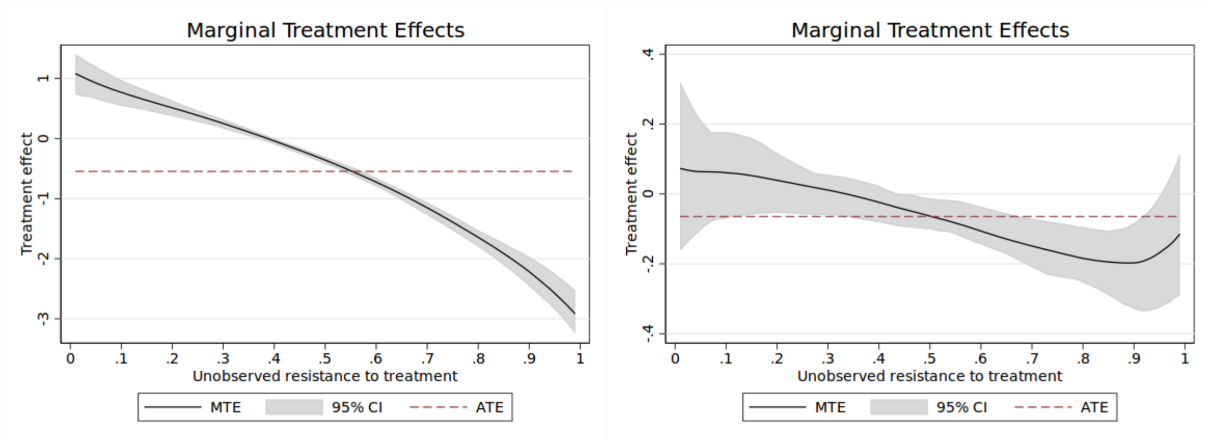
Figure 20: MTE curve – Polynomial



(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 50 repetitions

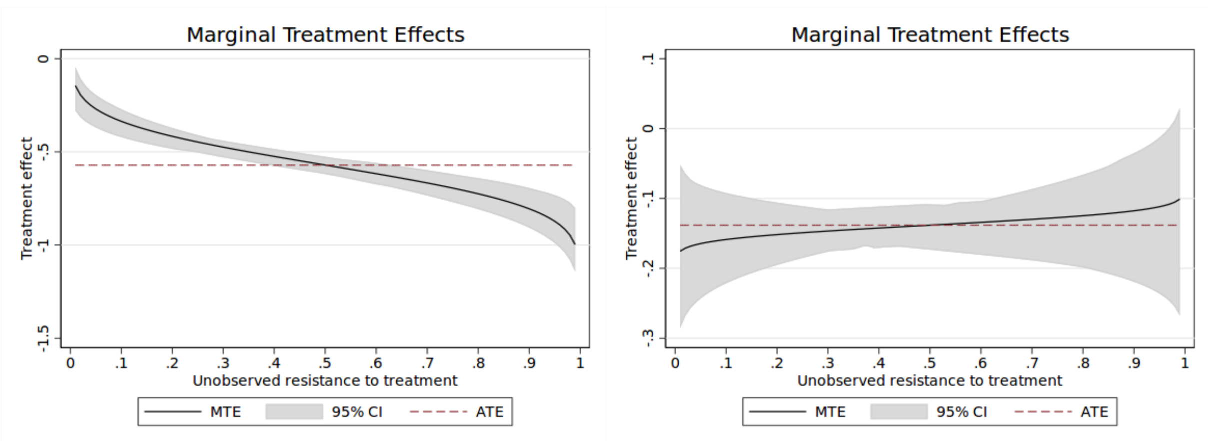
Figure 21: MTE curve – Semiparametric



(a) Length of stay (b) Adverse Event

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 50 repetitions.

Figure 22: MTE curve – Normal with area fixed effects



(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 50 repetitions. Include area fixed effects (not interacted with propensity score).

the untreated (ATU) is more negative than the effect on the treated (ATT). 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.

Table 8: Length of stay - Conventional estimates

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

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.

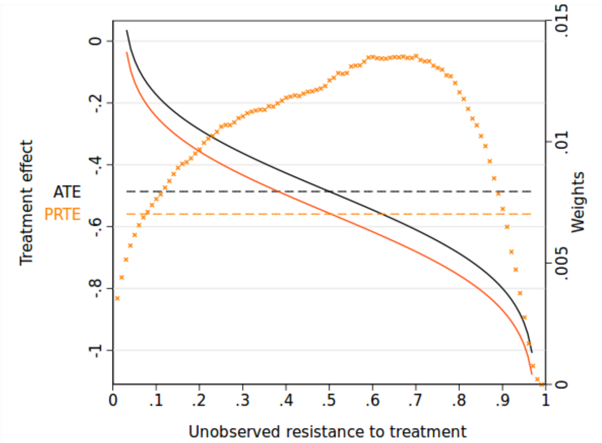
I compute this parameter for a counterfactual scenario in which I assign to lower skilled surgeons the same probability of using the robot as high skilled surgeons.

Table 9: Adverse Event – Conventional estimates

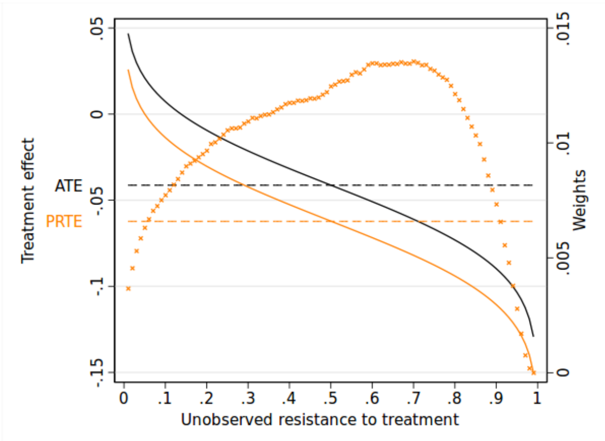
	(1) Normal	(2) Normal FE	(3) Polynomial	(4) Semiparametric
ATE	-0.073***	-0.138***	-0.067***	-0.065**
	(0.015)	(0.019)	(0.017)	(0.0211)
ATT	0.0623	-0.144***	0.0718*	0.043
	(0.032)	(0.019)	(0.033)	(0.039)
ATUT	-0.205***	-0.133***	-0.202***	-0.169***
	(0.038)	(0.039)	(0.043)	(0.049)
LATE	-0.067***	-0.136***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

Figure 23: Policy simulations - MTE and PRTE



(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. 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.

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, 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 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.

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 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.

References

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.

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

Banerjee, S. and Basu, A. (2021). Estimating endogenous treatment effects using latent factor models with and without instrumental variables. *Econometrics*, 9(1).

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.

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 Vytlacil, 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 healthcare. *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.

Cooper, Z., Gibbons, S., Jones, S., and McGuire, A. (2010). Does hospital competition improve efficiency? an analysis of the recent market-based reforms to the english nhs.

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., Dunglison, 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.

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.

Davies, B. (2022). Waiting for godot robot.

Deaton, A. (2003). Health, inequality, and economic development. *Journal of economic literature*, 41(1):113–158.

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.

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).

Gowrisankaran, G. and Town, R. J. (1999). Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics*, 18(6).

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.

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).

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.

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).

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)*.

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

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.

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.

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.

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.

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.

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).

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).

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., Dunglison, 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.

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

Appendix

Tables

Table 10: Covariates — 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 11: Probit regression - Selection equation

	(1)
	Robot=1
Z_{dist}	-0.009***
	(0.001)
Z_{days}	0.003***
	(0.001)
Distance to closest hospital	0.005***
	(0.005)
Closest is Teaching Hospital	0.169***
	(0.015)
AMI (Acute Myocardial)	-0.110*
	(0.044)
CHF (Congestive Heart)	-0.111
	(0.089)
PVD (Peripheral Vascular)	-0.097
	(0.059)
CEVD (Cerebrovascular	-0.142*
	(0.061)
Dementia	-0.002
	(0.210)
COPD (Chronic Obstructive Pulmonary)	-0.058**
	(0.022)
Rheumatoid Disease	-0.049
	(0.059)
PUD (Peptic Ulcer)	0.015
	(0.059)
Mild LD (Liver)	-0.009
	(0.076)
Diabetes	-0.124***
	(0.025)
Diabetes + Complications	-0.023
	(0.118)
HP/PAPL (Hemiplegia or Paraplegia)	0.021
	(0.147)
RD (Renal)	-0.185***
	(0.054)
Moderate/Severe LD (Liver)	-0.052
	(0.250)
Metastatic Cancer	-0.109*
	(0.050)
Age	0.091***
	(0.015)
Age squared	-0.002***
	(0.001)
White	-0.199***
	(0.016)
Skills	0.368***
	(0.017)
City indicator	-0.110***
	(0.016)
Year*month	Yes
Day of the week	Yes
Observations	50203

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: MTE - Baseline specification for length of stay and adverse event (Coefficients) - Normal model

	(1) Length of stay	(2) Adverse Event
Distance Closest Hospital	0.00227***	-0.000311
	(0.000480)	(0.000370)
Closest is Teaching Hospital	-0.0577***	0.00327
	(0.0119)	(0.00751)
AMI (Acute Myocardial)	0.0437	0.0234
	(0.0392)	(0.0260)
CHF (Congestive Heart)	0.150*	0.0127
	(0.0670)	(0.0447)
PVD (Peripheral Vascular)	0.0776	0.0122
	(0.0642)	(0.0321)
CEVD (Cerebrovascular	0.0940	0.0412
	(0.0623)	(0.0349)
Dementia	-0.121	-0.0665
	(0.195)	(0.113)
COPD (Chronic Obstructive Pulmonary)	0.0665***	0.0158
	(0.0179)	(0.0121)
Rheumatoid Disease	0.0112	-0.0158
	(0.0526)	(0.0282)
PUD (Peptic Ulcer)	0.0337	0.0381
	(0.0479)	(0.0355)
Mild LD (Liver)	0.147*	0.0925*
	(0.0728)	(0.0456)
Diabetes	0.0811***	0.0239
	(0.0218)	(0.0161)
Diabetes + Complications	0.137	0.0380
	(0.120)	(0.0677)
HP/PAPL (Hemiplegia or Paraplegia)	-0.0906	0.0504
	(0.128)	(0.0865)
RD (Renal)	0.110*	0.0355
	(0.0524)	(0.0293)
Moderate/Severe LD (Liver)	0.0928	0.211
	(0.284)	(0.220)
Metastatic Cancer	0.328***	0.0144
	(0.0426)	(0.0279)
Age	-0.0248*	-0.00729
	(0.0116)	(0.00653)
Age squared	0.000259**	0.0000668
	(0.0000945)	(0.0000530)
White	-0.0502***	0.00851
	(0.0128)	(0.00890)
Skills	-0.319***	-0.0319***
	(0.0110)	(0.00689)
City indicator	0.0358**	0.0232**
	(0.0113)	(0.00793)
Observations	49215	50203

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: MTE - Baseline specification for length of stay and adverse event (Coefficients interacted with propensity score) - Normal model

	(1) Length of stay	(2) Adverse Event
Distance Closest Hospital * Propensity score	-0.004***	0.001
	(0.001)	(0.001)
Closest is Teaching Hospital * Propensity score	0.039*	0.009
	(0.019)	(0.012)
AMI (Acute Myocardial) * Propensity score	-0.014	-0.006
	(0.055)	(0.039)
CHF (Congestive Heart) * Propensity score	-0.098	-0.055
	(0.123)	(0.071)
PVD (Peripheral Vascular) * Propensity score	-0.089	0.018
	(0.110)	(0.051)
CEVD (Cerebrovascular Disease) * Propensity score	-0.068	-0.066
	(0.101)	(0.053)
Dementia * Propensity score	0.217	0.0403
	(0.317)	(0.183)
COPD (Chronic Obstructive Pulmonary) * Propensity score	-0.055	0.005
	(0.030)	(0.017)
Rheumatoid Disease * Propensity score	-0.097	0.042
	(0.091)	(0.043)
PUD (Peptic Ulcer) * Propensity score	0.049	-0.015
	(0.082)	(0.053)
Mild LD (Liver) * Propensity score	-0.059	-0.147*
	(0.111)	(0.060)
Diabetes * Propensity score	0.012	-0.007
	(0.035)	(0.026)
Diabetes + Complications * Propensity score	-0.158	-0.023
	(0.196)	(0.111)
HP/PAPL (Hemiplegia or Paraplegia) * Propensity score	0.316	0.009
	(0.211)	(0.136)
RD (Renal) * Propensity score	-0.036	-0.026
	(0.093)	(0.047)
Moderate/Severe LD (Liver) * Propensity score	-0.259	-0.122
	(0.426)	(0.292)
Metastatic Cancer * Propensity score	-0.266***	-0.006
	(0.064)	(0.042)
Age * Propensity score	-0.029	0.001
	(0.020)	(0.011)
Age squared * Propensity score	0.001	-0.001
	(0.001)	(0.001)
White	0.028	0.009
	(0.019)	(0.012)
Skills * Propensity score	0.272***	0.026*
	(0.022)	(0.013)
City indicator * Propensity score	0.038*	-0.029*
	(0.019)	(0.012)
mills	-1.184***	-0.137***
	(0.072)	(0.036)
(Year*month) * Propensity score	Yes	Yes
Day of the week * Propensity score	Yes	Yes
Observations	49215	50203

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Conventional estimates

Ordinary least square regression

I estimate the relationship between robotic surgery and patients' outcomes using a linear regression model. In Table 14 the dependent variable is the logarithm of the patient length of stay in hospital. In Table 15 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 operated with robotic surgery and zero otherwise. The sample includes all patients that have undergone a radical prostatectomy in a NHS hospital in England. Table 14 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 15 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. YM indicates year-month controls, DOW indicates day of the week controls.

Table 14: Association of robotic approach and adverse event - OLS Regression

	(1)	(2)	(3)	(4)
Robot	-0.768***	-0.648***	-0.407***	-0.323***
	(0.057)	(0.056)	(0.065)	(0.046)
Patient characteristics	No	Yes	Yes	Yes
YM	No	No	Yes	Yes
DOW	No	No	Yes	Yes
Hospital FE	No	No	No	Yes
N	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

Table 15: Association of robotic approach and log length of stay - OLS Regression

	(1)	(2)	(3)	(4)
Robot	-0.0958***	-0.0809***	-0.0438***	-0.0639***
	(0.0105)	(0.00951)	(0.00958)	(0.0114)
Patient characteristics	No	Yes	Yes	Yes
YM	No	No	Yes	Yes
DOW	No	No	Yes	Yes
Hospital FE	No	No	No	Yes
N	61839	51424	51424	51424

Standard errors clustered at hospital level
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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 16. In Column 1 to 3, I show the first stage estimates from a Probit regression of an indicator of robotic approach on 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 the week controls. Table 17 shows the coefficients estimated for the simultaneous Probit model. In Panel A,

Table 16: Estimated impacts of robotic approach on log length of stay

	(1) FS	(2) FS	(3) FS	(4) IV	(5) IV	(6) IV
Z_{dist}	-0.127***		-0.093***			
	(0.003)		(0.003)			
Z_{days}		0.036***	0.027***			
		(0.001)	(0.001)			
Robot				-0.412***	-0.221***	-0.307***
				(0.023)	(0.023)	(0.019)
Patient characteristics	Yes	Yes	Yes	Yes	Yes	Yes
YM	Yes	Yes	Yes	Yes	Yes	Yes
DOW	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 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).

Robustness checks

Measuring skills with regression fixed effects

I test my baseline specification using regression fixed effects (rather than random effects) to measure surgeons skills. I estimate a logistic regression model where the dependent variable

Surgeons' experience

In Column 1 and 3 of Table 18 and Table 19, 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. 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.

Table 17: Estimated impacts of robotic approach on probability of adverse event

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

Robust standard errors.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Estimated coefficients: surgeon experience

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

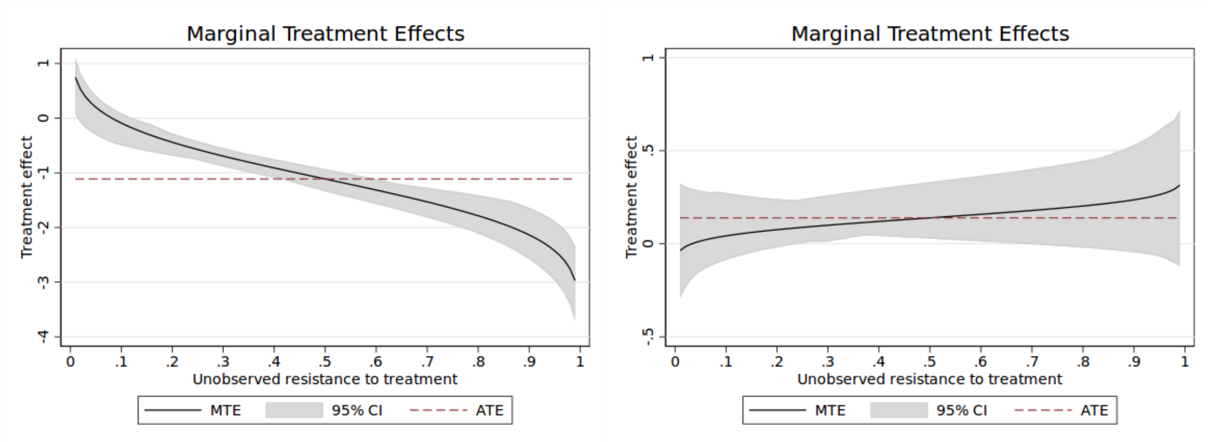
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Estimated coefficients: surgeon experience

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

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

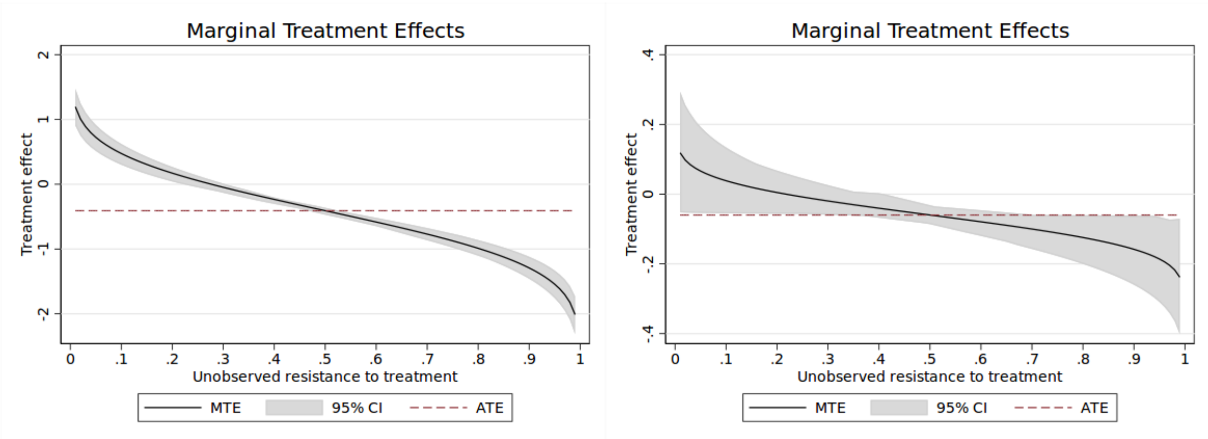
Figure 24: MTE curve – Normal with sample restriction by experience



(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 50 repetitions. Include area fixed effects (not interacted with propensity score). Sample restricted to surgeons working in 2003.

Figure 25: 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 50 repetitions. Include area fixed effects (not interacted with propensity score). Model includes dummies for the first year the surgeon is observed working.