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

Elena Ashtari Tafti

University College London[1]

17 October 2022

Abstract. In this paper, I investigate the potential of new technologies to reduce disparities in healthcare provision. Differences in providers' skills may cause variation in patient outcomes. The adoption of new technologies, like robots, may be a solution to this problem if technological gains are decreasing in users' skills or may exacerbate existing variation in performance otherwise. In England, the diffusion of surgical robots coincided with an improvement in average performance and a convergence in outcomes between high and lower skilled surgeons. I study whether this pattern can be attributed to the adoption of robots using the universe of inpatient admissions for prostate cancer surgery. To identify causal impacts, I exploit quasi-random variation in the geographic allocation of robots within a structural model, allowing for selection and heterogeneity in treatment effects. I find that robots improve patient outcomes, but their effects importantly depend on surgeons' skills. The robot has little impact on the performance of high skilled surgeons, while lower skilled surgeons gain the most from it. I also uncover a strong pattern of negative selection. 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 skills of 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 extremely grateful for the support and feedback I have received from David C. Chan. I am grateful to Oliver Ashtari Tafti, Nathaniel Breg, Alexander Clyde, Andrew Chesher, Lodovico de Vito, Alice Kuegler, Thomas P. Hoe, Thomas Lazarowicz, Mikkel Mertz, Lars Nesheim, Nikita Roketskiy, and George Stoye for their comments and feedback on this project.

1 Introduction

Most healthcare systems have at the core of their service provision the promotion of equity in access and quality of care (Jha et al., 2017). However, within many countries, differences in patient outcomes across areas and providers persist, even after controlling for patient risk (Skinner, 2011). Treatment utilization rates may explain part of this variation (Birkmeyer et al., 2013b; Tsugawa et al., 2017). But empirically health outcomes appear to be only marginally affected by it (Molitor, 2018). In fact, heterogeneity in healthcare providers' skills may be at the root of this unwarranted variation (Chandra and Staiger, 2020, 2007; Hull, 2018; Stoye, 2022). Differences in providers skills generate inequality and may exacerbate systematic disparities in access due to geographic location or social economic background (Finkelstein et al., 2016).

In this paper, I investigate the potential of robots to reduce variation in patient outcomes that arises from differences in surgeons skills. Across and within occupations, individuals differ substantially in their level of skills and healthcare providers, such as surgeons and doctors, are no different (Currie and MacLeod, 2017; Kolstad, 2013). The proliferation of scorecards and quality ratings within these professions is a clear indication that such differences are relevant and concern policymakers and patients alike. In England, the diffusion of robots coincided with an improvement in average surgical performance, and with a convergence in outcomes between high and lower skilled surgeons. I study whether this is attributable to the adoption of robotic surgery using administrative data on prostate cancer patients from the National Health Service (NHS). This is the most common type of cancer in men in the United Kingdom (UK) and its surgical treatment has been revolutionized by the adoption of robots (Hussain et al., 2014).

For the most part, the study of robots in economics has been confined to their effects on the labour market (Acemoglu and Restrepo, 2020; Humlum, 2019). The central idea is that robots compete against human labour in the production of different tasks, and as a consequence may reduce employment and wages. The subject of this literature are industrial robots; fully autonomous machines that can be programmed to perform manual tasks. In many applications, however, the robot is controlled by a user and is meant to aid rather than substitute workers to perform complex assignments. Surgical robots fall into this category. These robots are fully operated by surgeons and act as an extension of their users. The surgeon sits at a console that allows movements to be carried out through the machine and operate on the patient. The expectation was that, by replacing human hands with robotic arms, the patient would have a speedier and smoother recovery (Peters et al., 2018; Warren and Dasgupta, 2017).

I posit that, in the presence of skills heterogeneity, the extent to which a technology has the potential to reduce unwarranted variation in providers' performance depends on how the human and technological capabilities combine. Robotic technology may exacerbate existing variation in surgical performance, or may be a solution to this problem if its returns are decreasing in surgeons skills. In other words, the robot can reduce variation in outcomes if it helps *weaker* surgeons to become better, rather than better surgeons to become *stronger*. I envisage two possible scenarios. In the first, technological gains are *increasing* in users skills. Returns from using the technology are then concentrated among high skilled users, and variation in patient outcomes is likely to increase as a result of adoption. In the second scenario, the robot changes the way a task is performed so that providers' skills matter less. In this case, returns are *decreasing* in skills and the technology acts as an equalizing force on the market.

To measure improvements in surgeon's performance from using the robot, I focus on two patient outcomes: the speed of recovery (i.e. post-operative length of stay) and the occurrence of adverse events following surgery (i.e. post-operative morbidity). Adverse events include deaths, emergency readmissions, and complications from surgery specific to prostate cancer patients. The medical literature considers that - if any - the robot should have measurable benefits on these two margins (Higgins et al., 2017; Coelho et al., 2010; Lowrance et al., 2010; Nelson et al., 2007). Moreover, these are important cost drivers, which are often considered when evaluating whether a medical technology is worth adopting (Lotan, 2012).

I also use patient outcomes to measure surgical skills. In economics, skills are often proxied with education. This approach is not feasible when all individuals within a profession have the same level of education. Surgical skills have been found to map neatly in the level of morbidity that patients experience after surgery (Birkmeyer et al., 2013a). Clearly, the patients of high skill surgeons should experience, *ceteris paribus*, fewer negative events from surgery. To account for differences in patients pools, I adapt a risk adjustment methodology used by the Centers for Medicare Medicaid Services (CMS). Specifically, I construct a single risk-adjusted indicator of surgical skills reflecting mortality and readmissions rates as it is standard in the literature (Tsai et al., 2013). Moreover, my indicator also captures negative events from surgery specific to patients with prostate cancer. I use the three years preceding the introduction of robots nationally to measure skills. In this way, my indicator reflects patient outcomes when all operations were carried out with an 'open' method. Hence, it reflects the skills of the surgeon, in absence of any technological aid and will be exogenous to the adoption behaviour.

To this day, evidence that robotic surgery improves patients outcomes has been at best inconclusive (Coughlin et al., 2018; Yaxley et al., 2016; Robertson et al., 2013; Bolla et al., 2012). One reason for this is that the returns from using the robot are heterogeneous. Existing studies are based on small and selected samples (Neuner et al., 2012) and focus on correlation rather than causation (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, if the uptake of this technology is, as well, heterogeneous across the skills' distribution any naive correlation will speak more to the characteristics of the adopters rather than the technology itself. Part of my contribution is then to study the impact of technology adoption under jointly a selection problem and heterogeneous treatment effects. Indeed, there are two dimensions to the problem of assessing the impact of robots on surgeons' performance. First, the decision to use a robot is not random and likely to be the result of both surgeon and patient characteristics. Highly skilled surgeons may be more inclined to use the robot, while less skilled ones may refrain from doing so. At the same time, certain patients may demand to be treated with the new technology, while others may not. Second, the performance of the technology may also depend on the surgeon and the patient.For example, the robot may perform better when used by a skilled surgeon, but also may be better suited to operate on less complex patients. That is, treatment effects may be heterogeneous. Importantly, when treatment effects are heterogeneous, surgeons and patients may select, and be selected, based on their specific technological gains (Björklund and Moffitt, 1987). Simple comparisons or even regression-adjusted comparisons between robotic and traditional surgery would in this case provide misleading estimates of causal effects.

My econometric strategy directly tackles both issues. I use a structural approach introduced by Björklund and Moffitt (1987) and generalized by Heckman and Vytlacil (2005) that focuses on the marginal treatment effect (MTE). The MTE is the treatment effect for individuals at a particular margin of the so-called resistance to treatment. The building block of this approach is the generalized Roy model (Roy, 1951), where agents select into treatment based on their expected gains. Within this framework, a surgeon's performance depends on her skills, the patient's characteristics, and their interaction with the choice of treatment (i.e. robotic or traditional surgery). In turn, the surgeon's decision to use the robot is also modelled explicitly as a function of her skills, the characteristics of her patient, and a latent variable V (i.e. the resistance to treatment). Skills matter then for both the choice and the performance of surgeons, and the parameters of the model speak directly to the differences in performance between high and lower skilled surgeons with traditional (i.e. the untreated state) and robotic surgery (i.e. the treated state). Hence, the model allows me to directly observe whether the robot exacerbates or reduces differences in patient outcomes that arise from skills heterogeneity. In this framework, no restriction is imposed on the relationship between outcomes and resistance to treatment, so that selection may occur on unobserved characteristics as well. This is captured by the MTE curve. This curve, which I estimate, portrays the relationship between treatment effects and the latent V. Identification of the MTE requires no more assumptions than the standard instrumental variable approach (Vytlacil, 2002), but is more informative when effects are heterogeneous (Cornelissen et al., 2016). IV allows to identify the Local Average Treatment Effect (LATE) in the presence of heterogeneous treatment effect (Imbens and Angrist, 1994). The MTE can be thought as a local LATE, which however can be computed for the full distribution of V and whose aggregation allows therefore to compute all conventional treatment effects parameters (e.g. the average treatment effect, or ATE).

Identification of the model parameters requires at least one continuous instrument to be included in the selection equation. This should be independent of unobserved factors that influence selection, and outcomes in the treated or untreated state (Heckman and Vytlacil, 2005). I exploit the staggered adoption of robots over time to propose a novel instrumental variable. Acquisition of surgical robots in England has been managed by individual hospitals (Lam et al., 2021), with few adopting the technology as early as 2007, many between 2012 and 2017, and some never at all. The absence of any strategic national plan to regulate this process, resulted in an uneven distribution of robots and created differences in the availability of the technology overtime. Patients in different areas had different access to the technology, depending on whether their local hospitals adopted the robot. Some patients would be located close to a hospital that had adopted the technology, other patients may not. Moreover, as the uptake of robots grew over time, accessibility of the robot depended as well on the time of the operation.

Since patients usually go to their closest hospital, I construct an instrumental variable exploiting the difference between the date of cancer diagnosis and the date the nearest hospital to the patient adopted the robot. I argue diagnosis timing captures quasi-random variation in treatment assignment (i.e. being operated with a robot). This variable strongly predicts the probability of receiving robotic surgery, controlling for the year and month of operation. Relatively to men operated in the same period, but diagnosed days before, individuals whose diagnosis followed their closest hospital adoption of the robot are significantly more likely to be operated with it. Nevertheless, for this is instrument to be valid, diagnosis timing should impact patients outcomes exclusively through its effect on the probability of receiving robotic surgery. To provide evidence that this is the case, I test whether this instrument is correlated to patient outcomes in the period preceding the introduction of robots. If relative timing affects outcomes, other than through the probability of robotic surgery, it would surely emerge from this relationship. Diagnosis timing and patient outcomes are not correlated in the period pre-robots, which encourages me to believe that the exclusion restriction holds.

It is common practice in the MTE literature to use multiple instruments. This is done to expand the support over which the marginal treatment effect can be estimated (Carneiro et al., 2011). In the spirit of McClellan et al. (1994), I use as an additional instrument the patient's relative distance to a hospital offering robotic technology. A primary concern, when using relative distance as an instrument, is that this may be correlated with health in ways not accounted by the model (Hadley and Cunningham, 2004). However, because adoption is staggered, relative distance in my context varies over time for patients living in the same area. This allows for tighter handling of non-random selection than what is usually possible in papers employing this instrument (Cornelissen et al., 2018). Because my instrument varies across both geography and time, I can control for time-invariant differences in area's characteristics. This ensures that distance comparisons come only from patients who live in similar neighbourhoods. Still, relative distance may be correlated with individuals' health in other ways. I exploit the richness of my data to show that this is unlikely to be the case here. I use data on patients that suffered an acute myocardial infarction (AMI) to perform a falsification test. I test whether mortality for AMI patients depends on their relative distance to a robotic hospital in the year they suffered the AMI. If relative distance was related to individuals' health in ways not captured by the model, the instrument will correlate with AMI mortality. I find no evidence that hospital outcomes of AMI patients are correlated to their relative closeness to a robotic hospital.

I find that robotic surgery improves the surgeon's performance. The average treatment effect is statistically significant and negative, indicating that, on average, the robot reduces post-operative length of stay and morbidity. My estimates suggest the patients undergoing robotic surgery leave the hospital 0.3 days earlier (i.e. a ten percent reduction from an average of 2.9), and their probability of experiencing an adverse event is 0.06 percentage points lower relatively to those operated with traditional surgery.

However, my results show that these effects are highly heterogeneous. Technological gains strongly depend on the skills of the surgeon using the robot. Surgeons that were top performers before the adoption of the robot benefit the least from using the technology, while lower skilled surgeons appear to gain the most from it. Strikingly, the difference in effects between high and lower skill surgeons is almost as large as the average treatment effect I find in the population. Therefore, the robot exhibits *decreasing* returns in skills: it complements more strongly lower skill surgeons. The patients of high skill surgeons are 8 percentage points less likely to experience and adverse event from surgery relatively to those of lower skill surgeons in the untreated state. But this same patients, are only 1 percentage point less likely to experience these events with the robotic approach. A similar result emerges as well for the speed of recovery. When the robot is used, differences in patient outcomes between high and lower skill surgeons shrink, which suggests that the robot may have the potential to reduce unwarranted variation in surgical performance. This effect is the result of lower skill surgeons performing significantly more poorly in the untreated state, and the technology equalizing them to high skill surgeons in the treated state.

If The analysis also shows a strong pattern of negative selection on both observed and unobserved characteristics. In terms of observables, high skill surgeons use the technology more intensively, while lower skill surgeons use it less in spite of their higher returns. Surgeons appear in general to use the robot on younger and less complex patients, but at all levels high skill surgeons use the robot more. The pattern of selection I find here is mimicked, also, by the estimated MTE curve. The MTE curve is downward sloping, a higher resistance to treatment is associated with larger improvements in patient outcomes. If improving patient outcomes was the targeted objective, the technology should be used more intensively by lower skill surgeons because they have the highest returns from it. Hence, as a conclusive exercise, I simulate policies that may change the incentives to use robotic surgery in line with this idea.

This paper builds on several literatures. There is a vast strand that documents the observable heterogeneity driving the decision to use a technology. Part of this literature emphasizes the role of comparative advantage in adoption decision. A notable example is Suri (2011), which studies the adoption of hybrid maize in Kenya and shows that differences in the uptake of the technology can be explained by heterogeneous benefits and costs. Abaluck et al. (2016) and Chandra and Staiger (2020) bring this idea to the study of treatment decisions in healthcare. Abaluck et al. (2016) decomposes variation in the rates of imaging across doctors into heterogeneity in patients' benefits from testing and heterogeneity in physicians' tendency to test a given patient. Chandra and Staiger (2020) find that variation in the use of treatments may be explained by both comparative advantage and allocative inefficiency, suggesting that some provides are using too much or too little of them. My findings also speak to a large literature on the consequences of medical innovations (Newhouse, 1992; Cutler, 2007; Chandra and Skinner, 2012). Chandra and Staiger (2007), for example, shows that health care productivity depends on the heterogeneity of treatment effects across patients. I took inspiration from a series of papers that have collected reduced-form evidence on how robots affect firm performance and labour market outcomes (Acemoglu and Restrepo, 2020; Humlum, 2019), and the literature studying skilled biased technological change (SBTC) (Acemoglu, 2002; Card and DiNardo, 2002). Unlike Acemoglu and Restrepo (2020) and Humlum (2019), I study robots in abstraction from automation. Unlike the literature on SBTC I look at skills within and not across occupations.

My paper proceeds as follows. In Section 2, I describe surgical robots and their use for prostate cancer surgery. I use Section 3 and Section 4 to describe the data, and discuss how I measure technological gains and surgeons' skills. In Section 4, I also provide the empirical facts that motivated this work. I introduce my econometric model, and present the conditions required for identification and estimation of the parameters in Section 5. I provide a discussion of the definition and validity of the instrumental variables in Section 6. In Section 7 I show the results of the analysis, and I finally conclude in Section 8.

2 Robotic surgery and the treatment of prostate cancer

Although robots have found several applications in surgery, this paper focuses on robotic surgery for prostate cancer (or radical prostatectomy (RP)). I restrict my attention to this operation because the robot has played a notably large role in transforming how this is performed (Hussain et al., 2014). Prostate cancer is the most common type of cancer in men in the UK; that's 129 men are diagnosed with prostate cancer every day, and more than 11,500 die every year from it.[2] In the United States and Europe, the diffusion of robots for prostate cancer surgery has been incredibly rapid. In 2003, less than 1 percent of surgeons in the US performed this procedure robotically. Seven years later, already 86 percent of the 85,000 men who had prostate cancer surgery had a robot-assisted operation.[3] Eventually, by 2014, robotic surgery accounted for up to 90 percent of radical prostatectomies across the US.[4] This trend has been similar in England where, by 2014, the majority of cases (62.7 percent) were performed robotically (Marcus et al., 2017).

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 is the da Vinci surgical system. This is manufactured by the California-based company and market leader Intuitive. The robot has three components which are shown in Figure 1:

1. a viewing and control console that used by the surgeon,
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

[2]https://prostatecanceruk.org
[3]https://www.nytimes.com/2010/02/14/health/14robot.html?pagewanted=print$_r = 0\%E3\%80\%89$
[4]https://www.nature.com/articles/d41586-020-01037-w

incisions that are much smaller than what would be required for human hands but also to work at scales, where hand tremor 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 be carried out 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 while using the robot, thereby subjectively improving surgical performance (Tonutti et al., 2017).

Prior to 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 to clearly see the area of interest and operate. Other minimally invasive approaches, such as laparoscopy, had also been available prior to robotic surgery but had limited popularity because of the difficult position of the prostate. Throughout this paper, I will refer to all approaches that do not involve the use of robots as traditional surgery.

Among the greatest barriers to the adoption of 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 make significant changes to their practices. Robots usually necessitate a dedicated operating room, which is in many cases built for this purpose. 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. Nurses responsible for setting up the machine need specialized training, and any technical drawback during the operation is not only risky for the patient, but also prolongs operation time and generates inefficiencies for the hospital (Compagni et al., 2015).

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). From an oncological perspective, robotic surgery is equivalent to traditional surgery; they are both effective to remove cancer when this is confined to the prostate. Robotic surgery promised, however, 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. Some studies suggest that the robot has significant ergonomic advantages for the surgeons, which may justify the use of the technology (Jayant Ketkar et al., 2022). Regardless, if the robot does not improve how surgeons perform, the use of this expensive technology is hardly defensible from a societal perspective.

3 Data and institutional context

Health care in England 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 for treatment is made by the patient with the support and advice of their GP. Hospitals cannot refuse patients, but will schedule admissions and can 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). Adoption of surgical robots has occurred in the absence of guidelines, leaving to the individual provider the decision to adopt the technology as well as the development of best practices. There is currently no national account of the location and utilization of robots in the NHS, but a recent study suggests that at least 25 percent of hospitals own one (Lam et al., 2021).

The data I use comes from the Hospital Episodes Statistics (HES). This is a large administrative data set, which covers the universe of inpatient discharges from NHS hospitals in England. The data collect 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 are treated and the area where they live is also available from the data. Importantly, 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).

HES allows me to identify all radical prostatectomies occurring throughout NHS England from 2003 to 2017. My sample comprises all patients that have been diagnosed with prostate cancer, and that have been operated in a NHS hospital using the traditional or robotic technique. I have a total of 62,258 admissions for radical prostatectomy from 2003 to 2017, of which 25208 are performed with a robot. Table 1 summarizes the characteristics of patients for both the traditional and the robotic approach. Using HES, I can determine the date of the first robotic RP within each hospital, which I will consider the date of adoption of the technology for that hospital. Figure 3 presents the location of the hospitals adopting the robot for three windows of time; from 2006 to 2008, from 2009 to 2013, and from 2014 to 2015. In my identification strategy, I will exploit these differences in adoption timing.

From HES, it is possible to observe that prostate cancer surgery has become by far the most commonly performed robotic operation in England. Figure 4 displays 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. Only after 2013 robots are used in other specialities, but diffusion was significantly slower.

In England, the use of robotic surgery for RP grew from 5 percent in 2007 to 80 percent in 2017. Figure 5 plots the total number of RP by surgical approach from 2003 to 2017. The first notable use of robots is in 2007, and by the end of 2017 around 80 percent of RP are performed using a robot. The steady increase in the number of robotic operations is accompanied by a decrease in the number of traditional surgeries. This indicates a clear pattern of substitution toward the newer technology. 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. The increase in the number of RP can be ascribed to a significant increase in the number of prostate cancer of diagnoses, and a small increase in the use of surgery for such condition. Figure 6 displays the number of prostate cancer diagnoses and the share of patients opting for RP over time.

To identify technological gains, I investigate two patient's outcomes; length of stay (LOS) and the occurrence of adverse events from surgery (or post-operative morbidity). I focus on these two margins for three main reasons. First, 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). Second, LOS and post-operative morbidity are important cost drivers and, for this reason, should be considered when evaluating whether a medical technology is worth adopting (Lotan, 2012). Lastly, these are outcomes that I can easily and reliably measure from the data I have.

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

I exploit the panel dimension of the data to identify negative health events that are likely to be the result of the operation being poorly performed. I focus on three negative events: in-hospital deaths, 30 days emergency readmissions, and complications arising within two years from the 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. HES only covers interactions of patients with secondary care providers. A significant side effect of prostate cancer surgery are erectile dysfunctions, which are often addressed with medical interventions. I am not able to detect this dimension of performance with the data I have.

Table 1 summarizes both margins of surgical performance. The average post-operative LOS is 2.9 days, and more than 14 percent of individuals appear to have experienced an adverse event from surgery in my sample. Patients undergoing robotic surgery in my sample have on average lower LOS. Post-operative LOS for robotic surgery is 1.8 on average, while it is 3.8 for traditional surgery.

4 Measuring Skills

Skills are not observable and notoriously difficult to measure. The unit of 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 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 to measure surgeons' skills. In particular, I focus on post-operative morbidity which refers to adverse events and complications following surgery (Colombo et al., 2006) including postoperative death, readmission, and complications [5].

[5]Including urinary complications, and erectile dysfunction which I identify using OPCS code from patient's admissions following the surgery.

To compare outcome rates from different populations, I adapt a risk-adjustment methodology employed by CMS. This methodology was developed by the Yale New Haven Health System / Center for Outcomes Research and Evaluation (YNHHS/CORE). The objective of this exercise is to produce a single indicator of skills that represents a risk-standardized rate of outcomes. Following the CMS methodology, I compute the measure in two steps. In the first step, I estimate a regression model to adjust for differences in case mix, and to account for clustering of patients. The depend variable is a binary indicator of whether the patient has experienced any post-operative morbidity. I specify the model as:

$$\Pr(y_{ij} = 1 \mid x_i, j) = F(\alpha + \beta_j + \gamma x_i + \varepsilon_{ij}) \tag{1}$$

$$\beta_j \sim \mathcal{N}(0, \theta^2) \tag{2}$$

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \tag{3}$$

where $F(.)$ is the logistic function and y_{ij} is a binary variable that takes value 0 or 1. The remaining terms are: x_i a vector of (assumed exogenous) patient-level characteristics, α the intercept, β_j a surgeon-specific fixed effect, and ε_{ij} an error term capturing any over- or under-dispersion.

In the second step, I use the regression estimates from Equation (1) to compute a Standardized Risk Ratio (SRR) that measures surgeon's skills. The SRR is the ratio between the *predicted* and *expected* post-operative morbidity. The *expected* is the morbidity that would occur if a particular set of patients were treated by the average surgeon (i.e. the national average expected post-operative morbidity). The *predicted* is the equivalent number for a *specific* surgeon. I compute these terms as follows:

$$\text{predicted}_j = \sum_{i \in j} F\big(\widehat{\alpha} + \widehat{\beta}_j + \widehat{\gamma} x_i\big) \tag{4}$$

$$\text{expected}_j = \sum_{i \in j} F\big(\widehat{\alpha} + \widehat{\beta} x_i\big), \tag{5}$$

The post-operative morbidity standardized risk ratio (SRR) is:

$$\mathrm{SRR}_j = \frac{\mathrm{predicted}_j}{\mathrm{expected}_j}. \tag{6}$$

A SRR of 1 indicates that the level of post-operative morbidity for surgeon j is as expected given the pool of patients over which this is computed. An SRR above (below) 1 indicates that the surgeons is under- (over-) performing relative to the national average. In practice, as data on surgeons is sparse in HES, 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, I can fix the skill level and see how the performance of surgeons changes as the robot is introduced.

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 0.3, and the distribution is characterized by long tales to the right, suggesting that some surgeons perform particularly poorly.

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 with a SRR pre-robots below the 10th percentile (low post-operative morbidity), bottom surgeons are identified as those with a SRR above the 90th 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 2006, patients operated on by high-skilled surgeons experienced 5 days of post-operative LOS while low-skilled surgeons had an average post-operative LOS of 7 days. By 2012, this was down to 2 and 3 days respectively. A similar trend can be observed when inspecting the share of patients experiencing an adverse event from surgery. For highly skilled surgeons, 1 in 4 patients experienced an adverse event from surgery in 2003. For low-skilled surgeons, almost 80 percent of patients experience these events.

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 Horn et al. (2021). Horn et al. (2021) shows that that adopting a robot drives prostate cancer patients to the hospital.

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 patients are not randomly allocated to the robotic approach, and the choice of treatment may be endogenous to their characteristics. Moreover, individuals 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 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 a structural 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 resistance to treatment, which is an unobserved variable that influence selection. Identification of MTE is intuitively similar to the IV, and does not require any stronger assumption, but is more informative in the presence of heterogeneous effects Cornelissen et al. (2018). Heckman and Vytlacil (2005) shows that we can recover population level treatment effects as weighted averages of marginal treatment effects. The average treatment effect (ATE) is the unweighted average of the MTEs, and it is point identified for $0, 1 \in \text{supp } P(Z)$. The average treatment effect on the treated (ATT) is a weighted average of the MTEs where individuals with low values of resistance to treatment are given higher weights. The average treatment effect on the untreated (ATU) is a weighted average of the MTEs where individuals with high values of resistance to treatment are given higher weights.

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 $D \in [0, 1]$. Both outcomes depend on some observed characteristics X and an unobserved component which is additively separable,

$$Y_0 = h_0(X) + \epsilon_0 \tag{7}$$

$$Y_1 = h_1(X) + \epsilon_1 \tag{8}$$

ϵ_0 and ϵ_1 are error terms of mean zero conditional on X. Hence, $h_D(X) \equiv E[Y_D|X]$ for $D \in [0, 1]$. For each individual, depending on the choice of treatment, only one outcome is actually observable. The treatment choice is represented by an index threshold crossing model

$$D = 1[D^* \geq 0] \tag{9}$$

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

$$D^* = g(Z) - V \tag{10}$$

The Z vector may include some or all of the variables in X, but crucially includes a continuous instrumental variable that affects only the treatment status. V is what is called resistance to treatment. This a continuously distributed random variable representing all unobserved factors that make an individual less likely to choose $D = 1$. A key feature of this model is that 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.

The key assumptions associated with equations (7)-(10) are:

Assumption 1. $(\epsilon_0, \epsilon_1, V)$ are statistically independent of $Z|X$ (*Independence*).

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

The model, as presented it, combined with the two assumptions, 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).

To define the MTE, the decision rule is conventionally expressed in terms of the propensity score $P(Z)$, i.e. the probability that a patient is being assigned to the treatment given the set of observed covariates.

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

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. When the decision rule is expressed as $D = 1[P(Z) - U \geq 0]$ we can observed that Z affects the decision of treatment only through the propensity score $P(Z)$. The MTE, is then defined by the following conditional expectation:

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

The MTE is the average gain from treatment for individuals with characteristics $X = x$, and indifferent between treatments at the propensity score $P(Z) = u$. Since U is the quantile of V, the variation of $MTE(x,u)$ over values of u - the shape of the MTE curve - reflects how treatment effect varies with different quantiles of the unobserved resistance to treatment. The MTE can be non-parametrically identified, under Assumption 1 and 2, using the method of local instrumental variables (Heckman and Vytlacil, 1999). I will impose however two further assumptions that are commonly used in the applied literature.

Assumption 3. $E[U_1 - U_0 | X = x, U_d = u_d]$ does not depend on X;

Assumption 4. $\mu_j(X) = \beta_j X$ $\forall$ treatment option $j \in [0,1]$

Under Assumptions (3)), the MTE is additively separable in x and u and can be written as:

$$MTE(x,u) = x(\beta_1 - \beta_0) + E[\epsilon_1 - \epsilon_0 | X = x, U = u]$$

Assumption (4) imposes linearity in the expected values of Y_0 and Y_1 so that the expected outcome conditional on the value propensity score and X is:

$$E[Y|X = x, U_d = u_d] = x\beta_0 + x(\beta_1 - \beta_0)p + K(p) \tag{11}$$

where $K(p) \equiv E(\epsilon_1 - \epsilon_0 | U \leq p)$ is a non-linear function of the propensity score. $K(p)$ captures all the 'essential heterogeneity' in the outcome. As shown in Carneiro et al. (2011), the derivative of the outcome Y with respect to p identifies the MTE for individuals with $X = x$ and $U_d = p$:

$$\frac{\partial E[Y|X = x, P(Z) = p]}{\partial p} = x(\beta_1 - \beta_0) + \frac{\partial K(p)}{\partial p} = MTE[X = x, U = p] \tag{12}$$

The intuition is simple. Increasing the propensity score by a small amount shifts previously indifferent individuals into treatment and changes the observed outcome Y. 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. I will estimate this derivative using local instrumental variable method introduced by Heckman and Vytlacil (1999).

The true distribution of $K(p)$ is unknown, and the function could be non-linear. Thus, I will model the outcome both parametrically and non parametrically (partially-linear) in terms of the unobserved term.

Application of the MTE

In practice, I will have two margins over which to evaluate treatment effects. Namely, the patient length of stay in hospital and the probability of an averse event from surgery. I will model both outcomes as a linear function of patients characteristics, X, and surgeon skills:

$$Y_1 = \beta_1 X + \delta_1 Skills + \epsilon_1$$
$$Y_0 = \beta_0 X + \delta_0 Skills + \epsilon_0$$

So that the observed outcome can then be expressed as:

$$Y = Y_0 + D[\underbrace{(\beta_1 - \beta_0)X}_{\Delta_1} + \underbrace{(\delta_1 - \delta_0)Skills}_{\Delta_2} + \underbrace{U_1 - U_0}_{\Delta_3}] \tag{13}$$

and the individual specific treatment effect is the sum of Δ_1, Δ_2, and Δ_3. A negative treatment effect will reflect an increase in the performance of surgeons from using the robot (i.e. shorter length of stay and less adverse events). The choice of treatment, robotic or traditional surgery, will also be a function of surgeons skills and patients characteristics.

$$D = 1[D^* \geq 0] \tag{14}$$

The unobserved latent D^* will depend linearly on its components:

$$D^* = \beta_d X + \delta_d Skills + \gamma_d Z - V \tag{15}$$

Two patients with identical observed characteristics, and operated by a surgeon with equal skills, will be allowed to differ in terms of V. I will estimate the parameters in Equation 15 using a probit model. X will include a large set of patients clinical characteristics and demographic characteristic. I will include my measure of skills either as it is, or I will estimate the model using a binary variable for whether the surgeon as an SRR above the median, which I call a high-skilled indicator. I will estimate the parameters of Equation 11 using the method of local instrumental variables, and I will use these to compute the MTE as:

$$\widehat{MTE}(x,u) = (\hat{\beta}_1 - \hat{\beta}_0)X + (\hat{\delta}_1 - \hat{\delta}_0)Skills + \hat{K}'(u) \tag{16}$$

In my baseline specification I assume joint normality of $(\epsilon_0, \epsilon_1, V)$. In an alternative specification I let $K(p)$ be a polynomial of degree two in p. Finally, I also estimate MTE under no parametric assumption on the $K(p)$ conpenent. The parameter δ_d will inform us on whether skills matter for the use of the robot, and its relationship with $(\delta_1 - \delta_0)$ will describe the extent to which surgeons choose to use the robot based on their gains. Eventually, as I estimate the β and δ parameters, I will be able to speak to both the heterogeneity in gains and the extent to which selection occurs on observable and unobservable characteristics.

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 in a monotone way. Moreover, ideally, the instrument should have enough variation to generate a propensity score with full common support (Cornelissen et al., 2016).

I use the fact that robots have been acquired under no centralized strategy, leading to a staggered adoption, to distil out an exogenous variation in the probability of treatment.

In practice, I propose 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 randomize individuals into treatment. I will refer to this instrument with the name Z_{days}, and compute it for each patient as:

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

t is the date on which a hospital consultant first diagnosed the patient with prostate cancer. As diagnosis of prostate cancer requires a biopsy which is performed in hospital, the diagnosis date is identifiable in HES. T_R is the date on which the closest hospital to the patient performed its first robotic assisted prostatectomy. Therefore, Z_{days} is the number of days between the patient diagnosis and a proxy for the date the nearest hospital adopted robotic surgery.

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.

I expect this relative diagnosis timing to randomize individuals into different probabilities of receiving treatment. Patients diagnosed before adoption will be less likely to get treated than those diagnosed later, while patients diagnosed after will have a higher probability of being operated on with the robot. The key idea is that the date of adoption of the robot is random relatively to the individual health status, and hence unrelated to his potential outcomes. Consequentially, Z_{days} arguably affects 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 3 presents the results 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, i.e. three years prior to the first robotic radical prostatectomy. The coefficient on Z_{days} is not statistically significant. Column (4) to (6) show the coefficients estimated from a OLS regression of a binary indicator of adverse events on Z_{days} for increasingly richer specifications. The coefficient in column (4) is negative and statistically significant, but after controlling for patient characteristics, this correlation disappears. Overall, this is suggestive that the exclusion restriction is likely to be satisfied conditional on the covariates included in the model.

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.

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 from 2006 to 2010. Table 2 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. My conclusion are invariant to this specification.

I show how Z_{dist} is distributed in Figure 9. The average relative distance is patient 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 estimated 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. In Figure 12, 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 11, I show the average predicted probability evaluated as different values of this instrument. An individual whose value of Z_{dist} is 30 km has a probability of being treated of 0.4, doubling this distance reduces this probability by almost fifty percent.

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

7 Results

In this section, I present the results from the estimation of the structural model and the marginal treatment effects. First, I discuss the estimated parameters from the selection equation (i.e. Equation (10)), which I use to derive the propensity score $P(Z)$. The selection equation is estimated using a probit regression. Second, I present the parameters from Equation (16) estimated using the local instrumental variable method (Heckman and Vytlacil, 1999). These parameters speak to the differences in returns explained by the characteristics included in the model. In particular, $\delta_1 - \delta_0$ from Equation (16) will tell us whether the return from using the robot depend on the surgeon level of skills. However, the parameters are also informative on average differences in the untreated state attributable to observable surgeon and patient characteristics, which are captured by β_0 and δ_0. Third, I will show the MTE curve which I obtain under three different parametrizations of the model, which will allow me to discuss the pattern of selection on gains that is due to unobservable characteristics not captured in the model.

Lastly, I will present the estimates for the conventional treatment effect parameters and show a policy simulation.

Selection into robotic surgery

To estimate the selection equation, I use a probit model where I specify as explanatory variables the instruments, my measure of surgeon's skills, and a number of control variables encompassing patient clinical and demographic information, the distance to the patient closest hospital, and whether this is a teaching hospital.

In Table 5, I show the coefficients on skills from the selection equation. The dependent variable is a binary indicator for whether the patient has received 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 a high-skilled indicator. The high-skilled indicator takes value 1 if the surgeon is above the median level of skills, and zero otherwise.

To illustrate the magnitude of this relationship, in Figure 15, I show graphically how the probability of using the robot depends on skills. These are the marginal effects at different levels of my measure of skills, which I have normalized to be 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 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 16, I show how the predicted probability varies by age for surgeons above and below the median of skills. For both types of surgeons, the likelihood of using the robot diminishes with the age of the patient. But, at all age levels, high skilled surgeons are more likely to operate with the robot. The fact that higher skilled surgeons use the robot more intensively, conditional on patient characteristics, suggests either that they have higher returns from using the robot or that they are over using the technology for some reasons Chandra and Staiger (2020).

Skills and technological gains

The MTE framework allows estimating two sets of coefficients from the outcome equation (i.e. Equation 13). First, it allows estimating the relationship between the variables included in the model and the outcomes in the untreated state. This is the δ_0 vector from Equation 13. Second, I can estimate how the effect from using the robot changes depending on the observed characteristics included in the model. This is represented by the vector $\delta_1 - \delta_0$ in the Equation 13.

To fix ideas, I display below the two outcome equations, whose parameters I estimate. For clarity, I do not show the control variables in this representation. p represent the propensity score estimated from the selection equation using a probit regression.

$$E[log(LOS_{ij})|Skills = s, U_d = u_d] = \delta_0^L Skills_j + (\delta_1^L - \delta_0^L) Skills_j p + K^L(p) \tag{19}$$

$$E[\text{Adverse event}_{ij} = 1|Skills = s, U_d = u_d] = \delta_0^B Skills_j + (\delta_1^B - \delta_0^B) Skills_j p + K^B(p) \tag{20}$$

I estimate the coefficients under two specifications of the skills variable. I use the actual value of the SRR, a continuous measure of skills, or a high-skilled indicator (i.e. SRR below median). I display the coefficients in Table 6, and what emerges is the following.

First, surgical skills matter for both robotic surgery and traditional surgery, but much less when surgeons use the robot. The coefficient on skills is negative and statistically significant in the untreated state. With traditional surgery, high skilled surgeons' patients have better outcomes than the patients of lower skilled surgeons. The results suggest that, in the untreated state, the patients of high-skilled surgeons are 8 percentage points less likely to experience an adverse event from surgery relatively to those of low-skilled surgeons. Moreover, when operated with the traditional method, the patients of lower skilled surgeons stay in hospital almost twice as much as the patients of high skilled surgeons. However, the coefficient in the untreated state is positive and statistically significant, indicating the treatment effect is more negative the lower are the skills of the surgeons. Both post-operative morbidity and length of stay decrease from using the robot, but more significantly for lower skilled surgeons. The fact that the treatment effect is significantly more negative for lower skilled surgeons suggests limited complementarities with high skilled surgeons.

Second, the robot appear to have an equalizing effect on the performance of surgeons. Using the robot reduces substantially the difference in performance between high and lower skilled surgeons. For length of stay, differences in patients outcomes between low and high skilled surgeons shrink by 50 percent. For post-operative morbidity, with robotic surgery, the patients of lower skilled surgeons are only 1 percentage point less likely to experience an adverse event from surgery, relatively to those of high skilled surgeons. This result is important, it suggests that technology may help us reduce unwarranted variation in patient outcomes that comes from differences in skills. Comparing these coefficients to the estimates from the selection equation allows identifying whether surgeons of different quality select on the basis of 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.

Returns to treatment based on unobserved characteristics

Using the model parameters, I can estimate the MTE curve relating the returns from using the robot to the unobserved resistance to treatment. The latter summarized all surgeons and patients characteristics not included in the model that affect negatively affect the choice of using the robot. 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 is shown in Figure 17.

The MTE curve mimics the pattern of negative selection on observables. The relationship between the unobserved resistance to treatment V and the average unobserved gains from treatment is consistently negative for LOS, 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, have the lowest returns from the treatment. The shape of MTE curve for the probability of adverse events suggests a similar story, but we can't reject homogeneity on unboservable characteristics.

In Figure 18, I relax the assumption of joint normality and let the function $K(p)$ be approximated by a polynomial of p. Estimation in this case is achieved by a two-step procedure discussed in Heckman et al. (2006). For LOS, 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 by a partially linear regression of Y on X and $P(Z)$, and the estimation of $K(p)$ is achieved by a local polynomial regression. For a detailed discussion of the estimation methods available, refer to Heckman et al. (2006). Still, the MTE curve suggest negative selection for LOS. The MTE curve for the probability of an adverse event becomes instead significantly flatter and with a substantial negative deep around $U_D = 0$.

In turn, my results uncover a significant pattern of negative selection. The patients that have the most to gain from being operated with the robot are the least likely to receive robotic surgery. This pattern may be explained in a number of ways. The most likely explanations, between which I can't discriminate, are surgeons' differences in preferences and beliefs. For example, higher skilled surgeons may have a particular taste for technology, or may be overconfident that the robot improves patients' outcomes.

Conventional treatment effects and policy simulation

In Table 7 and Table 8, I show the treatment effects parameters, which I compute by appropriately integrating over the MTE curve. 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 findings discussed, the average treatment effect of the untreated (ATU) is more negative than the effect on the treated (ATT). This is the result of the fact that the patients least likely to receive robotic surgery are also those that would gain the most from it. Interestingly, the magnitude of these effects are small relatively to the effect that robotic surgery has on the difference between high and lower skilled surgeons.

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 unobservables in the model. Heckman and Vytlacil (2005) show how to compute the Policy Relevant Treatment Effect (PRTE) which is the mean effect from going to the baseline policy to an alternative policy per net person shifted in to treatment.

I compute this parameter for a counterfactual scenario in which I assign to lower skilled surgeons the same probability of using the robot as high skilled surgeons. Basically, I test the effect of using the robot if lower skilled surgeons were mandated to use the robot with the same intensity as high skilled surgeons. This policy simulation speaks to a hypothetical counterfactual scenario in which the 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 show what would happen 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 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 21 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 on what technologies can actually achieve. Using the case of robots in surgery, I showed that new technologies may help to reduce variation in the performance of workers. This is a particularly important finding in healthcare, where equity is a central concern of regulators and policymakers. But, can be applied more broadly to any context in which the provision of services, or the production of goods, 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. I show that although patients' outcomes may improve only marginally on average, robots have the potential to reduce unwarranted differences in patient outcomes that arise from heterogeneity in surgeons' skills. I have shown indeed that the robot helps lower skilled surgeons to perform almost as well as high skilled surgeons. This effect is by itself worth the adoption of these machines if we want to grant to all patients the same level of high quality care.

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. (2002). Technical change, inequality, and the labor market. *Journal of economic literature*, 40(1):7–72.

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

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.

Card, D. and DiNardo, J. E. (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of labor economics*, 20(4):733–783.

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).

Chandra, A. and Skinner, J. (2012). Technology growth and expenditure growth in health care. *Journal of Economic Literature*, 50(3):645–80.

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.

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.

Colombo, R., Naspro, R., Salonia, A., Montorsi, F., Raber, M., Suardi, N., Saccà, A., and Rigatti, P. (2006). Radical prostatectomy after previous prostate surgery: clinical and functional outcomes. *The Journal of urology*, 176(6):2459–2463.

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. (2007). The lifetime costs and benefits of medical technology. *Journal of Health Economics*, 26(6):1081–1100.

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

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).

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. (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, R. A. (2021). Technology adoption and market allocation: The case of robotic surgery. Technical report, National Bureau of Economic Research.

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)*.

Jha, V., Martin, D. E., Bargman, J. M., Davies, S., Feehally, J., Finkelstein, F., Harris, D., Misra, M., Remuzzi, G., Levin, A., et al. (2017). Ethical issues in dialysis therapy. *The Lancet*, 389(10081):1851–1856.

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.

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).

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.

Newhouse, J. P. (1992). Medical care costs: how much welfare loss? *Journal of Economic perspectives*, 6(3):3–21.

Peters, B. S., Armijo, P. R., Krause, C., Choudhury, S. A., and Oleynikov, D. (2018). Review of emerging surgical robotic technology. *Surgical endoscopy*, 32(4):1636–1655.

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.

Stoye, G. (2022). The distribution of doctor quality: evidence from cardiologists in england. *IFS Working Paper*.

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).

Tsai, T. C., Joynt, K. E., Orav, E. J., Gawande, A. A., and Jha, A. K. (2013). Variation in surgical-readmission rates and quality of hospital care. *New England Journal of Medicine*, 369(12):1134–1142.

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).

Warren, H. and Dasgupta, P. (2017). The future of robotics. *Investigative and clinical urology*, 58(5):297–298.

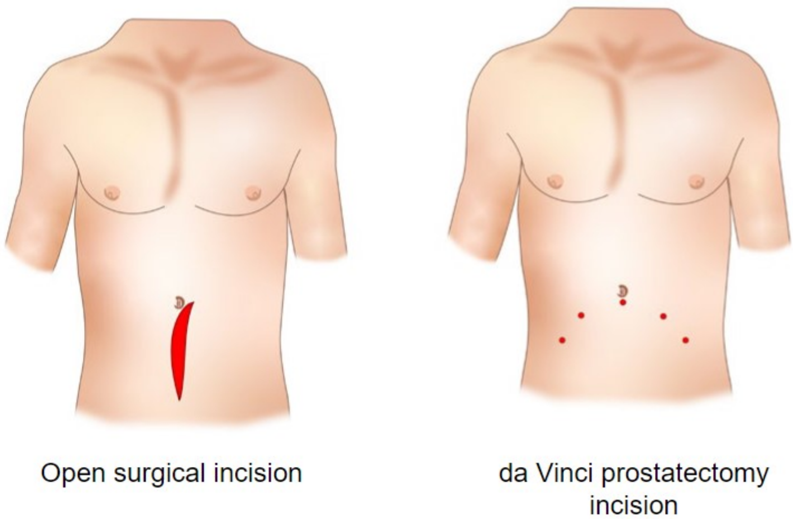
Yaxley, J. W., Coughlin, G. D., Chambers, S. K., Occhipinti, S., Samaratunga, H., Zajdlewicz, L., Dunglison, N., Carter, R., Williams, S., Payton, D. J., et al. (2016). Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled 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).

Figure 1: Picture of a Da Vinci surgical system

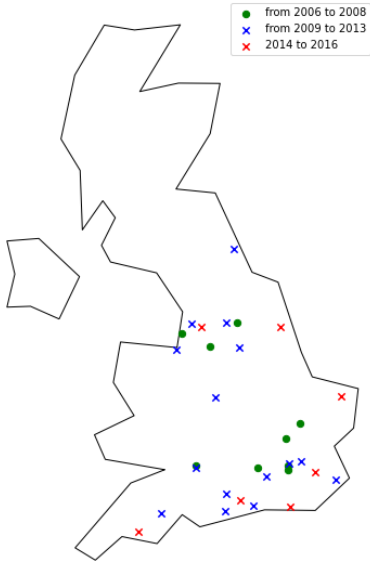
Note: Picture shows the Da Vinci Robot surgical system from Intuitive Inc. On the left the surgeon sitting at the console. Above the operating bed the robotic arms. On the right the vision cart.

Figure 2: Comparison of incisions



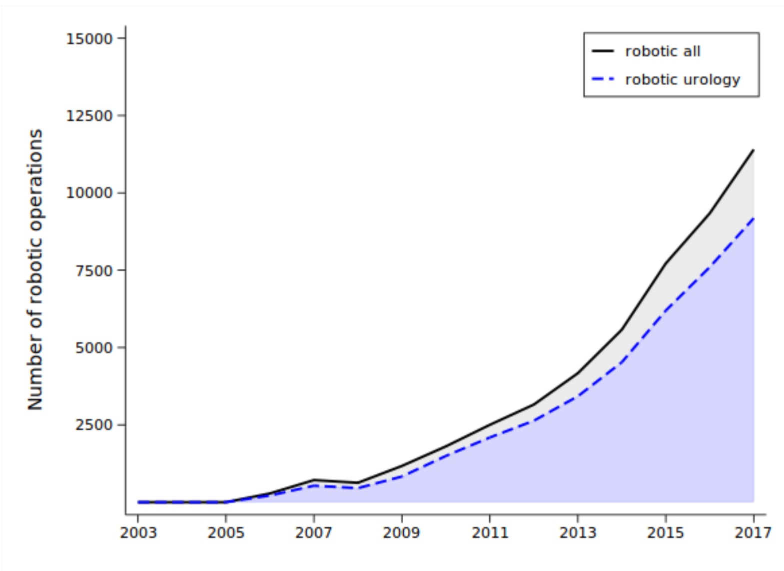
Note: Comparison of incisions required for traditional and robotic radical prostatectomy

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



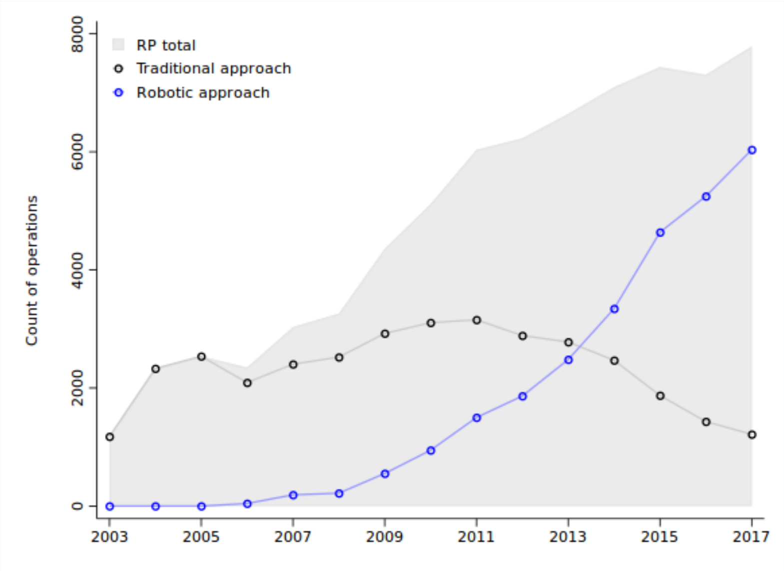
Note: Graph is produced using the HES data. The green crosses represent hospitals that are observed using for the first time a robotic operation code in 2006. The blue dots represent hospitals that are observed using for the first time a robotic operation code between 2007 and 2009. The red dots represent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

Figure 4: Diffusion of robotic surgery in the NHS



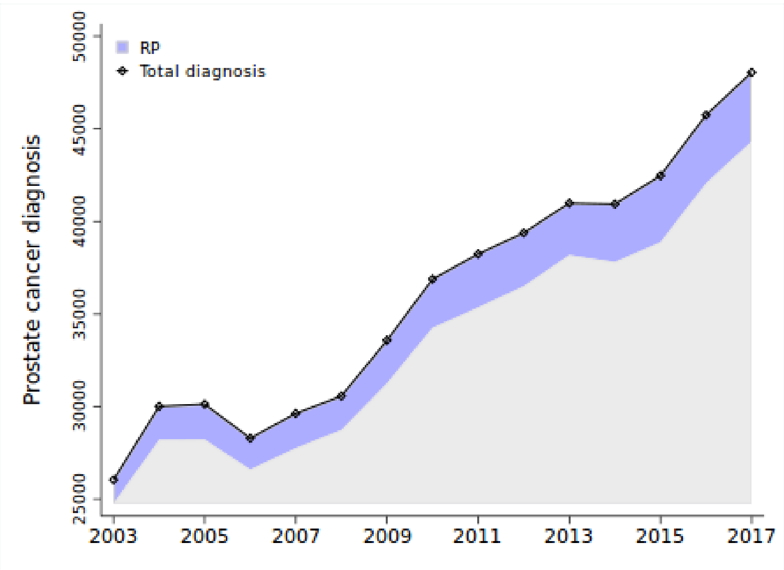
Note: The picture shows the number of robotic operations by year for urology compared to all other specialties in which robots are used. The data is from the Hospital Episodes Statistics.

Figure 5: Volume of robotic and traditional radical prostatectomies



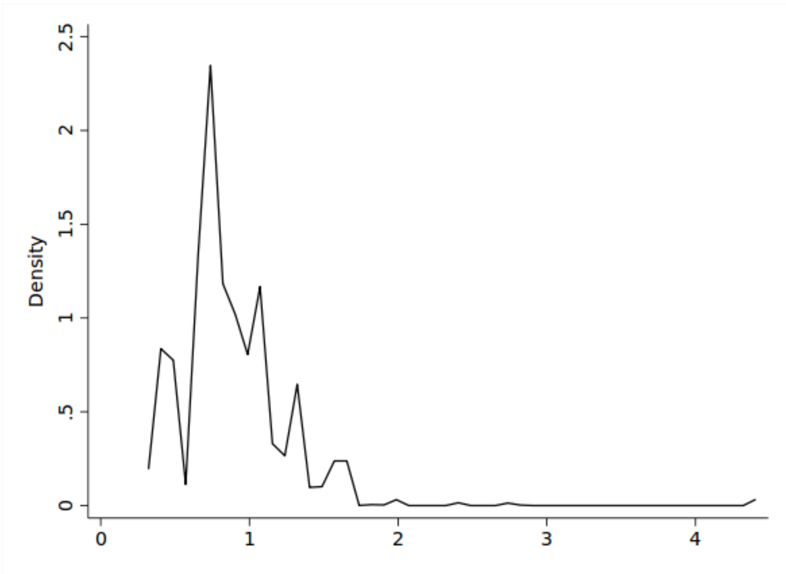
Note: Graph is produced using the HES data. The shaded gray area represents the total number of radical prostatectomies performed by NHS Hospitals in England. The black dots represent the number of radical prostatectomies performed using the open approach. The blue dots represent the number of radical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

Figure 6: Surgical interventions as a share of prostate cancer diagnosis



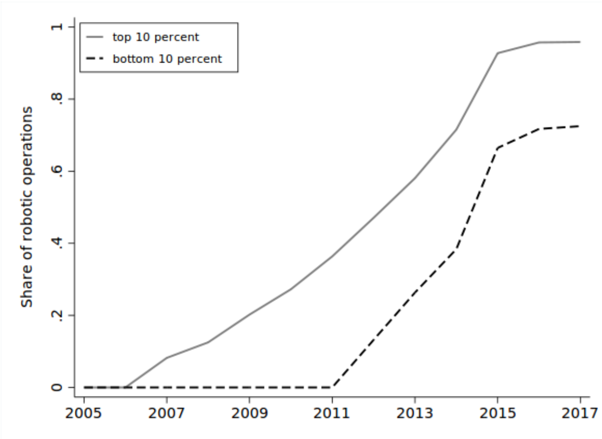
Note: Graph is produced using the HES data. The shaded blue area represents the number of radical prostatectomies performed by NHS hospitals in England. The shaded gray area represents the number of patients with prostate cancer that have undergone radio therapy treatment. The black line represents the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

Figure 7: Distribution of surgical skills

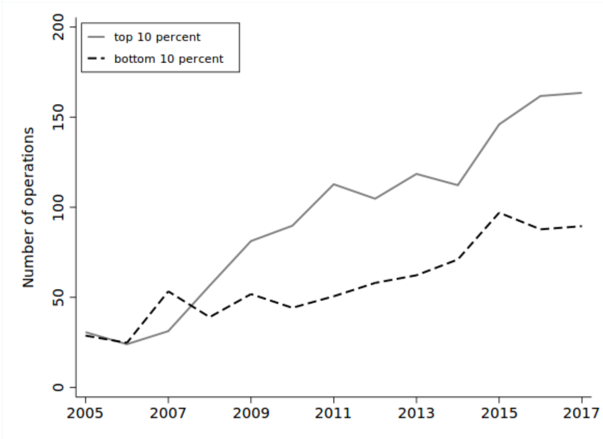


Note: Distribution of standardised risk ratios (SRR). Outcome is post-operative morbidity. SRR is computed as the ratio between predicted and expected morbidity. Predicted and expected post-operative morbidity are obtained by estimating a hierarchical logistic model accounting for patients' clinical and demographic characteristics. Estimates using all prostatectomy patients from 2005 to 2007.

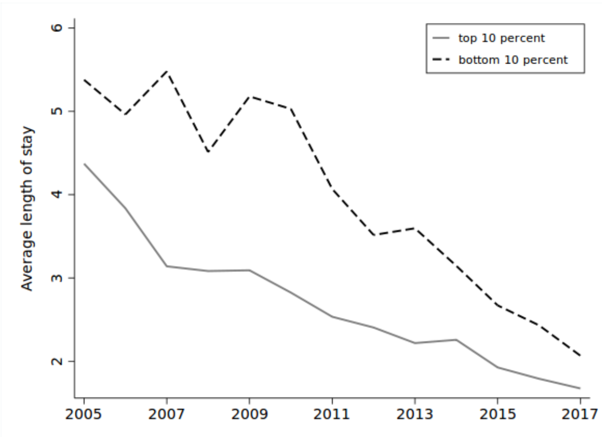
Figure 8: Key empirical facts



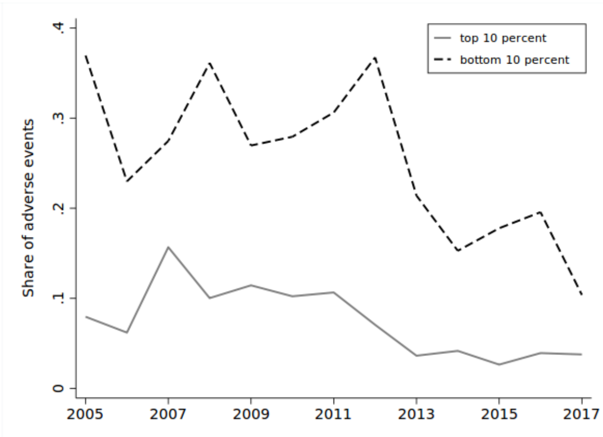
(a) Robotic rate



(b) Number of operations



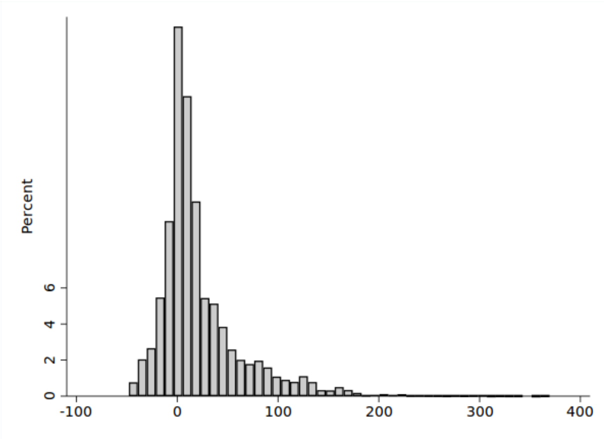
(c) Length of stay



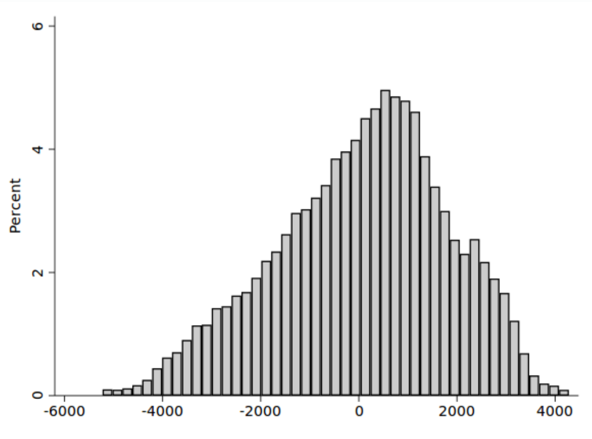
(d) Adverse events

Note: Surgeons are characterised as top quality if they have a SRR in the pre-robot period below the 10th percentile, they are characterised as bottom if they have a SRR above the 90th percentile. Panel (a) shows the level of robot use for two groups of surgeons. Raw robotic use is computed as the number of robotic operations per year over the total number of operations at the hospital level and then aggregated for the two groups. Panel (b) shows the number of operations performed on average by the two groups of surgeons over time. Panel (c) shows the average length of stay of patients in the two groups over time. Panel (d) shows the rate of adverse events for the two groups. The rate of adverse events is computed as the number of patients experiencing an adverse event from surgery over the total number of operations.

Figure 9: Variation of instrumental variables in sample data



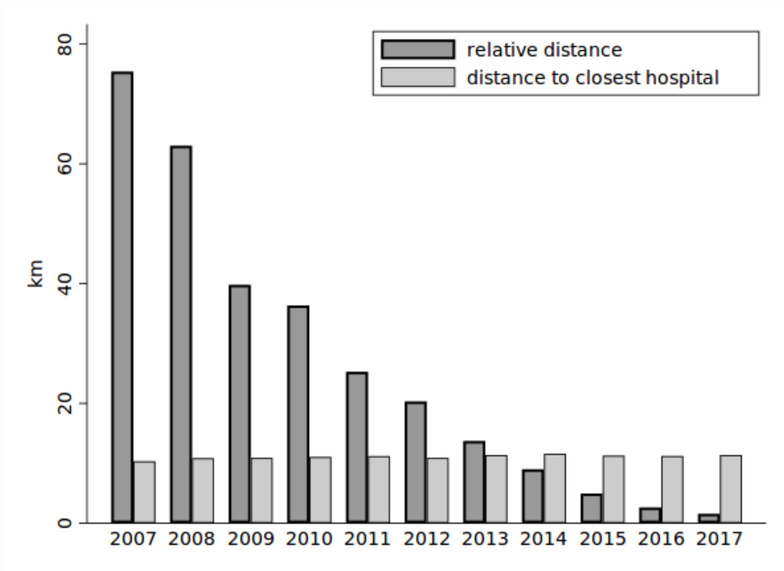
(a) Z_{dist} instrument



(b) Z_{days} instrument

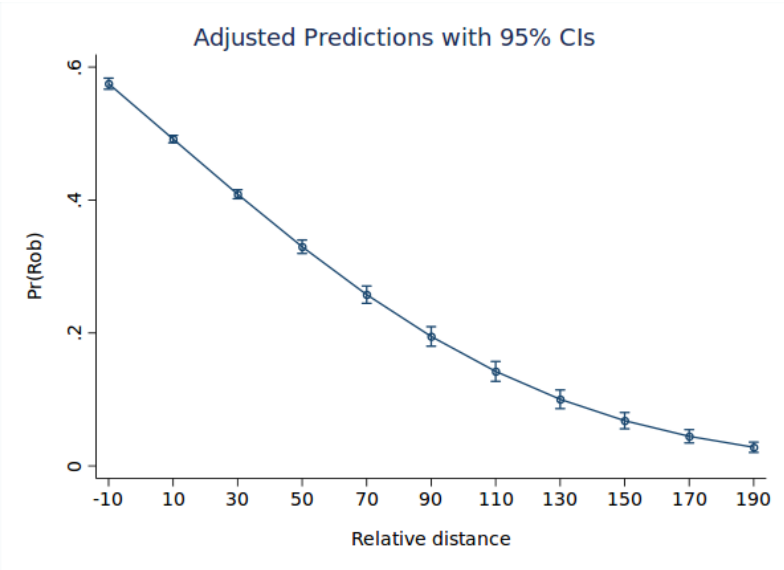
Note: Panel (a) plots the instrument Z_{dist} defined as the relative distance between the patients nearest hospital capable of offering robotic assisted radical prostatectomy and the closest hospital offering traditional radical prostatectomy. The distance is expressed in kilometers. Panel (b) plots 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.

Figure 10: Average relative distance to robotic hospital and to closest hospital



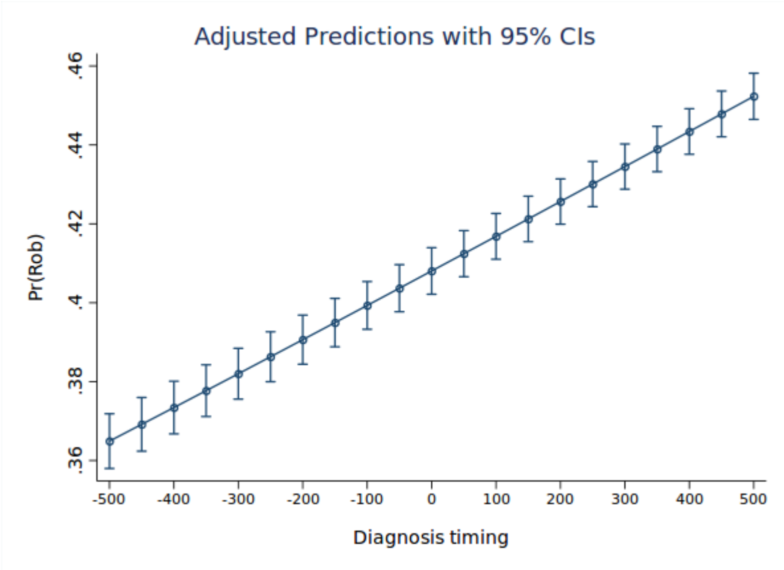
Note: Figure shows the average relative distance to hospital offering robotic surgery. Relative distance is computed as the difference between the individual distance to the closest hospital offering robotic technology and the distance to the closest hospital offering only traditional surgery. Distance to the closest hospital is allowed to be the distance to the closest robotic hospital or a hospital not offering robotic surgery. The individual 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.

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



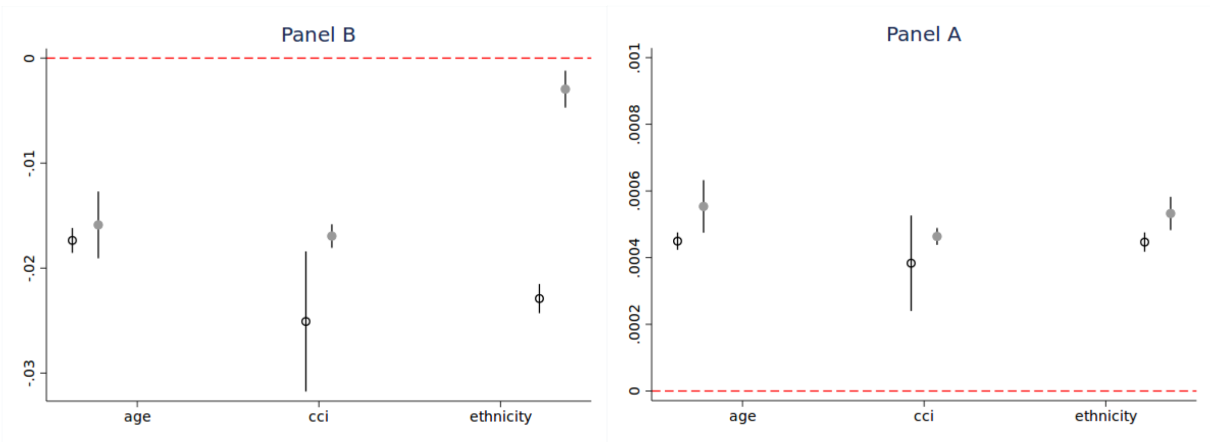
Note: Probit regression estimates, dependent variable robotic approach. Marginal probability estimated at different value of relative distance to hospital offering robotic approach. Covariates in the model at means, include demographic and clinical patient characteristics, distance to closest hospital, 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 12: 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.

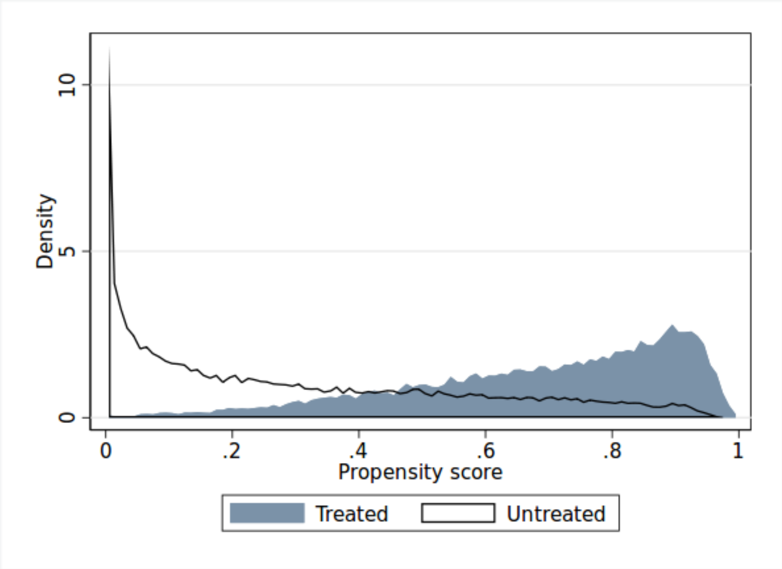
Figure 13: Test for monotonicity of the instruments



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

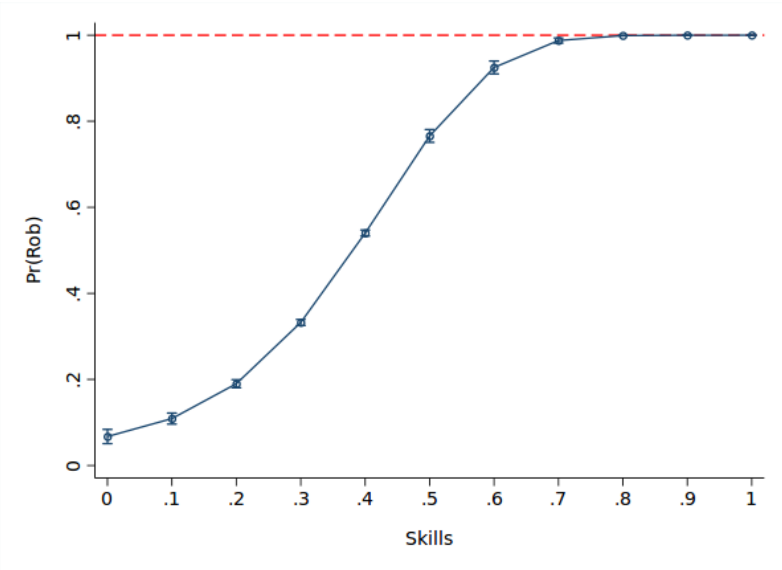
Note: OLS regression for subsets of the population. Age above and below 55. CCI above and below 2. Ethnicity white and all other ethnicity. Coefficients estimated using logistic regression. Dependent variable is a binary indicator of whether the individual has been operated using the robot. Demographic 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.

Figure 14: Common support



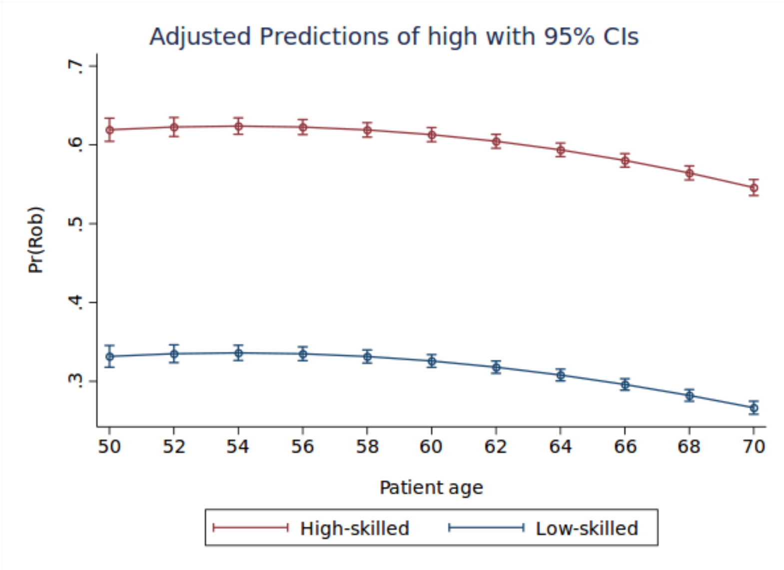
Note: Unconditional support jointly generated by instruments and covariates. Covariates in the model include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest hospital is a teaching hospital, instrument Z_{dist}, and Z_{days}. Model controls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

Figure 15: Estimated probability of robotic approach from selection equation - at values of standardised skills measure



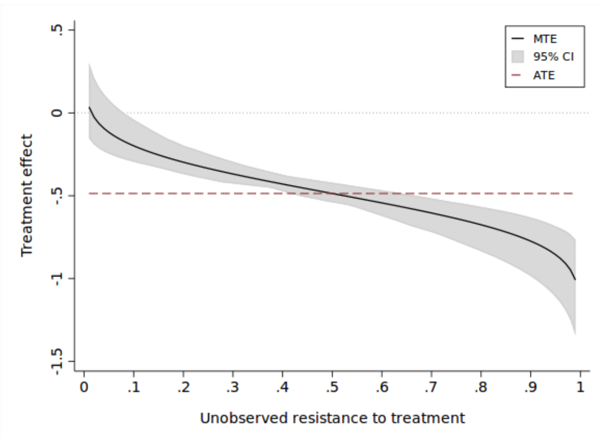
Note: Probit regression estimates, dependent variable robotic approach. Marginal probability estimated at different of the skills variable. Covariates in the model at means, include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest hospital 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. Skills measured with SRR normalised to be between 0 and 1. Delta method for standard errors.

Figure 16: Estimated probability of robotic approach from selection equation - at Age values by value of high-skilled indicator

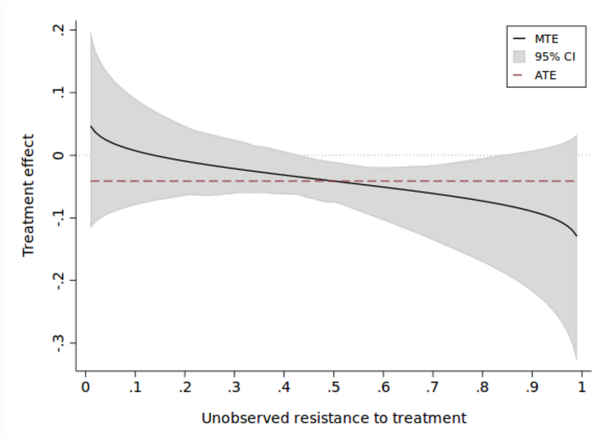


Note: Probit regression estimates, dependent variable robotic approach. Marginal probability estimated at different value of patient age. Covariates in the model at means, include demographic and clinical patient characteristics, distance to closest hospital, indicator for whether the closest hospital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

Figure 17: MTE curve - Normal



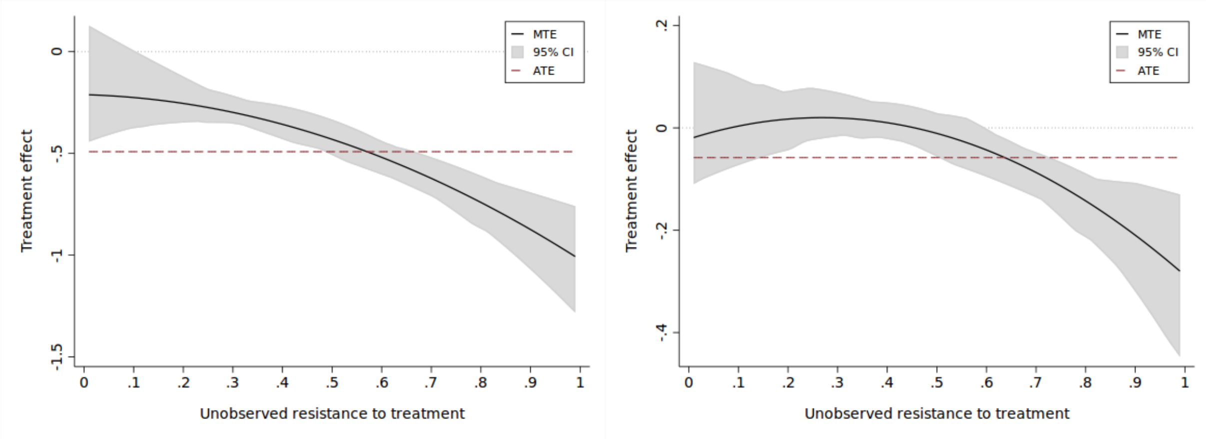
(a) LOS



(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

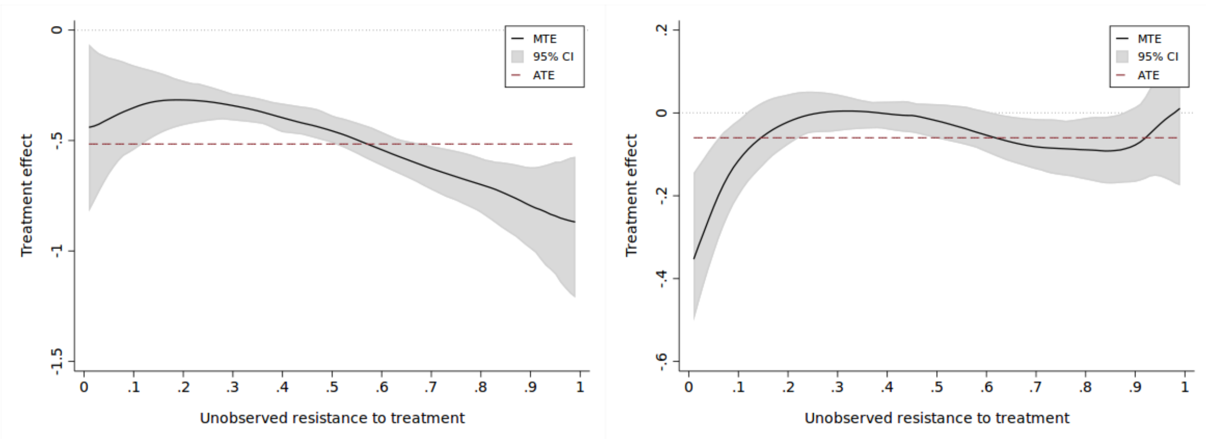
Figure 18: MTE curve - Polynomial



(a) LOS (b) Adverse Event

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

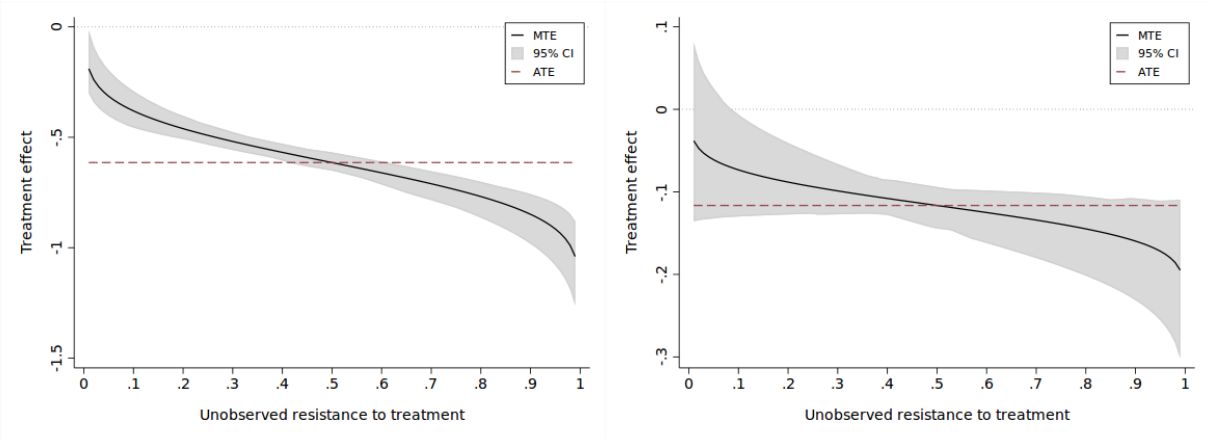
Figure 19: MTE curve - Semiparametric



(a) LOS (b) Adverse Event

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

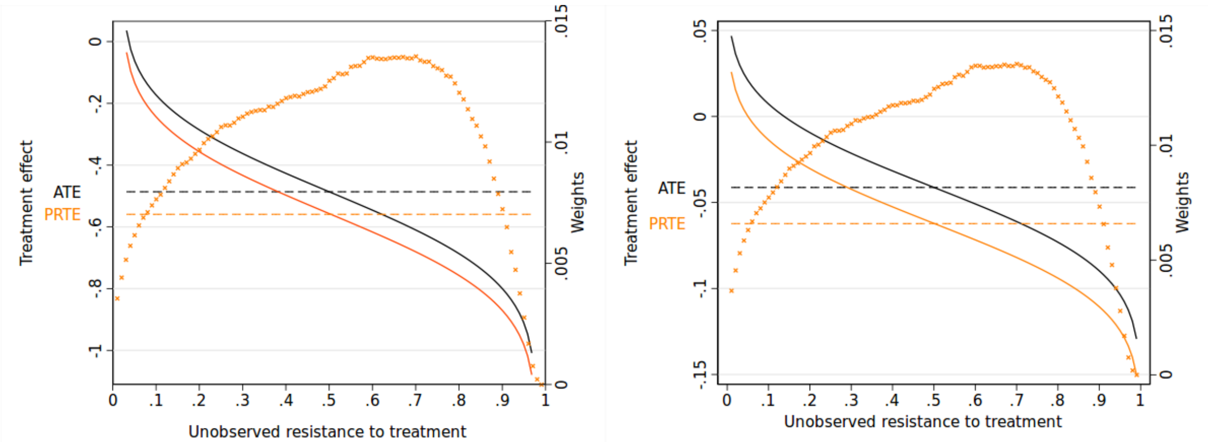
Figure 20: MTE curve - Normal with area fixed effects



(a) LOS (b) Adverse event

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

Figure 21: Policy simulations - MTE and PRTE



(a) LOS (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. In orange the estimated effects from policy simulation. Crosses indicated the weights.

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	

Table 2: Correlation between AMI patients mortality and Z_{dist}

	(1)	(2)	(3)
Z_{dist}	0.000449**	-0.000159	0.000314
	(0.000163)	(0.000185)	(0.000199)
Distance closest hospital		0.00269*	0.00111
		(0.00107)	(0.00130)
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 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
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.

Table 3: Correlation of surgical outcomes and Z_{days} - pre-robots

	Length of stay			Adverse event		
	(1)	(2)	(3)	(4)	(5)	(6)
Z_{days}	-0.000498	-0.000284	0.000726	-0.00117*	-0.000351	-0.000329
	(0.000477)	(0.000475)	(0.000495)	(0.000517)	(0.000508)	(0.000534)
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} mean	-2707	-2709	-2709	-2707	-2709	-2709
N	5566	5549	5549	5574	5557	5557

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
OLS regression. Coefficients and standard errors multiplied by 100. 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.

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***
	(0.0327)		(0.0319)	(0.0323)	(0.0334)	(0.0339)
Z_{days}		0.0574***	0.0413***	0.0405***	0.0418***	0.0243***
		(0.000476)	(0.000581)	(0.000593)	(0.000599)	(0.000696)
Marginal effects						
Z_{dist}	-0.651***		-0.307***	-0.316***	-0.319***	-0.272***
	(0.00842)		(0.0089)	(0.00896)	(0.00916)	(0.00892)
	(0.0327)		(0.0319)	(0.0323)	(0.0334)	(0.0339)
Z_{days}		0.0159***	0.0122***	0.0119***	0.0122***	0.00667***
		(0.0000669)	(0.000148)	(0.000152)	(0.000152)	(0.000185)
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 parenthesis. 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. Are controls include distance to 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 standardised risk ratio for post-operative morbidity (interpreted as the inverse of skills). Panel (A) shows the coefficients from the probit regression, Panel (B) the respective marginal average effects.

Table 5: Relevance of instruments - Probit regression dependent variable indicator of robotic surgery

	(1)	(2)	(3)
Coefficients			
Skills	1.188^{***}	1.624^{***}	
	(0.0247)	(0.0684)	
Skills^2		0.235^{***}	
		(0.0309)	
High-skilled indicator			0.739^{***}
			(0.0135)
Marginal effects			
Skills	0.309^{***}	0.321^{***}	
	(0.00609)	(0.00661)	
High-skilled indicator			0.199^{***}
			(0.00359)
Demographic controls	Yes	Yes	Yes
Clinical controls	Yes	Yes	Yes
Year-month	Yes	Yes	Yes
Day of the week	Yes	Yes	Yes
Area control	Yes	Yes	Yes
Robot (%)	0.49	0.49	0.49
Skills (SRR) mean	0.85	0.85	
High Skills (%)			0.51
N	52572	52572	52572

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Robust standard errors in parenthesis. 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. Are controls include distance to 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 are measured using the SRR, which is standardised risk ratio for post-operative morbidity (interpreted as the inverse of skills). High skills indicator takes value 1 for surgeons above the median of the distribution of the SRR.

Table 6: Heterogeneity in causal effects - Skills

	LOS (1)	Adverse Event (2)	LOS (3)	Adverse Event (4)
δ_0				
Skills	-0.302***	-0.167***		
	(0.0124)	(0.00842)		
High-skilled indicator			-0.220***	-0.0868***
			(0.0105)	(0.00704)
$\delta_1 - \delta_0$				
Skills	0.374***	0.0830***		
	(0.0327)	(0.0186)		
High-skilled indicator			0.321***	0.0701***
			(0.0216)	(0.0124)
Observations	51523	52572	51523	52572

* Standard errors bootstrapped with 100 repetitions $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 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 are measured using the SRR, which is the standardized risk ratio for post-operative morbidity (interpreted as the inverse of skills). High skills indicator takes value 1 for surgeons above the median of the distribution of the SRR. Estimated using 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}. Estimation of coefficients under the assumption of normality of unobserved components.

Table 7: Estimated treatment effects on Length of stay

	(1) Normal	(2) Normal FE	(3) Polynomial	(4) Semiparametric
ATE	-0.486***	-0.615***	-0.491***	-0.516***
	(0.0304)	(0.0229)	(0.0355)	(0.0341)
ATT	-0.189***	-0.421***	-0.171***	-0.255***
	(0.0443)	(0.0259)	(0.0462)	(0.0523)
ATU	-0.769***	-0.799***	-0.797***	-0.764***
	(0.0718)	(0.0446)	(0.0682)	(0.0670)
LATE	-0.454***	-0.629***	-0.456***	-0.492***
	(0.0199)	(0.0207)	(0.0248)	(0.0303)
Observations	51523	49379	51523	51523

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors bootstrapped with 10 repeitions. 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 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 8: Estimated treatment effects on Length of stay

	(1) Normal	(2) Normal FE	(3) Polynomial	(4) Semiparametric
ATE	-0.0412*	-0.116***	-0.0579***	-0.0602***
	(0.0170)	(0.0127)	(0.0147)	(0.0156)
ATT	0.0176	-0.0830***	0.0338	-0.0337
	(0.0284)	(0.0176)	(0.0255)	(0.0328)
ATU	-0.0978*	-0.149***	-0.146***	-0.0857*
	(0.0407)	(0.0208)	(0.0335)	(0.0387)
LATE	-0.0414**	-0.115***	-0.0480***	-0.0764***
	(0.0126)	(0.0132)	(0.0130)	(0.0147)
Observations	52572	50384	52572	52572

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors bootstrapped with 10 repeitions. 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 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.