Technology, skills, and performance: the case of robots in surgery

Elena Ashtari Tafti
University College London - Department of Economics
Institute of Fiscal Studies

I Introduction

In this paper, I explore how returns from the adoption of new technologies vary depending on the skills of the individual using it. Technology adoption is considered to be the engine that drives progress and economic growth. At the core of this belief, is the expectation that the technology will improve the way we do things. But often the anticipated benefits fail to materialize and debate erupts on whether the technology is actually beneficial (Hall, 2004).

I sustain that asking whether a new technology is better than the status quo is not the crux of the debate. Instead, we should ask ourselves which individuals may benefit from it, and which may not. I envisage two possible scenarios. In the first, technological gains are increasing in users skills. That is, the technology complements more strongly workers that have higher skills because it enhances their ability to perform. Returns from using the technology are then concentrated among these workers. In the second scenario, the technology changes the way a task in performed so that the skills of the worker matter less. In this case, returns are decreasing in skills and the technology acts as an equalizing force on the market by reducing workers heterogeneity. Either way, evaluating whether an innovation has the potential to improve performance requires considering how technological gains vary with users skills.

I bring this idea to the study of surgical robots. In healthcare, like in other settings, differences in providers' performance have preoccupied economists and policymakers (Skinner, 2011). Spending and treatment decisions are often believed to drive such differences Birkmeyer et al. (2013a); Tsugawa et al. (2017). In fact, heterogeneity in healthcare providers' skills may be at the root of this unwarranted variation (Chandra and Staiger, 2007; Hull, 2018; Stoye, 2022). Robotic technology can be a solution to this problem if its returns are decreasing in surgeons skills, or may exacerbate existing variation in surgical performance otherwise. In other words, the technology may improve the market as a whole if it can help 'weaker' surgeons to become better, rather than better surgeons to 'stronger'.

The study of robots in economics has been mostly confined to their effects on the labour market.

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 (Acemoglu and Restrepo, 2020). The subject of this literature are industrial robots; fully autonomous machines that can be programmed to perform manual tasks. In many applications, however, robots are controlled by a user and aid workers to perform complex assignments. This challenges the stereotype that robots are machines free from our human flaws, able to perform tasks precisely and consistently.

Surgical 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 robot. 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).

In England, the diffusion of robots coincided with an improvement in average surgical performance, and with a convergence in outcomes between high and low skill surgeons. I investigate whether these are attributable to the adoption of robotic surgery using administrative data from the National Health Service (NHS). I focus on estimating heterogeneity in the causal impact of robots on surgical performance for prostate cancer patients, the most common type of cancer in men in the United Kingdom (UK).

Part of my contribution is the recognition of a selection problem, together with the possibility that the effects of using the technology vary across both users and applications. 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. Certain patients may demand to be treated with the new technology, while others may not. Second, the performance of the technology may depend on the surgeon and the patient. That is, treatment effects may be heterogeneous. 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. Importantly, when treatment effects are heterogeneous, individuals may select 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). Ideally, we would like to observe for each patient his outcome with and without the robot. This is, of course, not possible as we only observe outcomes in either of the states. It is customary, then, to focus on some average difference in outcomes to draw inference on treatment effects. The MTE is the treatment effect for individuals at a particular margin of indifference. It is the average difference in outcomes for individuals that would be indifferent between traditional and robotic surgery at a particular value of the so-called resistance to treatment.

The building block of this approach is the generalized Roy (1951) model, where individuals select into treatment based on their expected gains. The econometrician explicitly models the decision to treat as a function of both observed and unobserved characteristics. The latter are summarized by a unique scalar V referred to as resistance to treatment (Zhou and Xie, 2019). I model the decision to use the robot as a function of surgeon's skills, patients characteristics, and the latent variable V. To test the hypothesis that surgeon's skills matter for technological gains, I allow surgical performance to depend on the choice of treatment (i.e. robotic or traditional surgery), patients' characteristics, surgeons' skills and their interactions. Hence, I allow for the possibility of selection and heterogeneous effects. Further, as no restriction is imposed on the relationship between outcomes and resistance to treatment, selection may occur on unobserved characteristics. This is captured by the MTE curve, which portrays the relationship between treatment effects and the latent V. For instance, a MTE decreasing (increasing) in resistance to treatment suggests positive (negative) selection based on unobservables.

Identification of the MTE curve requires no more assumptions than the standard instrumental variable approach, but it is more informative (Cornelissen et al., 2016). The MTE allows estimating all conventional treatment effects parameters (e.g. the average treatment effect, or ATE) but speaks to a broader set of issues. However, identification of this parameter 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). Few adopted 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 nationally and created differences in the availability of the technology overtime. Depending on their location, patients had to travel long distances or just needed to visit their closest hospital to be operated with the robot. Depending on the time of their diagnosis, patients could choose among few or many hospitals offering robotic surgery.

I exploit differences in diagnosis timing together with the fact that patients usually go to their closest hospital to construct this instrument. This is the number of days between the patient cancer diagnosis and the date the nearest hospital adopted the robot. I show that this variable is highly predictive of the probability of receiving robotic surgery. This result is robust to the inclusion of controls 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 robot adoption are significantly more likely to be operated with it. Nevertheless, for this is instrument to be valid, I need the relative diagnosis timing to impact outcomes exclusively through its effect on the probability of receiving robotic surgery. To provide evidence that this is the case, I test whether the instrument is correlated to prostate cancer patients' 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. I show that diagnosis timing and patients' outcomes are not correlated in the period before the introduction of 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 this parameter 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 the robotic technology. 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). A primary concern when using relative distance is that this may be correlated with health in ways not accounted by the model (Hadley and Cunningham, 2004). Because my instrument varies across geography and time, I can control for time-invariant differences in health across areas. 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. I use data on patients that suffered an acute myocardial infarction (AMI) to perform a falsification test. I compute for each of them their relative distance to a robotic hospital, for the year they suffered the AMI, and test whether this can explain variation in deaths. If relative distance was related to individuals' health in ways not captured by the model, this would surely emerge from this relationship. Reassuringly, hospital outcomes of AMI patients appear to be unrelated to their closeness to a robotic hospital.

I identify technological gains on two margins of surgical performance; the speed of recovery and the occurrence of adverse events from surgery (i.e. post-operative morbidity). I proxy the speed of recovery with the patient post-operative length of stay. Adverse events include deaths, emergency readmissions and complications from surgery specific to prostate cancer patients. I focus on these because 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 both important cost drivers and, for this reason, should be considered when evaluating whether a medical technology is worth adopting (Lotan, 2012).

To measure surgeons' skills, I develop a risk-adjusted measure that is informative of general and surgery specific surgical competences. This is based on standardized mortality and readmission ratios, which are routinely used to rank surgeons (Tsai et al., 2013). But, it also encompasses post-operative morbidity for patients with prostate cancer. The challenge is clearly to compare outcomes from surgeons that may have substantially different patients. I use a methodology developed used by Centers for Medicare Medicaid Services (CMS) to do this. I measure surgeons skills using the three years before the introduction of robots, when all operations were performed manually by surgeons.

My research indicates that robotic surgery improves over existing surgical techniques. On average, the robot reduces length of stay and the probability of an adverse event from surgery. I estimate the ATE to be close to -0.5 for an average length of stay of 3 days. For post-operative morbidity, this is instead approximately -0.05, relatively to an incidence of 14 percent in the individuals I observe.

However, my results show that these effects are highly heterogeneous. Technological gains strongly depend on the skills of the surgeon using the robot. Higher surgical skills in the period pre-robot translate into a reduction in the benefits from using this technology. Highly skilled surgeons have lower returns from using the robot, while low skilled surgeons appear to gain the most from it. The robot exhibits, therefore, decreasing returns in skills; it complements more low skilled surgeons. This suggests that the robot may have the potential to reduce unwarranted variation in surgical performance. In particular, it appears that the heterogeneity in treatment effects is the result of low-skilled surgeons performing significantly more poorly in the untreated state, and the technology equalizing them to high skilled surgeons in the treated state. For example, high skilled surgeons patients are 8 percentage points less likely to experience and adverse event from surgery when operated on with the traditional method. But, are only 1 percentage point less likely to experience these events, relatively to the patients of low skilled surgeons, with the robotic approach.

The analysis also shows a strong pattern of negative selection on both observed and unobserved characteristics. In terms of observables, higher quality surgeons are those using the technology more intensively, while lower skilled surgeons use it less in spite of their higher returns. This pattern hints at the fact that the adoption process may have been suboptimal and inefficient. Ideally, the technology should be used more intensively by low skilled surgeons. Hence, as a conclusive exercise, I simulate policies that may have created incentives to adopt in line with an efficient allocation of the technology.

The estimated MTE curve is also downward sloping, which further corroborates my findings. Patients at the lowest resistance to treatment have the least gains from being operated on with the technology. I provide a possible explanation that leverages differences in self-efficacy between high and low skilled surgeons. This explanation is in line with the idea that high (low) skilled surgeons are over (under) confident, and misguided in their belief that the technology can improve their performance.

Related literature [TO BE INCLUDED]

My paper proceeds as follows. In Section II, I describe surgical robots and their use for prostate cancer surgery. I use Section III and Section IV to describe the data, and discuss how I measure technological gains and surgeons' skills. In Section IV, 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 V. I provide a discussion of the definition and validity of the instrumental variables in Section VI. Lastly, in Section VII I show the results of the analysis, and I finally conclude in Section VIII.

II 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 men every year die from it.* In the United States and Europe, the diffusion of robots for prostate cancer surgery has been incredibly rapid. In 2003, less than 1 % 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

*https://prostatecanceruk.org

operation.[†] Eventually, by 2014, robotic surgery accounted for up to 90 % of radical prostatectomies across the US.[‡] This trend has been similar in England whereby 2014, the majority of cases (62.7%) 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 surgical robot's manipulator arm unit that includes three or four arms, depending on the model.

The instruments, including a video camera, are attached to the robotic arms and controlled by the surgeon. The robotic arms not only allow to work through 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.

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.

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

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 £1350000, with a median yearly maintenance cost of £492000. 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.

III 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 62258 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 6 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 3 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 5 displays the number of prostate cancer diagnoses and the share of patients opting for RP over time.

To identify technological gains, I investigate two margins of surgical performance; 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 patient 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.

IV 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. (2013b) 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. I group outcomes into two categories; general and surgery specific. General outcomes include deaths and readmissions within 30 days of hospital discharge. These outcomes are commonly used to rank surgeons across a variety of procedures. Surgery-specific outcomes include, instead, complications from surgery that are observed to affect prostate cancer patients (e.g. urinary complications). I develop a measure of skills that encompasses both categories and is therefore informative on both general and surgical specific competences. This is a risk adjusted measure of post-operative morbidity for prostate cancer patients.

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. 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. For my general measure of skills, y_{ij} takes value one if patient i ,operated by the surgeon j, has died or has been readmitted to hospital within 30 days of discharge, or if the patient has experienced post-operative morbidity within two years from the operation. 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 performance). The predicted outcome is the equivalent number for a *specific* surgeon. I compute these terms as follows:

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

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

I then compute the SRR as:

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

For example, a SRR of 1 indicates that the number of deaths and readmissions for surgeon j is as expected given the national average of surgeons treating similar patients. 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 descriptive facts on the relationship between robotic surgery, skills, and performance. I group surgeons into two categories; high and low-skilled surgeons. High-skilled surgeons are identified as those with a SRR below the 10th percentile, low skilled surgeons are identified as those with a SRR above the 90th percentile.

The first fact is that surgeons with higher skills appear to use the technology more intensively. These surgeons start using the robot before lower skilled surgeons, and by their second year of use, they operate on more than 20 percent of patients using the technology. It takes five more years for lower-skilled surgeons to use the technology at a similar rate. By the end of the sample period, both groups use the robot at a similar rate.

The second fact is that there has been a convergence in surgical performance between high and low-skilled surgeons over this period of time. I inspect two margins of surgical performance, length of stay and the probability of an adverse event from surgery. The average length of stay of prostate cancer patients has decreased substantially over the sample period. 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, both high and low skilled surgeons appear to have had an increase in the number of patients under their care. This is consistent with the increase in the number of prostate cancer diagnosis we observe in this period.

V Identification and estimation of MTE

My empirical strategy is tailored to the presence of heterogeneous treatment effects and the possibility of selection into treatment. Heterogeneity in treatment effects is likely to arise because individuals differ in their underlying health and thus in their outcomes, and treatment may be endogenous to such differences, with individuals selecting - or being selected - on the basis of their anticipated effects (Zhou and Xie, 2019). The most commonly used approach to deal with endogenous selection is the instrumental variable (IV) method where an external variable (i.e. the instrument) is used to distil out an exogenous variation in treatment then associated to the outcomes to estimate a causal effect (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 developed in Heckman and Vytlacil (2005) that builds on marginal treatment effects (MTE). MTEs are the average treatment effects for people with either a particular resistance to treatment or at a particular margin of indifference. With this method, I can estimate a continuum of treatment effects identified along the distribution of the individual unobserved characteristics that drive treatment decisions (Cornelissen et al., 2016).

The MTE has several advantages over the IV method. First, by relating individual effects to choices, MTE allows inferring the pattern of selection into treatment. Estimation of the MTE is therefore more informative than conventional IV estimators when treatment effects are heterogeneous (Cornelissen et al., 2016). Second, by identifying the whole range of individual treatment effects, this method allows to recover economically meaningful parameters which the IV method fails in some cases to identify (Heckman and Vytlacil, 2005). When treatment effects are heterogeneous, the IV estimator is only consistent for the Local Average Treatment Effect (LATE). This is an average effect for a subpopulation of "compliers" compelled to select into treatment by changes in the instrument (Imbens and Angrist, 1994). Compliers may be substantially different from the general population, and can change with the instrument used. MTE instead can be used to generate all conventional treatment parameters, such as the average treatment effect (ATE), the average treatment effect on the treated (ATT), or the LATE but also to calculate policy relevant treatment effect (PRTE), the likely effect of a policy that affects selection but not outcomes. Third, estimation of causal effects with IV is challenging when the instrument is more likely to be valid after conditioning on covariates. In fact, Imbens and Angrist (1994)' s interpretation applies to fully saturated models which may not be desirable in empirical applications, and whose complexity may hinder a straightforward interpretation of the estimates (Słoczyński, 2020). Lastly, when both the outcome and the endogenous variable are binary two-step estimation procedures do not produce consistent estimates, and identification is not guaranteed (Wooldridge, 2010).

Identification of MTE does not require any stronger assumption than the IV approach, but poses however a stronger burden on the instrument (Vytlacil, 2002). The instrumental variable of choice must be continuous and able to generate enough variation in the probability of treatment to identify for each individual their margin of indifference.

At the core of the MTE framework is a generalized Roy model of binary choice (Roy, 1951). There are two potential outcomes, Y_1 and Y_0, a treatment indicator D, and a vector of covariates X. The observed outcome is denoted by Y, and can be represented as;

$$Y = DY_1 + (1 - D)Y_0 \tag{7}$$

where

$$Y_0 = \mu_0(X) + U_0 \tag{8}$$

$$Y_1 = \mu_1(X) + U_1 \tag{9}$$

and $\mu_j(X) \equiv E[Y_j|X]$ $\forall$ treatment option $j \in [0, 1]$. U_0 and U_1 are error terms of mean zero conditional on X representing unobserved factors affecting Y_0 and, Y_1 respectively.

The individual's return from choosing $D = 1$ is given by:

$$Y_1 - Y_0 = \mu_1(X) - \mu_0(X) + U_1 - U_0$$

Heterogeneity in treatment effects in this model can arise therefore from both observed (X) and unobserved ($U_1 - U_0$) characteristics.

Treatment choice is represented by a latent variable model, where the decision rule is:

$$D = 1[D^* \geq 0] \tag{10}$$

A person chooses $D = 1$ whenever the latent propensity to take treatment $D^* \geq 0$; and chooses $D = 0$ otherwise. Treatment decisions, like treatment effects, depend on both observed and unobserved characteristics. Assuming that the selection equation is additively separable in its components it can be represented as:

$$D^* = \mu_D(Z) - V \tag{11}$$

The Z vector may include some or all of the components of X, and crucially it must also include one or more variables excluded from the outcome equation. V is a continuously distributed random variable representing unobserved, individual-specific resistance to treatment. Two patients with identical observed characteristics are allowed to differ in terms of V. A higher value of V reduces the probability that the patient is selected into treatment. Unobserved resistance to treatment is a key concept in the MTE framework; it ensures that all unobserved determinants of treatment are summarized by a single latent variable. Importantly, no restriction is imposed on the relationship between $(U_1 - U_0)$ and V, so that the unobserved gain is allowed to be correlated with the unobserved component that affects selection. This means that we explicitly allow for individuals to select into treatment on the basis of some anticipated treatment effect in a way that is not captured by the observed covariates included in the model. Variation of treatment effect by the latent variable captures all the unobserved treatment effect heterogeneity that may cause selection bias (Zhou and Xie, 2019).

In the MTE framework, the decision rule is conventionally expressed in terms of the propensity score $P(Z)$:

$$D = 1[P(Z) - U_d > 0] \tag{12}$$

where the latent variable $U_d \equiv F_V(V)$ is uniformly distributed on the interval $[0, 1]$, and it represents the quantiles of the distribution of the unobserved resistance to treatment V.

As individuals are only observed in the treated or the untreated state, it is impossible to estimate individual causal effect of the treatment. The treatment effects literature has focused therefore on different conditional expectations of the causal effects. Conditional on $X = x$ and the normalized latent variable $U_d = u_d$, the MTE is defined as the gain from treatment for an individual with characteristics $X = x$ indifferent between traditional and robotic surgery at the propensity score

$P(Z) = u_d$;

$$MTE(x, u) = \mu_1(X) - \mu_0(X) + E[U_1 - U_0|X = x, U_d = u_d] \tag{13}$$

The shape of the MTE function with respect to u_d illustrates how the average treatment effect changes as the unobserved resistance to treatment varies. For a positive outcome, a decreasing MTE function indicates positive selection into treatment, that is, individuals that are more likely to self-select into treatment have higher average treatment effects.

To identify and estimate MTEs, I impose the following assumptions:

(i.) $(U_0 - U_1, V)$ are statistically independent of Z conditional on X;

(ii.) $h(\cdot)$ is non-trivial function of Z conditional on X;

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

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

Assumption (i.) implies that the instrument Z is as good as random conditional on X. Assumption (ii.) implies that there must be a first stage in which the instrument Z induces variation in the probability of treatment after controlling for X. Vytlacil (2002) shows that (i.) and (ii.), together with the selection model, imply the independence and monotonicity conditions assumed in the LATE framework by Imbens and Angrist (1994), and that the independence and monotonicity conditions imply the selection model under no additional restrictions. Assumption (iii.) allows the MTE to be identified over the unconditional support of the propensity score (Carneiro et al., 2011). Assumption (iv.) 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:

$$E[Y|X = x, U_d = u_d] = \beta_0 x + (\beta_1 - \beta_0)xp + K(p) \tag{14}$$

where $K(p)$ is a non-linear function of the propensity score, and captures the 'essential heterogeneity' in the outcome that is correlated with the potential utility of each alternative.

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 $U_d = p$:

$$\frac{\partial E[Y|X = x, P(Z) = p]}{\partial p} = (\beta_1 - \beta_0)\, x + \frac{\partial K(p)}{\partial p} = MTE[X = x, U_d = p] \tag{15}$$

The intuition is simple. Increasing the propensity score by a small amount shifts previously indifferent individuals into treatment and increases the observed outcome Y by the share of individuals shifted times their treatment effect. 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. This clarifies the requirement of continuity for point identification. The continuous nature of the instrument allow us to vary the propensity score so that to hopefully identify the value for which all individuals are compliers, and thus their treatment effect.

Estimation of this derivative is achieved using the local instrumental variable (LIV) introduced by Heckman and Vytlacil (1999) or the separate approach. Both methods involve estimating an outcome model that includes the additively separable component $K(p)$. The true distribution of $K(p)$ is unknown, and the function could be non-linear. Thus, the outcome is alternatively modelled parametrically or non parametrically (partially-linear) in terms of the unobserved term.

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.

VI 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:

$$Z_{days} = t - T_R, \tag{16}$$

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;

$$Z_{dist} = D_R - D_T, \tag{17}$$

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 average effects are presented in Table 4 for increasingly richer specifications (multiplied by 100 to ease the presentation). Actually, Column (7) represents the estimated selection equation, which I will discuss in more details in Section VII.

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 and statistically significant coefficient in all specifications. This indicates that the more days passed after the closest hospital has adopted the robot, the more likely the patient is of getting the robotic surgery. In Figure 12, I show the average predicted probability evaluated as different values of this instrument. Again, the figure shows that the higher the value of Z_{days}, the higher is the predicted probability of getting the robotic approach. For example, an individual diagnosed two year 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 statistically significant negative coefficient, which indicates that the higher the 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 per cent.

Finally, the instruments should affect the probability of treatment in a monotone way. This means that no individual should switch out of treatment for a decrease in relative distance, or for an increase in diagnosis timing. 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 14. 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.

VII Results

There are three blocks of evidence that come from the estimation of MTEs. The first is the selection equation that includes the instruments and all variables believed to affect the choice of treatment. The second block includes the estimated outcome equation coefficients (i.e. Equation 14), which speak to differences in returns, but also in the untreated state, for the observed variables in the model. The last block is the estimation of the relationship between the unobserved resistance and the returns from treatment (i.e. the MTE curve). All these pieces must be discussed in conjunction to understand whether there is selection on gains, and to interpret - if any - selection on unobservable characteristics. I start discussing selection into treatment, and focus on the role of skills in determining whether the patient gets the robot. I move, then, to discussing heterogeneity in returns from surgeons' skills, and the type of technology-skill complementarity this suggests. Lastly, I discuss the pattern of selection on gains that is due to unobservable characteristics not captured in the model. For each of these steps, I present the assumptions I use for estimation and the alternative model specifications I test.

VII.A Selection into treatment

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. Column (7) from Table 4, shows the estimates from which I derive the propensity score. This specification generates a large common support for the propensity score, which ranges from 0 to 1. the unconditional support jointly generated by variation of both the instruments and the covariates, including my measure of surgical skills, is shown in Figure 15.

The partial derivative, or marginal effect, with respect to the SRR - the inverse of skills - tells us how a change in skills affects the expected probability of operating a patient with the robot. Surgical skills appear to be an important determinant of whether the patient will get robotic surgery. From Table 4, it is possible to observe that the SRR has a negative and statistically significant coefficient. The marginal effect is -0.3 indicating that a marginal increase in the SRR reduces the probability of using the robot by 0.3 percentage points. To ease the interpretation of this effect, I show in Table 5 the same probit regression for two transformations of my measure of skills. Namely, an indicator of high-skills and a standardized version of the SRR variable. The indicator of high-skills takes value one if the surgeon has a SRR below the median of the distribution (i.e. 0.8). In this case, the value of the margin is the difference in the probability of using the robot between high and low skilled surgeons. High-skills surgeons' average predicted probability of using the robot is 0.58 while for the rest is 0.38.

Higher skilled surgeons use the robot more intensively, regardless of the characteristics of the patients. In Figure 13, I show how the predicted probability varies by age for high and low skilled surgeons. For both surgeons groups, the likelihood of using the robot increased with the age of the patient, but at all age levels high skilled surgeons are more likely to operate with the robot.

Skills and technological gains

The MTE framework allows me to estimate two sets of coefficients from the outcome (i.e. Equation 14). First, it allows me to estimate the relationship between the variables included in the model and the outcomes in the untreated state. This is the β_0 vector from Equation 14. 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 $\beta_1 - \beta_0$ in the Equation 14.

To fix ideas, I display below the two outcome equations, whose parameters I estimate, in terms of surgical skills. For clarity, I do not show the control variables in this representation. p represent the propensity score estimated using a logistic regression.

$$E[log(LOS)_{ij}|X = x, U_d = u_d] = \beta_0^L Skills_j + (\beta_1^L - \beta_0^L) Skills_j p + K^L(p) \tag{18}$$

$$E[\text{Adverse event}_{ij} = 1|X = x, U_d = u_d] = \beta_0^B Skills_j + (\beta_1^B - \beta_0^B) Skills_j p + K^B(p) \tag{19}$$

I estimate the coefficients under three specifications of the skills variable. In the first, Table 6, I use the actual value of the SRR as my regressor of interest. In the second, Table 7, I use an indicator of high-skills (i.e. SRR below median). Third, in Table 8, I use a standardized version of the SRR. I do not expect the results to be substantially different, but I do this to ease the interpretation of the coefficients $\beta_0^L, \beta_0^B, \beta_1^L$ and β_1^B.

In all tables, I display the coefficients β_0^L and β_0^B in Column (1) and Column (2), and $\beta_1^L - \beta_0^L$ and $\beta_1^B - \beta_0^B$ in Columns (3) and (4).

From Table 6 Columns (1) and (2), it is possible to observe that surgeons'skills are positively related to log length of stay and the occurrence of adverse events from surgery. This is reassuring; my measure of surgical skills is indeed negatively related to the outcomes of the patients. Outcomes in the untreated state are also related to the patient clinical characteristics, such as the presence of renal disease or diabetes. More severe patients have longer length of stay and a higher probability of adverse events in the untreated state.

Columns (3) and (4) of Table 6 point to an equalizing effect of the robot. In the untreated state, we have observed that patients operated by low skilled surgeons have both higher length of stay and probability of experiencing an adverse event from surgery. However, their treatment effect from using the robot is more negative, indicating that these are the surgeons for which the robot has a stronger positive impact on performance. In other words, the robot reduces length of stay and the probability of an adverse event from surgery for strongly for patients of low skilled surgeons. This suggests that the returns from using the robot are decreasing in users skills. The robot appears to aid lower skilled surgeons, but does not improve as much the performance of high skilled ones. When I control in the model for area fixed effects as in Cornelissen et al. (2018) (i.e. Table 9), the difference between high and low skilled surgeons is sharper.

Table 7 helps to interpret the magnitude of these effects. As earlier, I define surgeons as high skills if they have an SRR below the median of the distribution (i.e. $0,8$). The results suggest that, in the untreated state, patients of a high-skilled surgeons are 8 percentage points less likely to experience an adverse event from surgery relatively to low skilled surgeons. At the same time, their treatment effect is about 7 percentage points lower. Using the robot reduces substantially the difference in performance between high and low skilled surgeons.

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; the surgeons that have the highest to gain are less likely to use the robot.

Returns to treatment based on unobserved characteristics

With the parameters from the selection equation and the estimates of $\beta_1 - \beta_0$, I can find the MTE curve for the two outcomes of interest. 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 16.

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 in effects based on unobservables.

In Figure 17, 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 much more precise estimates for 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 Table 10 and Table 11, 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. The robot reduces length of stay and the probability that the patients experience 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).

VIII Conclusion

[TO BE INCLUDED]

References

Acemoglu, Daron and Pascual Restrepo, “Robots and jobs: Evidence from US labor markets,” *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

Banerjee, Souvik and Anirban Basu, “Estimating Endogenous Treatment Effects Using Latent Factor Models with and without Instrumental Variables,” *Econometrics*, 2021, *9* (1).

Birkmeyer, John D, Bradley N Reames, Peter McCulloch, Andrew J Carr, W Bruce Campbell, and John E Wennberg, “Understanding of regional variation in the use of surgery,” *The Lancet*, 2013, *382* (9898), 1121–1129.

_ , Jonathan F Finks, Amanda O’reilly, Mary Oerline, Arthur M Carlin, Andre R Nunn, Justin Dimick, Mousumi Banerjee, and Nancy JO Birkmeyer, “Surgical skill and complication rates after bariatric surgery,” *New England Journal of Medicine*, 2013, *369* (15), 1434–1442.

Björklund, Anders and Robert Moffitt, “The Estimation of Wage Gains and Welfare Gains in Self-Selection Models,” *The Review of Economics and Statistics*, 1987, *69* (1).

Borghans, Lex, Francis Green, and Ken Mayhew, “Skills measurement and economic analysis: an introduction,” *Oxford Economic Papers*, 2001, pp. 375–384.

Card, David, Alessandra Fenizia, and David Silver, “The health impacts of hospital delivery practices,” Technical Report, National Bureau of Economic Research 2019.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil, “Estimating marginal returns to education,” *American Economic Review*, 2011, *101* (6).

Chandra, Amitabh and Douglas O Staiger, “Productivity spillovers in health care: evidence from the treatment of heart attacks,” *Journal of political Economy*, 2007, *115* (1), 103–140.

Coelho, Rafael F, Bernardo Rocco, Manoj B Patel, Marcelo A Orvieto, Sanket Chauhan, Vincenzo Ficarra, Sara Melegari, Kenneth J Palmer, and Vipul R Patel, “Retropubic, laparoscopic, and robot-assisted radical prostatectomy: a critical review of outcomes reported by high-volume centers,” *Journal of endourology*, 2010, *24* (12), 2003–2015.

Compagni, Amelia, Valentina Mele, and Davide Ravasi, "How early implementations influence later adoptions of innovation: Social positioning and skill reproduction in the diffusion of robotic surgery," *Academy of Management Journal*, 2015, *58* (1), 242–278.

Cooper, Zack, Stephen Gibbons, Simon Jones, and Alistair McGuire, "Does hospital competition improve efficiency? An analysis of the recent market-based reforms to the English NHS," 2010.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg, "From LATE to MTE: Alternative methods for the evaluation of policy interventions," *Labour Economics*, 2016, *41*, 47–60.

__ , __ , __ , and __ , "Who benefits from universal child care? Estimating marginal returns to early child care attendance," *Journal of Political Economy*, 2018, *126* (6), 2356–2409.

Davies, Benjamin, "Waiting for Godot Robot," 2022.

George, Evalyn I, COL Timothy C Brand, Jacques Marescaux et al., "Origins of robotic surgery: from skepticism to standard of care," *JSLS: Journal of the Society of Laparoendoscopic Surgeons*, 2018, *22* (4).

Hadley, Jack and Peter Cunningham, "Availability of safety net providers and access to care of uninsured persons," *Health services research*, 2004, *39* (5), 1527–1546.

Hall, Bronwyn H, "Innovation and diffusion," 2004.

Heckman, James J and Edward J Vytlacil, "Local instrumental variables and latent variable models for identifying and bounding treatment effects," *Proceedings of the national Academy of Sciences*, 1999, *96* (8).

__ and Edward Vytlacil, "Structural equations, treatment effects, and econometric policy evaluation," *Econometrica*, 2005, *73* (3).

__ , Sergio Urzua, and Edward Vytlacil, "Understanding instrumental variables in models with essential heterogeneity," *The review of economics and statistics*, 2006, *88* (3), 389–432.

Higgins, Rana M, Matthew J Frelich, Matthew E Bosler, and Jon C Gould, "Cost analysis of robotic versus laparoscopic general surgery procedures," *Surgical endoscopy*, 2017, *31* (1), 185–192.

Hull, Peter, "Estimating hospital quality with quasi-experimental data," *Available at SSRN 3118358*, 2018.

Hussain, Azhar, Amina Malik, Mohammed U Halim, and Aslankhani Mohammad Ali, "The use of robotics in surgery: a review," *International journal of clinical practice*, 2014, *68* (11), 1376–1382.

Imbens, Guido W and Donald B Rubin, "Estimating outcome distributions for compliers in instrumental variables models," *The Review of Economic Studies*, 1997, *64* (4).

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2).

Ketkar, Harsh Jayant, Seth Carnahan, and Brad N Greenwood, "Do New Job Tools Improve Women's Performance in Male-Dominated Fields? Evidence from Robotic Surgery," *Evidence from Robotic Surgery (March 28, 2022)*, 2022.

Lam, Kyle, Jonathan Clarke, Sanjay Purkayastha, and James M Kinross, "Uptake and accessibility of surgical robotics in England," *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2021, *17* (1), 1–7.

Lotan, Yair, "Is robotic surgery cost-effective: no," *Current opinion in urology*, 2012, *22* (1), 66–69.

Lowrance, William T, Elena B Elkin, Lindsay M Jacks, David S Yee, Thomas L Jang, Vincent P Laudone, Bertrand D Guillonneau, Peter T Scardino, and James A Eastham, "Comparative effectiveness of prostate cancer surgical treatments: a population based analysis of postoperative outcomes," *The Journal of urology*, 2010, *183* (4), 1366–1372.

Marcus, Hani J, Archie Hughes-Hallett, Christopher J Payne, Thomas P Cundy, Dipankar Nandi, Guang-Zhong Yang, and Ara Darzi, "Trends in the diffusion of robotic surgery: A retrospective observational study," *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2017, *13* (4), e1870.

McClellan, Mark, Barbara J McNeil, and Joseph P Newhouse, "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: analysis using instrumental variables," *Jama*, 1994, *272* (11).

Mookadam, Farouk and Heather M Arthur, "Social support and its relationship to morbidity and mortality after acute myocardial infarction: systematic overview," *Archives of internal medicine*, 2004, *164* (14), 1514–1518.

Nelson, Bradford, Melissa Kaufman, Gregory Broughton, Michael S Cookson, Sam S Chang, S Duke Herrell, Roxelyn G Baumgartner, and Joseph A Smith, "Comparison of length of hospital stay between radical retropubic prostatectomy and robotic assisted laparoscopic prostatectomy," *The Journal of urology*, 2007, *177* (3), 929–931.

Peters, Brian S, Priscila R Armijo, Crystal Krause, Songita A Choudhury, and Dmitry Oleynikov, "Review of emerging surgical robotic technology," *Surgical endoscopy*, 2018, *32* (4), 1636–1655.

Roy, Andrew Donald, "Some thoughts on the distribution of earnings," *Oxford economic papers*, 1951, *3* (2).

Skinner, Jonathan, "Causes and consequences of regional variations in health care," in "Handbook of health economics," Vol. 2, Elsevier, 2011, pp. 45–93.

Słoczyński, Tymon, "When Should We (Not) Interpret Linear IV Estimands as LATE?," *arXiv preprint arXiv:2011.06695*, 2020.

Stoye, George, "The distribution of doctor quality: evidence from cardiologists in England," *IFS Working Paper*, 2022.

Strother, Marshall C, Katharine F Michel, Leilei Xia, Kellie McWilliams, Thomas J Guzzo, Daniel J Lee, and David I Lee, "Prolonged length of stay after robotic prostatectomy: causes and risk factors," *Annals of Surgical Oncology*, 2020, *27* (5), 1560–1567.

Tonutti, Michele, Daniel S Elson, Guang-Zhong Yang, Ara W Darzi, and Mikael H Sodergren, "The role of technology in minimally invasive surgery: state of the art, recent developments and future directions," *Postgraduate Medical Journal*, 2017, *93* (1097).

Tsai, Thomas C, Karen E Joynt, E John Orav, Atul A Gawande, and Ashish K Jha, "Variation in surgical-readmission rates and quality of hospital care," *New England Journal of Medicine*, 2013, *369* (12), 1134–1142.

Tsugawa, Yusuke, Ashish K Jha, Joseph P Newhouse, Alan M Zaslavsky, and Anupam B Jena, "Variation in physician spending and association with patient outcomes," *Jama internal medicine*, 2017, *177* (5), 675–682.

Vytlacil, Edward, "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 2002, *70* (1).

Warren, Hannah and Prokar Dasgupta, "The future of robotics," *Investigative and clinical urology*, 2017, *58* (5), 297–298.

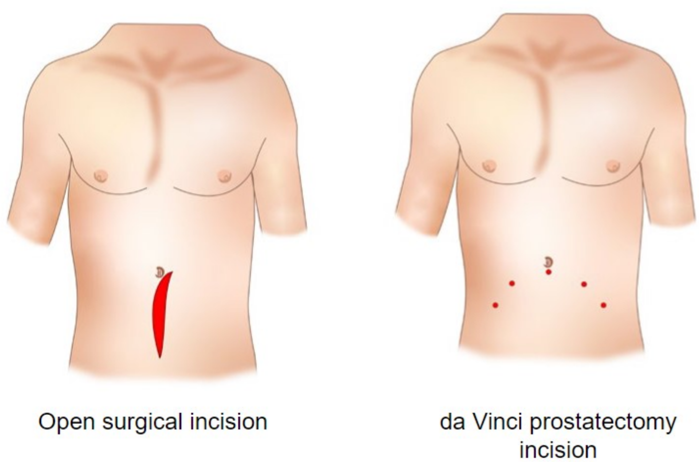
Wooldridge, Jeffrey M, *Econometric analysis of cross section and panel data*, MIT press, 2010.

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

Fig. 1 Picture of a Da Vinci surgical system

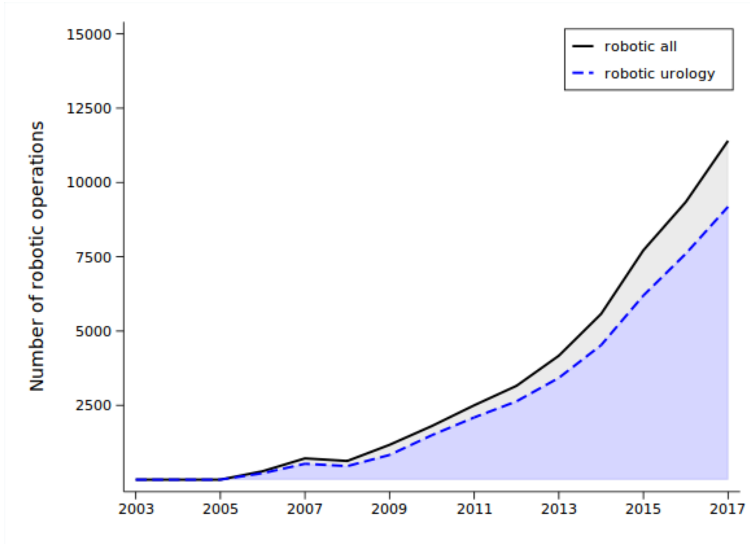
Picture shows the Da Vinci Robot surgical system from Intuitive Inc. On the left the surgeon sitting at the console. Above the operating bed the robotic arms. On the right the vision cart.

Fig. 2 Comparison of incisions



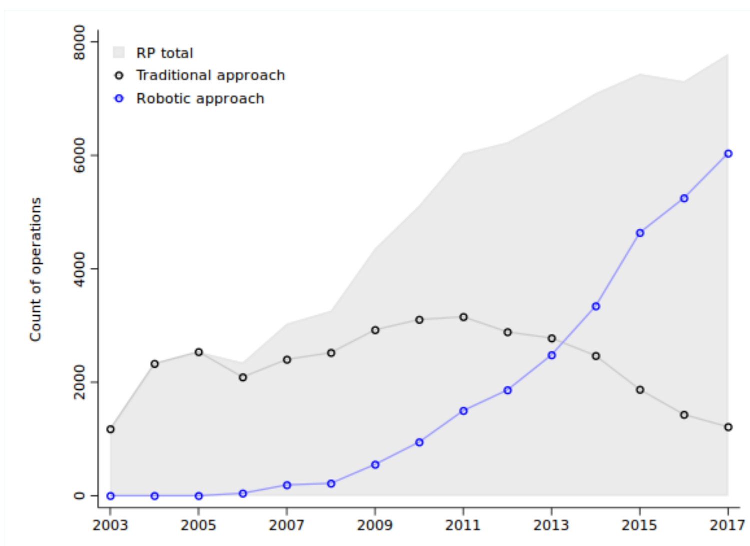
Comparison of incisions required for traditional and robotic radical prostatectomy.

Fig. 3 Diffusion of robotic surgery in the NHS



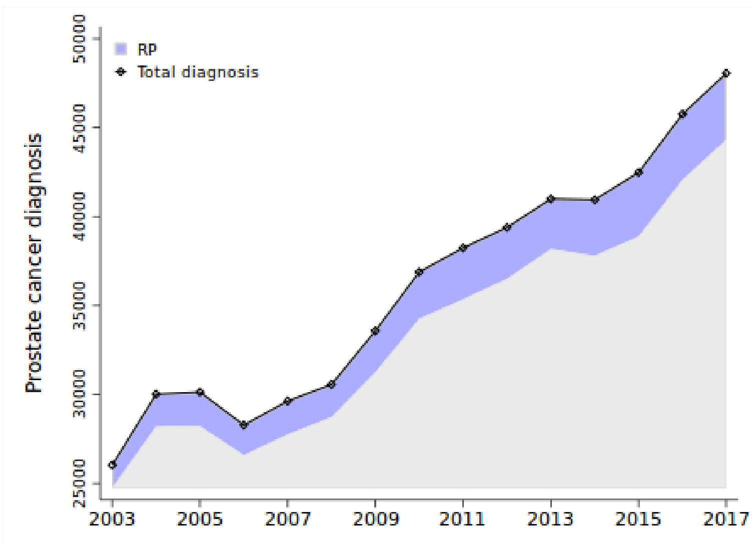
The picture shows the number of robotic operations by year for urology compared to all other specialties in which robots are used. The data is from the Hospital Episodes Statistics.

Fig. 4 Volume of robotic and traditional radical prostatectomies



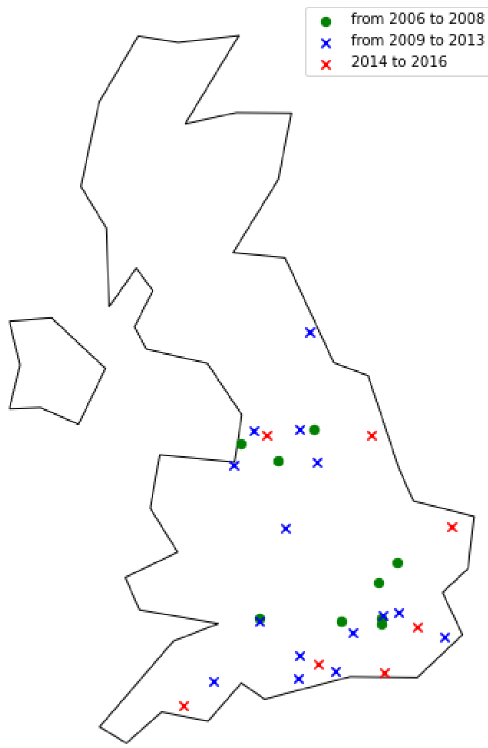
Note: Graph is produced using the HES data. The shaded gray area represents the total number of radical prostatectomies performed by NHS Hospitals in England. The black dots represent the number of radical prostatectomies performed using the open approach. The blue dots represent the number of radical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

Fig. 5 Surgical interventions as a share of prostate cancer diagnosis



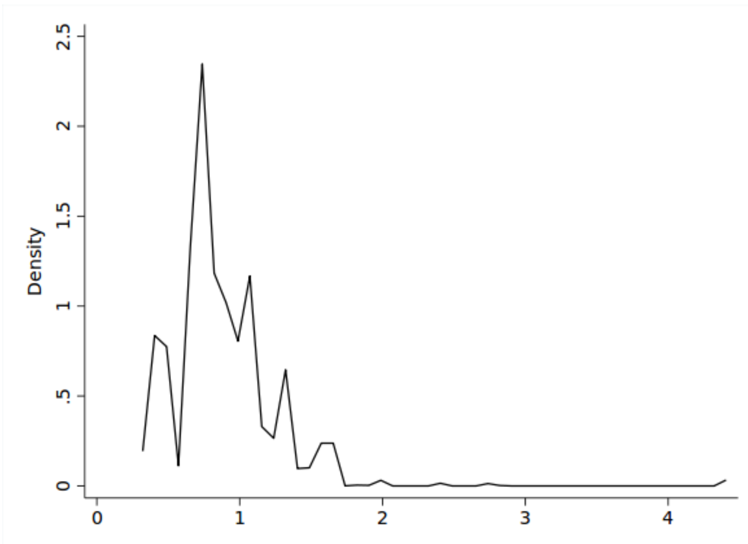
Note: Graph is produced using the HES data. The shaded blue area represents the number of radical prostatectomies performed by NHS hospitals in England. The shaded gray area represents the number of patients with prostate cancer that have undergone radio therapy treatment. The black line represents the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

Fig. 6 Hospital level diffusion of robotic surgery - Timing of adoption



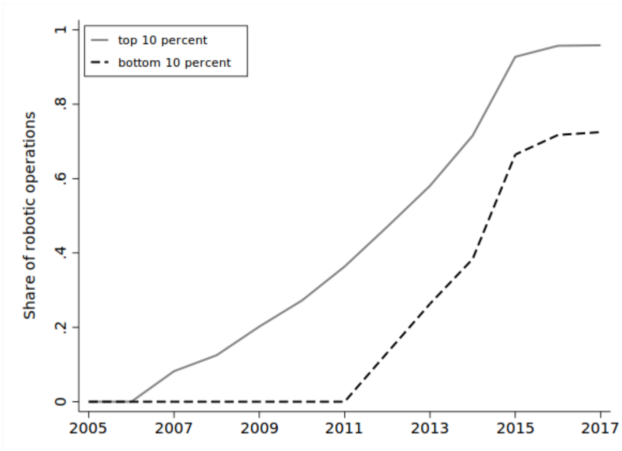
Note: Graph is produced using the HES data. The green crosses represent hospitals that are observed using for the first time a robotic operation code in 2006. The blue dots represent hospitals that are observed using for the first time a robotic operation code between 2007 and 2009. The red dots represent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

Fig. 7   Estimated standardized risk ratios - Measure of surgical skills

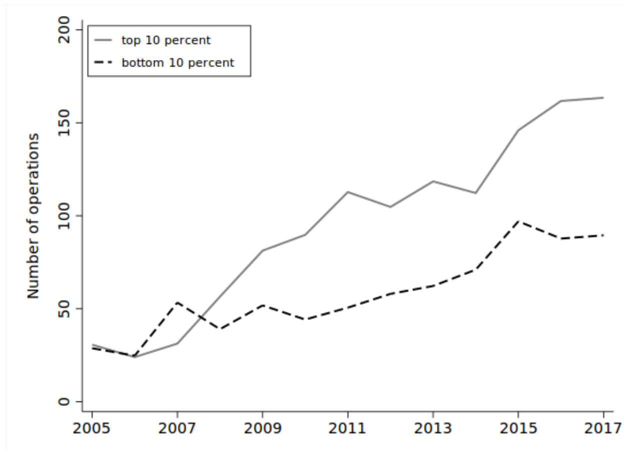


Note: Distribution of standardised risk ratios (SRR). Outcome is post-operative morbidity. SRR is computed as the ratio between predicted and expected morbidity. Predicted and expected post-operative morbidity are obtained by estimating a hierarchical logistic model accounting for patients' clinical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

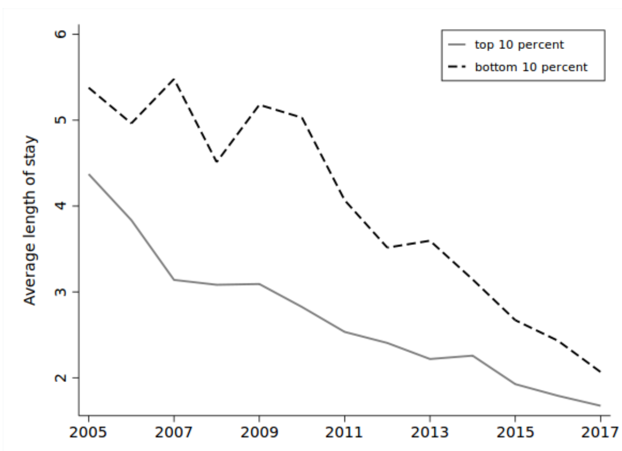
Fig. 8 Key empirical facts



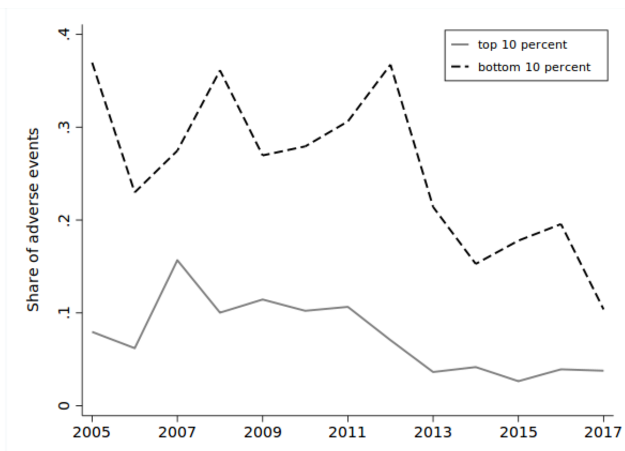
(a) Robotic rate



(b) Number of operations



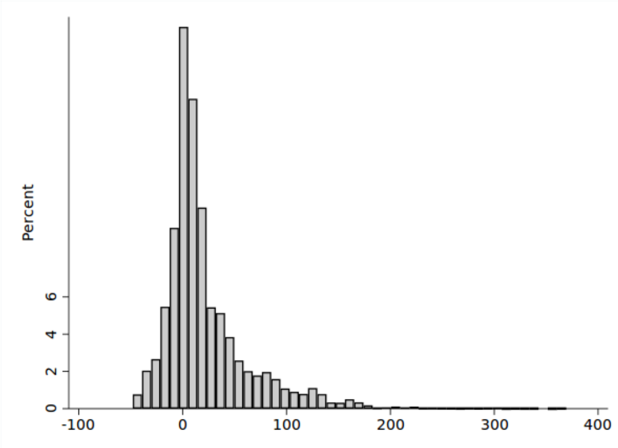
(c) Length of stay



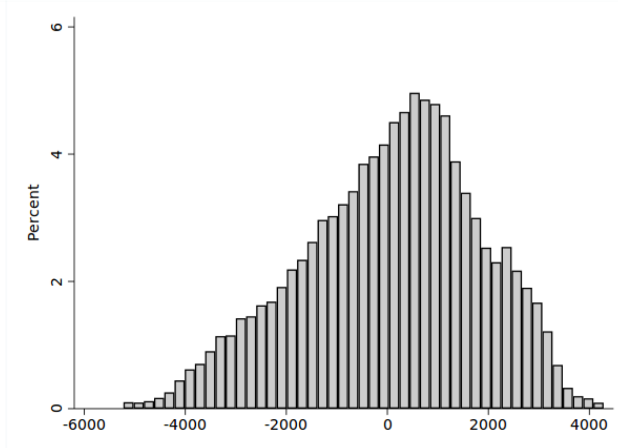
(d) Adverse events

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.

Fig. 9 Variation of instrumental variables in sample data



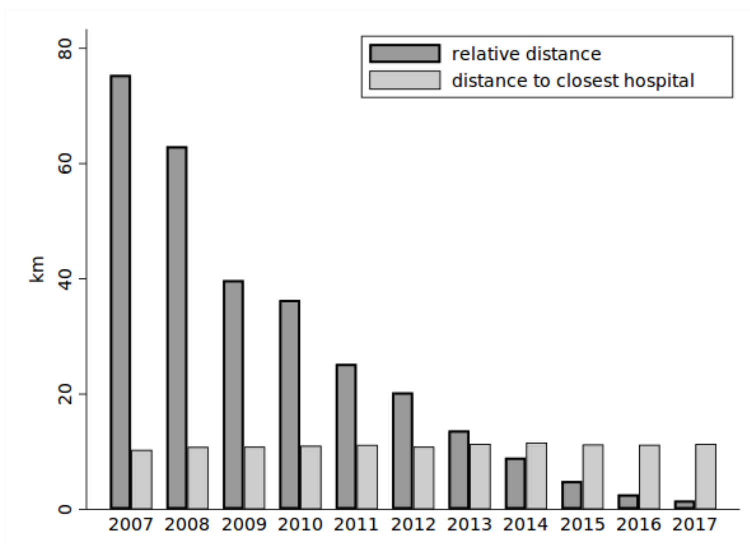
(a) Z_{dist} instrument



(b) Z_{days} instrument

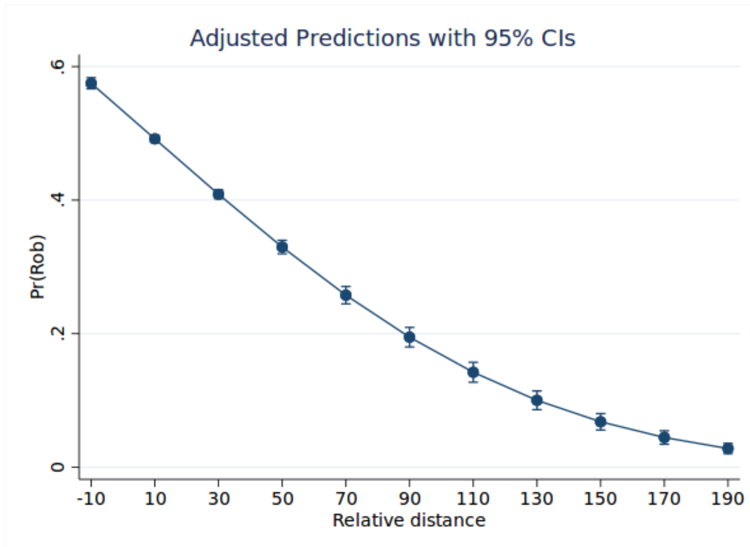
Note: Panel (a) plots the instrument Z_{dist} defined as the relative distance between the patients nearest hospital capable of offering robotic assisted radical prostatectomy and the closest hospital offering traditional radical prostatectomy. The distance is expressed in kilometers. Panel (b) plots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

Fig. 10 Average relative distance to robotic hospital and to closest hospital



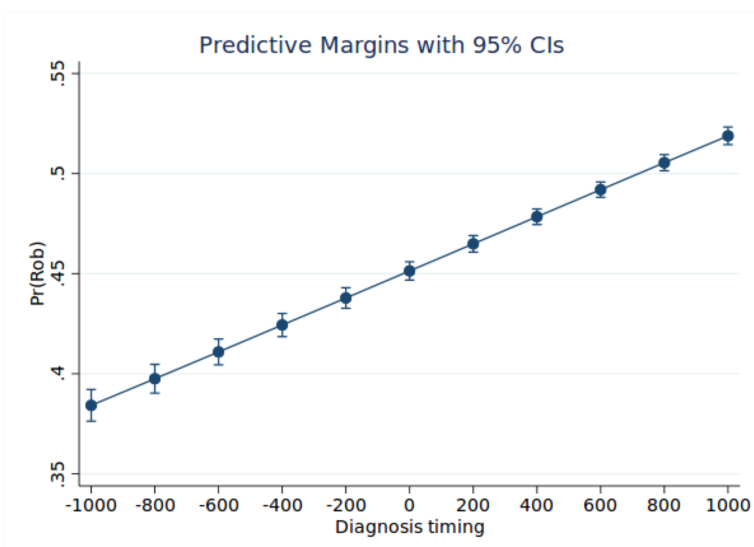
Note: Figure shows the average relative distance to hospital offering robotic surgery. Relative distance is computed as the difference between the individual distance to the closest hospital offering robotic technology and the distance to the closest hospital offering only traditional surgery. Distance 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.

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



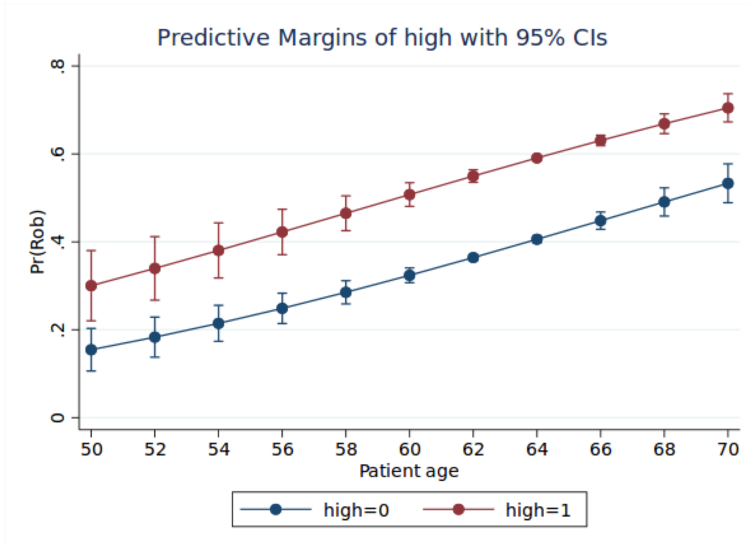
Note: Probit regression estimates, dependent variable robotic approach.

Fig. 12 Estimated probability of robotic approach from selection equation - at Z_{days} values



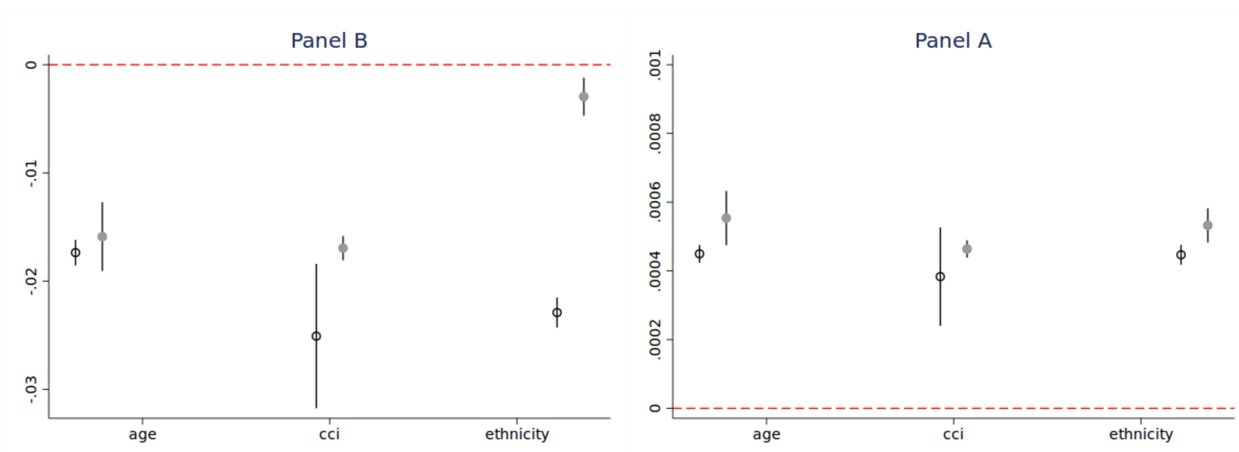
Note: Probit regression estimates, dependent variable robotic approach.

Fig. 13 Estimated probability of robotic approach from selection equation - at Age values for high and low skilled surgeons



Note: Probit regression estimates, dependent variable robotic approach.

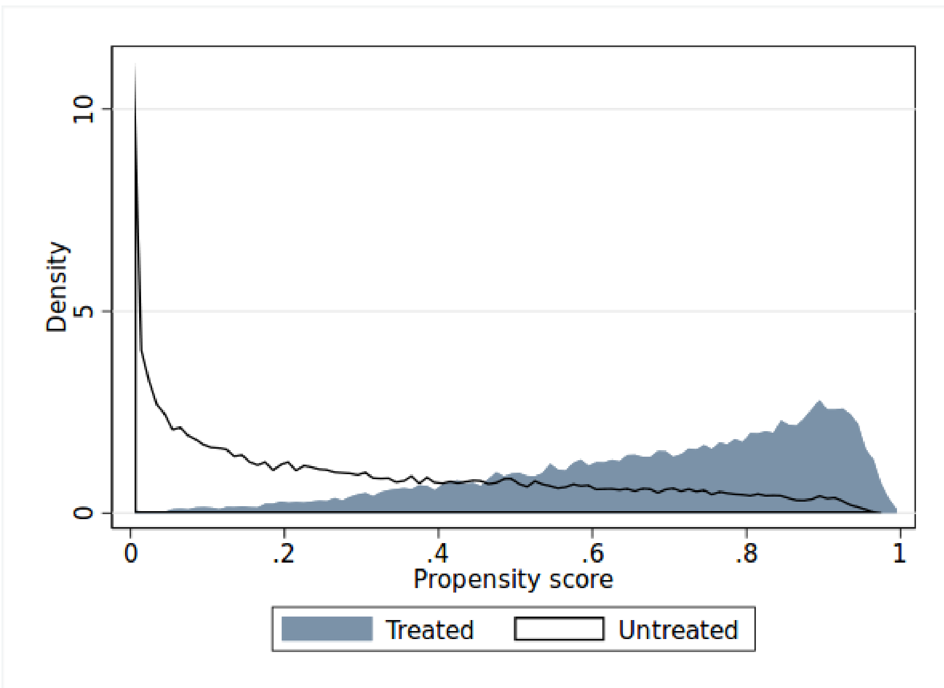
Fig. 14 Test for monotonicity of the instruments



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

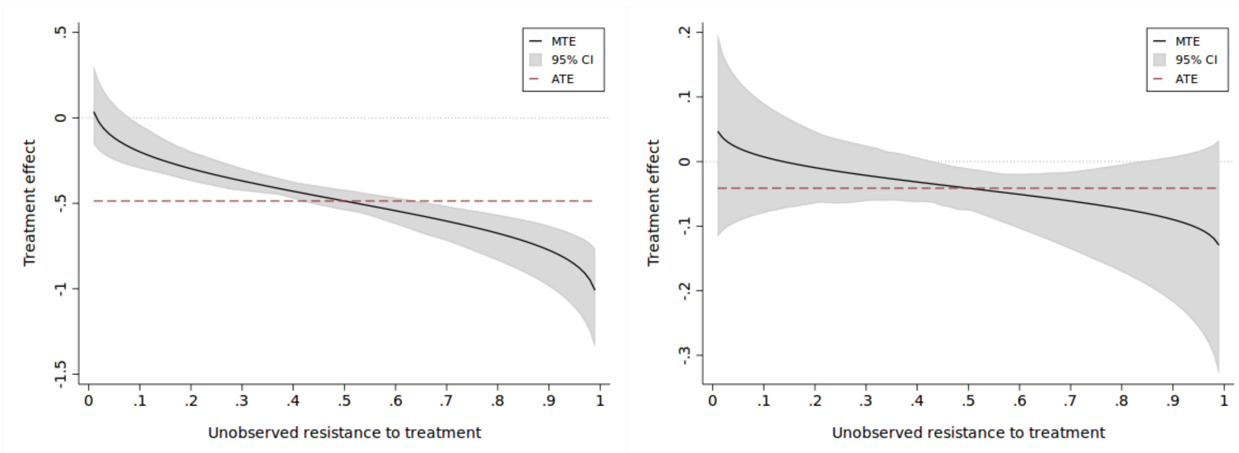
Note: OLS regression for subsets of the population. Age above and below 55. CCI above and below 2. Ethnicity white and all other ethnicity. Model controls are age, age squared, ethnicity, city indicator, ten comorbidity dummies (e.g. malignant neoplasm, diabetes), distance to closest hospital, indicator of whether the closest hospital is a teaching hospital, hospital quality per robots, and year fixed effects.

Fig. 15 Common support



Note: Model includes the measure of surgical skills pre-robot. Controls are 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, hospital quality pre-robots, and year, month and day of the week fixed effects.

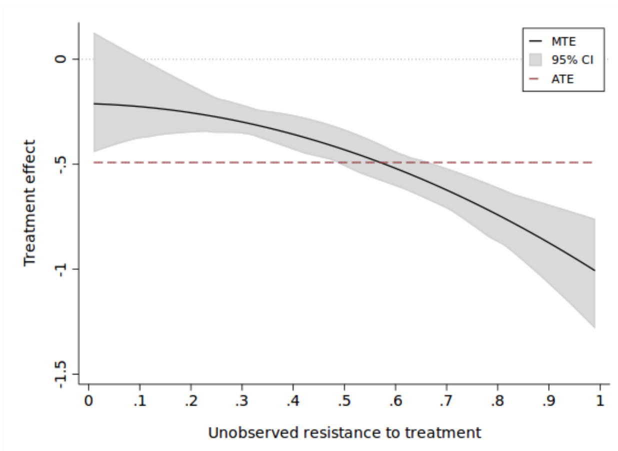
Fig. 16   MTE curve - Normal



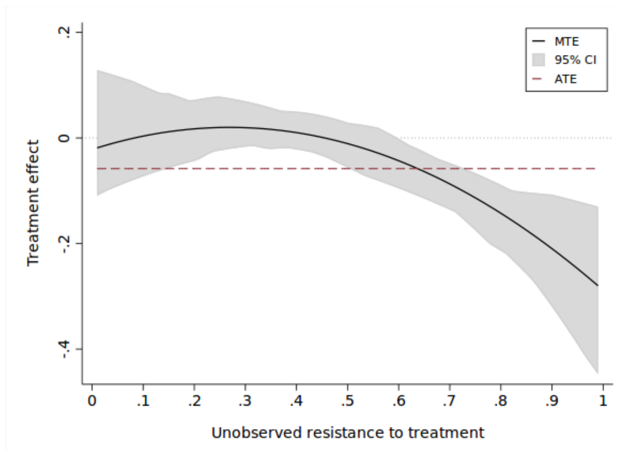
(a) LOS **(b) Adverse event**

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, parameterically 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, hospital quality per robots, and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions.

Fig. 17 MTE curve - Polynomial



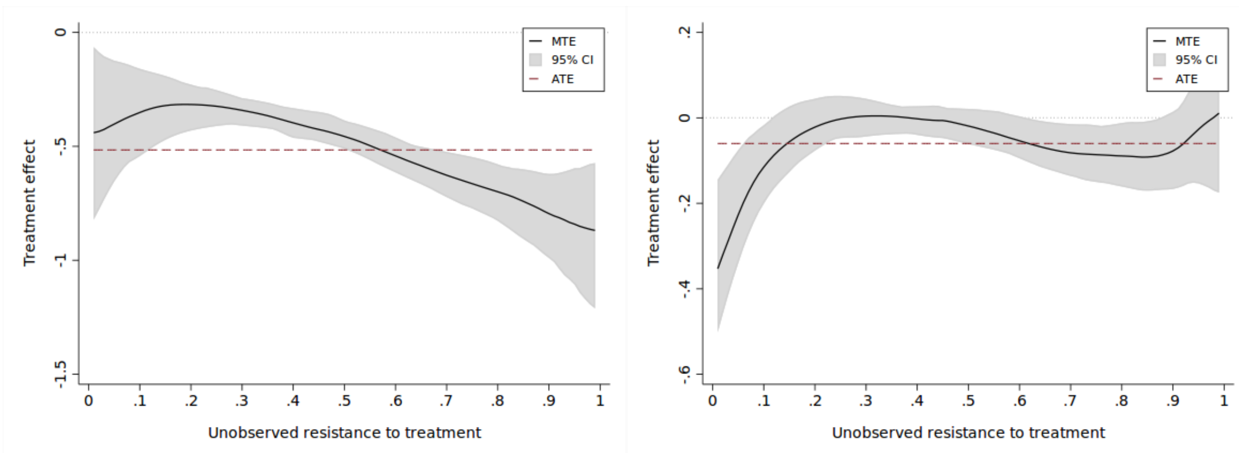
(a) LOS



(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, parameterically 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, hospital quality per robots, and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions.

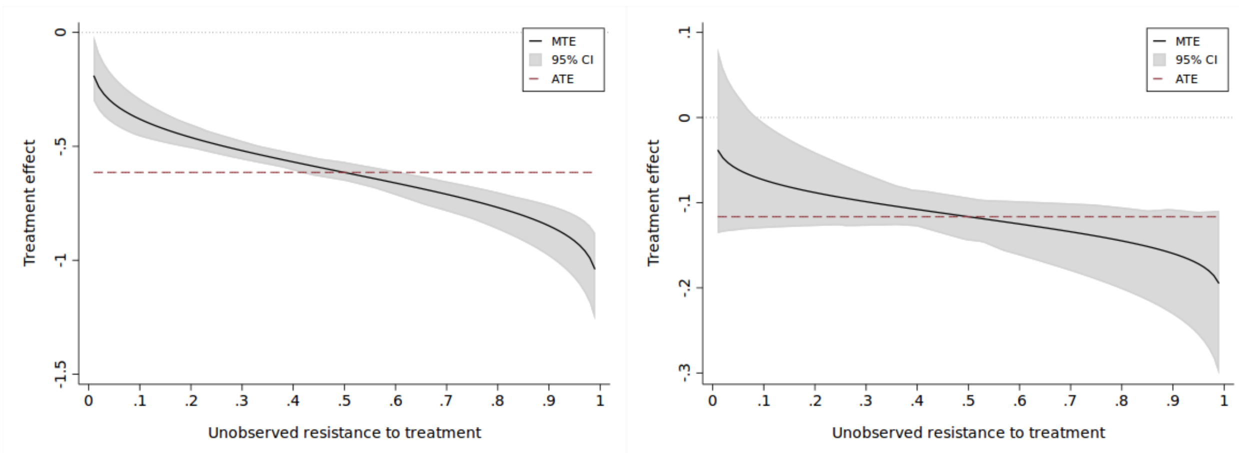
Fig. 18 MTE curve - Semiparametric



(a) LOS **(b) Specific skills**

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, semiparameterically. 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, hospital quality per robots, and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions.

Fig. 19 MTE curve - Normal with area fixed effects



(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, semiparameterically. 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, hospital quality per robots, and year, month and day of the week fixed effects. Standard errors are bootstrapped with 100 repetitions.

Table 1 Radical Prostatectomy Patients - Sample Summary Statistics - 2003/2017

	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 Test of exclusion restriction for Z_{dist} - Correlation between AMI patients outcomes and relative distance to robotic hospital

	(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)
Age			0.0919^{***}
			(0.00982)
Age sq.			-0.000226^{***}
			(0.0000661)
Year-month	No	Yes	Yes
Day of the week	No	Yes	Yes
Patient characteristics	No	No	Yes
Share of deaths	0.19	0.19	0.19
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 displayed are a total of ten comorbidity dummies (e.g. malignant neoplasm, diabetes). Sample of AMI patients from 2005 to 2009.

Table 3 Test of exclusion restriction for Z_{days} - Correlation of surgical outcomes and instrument 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)
Age		-1.68	-1.60		-1.84	-2.08
		(1.31)	(1.30)		(1.40)	(1.40)
Age sq.		0.0199	0.0191		0.0165	0.0184
		(0.0106)	(0.0105)		(0.0114)	(0.0114)
Patient characteristics	No	Yes	Yes	No	Yes	Yes
Year-month	No	No	Yes	No	No	Yes
Day of the week	No	No	Yes	No	No	Yes
ymean	1.7784	1.7780	1.7780	0.2435	0.2431	0.2431
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. Patient characteristics include 10 comorbidity dummies, ethnicity, sex, rural urban indicator. Sample of radical prostatectomy patients period 2003.

Table 4 Probit model - dependent variable indicator of robotic surgery

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (A)							
Z_{dist}	-1.95***		-1.04***	-1.07***	-1.10***	-0.990***	-1.05***
	(0.0327)		(0.0319)	(0.0323)	(0.0334)	(0.0339)	(0.0347)
Z_{days}		0.0574***	0.0412***	0.0405***	0.0417***	0.0243***	0.0226***
		(0.000476)	(0.000581)	(0.000593)	(0.000599)	(0.000696)	(0.000711)
SRR							-1.188***
							(0.0247)
Panel (B)							
Z_{dist}	-0.651***		-0.307***	-0.316***	-0.319***	-0.272***	-0.273***
Z_{days}		0.0159***	0.0122***	0.0119***	0.0122***	0.00667***	0.00587***
SRR							-30.9***
Demographic	No	No	No	Yes	Yes	Yes	Yes
Clinical	No	No	No	Yes	Yes	Yes	Yes
YM	No	No	No	Yes	Yes	Yes	Yes
DOW	No	No	No	Yes	Yes	Yes	Yes
Area	No	No	No	No	Yes	Yes	Yes
Share Rob	0.48	0.44	0.49	0.49	0.49	0.49	0.49
Z_{dist}	22	21	21	21	21	21	21
Z_{days}	389	68	389	387	387	387	387
N	53937	58906	52671	52572	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 Probit regression - Dependent variable indicator of robotic approach - Margins of SRR under different transformations

	(1)	(2)	(3)
SRR	-0.309***		
	(0.00609)		
SRR (Std.)			-0.115***
			(0.00227)
Indicator of high skills		0.199***	
		(0.00359)	
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. Controls for age, age squared, indicator for white ethnic profile, ten comorbidity variables, distance to closest hospital, indicator for closest hospital being teaching hospital, urban city indicator, year-month, day of the week controls. SRR is the standardised risk ratio for post-operative morbidity (interpreted as the inverse of skills). SRR (Std.) is the standardised version of the SRR. Indicator of high skills is a binary variable that takes value 1 if the SRR is below the median (i.e. 0.8).

Table 6 Heterogeneity in causal effects - Observable characteristics

	(1) LOS	(2) Adverse event	(3) LOS	(4) Adverse event
	β_0		$\beta_1 - \beta_0$	
SRR (Inverse of Skills)	0.302***	0.167***	-0.374***	-0.0830***
	(0.0138)	(0.00836)	(0.0342)	(0.0177)
Distance to closest hospital	0.00339***	-0.000394	-0.00597***	0.000461
	(0.000483)	(0.000343)	(0.000795)	(0.000487)
Indicator of teaching hospital	-0.0634***	0.00444	0.0138	-0.00529
	(0.0118)	(0.00883)	(0.0183)	(0.0125)
Age	-0.0111	-0.00560	-0.0526**	-0.00225
	(0.0107)	(0.00517)	(0.0182)	(0.00917)
Age sq.	0.000126	0.0000510	0.000402**	0.0000148
	(0.0000880)	(0.0000425)	(0.000149)	(0.0000755)
AMI	0.0407	0.0281	-0.0179	-0.0226
	(0.0354)	(0.0241)	(0.0561)	(0.0380)
COPD	0.0673***	0.0148	-0.0624*	0.00637
	(0.0150)	(0.0118)	(0.0256)	(0.0162)
Diabetes	0.0517**	0.0175	0.0664*	0.00226
	(0.0194)	(0.0123)	(0.0324)	(0.0184)
RD	0.0765	0.0322	0.0187	-0.0272
	(0.0475)	(0.0384)	(0.0832)	(0.0598)
Share of robotic operations	0.49	0.49	0.49	0.49
Z_{dist} mean	22.22	22.22	21.86	21.86
Z_{days} mean	367.08	367.08	387.00	387.00
N	49215	50203	49215	50203

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors bootstrapped with 100 repetitions.
Columns (1) and (3) display estimates from the outcome equation where the dependent variable is the logarithm of post-operative length of stay. Columns (2) and (4) display estimates from the outcome equation where the dependent variable is an indicator of adverse event from surgery. Coefficients of regressors not interacted with the propensity score measure effects on the outcome in the untreated state (Columns (1) and (2)). Coefficients of regressors interacted with the propensity score measure effects the difference of the effects between the treated and the untreated state (Columns (3) and (4)). Controls not displayed are ethnicity, a rural urban indicator, and there are a total of ten comorbidity dummies. Estimated using year, month, and day of the week fixed effects all interacted with the propensity score.

Table 7 Heterogeneity in causal effects - Observable characteristics - Indicator of skills

	(1) LOS	(2) Adverse event	(3) LOS	(4) Adverse event
	β_0		$\beta_1 - \beta_0$	
Highly-skilled indicator	-0.220***	-0.0868***	0.321***	0.0701***
	(0.0116)	(0.00728)	(0.0228)	(0.0136)
Distance to closest hospital	0.00357***	-0.000266	-0.00599***	0.000478
	(0.000469)	(0.000383)	(0.000811)	(0.000568)
Indicator of teaching hospital	-0.0978***	0.0000937	0.0772***	0.00462
	(0.0117)	(0.00848)	(0.0176)	(0.0114)
Age	-0.0101	-0.00674	-0.0547**	0.0000341
	(0.0114)	(0.00666)	(0.0209)	(0.00997)
Age sq.	0.000122	0.0000617	0.000410*	-0.00000576
	(0.0000937)	(0.0000544)	(0.000172)	(0.0000819)
AMI	0.0435	0.0269	-0.0215	-0.0194
	(0.0401)	(0.0238)	(0.0668)	(0.0351)
COPD	0.0671***	0.0183	-0.0597*	0.00189
	(0.0172)	(0.00958)	(0.0285)	(0.0160)
Diabetes	0.0527**	0.0233	0.0632*	-0.00807
	(0.0190)	(0.0136)	(0.0319)	(0.0198)
RD (Renal)	0.0942	0.0369	-0.00837	-0.0307
	(0.0513)	(0.0337)	(0.0861)	(0.0516)
Share of robotic operations	0.49	0.49	0.49	0.49
Z_{dist} mean	22.22	22.22	21.86	21.86
Z_{days} mean	367.08	367.08	387.00	387.00
N	49215	50203	49215	50203

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors bootstrapped with 100 repetitions.
Columns (1) and (3) display estimates from the outcome equation, where the dependent variable is the logarithm of post-operative length of stay. Columns (2) and (4) display estimates from the outcome equation, where the dependent variable is an indicator of adverse event from surgery. Coefficients of regressors not interacted with the propensity score measure effects on the outcome in the untreated state (Columns (1) and (2)). Coefficients of regressors interacted with the propensity score measure effects the difference of the effects between the treated and the untreated state (Columns (3) and (4)). Controls not displayed are ethnicity, a rural urban indicator, and there are a total of ten comorbidity dummies. Estimated using year, month, and day of the week fixed effects all interacted with the propensity score.

Table 8 Heterogeneity in causal effects - Observable characteristics - Standardized measure of skills

	(1) LOS	(2) Adverse event	(3) LOS	(4) Adverse event
	β_0		$\beta_1 - \beta_0$	
Standardized SRR (inverse of skills)	0.113***	0.0621***	-0.139***	-0.0309***
	(0.00503)	(0.00328)	(0.0143)	(0.00746)
Distance to closest hospital	0.00339***	-0.000394	-0.00597***	0.000461
	(0.000508)	(0.000341)	(0.000828)	(0.000515)
Indicator of teaching hospital	-0.0634***	0.00444	0.0138	-0.00529
	(0.0126)	(0.00892)	(0.0172)	(0.0125)
Age	-0.0111	-0.00560	-0.0526**	-0.00225
	(0.0106)	(0.00620)	(0.0177)	(0.00923)
Age sq.	0.000126	0.0000510	0.000402**	0.0000148
	(0.0000871)	(0.0000505)	(0.000144)	(0.0000757)
AMI	0.0407	0.0281	-0.0179	-0.0226
	(0.0394)	(0.0265)	(0.0622)	(0.0385)
COPD	0.0673***	0.0148	-0.0624*	0.00637
	(0.0163)	(0.0115)	(0.0274)	(0.0174)
Diabetes	0.0517*	0.0175	0.0664*	0.00226
	(0.0207)	(0.0130)	(0.0329)	(0.0194)
RD	0.0765	0.0322	0.0187	-0.0272
	(0.0484)	(0.0287)	(0.0795)	(0.0426)
Share of robotic operations	0.49	0.49	0.49	0.49
Z_{dist} mean	22.22	22.22	21.86	21.86
Z_{days} mean	367.08	367.08	387.00	387.00
N	49215	50203	49215	50203

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors bootstrapped with 100 repetitions.
Columns (1) and (3) display estimates from the outcome equation, where the dependent variable is the logarithm of post-operative length of stay. Columns (2) and (4) display estimates from the outcome equation, where the dependent variable is an indicator of adverse event from surgery. Coefficients of regressors not interacted with the propensity score measure effects on the outcome in the untreated state (Columns (1) and (2)). Coefficients of regressors interacted with the propensity score measure effects the difference of the effects between the treated and the untreated state (Columns (3) and (4)). Controls not displayed are ethnicity, a rural urban indicator, and there are a total of ten comorbidity dummies. Estimated using year, month, and day of the week fixed effects all interacted with the propensity score.

Table 9 Heterogeneity in causal effects - Observable characteristics - Area fixed effects

	(1) LOS	(2) Adverse event	(3) LOS	(4) Adverse event
	β_0		$\beta_1 - \beta_0$	
SRR (Inverse of Skills)	0.158***	0.146***	-0.349***	-0.181***
	(0.0151)	(0.00945)	(0.0288)	(0.0197)
Age	-0.0127	-0.00607	-0.0354**	0.000684
	(0.0103)	(0.00611)	(0.0133)	(0.00961)
Age sq.	0.000137	0.0000555	0.000258*	-0.0000102
	(0.0000839)	(0.0000500)	(0.000109)	(0.0000792)
AMI	0.0395	0.0149	-0.0435	-0.000957
	(0.0264)	(0.0238)	(0.0393)	(0.0290)
COPD	0.0378**	0.0186*	-0.0130	-0.000556
	(0.0121)	(0.00909)	(0.0212)	(0.0144)
Diabetes	0.0695***	0.0125	0.0130	0.0105
	(0.0152)	(0.0120)	(0.0223)	(0.0152)
RD	0.0189	0.0338	0.109	-0.0334
	(0.0430)	(0.0220)	(0.0627)	(0.0315)
Share of robotic operations	0.49	0.49	0.49	0.49
Z_{dist} mean	22.22	22.22	21.86	21.86
Z_{days} mean	367.08	367.08	387.00	387.00
N	49215	50203	49215	50203

[*] $p < 0.05$, [**] $p < 0.01$, [***] $p < 0.001$. Standard errors bootstrapped with 100 repetitions.
Columns (1) and (3) display estimates from the outcome equation where the dependent variable is the logarithm of post-operative length of stay. Columns (2) and (4) display estimates from the outcome equation where the dependent variable is an indicator of adverse event from surgery. Coefficients of regressors not interacted with the propensity score measure effects on the outcome in the untreated state (Columns (1) and (2)). Coefficients of regressors interacted with the propensity score measure effects the difference of the effects between the treated and the untreated state (Columns (3) and (4)). Controls not displayed are ethnicity, a rural urban indicator, and there are a total of ten comorbidity dummies. Estimated using year, month, and day of the week fixed effects all interacted with the propensity score.

Table 10 Estimated treatment effects of log 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

Standard errors in parentheses

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

Table 11 Estimated treatment effects on Adverse Events

	(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

Standard errors in parentheses

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