TECHNOLOGY, SKILLS, AND PERFORMANCE: THE CASE OF ROBOTS IN SURGERY

Elena Ashtari Tafti

In this paper, I investigate the potential of robots to improve surgeons' performance and study whether the returns from using this technology depend on the skills of the surgeon using it. Using data from the universe of inpatient admissions for prostate cancer surgery in England, I show that while technological advancements in many fields have typically favored more skilled individuals, here are the low-skill surgeons whose performance benefits the most. As differences in outcomes between highly and less skilled surgeons shrink when the technology is used, my analysis suggests that the robot may alleviate unwarranted disparities in the quality of care. However, I find evidence of suboptimal behavior on the provider's side. Even tho the low-skill surgeons have the highest returns, they use the technology the least, limiting the attainable benefits from technology adoption.

1 Introduction

Technological change has been observed to have important implications for the returns to skills across the economy. Current research suggests that advanced technologies such as robots and ICT have been skill-biased (Katz & Murphy 1992, Krusell et al. 2000, Autor et al. 2003, Acemoglu & Autor 2011). The losers in this process are typically those working in low-skill jobs, whose tasks can be easily automated and performed by machines.

In this paper, I take the idea of skill bias to a micro level and study the effects of technology adoption on the returns to skills within a single high-skill occupation. My focus is on the adoption of robots in surgery, where I investigate the potential of this technology to improve surgeons' performance and, as a consequence, patient outcomes, and study whether its returns depend on the actual skills of the surgeon using it.

As economists, we mostly think of robots competing against human labor in the production process (Acemoglu & Restrepo 2020, Humlum 2019). Tasks previously performed by workers are automated and executed by the machine more precisely and consistently. In many applications, however, the robot is meant to aid rather than substitute the worker. Surgical robots, for example, are fully operated by surgeons and act as an extension of their users. These machines, like many others, require explicit operator inputs to be effective, and for this reason, any conceivable outcome from adoption will inevitably depend on the interaction between the human and the technology.

Across and within occupations, individuals exhibit significant variation in their skills, and healthcare providers, such as surgeons and doctors, are no exception (Chan et al. 2022, Currie & MacLeod 2017, Kolstad 2013). While technological advancements in many fields have favored more skilled individuals (Acemoglu 2002, Autor et al. 1998), the existence and direction of a hypothetical skill bias in this context is ex-ante uncertain.

Operating the robot inevitably requires some skills, which may be more prevalent among surgeons who excelled with traditional techniques. At the same time, the robot eliminates the need for other skills, such as having steady hands, thereby reducing the performance advantage of conventionally superior surgeons. Whether one effect or the other prevails is of interest in itself, but it also comes with important consequences for policymaking. Differences in providers' skills are thought to generate inequality and aggravate systematic disparities in the access to public services (Finkelstein et al. 2016, Chandra & Skinner 2003, Deaton 2003). Robotic technology may exacerbate variation in surgical performance, or may be a solution to this problem if its returns are decreasing in surgeons' skills.

Using data from the universe of inpatient admissions for prostate cancer surgery in England, I find that robotic surgery improves surgeons' performance. The robot reduces post-operative length of stay and morbidity across patients. However, my analysis reveals that these effects are highly heterogeneous, and technological gains significantly depend on the surgeon's skills. High-skill surgeons benefit the least from the technology, while low-skill surgeons gain the most. Technically, the robot appears to have *decreasing* returns in skills, with a bias that favors the low-skill surgeons. As differences in patient outcomes between highly and less skilled surgeons shrink when the technology is used, my analysis suggests that the robot may reduce variation in patient outcomes. This effect appears to ensue from low-skill surgeons performing significantly more poorly without any technological aid and the technology equalizing them to the high-skill surgeons.

Identifying the existence of skill bias at a micro level comes with its own empirical challenges. Part of my contribution is to study the impact of this new technology in the presence of both heterogeneous treatment effects and a selection problem. To this day, medical evidence that robotic surgery improves patient outcomes, relative to the more invasive alternative (i.e., open surgery), has been at best inconclusive (Coughlin et al. 2018, Yaxley et al. 2016, Robertson et al. 2013, Bolla et al. 2012). Existing studies are based on small and selected samples (Neuner et al. 2012) and are not designed to identify causal effects (Ho et al. 2013). If the potential of robotic surgery to improve performance depends on surgical skills, small sample studies will reflect only part of the picture. Moreover, if the uptake of this technology is also heterogeneous across the skills distribution, any naive correlation will speak more to the characteristics of the adopters rather than the technology itself. Importantly, when treatment effects are heterogeneous, surgeons and patients may choose the robot based on their specific technological gains. Regression-adjusted comparisons between robotic and traditional surgery would, in this case, provide misleading estimates if adoption is informed by unobserved factors that influence selection (Suri 2011).

To identify causal effects, I leverage an approach introduced by Björklund & Moffitt (1987) and generalized in Heckman & Vytlacil (1999) and Heckman & Vytlacil (2005) that concentrates on the Marginal Treatment Effect (MTE). In this context, the MTE is the average effect of robots on the outcome of a surgeon's patient equally likely to be operated with robotic and traditional surgery (i.e., at the same margin of indifference). I show that the MTE nests my parameter of interest, and its estimation allows me to directly identify the causal effects of robots on patient outcomes and how these depend on observable surgical skills.

Identification of causal effects in the MTE framework requires, in most cases, no stronger assumptions than standard instrumental variable methods but poses a more substantial burden on the instrument. This method usually requires at least one valid and continuous instrumental variable with large support (Mogstad et al. 2018, Heckman & Vytlacil 2005). In England, the acquisition of surgical robots has been managed by individual hospitals, resulting in an uneven distribution of robots geographically and creating differences in the availability of the technology across areas and over time (Lam et al. 2021). Building on the work of Gowrisankaran & Town (1999), McClellan & Newhouse (1997) and McClellan et al. (1994), I argue that variation in a patient's relative distance to a robot-equipped hospital is sufficient to identify my parameter of interest under the assumption that patients do not sort across providers based on their unobserved health returns from the technology.

An important constraint I face is that skills are not directly observable, and in this context, using education as a proxy is unfeasible. I revert to employing a single risk-adjusted indicator of surgeons' postoperative morbidity to measure skills. The indicator tells me whether the surgeon counts more complications and similar adverse events than what would be expected given their patients' observable characteristics. Because I anticipate the robot will impact surgeons' performance, I estimate this indicator using data from the years preceding the national introduction of this technology. In fact, the indicator is measured when all operations were carried out without technological aid and is not affected by the surgeons' adoption behavior. To ensure that the skill measure I derive is not affected by unobserved factors that affect selection across high and low-skill surgeons, I employ three alternative strategies and test the robustness of my results to each.

Having established that surgeons exhibit significant variation in their skills, I focus on two patient outcomes to identify the effect of the robot on surgeons' performance. These are the speed of recovery (i.e., postoperative length of stay) and the occurrence of adverse events from surgery (i.e., postoperative morbidity, which I also use to estimate skills). Both matter to physicians, patients, and policymakers (Lotan 2012), and robotic surgery should have a measurable effect on them as it increases precision and requires smaller incisions (Higgins et al. 2017, Coelho et al. 2010, Lowrance et al. 2010, Nelson et al. 2007).

This study builds on several branches of the economics literature, yet nonetheless, is novel from many perspectives. First and foremost, this paper draws on the literature on the effects of technology on the labor market. The idea that technology complements either high or low-skill workers has been incorporated in many notable contributions discussing the evolution of wages and earnings in the US (Carneiro & Lee 2011, Acemoglu & Autor 2011, Autor et al. 2008, 1998, Katz & Murphy 1992). I bring the concept of skill bias to a new setting where individuals' skills vary within the same occupation. This is in stark contrast to the literature in this area, which generally thinks about the interaction of education and technology (Goldin & Katz 2018). In my context, all workers have the same education level, but the technology still complements some more than others. This provides a new perspective on the implications of technology adoption, which may have distributional effects across and within occupations.

The idea that technology makes certain skills redundant while also creating new tasks has been argued in a voluminous and influential branch of the economics literature (Acemoglu & Restrepo 2019, 2018, Acemoglu & Autor 2011, Autor et al. 2003, Acemoglu & Zilibotti 2001, Zeira 1998). Autor et al. (2003), for example, discusses how the computer has replaced workers in cognitive and manual tasks and complemented them in nonroutine tasks. This approach has significantly influenced how I conceptualize robots and surgeons working together. My contribution is to show that this trade-off, which determines who benefits from technology adoption, also works at a micro level.

An ever-growing part of this field studies robots in particular (Acemoglu et al. 2023, Acemoglu & Restrepo 2022, Koch et al. 2021, Acemoglu & Restrepo 2020, Acemoglu et al. 2020, Humlum 2019, Graetz & Michaels 2018). The empirical focus is primarily on automated industrial robots and their industry-level effects on employment and wages. Few scholars have studied surgical robots. Using US data Horn et al. (2022) shows that adopting a robot can increase the demand for a hospital. Maynou et al. (2021) describes a similar pattern for the UK and show that adopting robots correlates with reduced readmissions and length of stay. Maynou et al. (2022) discusses how using robots for prostate cancer patients affected their diffusion in other specialties in the UK. To the best of my knowledge, this is the first paper to incorporate considerations about skill complementarities in relation to robots at the micro level that occurs within a high-skill occupation.

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

Lastly, the results of this paper are consistent with the nascent literature on the effects of generative AI. Brynjolfsson et al. (2023), Choi & Schwarcz (2023), Noy & Zhang (2023), and Peng et al. (2023) all find that AI compresses the productivity distribution, with lower-skill workers benefiting the most. The intuition behind the results of these studies is closely aligned with the empirical evidence I find in this paper. AI enhances the performance of low-skill workers by mimicking the behavior of high-skill workers. Similarly, my findings suggest that the robot improves the performance of low-skill surgeons by reproducing the skills of a high-skill surgeon, such as having steady hands.

The paper proceeds as follows: Section 2 describes surgical robots and their use for prostate cancer surgery; Section 3 describes the data and institutional setting; Section 4 explains how I measure surgeons' skills; Section 5 presents the econometric challenge, model, and the conditions required to identify and estimate the parameter of interest; Section 6 summarizes the results; Finally, I offer my conclusions in Section 7.

2 Robotic Surgery for Prostate Cancer

2.1 Robotic Surgery

The use of robotics in surgery was hypothesized as far back as 1967, but it took nearly 30 years for the National Aeronautics and Space Administration (NASA) to complete the first functional surgical robot (George et al. 2018). Now, the only type of surgical robot currently available in the US and the UK is the da Vinci surgical system manufactured by the California-based company Intuitive.

The da Vinci robot has three components, which I show in Figure I:

1. a viewing and control console that the surgeon uses,
2. a manipulator arm unit that includes three or more arms,
3. a vision cart that connects all the system elements and facilitates communication between the surgeon console and robotic arms.

The surgeon sits at the console and controls the robotic arms using her own hands. The console consists of finger loops, joysticks, and foot pedals that allow the surgeon's movements to go through the robot. The 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. The surgical instruments, including a video camera, are attached to the robotic arms. The robotic arms allow working through incisions much smaller than those typically required and at a scale where hand tremors pose fundamental limitations (Tonutti et al. 2017). By increasing articulation, integrating tremor filtering, and simulating tactile feedback, the system amplifies the surgeon's dexterity and eye-hand coordination, leading to a subjective enhancement in surgical performance (Tonutti et al. 2017).

2.2 Robotic Assisted Radical Prostatectomy

Although robots have several applications in surgery, I focus on robotic surgery for prostate cancer patients (or radical prostatectomy (RP) patients).

Prostate cancer ranks as the second most frequently diagnosed cancer in men globally, with surgery being among the common treatment options available.[1] Because the prostate is hard to access and is close to many blood vessels and important nerves, surgeons' skills play a fundamental role in determining patient outcomes from this surgery.[2]

Urology is at the forefront of robotic surgery in most countries, and RP is by far the most commonly performed procedure robotically, both in absolute and relative terms. In the US,

[1] World Health Organization
[2] Prostate Cancer Foundation

for example, the diffusion of robots for prostate cancer surgery has been incredibly rapid. In 2003, less than 1 percent of surgeons performed this procedure robotically. Seven years later, 86 percent of the 85,000 men who had prostate cancer surgery had a robot-assisted operation. Eventually, by 2014, robotic surgery accounted for up to 90 percent of radical prostatectomies across the US.[3] This trend has been similar in England where, by 2018, the majority of cases (88% percent) were performed robotically (Maynou et al. 2022).

The robot has played a notable role in transforming how surgeons perform RP. Before robots, prostate cancer surgery was usually performed with an 'open' method. In the 'open' method, the surgeon makes a single large incision that allows the area of interest to be seen and operated.[4] From an oncological perspective, robotic surgery is equivalent to traditional surgery; they are both practical to remove cancer when it is confined to the prostate. However, robotic surgery promised to reduce blood loss, pain, scarring, infections, and average length of stay (among others) by replacing the practice of cutting patients open with a technique that involved few small incisions (see Figure II) and minuscule tools.

Among the most significant barriers to adopting robotic surgery in urology and elsewhere are the high costs of purchasing and maintaining robots (Marcus et al. 2017). The acquisition cost of a da Vinci robot ranges between \$0.5 and \$2.5 million, depending on model, configuration, and location (Eckhoff et al. 2023). This does not include the annual service fee of up to \$190,000 or the reoccurring cost of instruments and accessories (Eckhoff et al. 2023). On top of that, robots usually require a dedicated operating room, often built for this purpose, and both surgeons and nurses also need specialized training. Operating using the console requires significant coordination between the head surgeon and the assistant working at the bedside. Any technical drawback during the operation is risky for the patient but also prolongs operation time and generates inefficiencies for the hospital (Compagni et al. 2015).

[3] Crew, B., *Worth the cost? A closer look at the da Vinci robot's impact on prostate cancer surgery*, Nature. Retrieved on 22 April 2020 from https://www.nature.com/articles/d41586-020-01037-w

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

To this day, conclusive evidence on whether the robot is superior at performing RP is relatively scarce, and doubts remain on whether the supposed benefits outweigh the costs of this technology (Davies 2022).

3 Data and Institutional Setting

3.1 Data

The data I use come from the Hospital Episodes Statistics (HES), an administrative data set covering the universe of inpatient discharges from the English National Health Service (NHS).

From HES, I collect data on all radical prostatectomy patients operated by an NHS hospital between April 2004 and April 2018. For each of them, HES records the method used to perform the operation (e.g., traditional or robotic), the consultant in charge, and the hospital where this was done. This allows me to track how the technology diffuses across hospitals, surgeons, patients, and over time.[5]

HES also provides detailed patient demographic and clinical information, including age, sex, ethnicity, admission date, discharge date, and up to 20 recorded diagnoses and operation codes. Geographical information, such as the patient's area of residence, is also available.

Patients maintain their identification codes over time, enabling me to compile their hospital records pre-RP and accurately monitor their post-operative outcomes. By inspecting the patient hospital history (i.e., admission before the operation of interest), I identify 15 risk variables deemed by the medical literature to be important parameters of patients' health (e.g., heart diseases, diabetes) and likely to influence their outcomes from surgery. From the data, I also compute the number of admissions before the patient's surgery, the number of diagnoses at the time of operation, the time that has passed since their cancer diagnosis, and

[5] I have access to HES records from 2000 to 2018. Information on the consultant in charge of the operation is available from 2004. Information on the operation method is available since

the time they had to wait since their initial referral for the operation (i.e., waiting time).

In Table I, I summarize the characteristics of the patients in my sample separately for the traditional and robotic approaches. Table II provides the difference in means between robotic and traditional surgery patients controlling for the year of operation to account for changes in patient compositions over time. Robotic surgery patients appear to be younger and significantly healthier (i.e., lower incidence of risk variables). The hospital history variables also point to the robotic patients being less risky, with fewer prior hospital admissions and fewer diagnoses at the time of operation. Overall, there appears to be a significant degree of positive selection into robotic surgery, with healthier patients being operated on using this technology.

3.1.1 Patient Outcomes as Skill and Performance Measures

The information in HES allows me to reliably measure two patient outcomes. These are the patient length of stay in the hospital (LOS) and the occurrence of adverse events from surgery (i.e., negative post-operative outcomes or post-operative morbidity).

I rely on the occurrence of adverse events from surgery to differentiate between high- and low-skill surgeons before the introduction of robots. My objective is to establish a clear distinction in skill level, and this outcome is the most salient indicator of surgical skills. I then identify the effect of the robot on adverse events and length of stay to evaluate whether the technology has affected surgical performance.

I focus on these outcomes for three reasons. Undoubtedly, patients desire to spend fewer days in the hospital and minimize complications from surgery. If robotic surgery would improve these outcomes, patients would clearly benefit from it. LOS and adverse events are also important cost drivers to the system and are often considered when evaluating whether a technology is worth adopting (Lotan 2012). Lastly, the medical literature considers that — if any — robotic technology should have measurable benefits on these two margins (Higgins et al. 2017, Coelho et al. 2010, Lowrance et al. 2010, Nelson et al. 2007). Robotic surgery allows operating using small, compact tools that fit into narrow incisions. The procedure is consequently less invasive and should, therefore, increase the speed of recovery (or reduce the length of stay). Further, because these tools allow for higher precision, the incidence of adverse events, especially surgical complications, should diminish.

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

Adverse health events serve as a critical indicator of the quality of the procedure performed. I consider a broad range of events to be part of this category, including in-hospital deaths (within 30 days), emergency readmissions (within 30 days), and the occurrence of surgical complications. This last category is specific to RP patients. It includes urinary complications, and erectile dysfunctions that require further surgical interventions[6], and blood transfusions. Urinary complications and erectile dysfunction, measured within two years of the operation, are common side effects of prostate cancer surgery and are often employed to measure surgical performance for this operation (Hugosson et al. 2011, Hu et al. 2003, Fowler Jr et al. 1995). Blood transfusions are typically associated with surgical complications and are frequently used as a quality indicator (Porcaro et al. 2022).

Table I summarizes the outcomes by surgical approach. The average post-operative length of stay in hospital is 4 days for traditional surgery and 1.9 for robotic surgery. Urinary complications are the most common category of adverse events, while death and readmissions

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

are extremely rare. Adverse events are almost twice as likely to occur with traditional surgery than with the robot (19% vs 8% with robotic surgery). These differences are also statistically significant when comparing patients operated in the same year as in Table II.

3.2 Institutional Setting

The NHS is the second-largest single-payer healthcare system in the world. Hospitals in the NHS provide care to patients and are reimbursed by the government under nationally agreed tariffs. Access to planned or elective procedures is rationed through waiting times and requires an initial referral from a General Practitioner (GP).

A patient in the NHS suspected of having prostate cancer would be referred by their GP to a specialist center for diagnosis. Following diagnosis, patients receive comprehensive information about all treatment options, including surgical intervention if the tumor is confined to the prostate. Both robotic and traditional surgical approaches are typically presented and discussed, including the advantages and disadvantages of each technique. If the patient's local provider does not offer robotic surgery, they can request a referral to a hospital where it is available (Coulter 2010). Hospitals cannot refuse patients but may schedule admissions or cancel treatments if there is a lack of capacity.

In the context of RP surgeries, robotic surgery began gaining popularity in the English NHS around 2007. Figure IV illustrates both the total number and the proportion of RPs performed by the two different surgical approaches. Notably, the use of robotic surgery increased from 5 percent in 2007 to 80 percent in 2017.

Since 2006, the NHS constitution has guaranteed patients the right to choose where they receive treatment, ensuring fair access to resources for all. However, this principle has not consistently ensured the same level of accessibility in the context of robotic surgery. The adoption of the Da Vinci system occurred in a scattered and uncoordinated manner, leading to significant variation in the availability of robots across regions and over time (Lam et al. 2021). Figure V shows the location of hospitals adopting the robot by time of adoption, as identified from the HES data. Twenty-three hospitals adopted the robot before 2011, another 23 between 2011 and 2014, and seven adopted it after 2017. Meanwhile, 100 hospitals offering RP surgery never adopted the robot during the observation period. The decision to adopt this technology was left to the individual hospital; surgeons were not mandated to learn how to use it, and it took nearly 14 years to develop any clinical guidelines and a national plan for the rollout of the robot (Maynou et al. 2022). This delay was largely attributed to uncertainty regarding the machine's benefits.

4 Measuring Surgical Skills

Skills are not directly observable and notoriously difficult to measure. The measurement most commonly called upon in economics is some indicator of educational attainment (Borghans et al. 2001). Still, 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 individual's skills. For example, Birkmeyer et al. (2013) shows a clear relationship between surgical skills and patient outcomes.

When using outcomes to compare surgeons, it's essential to recognize the influence that risk factors and other patient characteristics play. In fact, some form of risk adjustment is necessary to account for the fact that different surgeons operate on different patient populations. The procedure I use to produce a single risk risk-adjusted indicator of skills proceeds in two steps.[7] In the first step, I estimate a random coefficient model with a surgeon random intercept. In the second step, I use the regression estimates to compute a surgeon-level Standardized Risk Ratio (SRR) of postoperative outcomes, which I use as my skills measure. This approach helps to mitigate differences in health and other risk factors that impact observed outcomes across surgeons and generates a comparable indicator of skills that I can use to categorize surgeons.

[7]This methodology is inspired mainly by the work of Horwitz et al. (2014) for the Centers for Medicare & Medicaid Services (CMS)

4.1 Risk Adjustment Methodology

The surgeon random intercept model allows me to quantify the variation in outcomes across surgeons that the observable patient characteristics cannot explain. This is the surgeon's specific contribution to the patient's outcome, which should reflect their skills.

Let Y_{ij} for patient i operated by surgeon j denote a binary outcome equal to one if the patient experiences an adverse event from surgery. X_{ij} denotes a set of risk factors identified by the medical literature to influence the outcome of patient i which are observed in the data. I assume that the outcome is related linearly to the covariates via a logistic function $F(\cdot)$ with its conditional distribution assumed to be Bernoulli:

$$\begin{aligned} Prob\ (Y_{ij} = 1) &= F\ (\alpha_j + \beta X_{ij}) \\ \alpha_j &= \mu + \omega_j \\ \omega_j &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{1}$$

Here, α_j represents a surgeon-specific intercept, μ is the adjusted average outcome across all surgeons, and ω_j is the random component allowed to result from an unknown process.

The estimation's objective is σ^2, the variance of the surgeon random effect distribution or the between surgeons variance. This component represents the surgeons' unique contribution to the overall variance in patient outcomes. From σ^2, α_j, the empirical Bayes estimate of the individual surgeon effect, can be obtained using the Bayes posterior means method.

4.2 Standardized Risk Ratio

I use the model estimated parameters to compute a surgeon-specific ratio of *predicted* and *expected* patients' postoperative morbidity or adverse events (Horwitz et al. 2014).

Let M_j denote the set of patients operated by surgeon j. The *predicted* number of adverse events for surgeon j follows from the conventional definition of prediction. It is calculated as the sum of the predicted probability of the event for each patient $i \in M_j$ using the model estimated parameters $\widehat{\beta}$ and the Empirical Bayes estimate $\widehat{\alpha}_j$.

$$\widehat{Y}_j = \sum_{i \in M_j} logit^{-1}\big(\widehat{\alpha}_j + \widehat{\beta} X_{ij}\big) \tag{2}$$

The *expected* number of adverse events is more nuanced. This is calculated as the sum of the predicted probability for each patient $i \in M_j$, ignoring the estimated surgeon-specific random effect $\widehat{\alpha}_j$. That is the probability of an adverse event given the estimated parameters, but where σ is set to zero. Or, equivalently, the probability of an adverse event when the dispersion in α_j is zero.

$$\widetilde{Y}_j = \sum_{i \in M_j} logit^{-1}\big(\widehat{\mu} + \widehat{\beta} X_{ij}\big) \tag{3}$$

The SRR of surgeon j is then the ratio of these two estimates:

$$\text{SRR}_j = \frac{\widehat{Y}_j}{\widetilde{Y}} \tag{4}$$

The ratio summarizes whether a surgeon is doing better or worse than we would expect given her pool of patients. A value of 1 indicates that surgeon j's level of adverse events is as expected, given her patients. A value above (below) 1 suggests that the surgeon is under- (over-) performing relative to an average surgeon with a comparable population of patients. I use the SRR as a continuous skills measure in the analysis, with lower values indicating higher skills. Alternatively, I create a binary indicator of skills that takes value 1 if the SRR is below the median of its distribution and 0 otherwise.

4.3 Measure Validity and Estimation

The validity of the SRR as a measure of skills relies on the model effectively managing omitted variables and unobserved heterogeneity.

First and foremost, I want the measure not to be affected by the surgeon's technology use. For this reason, I estimate the SRR and the model behind it using data from 2004 to 2007, a period before the diffusion of robots, as observed in my sample. By focusing on when all operations are performed with the traditional method, the outcomes are not endogenous to the surgeon's technology use and reflect the surgeon's stock of skills ahead of the introduction of robots.

Generally, I also want to exclude that the heterogeneity I attribute to the surgeon's random component arises from patient composition or hospital quality. I produce three versions of the SRR that reflect different restrictions on the variability in outcomes I allow to be ascribed to the surgeon's random component. Eventually, I will use all of them to test the robustness of my results to my design choices.

To account for unobservable heterogeneity in the patient population across surgeons, I include a large set of observable patient risk factors and demographic characteristics in the risk adjustment model. Further, I control for the area where the patient lives. This substantially reduces the estimated between-surgeon variance parameter, which drops by about 40 percent, as shown in Table III. Indeed, much of the variation at the tails of the distribution stemmed from unobserved heterogeneity within the RP patient population (see Figure III). I call the SRR obtained using postal area fixed effects srr_{POST}.

I need to acknowledge that patient outcomes may also be shaped by the hospital, its environment, and the quality of the care team (e.g., nurses and healthcare assistants). For this reason, I alternatively control for the hospital in which the surgeon operates. Many hospitals count multiple surgeons so that we can estimate the surgeon's random effect separately from the hospital's fixed effect. This reduces the variability in outcomes explained by the surgeon-specific random effect by 80 percent. This notable shift in the surgeon variance should be considered as the compound effect of accounting for both time-invariant differences in hospital quality and the characteristics of the patient population treated. Nevertheless, the estimated variance remains statistically significant, suggesting that hospital quality and patient characteristics account for only a portion of this variability. I call the SRR obtained using hospital fixed effects srr_{HOSP}.

If more severely ill patients are admitted to hospitals with higher quality surgeons, but my model fails to capture their level of illness accurately, my measure would be incorrect. Including hospital controls captures differences in outcomes that may result from this type of patient selection across healthcare providers. Still, it may be that within a hospital, more severely ill patients are assigned to the most skilled doctor. As a further check, I estimate the model using data on patients admitted via the emergency department (ED). The nature of emergency admissions allows me to exclude this type of selection conclusively: emergency patients are operated by whoever is on call and free at that time. In this case, however, I estimate the model using all surgical urology patients because RP is rarely done in an emergency setting. I call the SRR obtained using hospital fixed effects and estimated using ED patients srr_{ED}. Variation across surgeons in this sample is substantially lower, partly indicating that selection is less salient. This may, however, also result from the operations performed in emergency being less complex and, therefore, less likely to show off the surgeon's skills.

5 Econometric Strategy

In this Section, I present my parameter of interest and explain the assumptions I need to identify it within the MTE framework. I then describe the variable I use as an instrument and discuss its validity.

5.1 Parameter of Interest

Understanding whether the returns from using the robot depend on the surgeon's skills involves a significant challenge in terms of identification. On the one hand, I have endogeneity driven by the patient-surgeon's choice of technology. On the other hand, I have to consider that outcomes may systematically differ along the skill distribution because of patients' characteristics.

To formalize the problem, I assume there are two types of surgeons, high and low-skill, and abstract from any control variables. Ideally, I would want to estimate the following equation:

$$Y = \alpha + \beta D_R + \gamma D_H + \tau D_R D_H + \epsilon \tag{5}$$

where Y is the observed patient outcome (e.g., mortality), D_R is an indicator variable that equals one if a robot has been used, D_H indicates whether a high-skill surgeon operated on the patient, and ϵ is an idiosyncratic error term.

The parameter of interest is τ, the interaction between the skills and the robot, which captures the complementarity between the technology and the human capability. This parameter is identified by the following difference in average treatment effects:

$$\tau = E[\Delta_H - \Delta_L] \tag{6}$$

where $\Delta_H \equiv Y_H^R - Y_H^T$ is the difference in outcomes between robotic (R) and traditional surgery (T) for a high skill surgeon (H), and $\Delta_L \equiv Y_L^R - Y_L^T$ is the same difference for a low skill surgeon (L).

5.2 Identification Challenge

Estimating τ requires finding a sample equivalent of the difference in outcomes between robotic and traditional surgery, Δ_S, for each skill level $S \in (H, L)$.

If patients were randomly allocated to surgical approaches (traditional or robotic) and surgeons (high and low-skill), the following equality would hold :

$$\begin{aligned} \Delta_S \equiv E[Y_S^R - Y_S^T] &= E[Y_S^R | R_S] - E[Y_S^T | T_S] \\ &= E[Y|R_S] - E[Y|T_S] \ \forall \ S \in (H, L) \end{aligned} \tag{7}$$

In Equation 7, R_S and T_S indicate the type of surgery, robotic or traditional, performed by a surgeon of skills S. Because of random assignment, the outcome is independent of whether the patient actually receives one or the other surgical approach and an OLS regression would be enough to get an unbiased estimate of my parameter of interest from Equation 5.

In my context, however, treatment assignment is non-random, and patients may receive one or the other surgical approach because of their characteristics, some of which may be unobservable.[8] The possibility of systematic differences between robotic and traditional surgery patients may result in the OLS being biased:

$$E[Y|R_S] - E[Y|T_S] = E[\Delta_S|R_S] + \underbrace{E[Y_S^T|R_S] - E[Y_S^T|T_S]}_{\text{treatment selection bias}} \ \forall\, S \in (H, L) \tag{8}$$

Moreover, even in the absence of selection into the robotic treatment, comparing treatment effects across different surgeon types would generally be invalid, as surgeons of different skills operate on potentially different patients, i.e., the treated population would not be comparable across skills.

$$E[\Delta_H|R_H] - E[\Delta_L|R_L] = E[\Delta_H - \Delta_L|R_H] + \underbrace{E[\Delta_L|R_H] - E[\Delta_L|R_L]}_{\text{skill selection bias}} \tag{9}$$

In this case, the selection bias will depend on the distribution of patients across surgeons. For example, if high-skill surgeons operate on a population where the robot is particularly effective, we would overstate the complementarity between the robot and surgical skills. On the contrary, if this was true for the low-skill surgeons, we would underestimate the importance of skills for the technology returns.

[8] In this discussion, I refer to the robotic approach as the treatment. The untreated state refers to a state where the patient is operated with traditional surgery.

5.3 Identification Strategy

To deal with this complex identification problem, I model the patient-surgeon decision between alternative treatments as in Roy (1951). This is not the surgeon's decision to adopt the robot but should be rather thought of as the result of a consultation between the patient and the surgeon regarding the treatment option that best suits the patient. Within this model, I impose a standard conditional independence assumption between the observed and unobserved components that determine the returns from treatment, which allows me to compare treatment effects across skills. Under this assumption, I show that my parameter of interest is a trivial component of the MTE of receiving robotic relative to traditional surgery.

5.3.1 Discrete Treatment Choice

I take the choice of treatment (robot vs. traditional surgery) to be the result of a joint decision of the patient and the surgeon, which I represent in a generalized Roy model for discrete choices (Roy 1951, Heckman & Vytlacil 2007).

The model consists of two potential outcomes, Y^R and Y^T, and a binary indicator R that summarizes the treatment status. For notational simplicity, I omit the patient and surgeon identifier. Y^R denotes the potential outcome if the individual was treated with the robot ($R = 1$), and Y^R denotes the potential outcome if the individual was treated with traditional surgery ($R = 0$). The observed outcome Y can be written in switching regression form (Quandt 1958):

$$Y = R\, Y^R + (1 - R)\, Y^T \tag{10}$$

The potential outcomes, for which I impose a linear structure, depend on the surgeon's skills, S, and patient characteristics, X. For now, I take S to be a continuous variable.

$$\begin{aligned} Y^R &= \tau^R S + \mu^R X + U^R \\ Y^T &= \tau^T S + \mu^T X + U^T \end{aligned} \tag{11}$$

Combining Equations 12 and 11 we get that:

$$Y = \tau^T S + \mu^T X + R \underbrace{\left(\tau^R - \tau^T\right) S + \left(\mu^R - \mu^T\right) X + \left(U^R - U^T\right)}_{\Delta} + U^T \tag{12}$$

where the individual treatment effect Δ depends on both observed, S and X, and unobserved, U^T and U^R, components.

Patients and surgeons compare the outcomes of each treatment alternative and opt for the one with the highest surplus. As in Roy (1951), this process is represented by latent index crossing model, where the latent part in this selection equation is interpreted as the expected net utility of treatment (Vytlacil 2002).

$$R = 1[\gamma(Z) > V] \text{ where } Z = (S, X, Z^*) \tag{13}$$

Whether the patient is operated on with the robot may depend on all the same observable factors likely to influence the outcomes, including the surgeon's skills. The commonality of variables with the outcome equation explicitly allows treatment choice to be based on the patient's expected returns from treatment, which may vary depending on the surgeon. Z also contains at least one excluded continuous variable Z^* that affects the patient's probability of getting the treatment but does not directly impact the outcomes. I will discuss Z^* in Section 5.4.

V is an unobservable negative shock to the latent index, typically called unobserved resistance to or negative preference for treatment (Zhou & Xie 2019). V is assumed to have a continuous distribution so that we can rewrite Equation 13 as $R = 1[P(Z) > U^D]$, where U^D represents the quantiles of the distribution of V and $P(Z) \equiv Pr(R = 1|Z)$ is the propensity score. The individual gets the treatment whenever the propensity score $P(Z)$ exceeds U^D, which, by construction, has a uniform distribution in the population (Carneiro et al. 2011).

I do not restrict the relationship between U^R, U^T, and V, so the treatment choice may be endogenous to the unobserved components. However, I assume additive separability between observed and unobserved heterogeneity in treatment effects (Brinch et al. 2017).

Assumption 1 *Additive Separability,* $\forall\ S \in (H, L)$

$$E[U^R|S, X, U^D] = E[U^R|U^D]$$
$$E[U^T|S, X, U^D] = E[U^T|U^D]$$

Assumption 1 allows treatment effects to vary by the observable, S and X, and unobservable component, U^R and U^T, but not the interaction of the two terms. Although restrictive, this assumption is much looser than what is usually assumed by most of the instrumental variable literature (Brinch et al. 2017). This assumption excludes the possibility that selection across surgeons is systematically correlated with the unobserved returns from robotic surgery and allows for a comparison of treatment effects across surgeons' skills, conditional on the individual quantile of resistance to treatment U^D.

5.3.2 Marginal Treatment Effect

Under Assumption 1, and the structure of the Roy model, my parameter of interest τ is a trivial element of the Marginal Treatment Effect (Heckman & Vytlacil 1999, 2001, 2005, 2007).

$$\begin{aligned} MTE(s, x, u) &= E[Y^R \ - Y^T|\ S = s, X = x, U^D = u] \\ &= (\tau^R - \tau^T)\ s + (\mu^R - \mu^T)\ x + E[U^R - U^T \mid U^D = u] \qquad (14) \\ &= \Delta\tau s + \Delta\mu x + E[U^R - U^T \mid U^D = u] \end{aligned}$$

The MTE is the expected treatment effect conditional on observed covariates and the normalized latent variable U^D. For each quantile of resistance to treatment in the sample, some individuals would be treated, and some would not. The MTE arises then from comparing outcomes between treated and untreated individuals with equivalent observable characteristics and level of resistance to treatment.

This parameter, initially developed by Björklund & Moffitt (1987) and popularized by Carneiro et al. (2011), can be estimated by Local IV (LIV) (Carneiro et al. 2011). The intuition is simple. Increasing the propensity score by a small amount shifts previously indifferent individuals into treatment and induces a change in the observed outcome that can be attributed to the treatment effect, that is the marginal treatment effect. Alternatively, as suggested in Heckman & Vytlacil (2007), the MTE can be estimated using the separate approach, which has the benefit of estimating all the parameters of both the potential outcomes so that I can plot these over the distribution of the unobserved resistance to treatment. I follow this method and estimate the MTE using the separate approach under a normality restriction on the unobserved components (Andresen 2018).

5.4 Exogenous variation in treatment probability

The MTE framework requires at least one continuous variable included in the selection equation but excluded from the outcome equation (Heckman & Vytlacil 2005). The absence of a centralized strategy for the acquisition of robots in England resulted in significant disparities in access to robotic surgery, both across geography and over time, which I leverage to create this instrument.

5.4.1 Instrument Definition

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) employed the relative distance from a mother's home 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. The idea is that geographic distance randomizes patients to different likelihoods of receiving treatment but, unlike absolute distance, is not expected to be correlated with health or other factors affecting patient outcomes and treatments.

Inspired by this body of work, I use the differential distance from the patient's residence to a hospital capable of providing robotic surgery as an instrument. I refer to the differential distance instrument as Z_{dist}, and I compute it for each patient as:

$$Z_{dist} = d_R - d_T, \tag{15}$$

d_R is the geographic distance in kilometers between the patient and the nearest hospital with a robot when the patient is operated, and d_T is the geographic distance between the patient and the nearest hospital.

Data on where a patient lives in HES is limited to the postal area, but HES includes information on the patient's GP. Hence, I use the patient's GP's postcode to proxy for his location. In England, individuals must register with a GP to obtain a referral necessary to access non-emergency services from hospitals. Patients can only register at GP practices near their home address, so the GP's postcode is a good proxy for the patient's location.

5.4.2 Instrument Validity

As in Imbens & Rubin (1997), the instrument should only affect patients' outcomes through its effect on the probability of receiving robotic surgery. Firstly, I want to exclude that patients living relatively closer to a robotic hospital are just in better health than those living far away. Secondly, I want to rule out the possibility of 'correlated beneficial care', where the availability of the treatment of interest may be correlated with other hospital practices or characteristics that can affect health outcomes independently (Card et al. 2019, Doyle Jr et al. 2015, McClellan et al. 1994).

Unobserved Patient Heterogeneity: To mitigate the first concern, I include in the selection equation an extensive set of observable patient characteristics, but I also directly control for the area where the patient lives. In this way, I only leverage the instrument variation within narrowly defined geographic locations. This represents a significant step forward compared to previous studies utilizing this instrument, for which typically relative distance is constant in time.

As illustrated in Figure VI, the average relative distance for my sample of prostate cancer patients was 61 kilometers in 2007, declining to 25 in 2012, and further diminishing to 7 by 2017. This trend arises from hospitals' gradual adoption of robotic technology, which induces changes in relative distance for individuals residing in the same area depending on the date of their cancer diagnosis.

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). 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 & Arthur 2004). This means that AMI mortality can serve as a proxy for both individuals' health and physical well-being and for economic and social risk factors.

Table V presents the results for this test. The coefficients are estimated 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 admitted via the emergency department from 2009 to 2014. As in Card et al. (2019), I control for the distance from the nearest hospital to the patient's residence to account for disparities in care accessibility. Standard errors are clustered at the patient postal area level. Relative distance to a robotic hospital is not statistically significant for AMI patients outcomes regardless of the set of controls in the model, suggesting no systematic association with any health determinant. This remains true regardless of whether I use relative distance as a continuous predictor or as a dummy variable.

Correlated Beneficial Care: To address the possibility of correlated beneficial care, I show that a patient's relative distance to a hospital offering robotic surgery is only predictive of surgical outcomes when the hospital has, in fact, adopted the robot. In other words, I show that being relatively closer to a hospital that will eventually adopt the robot cannot predict an RP patient outcome from surgery. If robotic adoption was correlated with other beneficial practices, the patient relative distance to a hospital that adopts in $t+n$ should affect the outcomes of patients operated in t as long as n is small enough.

I implement this test as follows. For each RP patient operated in $t \leq 2010$, I compute Z_{dist}^{2010} that is their relative distance to a hospital offering robotic surgery n periods after. Then, I estimate the relationship between Z_{dist}^{2010} and the outcomes of patients in 2006, 2007, 2008, 2009, and 2010. Notice that for the patients in 2010, this is effectively a reduced form estimate of the relationship between robotic surgery and patient outcomes. I take $t \leq 2010$ as my test years because robotic surgery then was relatively scarce, and we can find many patients for which there was a significant change in relative distance in 2010. However, these years are close enough to 2010, so practices would not have changed significantly.

I present the results in Table VI. I find no statistically significant relationship between Z_{dist}^{2010} and the outcomes of patients in the preceding years. On the contrary, for 2010, the coefficient is positive and statistically significant, indicating for a unit increase in relative distance to a hospital that actually offers robotic surgery, we expect an increase in postoperative length of stay of 0.2 percent. Similarly, a unit increase in relative distance results in a 0.6 percent increase in the odds of experiencing an adverse event from surgery for the 2010 sample.

5.4.3 Relevance, Monotonicity, and Common Support

I establish that the instrument is relevant in Figure VII. In the left panel, I show a non-parametric representation of the relationship between the endogenous variable, robotic surgery, and the instrument, relative distance, conditional on year and postal area controls. In the right panel, I plot the predicted probability of robotic surgery from a logistic regression, whereas on the right side, I have relative distance, year, and postal area dummies.

What emerges is a clear negative correlation between the probability of robotic surgery and the instrument. The model predicts that when the patient's nearest hospital offers robotic surgery, he has a 60 percent probability of being operated with the robot. Having to travel 10 km more reduces this probability by 6 percentage points. For a relative distance above 100 km, the patient has a 10 percent probability of getting the robot. On average, a 1 km increase in the patient's relative distance decreases the probability of having robotic surgery by 0.004 percentage points.

For the instrument to be valid, there should also be no defiers (Imbens & Rubin 1997). In our context, it is actually unlikely that an individual would be more prone to receive traditional surgery when relatively closer to a hospital offering the robot. To corroborate that this is actually the case, I estimate the relationship between robotic surgery and the instrument for different subgroups of patients. I do this for individuals above and below the median age, the median Charlson Comorbidity Index (CCI), white individuals, and those from other ethnic backgrounds. I present the coefficients estimated using a logistic regression in Table VII. Reassuringly, Z_{dist} always has a negative coefficient, indicating that increasing the relative distance to a robotic hospital weakly decreases the likelihood of undergoing robotic surgery regardless of the group of patients I focus on.

Finally, the instrument should induce sufficient variation across observable characteristics to generate a propensity score, $P(Z)$, with full common support (i.e., support in the unit interval for treated and untreated individuals). This requirement is specific to the MTE framework and is needed to ensure the possibility of computing a reasonable estimate of the Average Treatment Effect (ATE). In Figure VIII, I present the unconditional support jointly generated by the instrument and covariates. The instruments create a common support in the estimated propensity score that spans virtually the full unit interval.

6 Results

6.1 Skills and technology gains

In Table VIII and IX, I give the estimates based on Equation 14 for my three alternative measures of surgical skills (i.e., the SRR), in both a continuous and binary version. The SRR, derived using the risk adjustment methodology from Section 4, should be understood as negatively correlated with skills, meaning that high-skill surgeons will have a lower SRR. The binary version uses the SRR to establish a cutoff between high and low-skill surgeons. Surgeons with an SRR above the median are considered low-skill, while the rest are considered high-skill.

In all specifications, I control for the patient demographic characteristics, 15 risk variables, hospital history, the patient postal area, and year of operation fixed effects. The controls are included both as stand-alone variables and interacted with the propensity score $P(Z)$ to allow for heterogeneity in treatment effects. The exception is the postal area, which I restrict to have the same effect in the treated and untreated state. The indicator of whether the patient undergoes robotic surgery is instrumented using Z_{dist} while also controlling for the distance of the nearest hospital to the patient residence.

The performance measure in Table VIII is a binary indicator of adverse events. Panel (A) presents the relationship between surgical skills and patient outcomes when the operation is performed with traditional surgery (i.e., τ^T in Equation 14). Panel (B) shows the coefficients on the variables interacted with the propensity score, which tell us how the treatment effects differ by skill level ($\Delta\tau$ in Equation 14) and speak directly to the possibility of skill bias.

In Panel (A) of Table VIII, the results suggest a clear positive relationship between the surgeon's skills, measured with the SRR, and performance in the untreated state. With traditional surgery, the surgeons I identify as high-skill with my risk-adjusted indicator perform better than the rest. For both srr_{POST} and srr_{HOSP}, a one standard deviation increase in the skill measure is associated with about a 3 percentage points change in adverse events. The coefficient on srr_{ED} is also positive but not statistically significant.

In Panel (B) of Table VIII, the results indicate that high-skill surgeons do not benefit significantly from using the robot, whereas low-skilled surgeons experience the most notable improvement in their performance. A higher SRR (lower skills) is associated with a greater performance improvement from using the robot. The coefficients on the continuous skill measure interacted with the propensity score are negative and statistically significant for srr_{POST} and srr_{HOSP}.

Figure IX provides a more intuitive representation of this result. In the figure, I plot the coefficient for the binary measures in the treated state next to the coefficient interacted with the propensity score. For the coefficient on ($\mathbb{1}[srr_{POST} < p(50)]$) in the untreated state, a switch from a high-skill to a low-skill surgeon induces a change in the probability of an adverse event from surgery of 6 percentage points. The difference in treatment effects with this skill measure is estimated to be about the same magnitude but positive so as to favor the low-skill surgeons. This implies almost no difference in the treated state (robotic surgery) between high and low-skill surgeons. The same result holds true when I focus on ($\mathbb{1}[srr_{HOSP} < p(50)]$). According to this skill measure, there is a 5 percentage point difference between low and high-skill surgeons in the untreated state. This difference is close to 0 in the treated state because of a difference in treatment effects of 0.4.

The same results hold in Table IX when I focus on postoperative length of stay as my performance measure. The coefficient interacted with the propensity score is negative for the continuous skill measure, indicating that low-skill surgeons gain the most from using the robot as the estimated difference in treatment effects implies a much stronger reduction in postoperative length of stay relative to their high-skill counterparts. The coefficient on the binary skill measure is consistent with this idea, although not statistically significant in all specifications.

Overall, the findings suggest a leveling effect of the technology. While significant differences exist between high and low-skill surgeons in traditional surgery, these gaps notably diminish with the robotic technology. This is true for both outcomes and irrespective of the skill measure I use in the analysis. The benefits derived from using the robot primarily accrue to low-skill surgeons, as shown by the positive difference in treatment effects. The gap between high and low-skill surgeons is much smaller when using the robot. This points to limited complementarities between the robotic technology and the skills required to perform traditional surgery successfully. There is no skill premium here, and the technology is biased toward the low-skill.

6.2 Selection equation

I have established that low-skill surgeons benefit the most from using the technology. Their returns from using the robot are higher both in terms of adverse events and postoperative length of stay. The selection equation suggests, however, that their use of the robot is much lower than that of their high-skill counterparts.

Table X presents the estimates from Equation 13, i.e., the selection equation. The dependent variable is a binary indicator of whether the patient has been operated with robotic surgery. The model controls for the patient demographic characteristics, 15 risk variables, hospital history, year, and postal area fixed effects. The instrument Z_{dist} and the patient distance to the nearest hospital are also included.

The estimates show that surgical skills significantly determine whether a patient undergoes robotic surgery. The relationship is statistically significant, but the odds ratio (i.e., the exponentiated coefficient) for the SRR is below 1, indicating that the likelihood of using the robot decreases as this variable increases. Conversely, for the binary skill measure, the odds ratio is above 1, indicating that patients of high-skill surgeons ($\mathbb{1}[srr_{HOSP} < p(50)]$) have higher odds of being operated on with the robot.

This result implies negative selection on gains, whereby those who have the highest return from the treatment, the patients of the low-skilled surgeons, are less likely to get it. Relative to their returns, low-skill surgeons appear to underuse the robot, limiting this technology's potential impact.

6.3 Returns to treatment based on unobserved characteristics

The estimated MTE curves under the assumption of joint normality of the error terms are shown in Figure X and XI. I do not find significant heterogeneity in effects, based on unobservable characteristics, for the adverse event outcome. The estimated marginal treatment effects, plotted at the mean of X in the sample, show a flat curve over the quantiles of resistance to treatment. For postoperative length of stay, I find a pattern of positive selection of gains. The individuals at the lowest quantile of resistance to treatment have the largest reduction in length of stay when using the robot. The separate approach shows that The estimated expected outcome with the robot is a flat line over the resistance to treatment; the robot is equalizing the outcomes of more and less resistant patients in the treated state.

6.4 Conventional and policy relevant treatment effects

I show the conventional treatment effects parameters in Table VIII and IX, which I compute by appropriately integrating over the MTE curve. The introduction of the robot significantly enhances the performance of surgeons. The Average Treatment Effect (ATE) is negative and statistically significant, indicating that, on average, the robot reduces postoperative morbidity by 10 percentage points and shortens the postoperative length of stay by about 45 percent. For adverse events, the Average Treatment on the Treated (ATT), the Average Treatment on the Untreated (ATUT), and the Local Average Treatment Effect (LATE) are all very similar to the ATE. Regarding postoperative length of stay, the ATT is more negative than the ATE, which is consistent with the idea of positive selection into treatment. The LATE parameter is even more negative, suggesting that individuals persuaded into robotic surgery by a change in the instrument experience substantial benefits from the treatment.

I have shown that the low-skill surgeons have the highest returns but use the technology the least. Hence, a natural question is how much are we losing from this behavior? As a conclusive exercise, I exploit the structure of the model to conduct a policy simulation that allows me to answer this question. Following Heckman & Vytlacil (2005), Carneiro et al. (2011), I consider a class of policies that change $P(Z)$, the probability that the patient is operated with the robot, but that do not affect the potential outcomes or the unobservable characteristics in the model. Heckman & Vytlacil (2005) show how to compute the Policy Relevant Treatment Effect (PRTE), the mean effect of going from the baseline policy to an alternative policy per net person shifted into treatment.

I estimate the PRTE for a counterfactual scenario in which I assign low-skill surgeons the same probability of using the robot as the high-skill surgeons. Basically, I evaluate the average performance effect of the robot if the low-skill surgeons were mandated to use the technology with the same intensity as the high-skill ones. This policy simulation speaks to a hypothetical counterfactual in which the barriers that limit the use of the robot by low-skill surgeons were lifted. Suppose surgeons use the robot on a patient only when the expected benefit from treatment is above a certain threshold, aligning with the principles of the Roy model of treatment choice. However, less skilled surgeons either misconstrue the benefits or apply an incorrect threshold, resulting in the underutilization of the treatment. Then, this policy counterfactual shows what would happen to the average treatment effect if the low-skill surgeons had the same threshold to use the robot as the high-skill surgeons.

The results of this exercise are shown in Table XI for both margins of performance. The PRTE is always more negative than the ATE, indicating that inducing low-skill surgeons to use the robot more intensively would generate larger gains from the adoption of robots. This is particularly evident for adverse events. For example, in Figure XII, I show the shift that would occur in the MTE curve from this policy when using srr_{POST} as my skill measure. The PRTE is almost five percentage points lower than the ATE I have estimated. This additional reduction in adverse events is lost because of the underutilization of the technology by the low skill surgeons.

7 Discussion and Conclusion

In this paper, I presented evidence that robotic surgery has a positive effect on surgeons' performance. Patient post-operative length of stay and morbidity diminish when the robot is used relative to traditional techniques. I have shown, however, that these effects are highly heterogeneous and significantly depend on the skills of the surgeon using it. The low-skill surgeons benefit the most from using the robot, while high-skill surgeons benefit the least. Thus, the technology appears to be skill-biased.

My interpretation is that the skill bias arises because the robot effectively takes over specific challenging tasks that were traditionally carried out by surgeons. For example, the task of holding the instruments is now left to the robot. By substituting the surgeon in these critical functions, the robot levels the playing field, enabling low-skill surgeons to achieve a performance closer to that of their high-skill counterparts. The relative advantage of high-skill surgeons diminishes, as the technology takes over the very tasks that previously set them apart.

However, even with the robot, significant differences persist between high and low-skill surgeons. They utilize the technology differently. Low-skill surgeons do not use the technology to its full potential and actually underuse it despite the significant benefits it offers to them. This behavior could stem from incorrect beliefs about the returns, a lack of comfort with the technology, insufficient training, or reluctance to fully integrate it into their practice for other reasons. Consequently, the benefits of this technology are not fully realized.

In a broader sense, the results of this paper suggest that the adoption of this technology in surgery will have important consequences for the future of work in this profession. The possibility of operating with the robot will likely change the profile of the individuals working in this job, as the skills traditionally required from surgeons appear to matter less for their performance. Importantly, this shift may also alter the trajectory of a surgeon's career. While a surgeon's professional journey historically hinged on their physical prowess and was constrained by the inevitable decline that accompanies aging, this dynamic may no longer hold true. Only time and research will tell.

References

Abaluck, J., Agha, L., Kabrhel, C., Raja, A. & Venkatesh, A. (2016), 'The determinants of productivity in medical testing: Intensity and allocation of care', *American Economic Review* **106**(12), 3730–64.

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

Acemoglu, D. & Autor, D. (2011), Skills, tasks and technologies: Implications for employment and earnings, *in* 'Handbook of labor economics', Vol. 4, Elsevier, pp. 1043–1171.

Acemoglu, D., Koster, H. R. & Ozgen, C. (2023), Robots and workers: Evidence from the netherlands, Technical report, National Bureau of Economic Research.

Acemoglu, D., Lelarge, C. & Restrepo, P. (2020), Competing with robots: Firm-level evidence from france, *in* 'AEA papers and proceedings', Vol. 110, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 383–388.

Acemoglu, D. & Restrepo, P. (2018), 'The race between man and machine: Implications of technology for growth, factor shares, and employment', *American economic review* **108**(6), 1488–1542.

Acemoglu, D. & Restrepo, P. (2019), 'Automation and new tasks: How technology displaces and reinstates labor', *Journal of economic perspectives* **33**(2), 3–30.

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

Acemoglu, D. & Restrepo, P. (2022), 'Tasks, automation, and the rise in us wage inequality', *Econometrica* **90**(5), 1973–2016.

Acemoglu, D. & Zilibotti, F. (2001), 'Productivity differences', *The Quarterly Journal of Economics* **116**(2), 563–606.

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

Autor, D. H., Katz, L. F. & Kearney, M. S. (2008), 'Trends in us wage inequality: Revising the revisionists', *The Review of economics and statistics* **90**(2), 300–323.

Autor, D. H., Katz, L. F. & Krueger, A. B. (1998), 'Computing inequality: have computers changed the labor market?', *The Quarterly journal of economics* **113**(4), 1169–1213.

Autor, D. H., Levy, F. & Murnane, R. J. (2003), 'The skill content of recent technological change: An empirical exploration', *The Quarterly journal of economics* **118**(4), 1279–1333.

Birkmeyer, J. D., Finks, J. F., O'reilly, A., Oerline, M., Carlin, A. M., Nunn, A. R., Dimick, J., Banerjee, M. & Birkmeyer, N. J. (2013), 'Surgical skill and complication rates after bariatric surgery', *New England Journal of Medicine* **369**(15), 1434–1442.

Björklund, A. & Moffitt, R. (1987), 'The estimation of wage gains and welfare gains in self-selection models', *The Review of Economics and Statistics* **69**(1).

Bolla, M., van Poppel, H., Tombal, B., Vekemans, K., Da Pozzo, L., De Reijke, T. M., Verbaeys, A., Bosset, J.-F., Van Velthoven, R., Colombel, M. et al. (2012), 'Postoperative radiotherapy after radical prostatectomy for high-risk prostate cancer: long-term results of a randomised controlled trial (eortc trial 22911)', *The Lancet* **380**(9858), 2018–2027.

Borghans, L., Green, F. & Mayhew, K. (2001), 'Skills measurement and economic analysis: an introduction', *Oxford Economic Papers* pp. 375–384.

Breg, N. (2022), Medical technologies with comparative advantages on different dimensions: Evidence from hysterectomy.

Brinch, C. N., Mogstad, M. & Wiswall, M. (2017), 'Beyond late with a discrete instrument', *Journal of Political Economy* **125**(4), 985–1039.

Brynjolfsson, E., Li, D. & Raymond, L. R. (2023), Generative ai at work, Technical report, National Bureau of Economic Research.

Card, D., Fenizia, A. & Silver, D. (2019), The health impacts of hospital delivery practices, Technical report, National Bureau of Economic Research.

Carneiro, P., Heckman, J. J. & Vytlacil, E. J. (2011), 'Estimating marginal returns to education', *American Economic Review* **101**(6).

Carneiro, P. & Lee, S. (2011), 'Trends in quality-adjusted skill premia in the united states, 1960–2000', *American Economic Review* **101**(6), 2309–2349.

Chan, D. C., Gentzkow, M. & Yu, C. (2022), 'Selection with variation in diagnostic skill: Evidence from radiologists', *The Quarterly Journal of Economics* **137**(2), 729–783.

Chandra, A. & Skinner, J. S. (2003), 'Geography and racial health disparities'.

Chandra, A. & Staiger, D. O. (2007), 'Productivity spillovers in health care: evidence from the treatment of heart attacks', *Journal of political Economy* **115**(1), 103–140.

Chandra, A. & Staiger, D. O. (2020), 'Identifying sources of inefficiency in healthcare', *The quarterly journal of economics* **135**(2), 785–843.

Choi, J. H. & Schwarcz, D. (2023), 'Ai assistance in legal analysis: An empirical study', *Available at SSRN 4539836* .

Coelho, R. F., Rocco, B., Patel, M. B., Orvieto, M. A., Chauhan, S., Ficarra, V., Melegari, S., Palmer, K. J. & Patel, V. R. (2010), 'Retropubic, laparoscopic, and robot-assisted radical prostatectomy: a critical review of outcomes reported by high-volume centers', *Journal of endourology* **24**(12), 2003–2015.

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

Cooper, Z., Gibbons, S., Jones, S. & McGuire, A. (2010), Does hospital competition improve efficiency? an analysis of the recent market-based reforms to the english nhs, Cep discussion papers, Centre for Economic Performance, LSE.
URL: *https://EconPapers.repec.org/RePEc:cep:cepdps:dp0988*

Coughlin, G. D., Yaxley, J. W., Chambers, S. K., Occhipinti, S., Samaratunga, H., Zajdlewicz, L., Teloken, P., Dunglison, N., Williams, S., Lavin, M. F. et al. (2018), 'Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: 24-month outcomes from a randomised controlled study', *The Lancet Oncology* **19**(8), 1051–1060.

Coulter, A. (2010), 'Do patients want a choice and does it work?', *Bmj* **341**.

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

Davies, B. (2022), 'Waiting for godot robot', *The Journal of Urology* **208**(4), 784–785.

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

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

Eckhoff, J. A., Müller, D. T., Brunner, S. N., Fuchs, H. F. & Meireles, O. R. (2023), 'Do the costs of robotic surgery present an insurmountable obstacle? a narrative review', *International Journal of Abdominal Wall and Hernia Surgery* **6**(2), 71–76.

Finkelstein, A., Gentzkow, M. & Williams, H. (2016), 'Sources of geographic variation in health care: Evidence from patient migration', *The quarterly journal of economics* **131**(4), 1681–1726.

Fowler Jr, F. J., Barry, M. J., Lu-Yao, G., Wasson, J., Roman, A. & Wennberg, J. (1995), 'Effect of radical prostatectomy for prostate cancer on patient quality of life: results from a medicare survey', *Urology* **45**(6), 1007–1015.

George, E. I., Brand, C. T. C., Marescaux, J. et al. (2018), 'Origins of robotic surgery: from skepticism to standard of care', *JSLS: Journal of the Society of Laparoendoscopic Surgeons* **22**(4).

Goldin, C. & Katz, L. F. (2018), The race between education and technology, *in* 'Inequality in the 21st Century', Routledge, pp. 49–54.

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

Graetz, G. & Michaels, G. (2018), 'Robots at work', *Review of Economics and Statistics* **100**(5), 753–768.

Heckman, J. J. & Vytlacil, E. (2001), 'Policy-relevant treatment effects', *American Economic Review* **91**(2), 107–111.

Heckman, J. J. & Vytlacil, E. (2005), 'Structural equations, treatment effects, and econometric policy evaluation', *Econometrica* **73**(3).

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

Heckman, J. J. & Vytlacil, E. J. (2007), 'Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation', *Handbook of econometrics* **6**, 4779–4874.

Higgins, R. M., Frelich, M. J., Bosler, M. E. & Gould, J. C. (2017), 'Cost analysis of robotic versus laparoscopic general surgery procedures', *Surgical endoscopy* **31**(1), 185–192.

Ho, C., Tsakonas, E., Tran, K., Cimon, K., Severn, M., Mierzwinski-Urban, M., Corcos, J. & Pautler, S. (2013), 'Robot-assisted surgery compared with open surgery and laparoscopic surgery: clinical effectiveness and economic analyses'.

Horn, D., Sacarny, A. & Zhou, A. (2022), 'Technology adoption and market allocation: The case of robotic surgery', *Journal of Health Economics* **86**, 102672.

Horwitz, L. I., Partovian, C., Lin, Z., Grady, J. N., Herrin, J., Conover, M., Montague, J., Dillaway, C., Bartczak, K., Suter, L. G. et al. (2014), 'Development and use of an administrative claims measure for profiling hospital-wide performance on 30-day unplanned readmission', *Annals of internal medicine* **161**, S66–S75.

Hu, J. C., Gold, K. F., Pashos, C. L., Mehta, S. S. & Litwin, M. S. (2003), 'Role of surgeon volume in radical prostatectomy outcomes', *Journal of Clinical Oncology* **21**(3), 401–405.

Hugosson, J., Stranne, J. & Carlsson, S. V. (2011), 'Radical retropubic prostatectomy: a review of outcomes and side-effects', *Acta oncologica* **50**(sup1), 92–97.

Humlum, A. (2019), 'Robot adoption and labor market dynamics', *Princeton University* .

Imbens, G. W. & Rubin, D. B. (1997), 'Estimating outcome distributions for compliers in instrumental variables models', *The Review of Economic Studies* **64**(4).

Jayant Ketkar, H., Carnahan, S. & Greenwood, B. N. (2022), 'Do new job tools improve women's performance in male-dominated fields? evidence from robotic surgery', *Evidence from Robotic Surgery (March 28, 2022)* .

Katz, L. F. & Murphy, K. M. (1992), 'Changes in relative wages, 1963–1987: supply and demand factors', *The quarterly journal of economics* **107**(1), 35–78.

Koch, M., Manuylov, I. & Smolka, M. (2021), 'Robots and firms', *The Economic Journal* **131**(638), 2553–2584.

Kolstad, J. T. (2013), 'Information and quality when motivation is intrinsic: Evidence from surgeon report cards', *American Economic Review* **103**(7), 2875–2910.

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V. & Violante, G. L. (2000), 'Capital-skill complementarity and inequality: A macroeconomic analysis', *Econometrica* **68**(5), 1029–1053.

Lam, K., Clarke, J., Purkayastha, S. & Kinross, J. M. (2021), 'Uptake and accessibility of surgical robotics in england', *The International Journal of Medical Robotics and Computer Assisted Surgery* **17**(1), 1–7.

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

Lowrance, W. T., Elkin, E. B., Jacks, L. M., Yee, D. S., Jang, T. L., Laudone, V. P., Guillonneau, B. D., Scardino, P. T. & Eastham, J. A. (2010), 'Comparative effectiveness of prostate cancer surgical treatments: a population based analysis of postoperative outcomes', *The Journal of urology* **183**(4), 1366–1372.

Marcus, H. J., Hughes-Hallett, A., Payne, C. J., Cundy, T. P., Nandi, D., Yang, G.-Z. & Darzi, A. (2017), 'Trends in the diffusion of robotic surgery: A retrospective observational study', *The International Journal of Medical Robotics and Computer Assisted Surgery* **13**(4), e1870.

Maynou, L., Mehtsun, W. T., Serra-Sastre, V. & Papanicolas, I. (2021), 'Patterns of adoption of robotic radical prostatectomy in the united states and england', *Health Services Research* .

Maynou, L., Pearson, G., McGuire, A. & Serra-Sastre, V. (2022), 'The diffusion of robotic surgery: Examining technology use in the english nhs', *Health Policy* **126**(4), 325–336.

McClellan, M., McNeil, B. J. & Newhouse, J. P. (1994), 'Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: analysis using instrumental variables', *Jama* **272**(11).

McClellan, M. & Newhouse, J. P. (1997), 'The marginal cost-effectiveness of medical technology: a panel instrumental-variables approach', *Journal of Econometrics* **77**(1), 39–64.

Mogstad, M., Santos, A. & Torgovitsky, A. (2018), 'Using instrumental variables for inference about policy relevant treatment parameters', *Econometrica* **86**(5), 1589–1619.

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

Nelson, B., Kaufman, M., Broughton, G., Cookson, M. S., Chang, S. S., Herrell, S. D., Baumgartner, R. G. & Smith, J. A. (2007), 'Comparison of length of hospital stay between radical retropubic prostatectomy and robotic assisted laparoscopic prostatectomy', *The Journal of urology* **177**(3), 929–931.

Neuner, J. M., See, W. A., Pezzin, L. E., Tarima, S. & Nattinger, A. B. (2012), 'The association of robotic surgical technology and hospital prostatectomy volumes: increasing market share through the adoption of technology', *Cancer* **118**(2), 371–377.

Noy, S. & Zhang, W. (2023), 'Experimental evidence on the productivity effects of generative artificial intelligence', *Science* **381**(6654), 187–192.

Peng, B., Galley, M., He, P., Cheng, H., Xie, Y., Hu, Y., Huang, Q., Liden, L., Yu, Z., Chen, W. et al. (2023), 'Check your facts and try again: Improving large language models with external knowledge and automated feedback', *arXiv preprint arXiv:2302.12813* .

Porcaro, A. B., Rizzetto, R., Amigoni, N., Tafuri, A., Shakir, A., Tiso, L., Cerrato, C., Antoniolli, S. Z., Lacola, V., Gozzo, A. et al. (2022), 'Severe intraoperative bleeding predicts the risk of perioperative blood transfusion after robot-assisted radical prostatectomy', *Journal of Robotic Surgery* **16**(2), 463–471.

Quandt, R. E. (1958), 'The estimation of the parameters of a linear regression system obeying two separate regimes', *Journal of the american statistical association* **53**(284), 873–880.

Robertson, C., Close, A., Fraser, C., Gurung, T., Jia, X., Sharma, P., Vale, L., Ramsay, C. & Pickard, R. (2013), 'Relative effectiveness of robot-assisted and standard laparoscopic prostatectomy as alternatives to open radical prostatectomy for treatment of localised prostate cancer: a systematic review and mixed treatment comparison meta-analysis', *BJU international* **112**(6), 798–812.

Roy, A. D. (1951), 'Some thoughts on the distribution of earnings', *Oxford economic papers* **3**(2).

Strother, M. C., Michel, K. F., Xia, L., McWilliams, K., Guzzo, T. J., Lee, D. J. & Lee, D. I. (2020), 'Prolonged length of stay after robotic prostatectomy: causes and risk factors', *Annals of Surgical Oncology* **27**(5), 1560–1567.

Suri, T. (2011), 'Selection and comparative advantage in technology adoption', *Econometrica* **79**(1), 159–209.

Tonutti, M., Elson, D. S., Yang, G.-Z., Darzi, A. W. & Sodergren, M. H. (2017), 'The role of technology in minimally invasive surgery: state of the art, recent developments and future directions', *Postgraduate Medical Journal* **93**(1097).

Vytlacil, E. (2002), 'Independence, monotonicity, and latent index models: An equivalence result', *Econometrica* **70**(1).

Yaxley, J. W., Coughlin, G. D., Chambers, S. K., Occhipinti, S., Samaratunga, H., Zajdlewicz, L., Dunglison, N., Carter, R., Williams, S., Payton, D. J. et al. (2016), 'Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study', *The Lancet* **388**(10049), 1057–1066.

Zeira, J. (1998), 'Workers, machines, and economic growth', *The Quarterly Journal of Economics* **113**(4), 1091–1117.

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

Tables

Table I: Characteristics of radical prostatectomy patients

	(1) Traditional				(2) Robotic			
	mean	sd	min	max	mean	sd	min	max
Demographic Characteristics								
Age	63.570	6.847	16	98	62.865	6.617	20	82
White	0.740	0.439	0	1	0.644	0.479	0	1
Rural-urban indicator	5.388	0.933	1	9	5.434	1.047	1	9
Risk Factors								
Acute myocardial infarction	0.020	0.140	0	1	0.024	0.153	0	1
Congestive heart failure	0.007	0.081	0	1	0.006	0.078	0	1
Peripheral vascular disease	0.012	0.110	0	1	0.012	0.110	0	1
Cerebrovascular disease	0.010	0.101	0	1	0.011	0.104	0	1
Dementia	0.001	0.033	0	1	0.001	0.029	0	1
Chronic obstructive pulmonary disease	0.089	0.285	0	1	0.105	0.307	0	1
Rheumatoid disease	0.011	0.103	0	1	0.013	0.113	0	1
Peptic ulcer disease	0.010	0.102	0	1	0.014	0.116	0	1
Mild liver	0.006	0.075	0	1	0.011	0.102	0	1
Diabetes	0.074	0.262	0	1	0.082	0.275	0	1
Diabetes + complications	0.003	0.056	0	1	0.003	0.057	0	1
Hemiplegia or paraplegia	0.002	0.041	0	1	0.002	0.044	0	1
Renal disease	0.016	0.125	0	1	0.016	0.125	0	1
Moderate/severe liver disease	0.001	0.026	0	1	0.001	0.029	0	1
Metastatic cancer	0.013	0.114	0	1	0.019	0.135	0	1
AIDS	0.000	0.011	0	1	0.001	0.024	0	1
Hospital History								
Year of operation	2010		2004	2018	2014		2006	2018
Number hospital admissions at time of operation	2.614	4.057	0	370	3.152	4.736	0	279
Number of diagnosis at time of operation	2.636	1.907	1	20	2.948	1.977	1	20
Days since diagnosis at time of operation	80.477	247.694	0	4862	115.708	337.408	0	5328
Waiting time	44.196	37.195	1	1123	36.446	30.809	1	1072
Patient Outcomes								
In hospital death (30-days)	0.001	0.031	0	1	0.000	0.018	0	1
Emergency Readmission (30-days)	0.001	0.038	0	1	0.001	0.025	0	1
Urinary complication (2-years)	0.170	0.376	0	1	0.078	0.268	0	1
Erectile dysfuction (2-years)	0.016	0.125	0	1	0.010	0.099	0	1
Blood transfusion	0.009	0.095	0	1	0.000	0.012	0	1
Any adverse event	0.191	0.393	0	1	0.086	0.281	0	1
Length of stay (LOS)	4.406	3.620	0	124	1.935	1.784	0	60
Postoperative LOS	3.939	3.324	0	123	1.801	1.722	0	60
N	38389				29278			

Note: The sample includes all RP Patients operated by NHS hospitals in England from April 2004 to April 2017. Patient risk factors are identified using all inpatient admissions preceding the admission for RP surgery.

Table II: Balancing of characteristics between traditional and robotic surgery patients

	(1) Traditional	(2) Robotic	(3) $\hat{\Delta}$	(4) β	(5) se
Demographic Characteristics					
Age	63.570	62.865	-0.074	-1.511***	(0.169)
White	0.740	0.644	-0.148	-0.107***	(0.025)
Rural-Urban Indicator	5.388	5.434	0.033	0.025	(0.044)
Risk Factors					
Acute myocardial infarction	0.020	0.024	0.019	-0.009***	(0.002)
Congestive heart failure	0.007	0.006	-0.005	-0.004***	(0.001)
Peripheral vascular disease	0.012	0.012	0.000	-0.005***	(0.001)
Cerebrovascular disease	0.010	0.011	0.005	-0.004**	(0.001)
Dementia	0.001	0.001	-0.006	-0.001**	(0.000)
Chronic obstructive pulmonary disease	0.089	0.105	0.039	-0.016***	(0.003)
Rheumatoid disease	0.011	0.013	0.014	-0.003**	(0.001)
Peptic ulcer disease	0.010	0.014	0.021	-0.002**	(0.001)
Mild liver disease	0.006	0.011	0.039	-0.000	(0.001)
Diabetes	0.074	0.082	0.021	-0.016***	(0.004)
Diabetes + complications	0.003	0.003	0.002	-0.001*	(0.001)
Hemiplegia or Paraplegia	0.002	0.002	0.005	-0.000	(0.000)
Renal disease	0.016	0.016	0.000	-0.011***	(0.002)
Moderate/severe liver disease	0.001	0.001	0.004	-0.000	(0.000)
Metastatic Cancer	0.013	0.019	0.030	-0.003	(0.002)
AIDS	0.000	0.001	0.017	0.000	(0.000)
Hospital History					
Number hospital admissions at time of operation	2.614	3.152	0.086	-0.577***	(0.094)
Number of diagnosis at time of operation	2.636	2.948	0.114	-0.386***	(0.068)
Days since diagnosis at time of operation	80.477	115.708	0.084	0.216	(7.977)
Waiting time	44.196	36.446	-0.160	-5.516***	(1.380)
Patient Outcomes					
In hospital death (30-days)	0.001	0.000	-0.017	-0.000	(0.000)
Emergency Readmission (30-days)	0.001	0.000	-0.018	-0.000	(0.000)
Urinary complication (2-years)	0.170	0.078	-0.200	-0.046***	(0.006)
Erectile dysfunction (2-years)	0.016	0.010	-0.038	-0.004	(0.003)
Blood Transfusion	0.009	0.000	-0.094	-0.002**	(0.001)
Any adverse event	0.191	0.086	-0.217	-0.052***	(0.006)
Length of stay (LOS)	4.406	1.935	-0.612	-1.256***	(0.101)
Post-operative LOS	3.939	1.801	-0.571	-1.257***	(0.097)

Note: The sample includes all RP Patients operated by NHS hospitals in England from April 2004 to April 2017. Colum (3) shows the normalized difference between traditional and robotic surgery. Column (4) shows the coefficient from a regression of the row variable on an indicator of robotic approach with year fixed effects. Standard errors from the regression are shown in Column (4) and clustered at the patient postal area level.

Table III: Estimated between surgeon variance component from risk adjustment model

	RP Patients			Urology Emergency Patients		
	(1)	(2)	(3)	(4)	(5)	(6)
Between Surgeons Variance $\widehat{\sigma}^2$	0.592***	0.362***	0.114**	0.099***	0.080***	0.046**
	(0.095)	(0.036)	(0.035)	(0.019)	(0.017)	(0.014)
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes
Postal Area Fixed Effects	No	Yes	No	No	Yes	No
Hospital Fixed Effects	No	No	Yes	No	No	Yes
N	12486	12273	12477	124372	123335	124356

Note: Sample includes patients operated by NHS hospitals in England from 2004 to 2007. Columns 1, 2, and 3 include only RP patients. Columns 4, 5, and 6 include all urology patients admitted for surgery from the emergency department. The dependent variable in the risk adjustment model is an indicator of adverse events from surgery. For urology patients, this includes only 30 days in-hospital deaths and emergency readmissions. For RP patients, adverse events also include complications from surgery arising within 2 years of the operation. Baseline controls include patient demographic characteristics, 15 risk variables, hospital history variables, and year-fixed effects. The variance component is estimated using a multilevel mixed effects logistic regression. The conditional distribution of the response given the surgeon random effect is assumed to be Bernoulli, with the probability of adverse outcome determined by the logistic cumulative distribution function. Robust standard errors in parenthesis. * Statistically significant 0.10 level, ** at 0.05 level,*** at 0.01 level.

Table IV: Summary of alternative measures of skills

	sample	controls	mean	sd	min	max
srr_{POST}	RP patients	postal area fixed effects	0.941	0.308	0.366	2.875
srr_{HOSP}	RP patients	hospital fixed effects	0.959	0.164	0.600	1.804
srr_{ED}	Emergency patients	hospital fixed effects	1.012	0.086	0.752	1.275

Note: SRR estimated using mixed effects logistic regression with surgeon random intercept. All models include controls for patient demographic characteristics, 15 risk variables, and year-fixed effects. For RP patients, controls also include patient hospital history. For emergency urology patients, controls also include sex. The model's outcome for RP patients is a binary indicator of adverse events, including complications from surgery within 2 years, death, and emergency readmissions within 30 days of discharge. The model's outcome for emergency urology patients is a binary indicator of adverse events, including death and emergency readmissions within 30 days of discharge. All models are estimated using data from 2004 to 2007.

Table V: In-hospital death for AMI patients and relative distance to robotic hospital

	(1)	(2)	(3)	(4)	(5)	(6)
Z_{dist}	-0.0004	-0.001	-0.0004			
	(0.001)	(0.001)	(0.001)			
$Z_{dist} > 25$ km				0.005	-0.018	0.021
				(0.037)	(0.038)	(0.032)
Distance to nearest hospital	-0.003**	-0.003*	-0.001	-0.003**	-0.003*	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	No	Yes	Yes	No	Yes	Yes
Postal Area Fixed Effects	No	No	Yes	No	No	Yes
Mean in-hospital death	0.05	0.05	0.05	0.05	0.05	0.05
Mean Z_{dist}	38.07	38.07	38.07	38.07	38.07	38.07
N	385431	385431	385401	385431	385431	385401

Note: Sample includes patients admitted from the emergency department with AMI diagnosis code in England from April 2009 to April 2014. Baseline controls include age, sex, indicator of ethnically white, indicator of rural or urban area of residence, and day of the week. Columns (1), (2), and (3) have relative distance as a continuous variable. Columns (4), (5), and (6) compare outcomes of AMI patients above and below the median relative distance (25 km). The estimates are from a logistic regression with standard errors clustered at the patient postal area level. * Statistically significant 0.10 level, ** at 0.05 level,*** at 0.01 level.

Table VI: Patient outcomes and relative distance to robotic hospital in 2010

	(1) 2006	(2) 2007	(3) 2008	(4) 2009	(5) $\leq$ 2009	(6) 2010
Independent variable Z_{dist}^{2010}						
Panel A: Postoperative length of stay						
	0.001	0.000	0.000	0.002	0.0004	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Panel B: Adverse event from surgery						
	0.002	-0.0004	-0.001	0.004	0.001	0.006***
	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)
N	1399	1258	1111	815	7536	4384

Note: Sample includes RP patients operated by NHS hospitals in England. Panel A shows the estimates from a linear regression of log post-operative length of stay on Z_{dist} as of 2010. Panel B shows the estimates from a logistic regression of a binary indicator of an adverse event from surgery on Z_{dist} as of 2010. The columns indicate the patient's year of operation. Controls include age, ethnicity, and 15 comorbidity variables. Standard errors are clustered at the patient postal area level. * Statistically significant 0.10 level, ** at 0.05 level,*** at 0.01 level.

Table VII: Test for monotonicity of relative distance instrument

	$Age < 64$	$Age \geq 64$	$CCI < 2$	$CCI \geq 2$	$White = 1$	$White = 0$
Z_{dist}	-0.003***	-0.003***	-0.003***	-0.003***	-0.002***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	31169	27284	47786	10667	17057	41396

Note: The sample includes all RP Patients operated by NHS hospitals in England from April 2007 to April 2017. Estimates from logistic regression where the dependent variable is an indicator of the robotic approach. Controls include patient postal area and year fixed effects. Standard errors are clustered at the patient postal area level. * Statistically significant 0.10 level, ** at 0.05 level,*** at 0.01 level.

Table VIII: Estimated effects — Dependent variable indicator of adverse event from surgery

	(1)	(2)	(3)	(4)	(5)	(6)
Panel (A): τ^T						
srr_{POST}	0.119***					
	(0.018)					
srr_{HOSP}		0.174***				
		(0.025)				
srr_{ED}			0.0550			
			(0.060)			
$\mathbb{1}[srr_{POST} < p(50)]$				-0.0648***		
				(0.009)		
$\mathbb{1}[srr_{HOSP} < p(50)]$					-0.0483***	
					(0.009)	
$\mathbb{1}[srr_{ED} < p(50)]$						0.0138
						(0.011)
Panel (B): $\Delta\tau$						
$srr_{POST} * P(Z)$	-0.110***					
	(0.020)					
$srr_{HOSP} * P(Z)$		-0.121***				
		(0.036)				
$srr_{ED} * P(Z)$			0.000661			
			(0.063)			
$\mathbb{1}[srr_{POST} < p(50)] * P(Z)$				0.0601***		
				(0.014)		
$\mathbb{1}[srr_{HOSP} < p(50)] * P(Z)$					0.0437***	
					(0.013)	
$\mathbb{1}[srr_{ED} < p(50)] * P(Z)$						-0.0223
						(0.013)
Panel (C)						
ATE	-0.104**	-0.116***	-0.112***	-0.108**	-0.107***	-0.110**
	(0.032)	(0.032)	(0.030)	(0.036)	(0.031)	(0.034)
ATT	-0.107**	-0.126***	-0.116***	-0.112**	-0.115***	-0.114**
	(0.034)	(0.035)	(0.031)	(0.038)	(0.034)	(0.035)
ATUT	-0.102***	-0.109***	-0.109***	-0.105**	-0.102***	-0.106**
	(0.031)	(0.031)	(0.030)	(0.036)	(0.031)	(0.034)
LATE	-0.105*	-0.126**	-0.116***	-0.0973	-0.120**	-0.115**
N	31189	31189	39471	31189	31189	39471

Note: Estimates from the outcome equation where the dependent variable is a binary indicator of adverse event from surgery. In Panel (A) the coefficients measure the effects of skills on the outcome in the untreated state (τ_0 in Equation 14). In Panel (B) the coefficients interacted with the propensity score measure the difference in effects between the treated and untreated state ($\Delta\tau$ in Equation 14). Controls not displayed include age, 15 risk variables, indicator of ethnically white, hospital history variables, distance to the closest hospital, year, and patient postal area fixed effects. The year and the postal area fixed effects are not interacted with the propensity score. The propensity score is estimated using a probit regression of the treatment indicator (robot=1) on the control and the instrument Z_{dist}. Bootstrapped standard errors clustered at the patient postal area level are reported in parentheses. * Statistically significant 0.10 level, ** at 0.05 level,*** at 0.01 level.

Table IX: Estimated effects — Dependent variable postoperative length of stay (log)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel (A): τ^T						
srr_{POST}	0.121**					
	(0.041)					
srr_{HOSP}		0.221**				
		(0.069)				
srr_{ED}			0.622**			
			(0.208)			
$\mathbb{1}[srr_{POST} < p(50)]$				-0.107***		
				(0.027)		
$\mathbb{1}[srr_{HOSP} < p(50)]$					-0.0972***	
					(0.029)	
$\mathbb{1}[srr_{ED} < p(50)]$						-0.0887**
						(0.030)
Panel (B): $\Delta\tau$						
$srr_{POST} * P(Z)$	-0.128					
	(0.070)					
$srr_{HOSP} * P(Z)$		-0.0740				
		(0.111)				
$srr_{ED} * P(Z)$			-0.883***			
			(0.252)			
$\mathbb{1}[srr_{POST} < p(50)] * P(Z)$				0.0810*		
				(0.037)		
$\mathbb{1}[srr_{HOSP} < p(50)] * P(Z)$					0.0762	
					(0.042)	
$\mathbb{1}[srr_{ED} < p(50)] * P(Z)$						0.114**
						(0.040)
Panel (C)						
ATE	-0.454***	-0.476***	-0.418***	-0.475***	-0.463***	-0.433***
	(0.083)	(0.080)	(0.076)	(0.090)	(0.080)	(0.072)
ATT	-0.498***	-0.527***	-0.469***	-0.519***	-0.513***	-0.487***
	(0.092)	(0.090)	(0.084)	(0.104)	(0.091)	(0.080)
ATUT	-0.425***	-0.441***	-0.377***	-0.445***	-0.429***	-0.388***
	(0.079)	(0.075)	(0.069)	(0.082)	(0.074)	(0.067)
LATE	-0.546***	-0.573***	-0.507***	-0.550***	-0.580***	-0.508***
	(0.105)	(0.100)	(0.085)	(0.114)	(0.110)	(0.079)
N	30637	30637	38817	30637	30637	38817

Note: Estimates from the separate method for the outcome equation where the dependent variable is log post-operative length of stay. In Panel (A) the coefficients measure the effects of skills on the outcome in the untreated state (τ^T in Equation 14). In Panel (B) the coefficients interacted with the propensity score measure the difference in effects between the treated and untreated state ($\Delta\tau$ in Equation 14). Controls not displayed include age, 15 risk variables, indicator of ethnically white, hospital history variables, distance to the closest hospital, year, and patient postal area fixed effects. The year and the postal area fixed effects are not interacted with the propensity score. The propensity score is estimated using a probit regression of the treatment indicator (robot=1) on the control and the instrument Z_{dist}. Bootstrapped standard errors clustered at the patient postal area level are reported in parentheses. * Statistically significant 0.10 level, ** at 0.05 level,*** at 0.01 level.

Table X: Estimates from selection equation

	(1)	(2)	(3)	(4)	(5)	(6)
Z_{dist}	-0.033***	-0.033***	-0.034***	-0.034***	-0.033***	-0.034***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)
srr_{POST}	-1.616***					
	(0.400)					
srr_{HOSP}		-1.439				
		(0.755)				
srr_{ED}			-3.696*			
			(1.485)			
$\mathbb{1}[srr_{POST} < p(50)]$				1.384***		
				(0.221)		
$\mathbb{1}[srr_{HOSP} < p(50)]$					0.444*	
					(0.180)	
$\mathbb{1}[srr_{ED} < p(50)]$						0.917***
						(0.245)
Share robotic surgery	0.403	0.403	0.403	0.403	0.403	0.403
Mean Z_{dist}	29.355	29.355	29.355	29.355	29.355	29.355
Mean skill measure	0.942	0.958	1.011	0.643	0.549	0.555
N	31189	31189	31189	31189	31189	31189

Note: Probit regression estimates exponentiated coefficients. Controls include patient demographic characteristics, 15 risk variables, hospital history variables, patient postal area, distance to the nearest hospital, and year-fixed effects. Standard errors are clustered at the patient postal area level. * Statistically significant 0.10 level, ** at 0.05 level,*** at 0.01 level.

Table XI: Estimated ATE and PRTE

	Adverse Event from surgery			Postoperative length of stay		
	srr_{POST}	srr_{HOSP}	srr_{ED}	srr_{POST}	srr_{HOSP}	srr_{ED}
ATE	-0.116***	-0.104***	-0.112***	-0.454***	-0.476***	-0.418***
	(0.014)	(0.014)	(0.013)	(0.021)	(0.021)	(0.019)
PRTE	-0.131***	-0.136***	-0.113***	-0.495***	-0.486***	-0.438***
	(0.015)	(0.015)	(0.013)	(0.022)	(0.022)	(0.020)
N	30637	30637	38817	31189	31189	39471

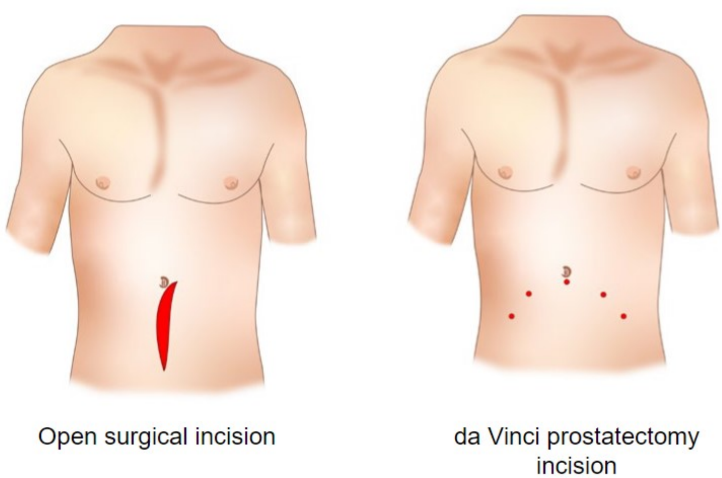
Note: Standard errors clustered at the patient postal area level are reported in parentheses. * Statistically significant 0.10 level, ** at 0.05 level,*** at 0.01 level.

Figures

Figure I: Picture of a Da Vinci surgical system

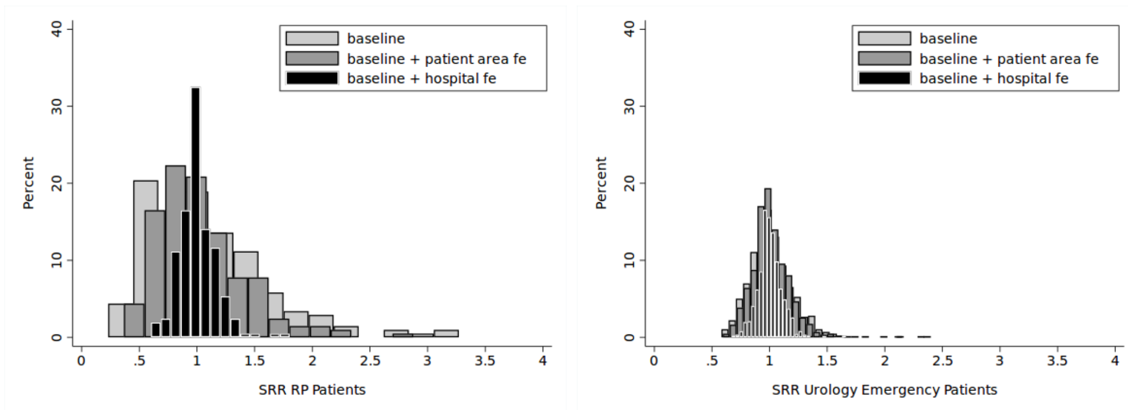
Note: Picture shows the Da Vinci Robot surgical system from Intuitive Inc. On the left the surgeon sitting at the console. Above the operating bed the robotic arms. On the right the vision cart.

Figure II: Comparison of incisions



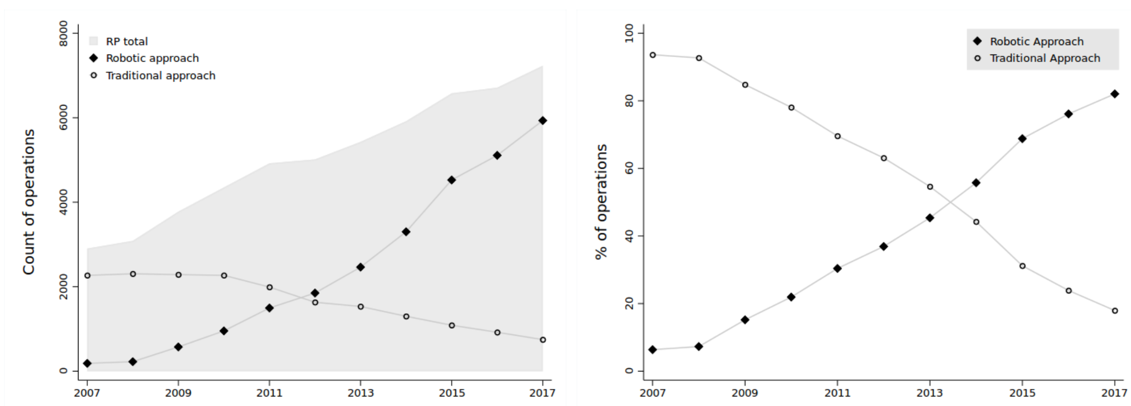
Note: Comparison of incisions required for traditional and robotic radical prostatectomy

Figure III: Comparison of skill measures



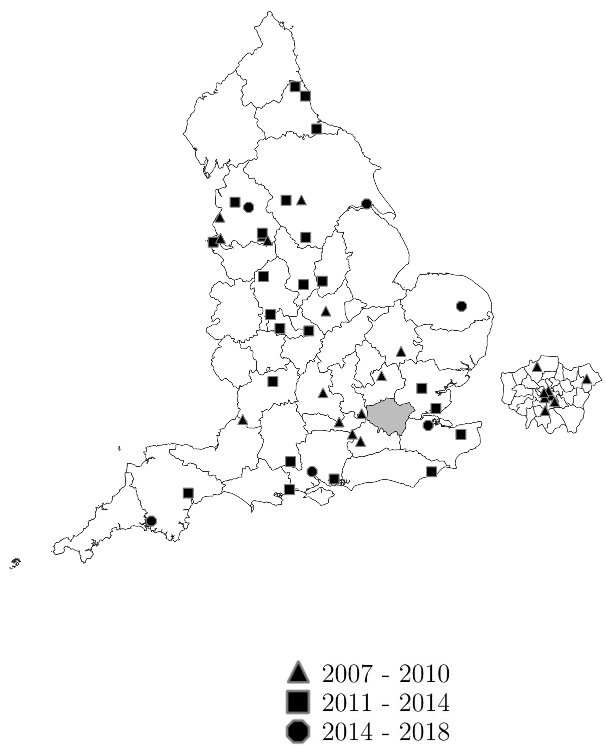
Note: The panel on the left shows the distribution of SRR computed using RP patients from 2004 to 2007 under three different model specifications for the risk adjustment. The panel on the right shows the distribution of SRR computed using emergency urology patients from 2004 to 2007 under three different model specifications for the risk adjustment.

Figure IV: Uptake of robotic radical prostatectomy in England



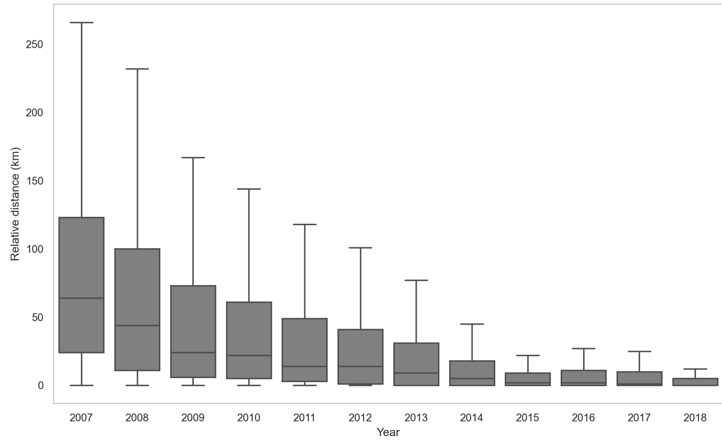
Note: The Figure plots the total number (left) and share (right) of RP by surgical approach from 2007 to 2017. The data shows that the use of robotic surgery for RP in England grew from 5 percent in 2007 to 80 percent in 2017. The steady increase in the number of robotic operations coincided with a decrease in the number of traditional surgeries. Traditional approach includes laparoscopies.

Figure V: Hospital adoption of robotic surgery for RP patients



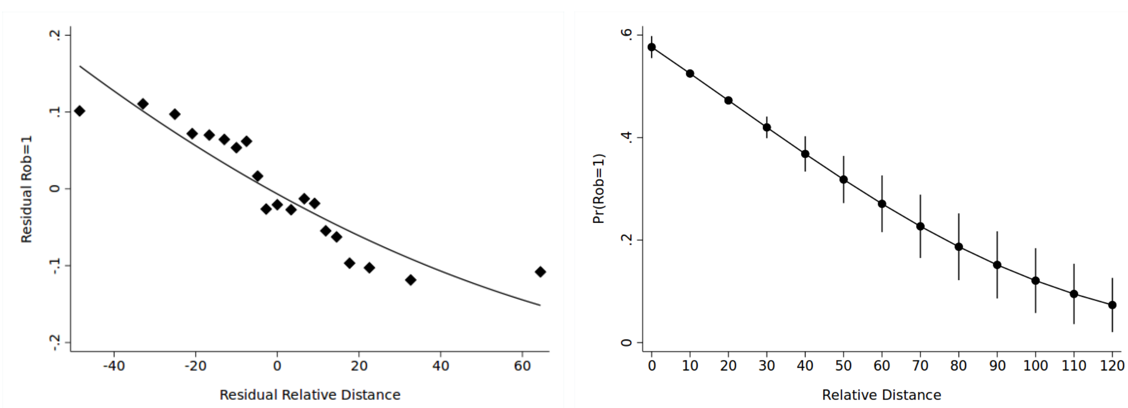
Note: The figure shows the location of trusts observed using a robot for RP surgery in the observation period. The trusts are divided into bins 2007-2010, 2011-2014, 2015-2018 according to the time they are first observed using the robot.

Figure VI: Relative distance (km) distribution over time



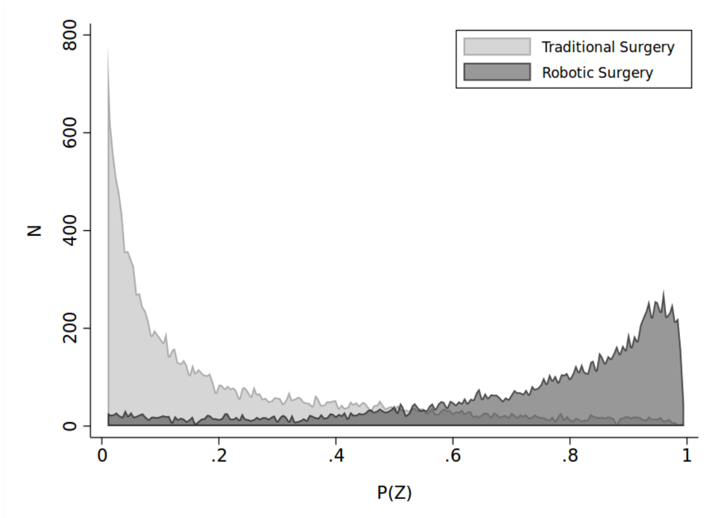
Note: The figure shows the median and interquartile range of relative distance by year. Outside values are excluded. Data includes all patients undergoing a radical prostatectomy in England for the period April 2007 to April 2018.

Figure VII: Relationship between relative distance and robotic approach



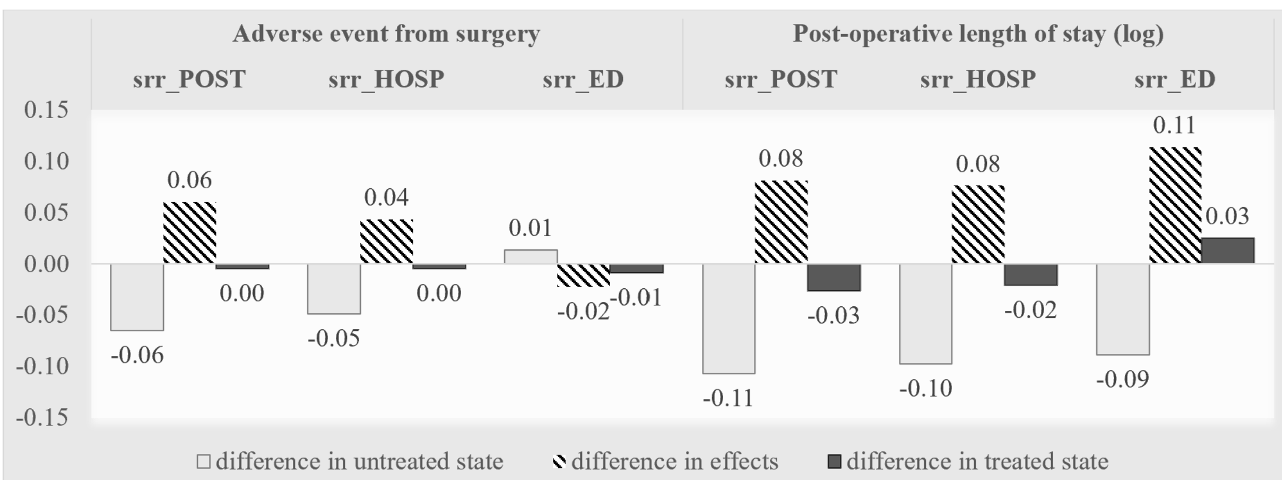
Note: The panel on the left shows a binscatter plot where the dependent variable is an indicator of whether the patient is operated on with the robot, and the independent variable of interest is the patient's relative distance to a hospital offering robotic surgery. The variables are residualised based on the year of operation and the patient's postal area. The panel on the right shows the average predicted probability of having robotic surgery at different values of relative distance. The margins are calculated from a logistic regression, where the controls are the year of operation and the patient's postal area. Standard errors are clustered at the postal area level. The sample includes all RP patients operated on by NHS hospitals in England from April 2007 to April 2018.

Figure VIII: Common support representation



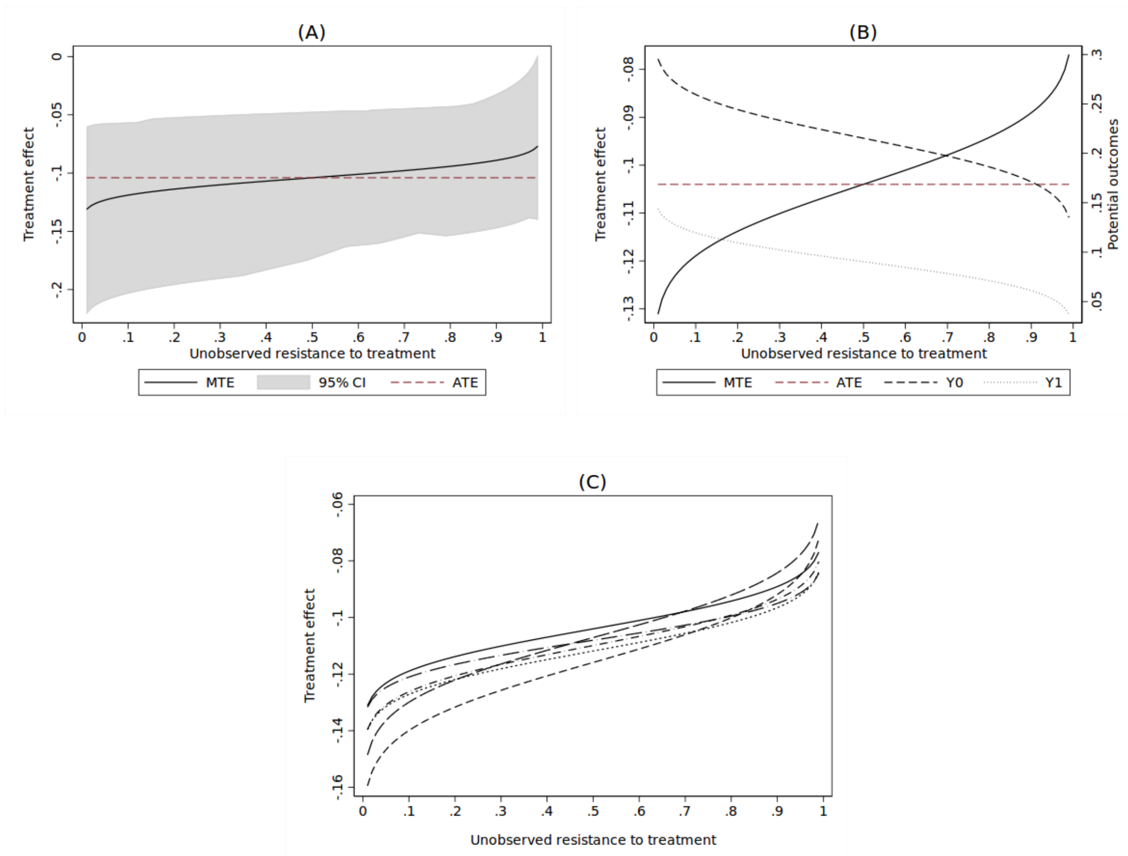
Note: Unconditional support jointly generated by instrument and covariates. Covariates in the model include patient demographic characteristics, 15 risk variables, hospital history variables, distance to the nearest hospital, patient postal area, and year-fixed effects.

Figure IX: Representation of effects for binary indicator of skills



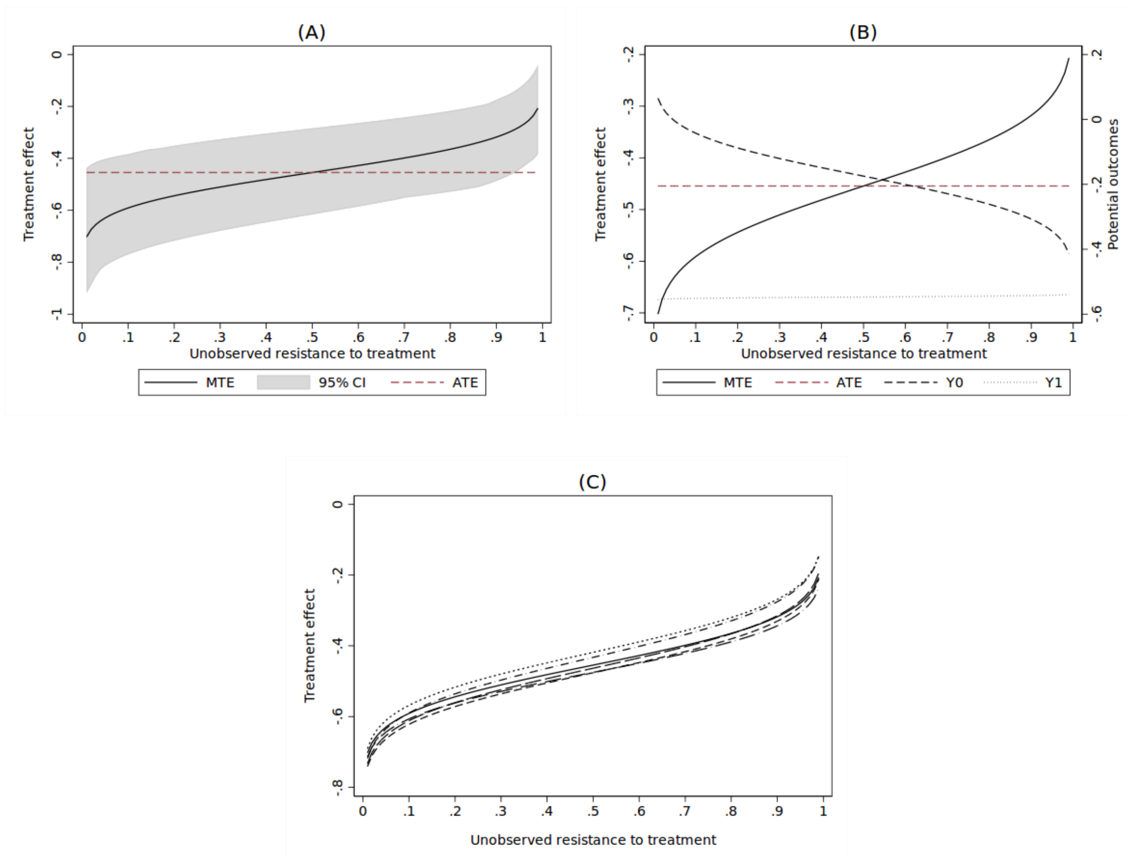
Note: The figure shows the coefficients obtained from estimating the MTE for my three alternative skills measures and both patient outcomes. The gray bar shows the difference in the untreated state (traditional surgery) between high- and low-skilled surgeons. This is the coefficient on the indicator of skills in the outcome equation (Equation 14). Low-skilled surgeons are defined as those with a SR above the median. The dashed bar represents the difference between high- and low-skilled surgeons in treatment effect. This is the coefficient on the indicator of skills interacted with the propensity score in the outcome equation (Equation 14). The black bar shows the implied difference in the treated state between high and low skilled surgeons.

Figure X: Marginal treatment effect curve - Indicator of adverse event from surgery



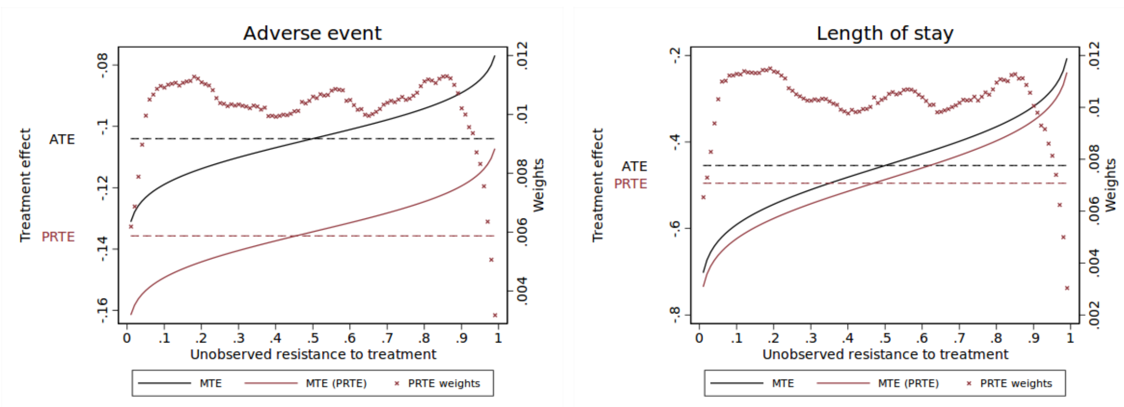
Note: Marginal treatment effects curve of robotic surgery from the separate method using srr_{POST} as the skill measure in the model for (A) and (B). In (C), I compare the MTE curve for my skill measures in both continuous and binary versions. The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p, parametrically under the assumption of $K(p)$ is normal. All specifications use the instrument Z_{dist} as the excluded variable, and control for patient demographic characteristics, 15 risk variables, hospital history variables, distance to the nearest hospital, patient postal area, and year-fixed effects. Area fixed effects are not interacted with propensity score. Standard errors are bootstrapped with 100 repetitions and clustered at the postal area level.

Figure XI: Marginal treatment effect curve — Postoperative length of stay (log)



Note: Marginal treatment effects curve of robotic surgery from the separate method using srr_{POST} as the skill measure in the model for (A) and (B). In (C), I compare the MTE curve for my skill measures in both continuous and binary versions. The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Gray bands are 95% confidence intervals. Unobserved heterogeneity, modeled as a function of the propensity score, p, parametrically under the assumption of $K(p)$ is normal. All specifications use the instrument Z_{dist} as the excluded variable, and control for patient demographic characteristics, 15 risk variables, hospital history variables, distance to the nearest hospital, patient postal area, and year-fixed effects. Area fixed effects are not interacted with propensity score. Standard errors are bootstrapped with 100 repetitions and clustered at the postal area level.

Figure XII: Policy relevant treatment effect



Note: Estimates of marginal treatment effects of robotic surgery, as opposed to traditional surgery, on the probability of an adverse event and log postoperative length of stay. The horizontal axis in each plot is the percentile on the distribution of unobserved resistance to robotic choice. Unobserved heterogeneity is modeled as a function of the propensity score, p, parametrically under the assumption of joint normality in the separate approach. The estimated effects of the policy simulation are in orange. Crosses indicate the weights