Technology, Skills, and Performance:

The Case of Robots in Surgery

Elena Ashtari Tafti
Ludwig Maximilian University of Munich

December 7, 2023

NBER Economics of Health, Fall 2023

Motivation

- **Usually we think of machines/capital that substitute labor**
  - 'Labor will become less and less important...More and more workers will be replaced by machines.', *Leontief (1952)*
  - especially true when it comes to robots *Acemoglu and Restrepo (2020), Humlum (2019), Acemoglu and Restrepo (2018)*

- **Increasingly, the machine actually acts as an extension of the user**
  - *e.g.* chatbot for professional writing tasks and coding, human-vehicle interaction in automated driving
  - these technologies hold promise to improve the way workers perform tasks
  - but, requires the user to know and use the machine appropriately

- **Joint production of human and machine**
  - How does the technology affect workers' performance?
  - How do these workers' characteristics affect the technology's adoption and performance?
  - How is the distribution of performance affected?

This paper

This paper

- **Individual and robot working together**
  - no replacement effect
  - complex task and highly trained individuals
  - the focus is on surgical performance/ patient outcomes

- **Who benefits from using the technology?**
  - treatment effects heterogeneity
  - the ‘robot’ treatment
  - surgeon’s skills as the key source of heterogeneity

- **Instrumental Variable**
  - to account for essential heterogeneity, *Heckman, Urzua, and Vytlacil (2006)*
  - extended relative distance instrument *McClellan, McNeil, Newhouse (1994)*
  - taking advantage of scattered adoption of robots

- **Marginal Treatment Effects**
  - recovers a distribution of treatment effects
  - building block of conventional causal parameters
  - policy simulations to explore alternative allocations of the technology

Why healthcare?

- **Technology's performance and providers' behavior consequential for people's lives**

- **Technology is financed with public income and under government scrutiny**
  - *e.g. UK government has pledged £2.4bn to implement AI, robotics*
  - Significant focus on cost vs. benefits to determine whether a technology is approved *e.g. NICE's technology appraisal (TA) in the UK, Certificate of Need laws that affect hospital capital in the US*

- **Unwarranted differences in outcomes for observationally similar patients**
  - uneven distribution of inputs across places and providers of services *e.g.*, *Finkelstein, Gentzkow, and Williams (2021), Chandra, Kakany, and Sacarny (2020)*
  - difference in providers' skills can exacerbate this phenomenon, *e.g. Chan, Gentzkow, and Yu (2021) Currie and MacLeod (2017) Chandra and Staiger (2007)*
  - "Machines" may guarantee a consistent and effective delivery, *Weber (1921)*

Preview of results

- **The robot improves the performance of surgeons**
  - focus is prostate cancer surgery
  - lower post-operative morbidity, *-10 percentage points*
  - lower post-operative length of stay in hospital, *more than half a day*

- **Large degree of heterogeneity in effects by surgeon's skills**
  - lower skilled surgeons have the highest returns, *difference in effects of 11 percentage points for morbidity*
  - robots reduce the gap in performance between high and low-skilled
  - results robust to various measures of pre-robot performance

- **But, negative selection on gains**
  - lower skilled have the highest returns but use the robot the least, *high skilled surgeons are 58 percentage points more likely to use the robot*
  - evidence of barriers that limit the potential benefit
  - policy simulation to estimate the missed gains from selection, *for morbidity additional 6 percentage points reduction on average*

Outline

1. **Empirical Setting**
2. **Data**
3. **Empirical Strategy**
    - Measuring Surgical Skills
    - Instrumental Variable
    - Marginal Treatment Effect
4. **Results**
    - Outcomes
    - Selection
    - Policy Simulation
5. **Conclusive Remarks**

Empirical Setting

The Technology: Surgical Robot

- less invasive procedure, and higher precision
- *'But often, technology spreads long before investigators know whether it is worthwhile.'*, *The New York Times on robotic surgery*
- Robot is costly to purchase and to maintain (£1.7 million pounds fixed and £140,000/year for maintenance)

The Operation: Prostate Cancer Surgery

- most common cancer in men in the United Kingdom
- standard but complex operation
- documented variation in surgeon's skills
- most common robotic procedure (about 80%)

The Market: English National Health Service

- healthcare is free at point of use
- after 2007 individuals are free to choose the hospital they want to visit
- decision to buy the robot left to the individual provider, *Lam and Clarke (2021)*

Data

- **Hospital records from all public hospital in England**
  - Hospital Episodes Statistics (HES)
  - 2004 to 2017
  - $\sim$ 60000 radical prostatectomies (RP)

- **Patient clinical and demographic data**
  - age, ethnicity, area of residence
  - diagnosis, treatment, and operations
  - hospital and surgeon identifiers

- **Information on surgical approach (robotic vs traditional)**
  - allows tracking diffusion across hospitals and over time show it
  - three years pre-robots to evaluate surgeons in the absence of the technology

- **Two dimensions of performance over which to evaluate the robot**
  - directly linked to surgeons ability
  - post-operative length of stay in hospital
  - post-operative morbidity (*e.g.*, complications from surgery)

Measuring Skills Pre-Robots

- I need to measure skills to evaluate whether the effects of robots depend on them
    - in this context, using education level is not an option
    - challenging as different surgeons may operate on different populations

Measuring Skills Pre-Robots

- I need to measure skills to evaluate whether the effects of robots depend on them
  - in this context, using education level is not an option
  - challenging as different surgeons may operate on different populations

- Using data **pre-robots** from 2004 to 2007, I risk adjust surgeons outcomes
  - idea is to relate surgeon j's predicted to expected post-morbidity experienced by prostate cancer patients

Measuring Skills Pre-Robots

- I need to measure skills to evaluate whether the effects of robots depend on them
  - in this context, using education level is not an option
  - challenging as different surgeons may operate on different populations

- Using data **pre-robots** from 2004 to 2007, I risk adjust surgeons outcomes
  - idea is to relate surgeon j's predicted to expected post-morbidity experienced by prostate cancer patients
  - **random coefficient regression** with a surgeon intercept $\omega_j \sim \mathcal{N}(0, \tau^2)$

$$\begin{aligned} Pr(Y_{ij} = 1) = & \quad F(\alpha_j + \beta X_{ij}) \\ \alpha_j = & \ \mu + \omega_j \end{aligned}$$

  - **standardized risk ratio** as in Horwitz et al. (2014) will measure skills pre-robot

$$\widehat{\textbf{Skills}}_j = \frac{\sum_{i \in j} f\left(\widehat{\alpha}_j + \widehat{\beta} X_{ij}\right)}{\sum_{i \in j} f\left(\widehat{\mu} + \widehat{\beta} X_{ij}\right)}$$

- X_{ij} includes patient characteristics and surgeon's volume together with time-fixed effects

Measuring Skills Pre-Robots

- I need to measure skills to evaluate whether the effects of robots depend on them
  - in this context, using education level is not an option
  - challenging as different surgeons may operate on different populations

- Using data **pre-robots** from 2004 to 2007, I risk adjust surgeons outcomes
  - idea is to relate surgeon *j*'s predicted to expected post-morbidity experienced by prostate cancer patients
  - **random coefficient regression** with a surgeon intercept $\omega_j \sim \mathcal{N}(0, \tau^2)$

$$Pr(Y_{ij} = 1) = F(\alpha_j + \beta X_{ij})$$
$$\alpha_j = \mu + \omega_j$$

  - **standardized risk ratio** as in *Horwitz et al. (2014)* will measure skills pre-robot

$$\widehat{\textbf{Skills}}_j = \frac{\sum_{i \in j} f\left(\widehat{\alpha}_j + \widehat{\beta} X_{ij}\right)}{\sum_{i \in j} f\left(\widehat{\mu} + \widehat{\beta} X_{ij}\right)}$$

- X_{ij} includes patient characteristics and surgeon's volume together with time-fixed effects

- Is this enough to account for unobserved heterogeneity?
  - in this period, no choice of provider for the patient
  - further validated using outcomes from ER urology patients

Heterogeneity in skills pre-robots

Figure 1: Distribution of skills

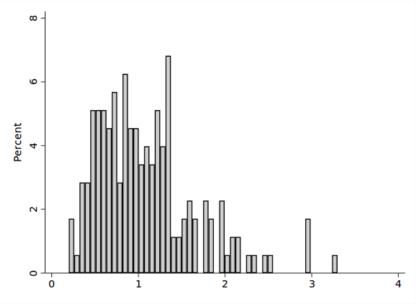
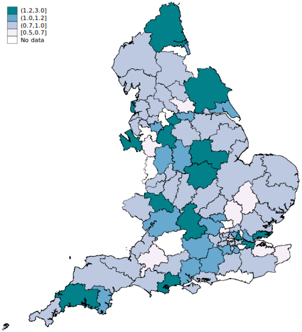


Figure 2: Geographic distribution



Note: Heterogeneity in surgical skills proxied by standardized risk ratio. Ratio above 1 indicates the surgeons is under performing. Ratio below 1 indicates the surgeon is over performing. Data for estimation is HES inpatient data for England from 2004 to 2007.

Identification: set up

- Treatment effect of using the robot

$$Y_{ij} = Robot\, Y_{ij}^1 + (1 - Robot)\, Y_{ij}^0 = Y_{ij}^0 + \underbrace{(Y_{ij}^1 - Y_{ij}^0)}_{\Delta} Robot$$

$$\Delta = \underbrace{X_{ij}(\beta_1 - \beta_0) + Skills_j(\delta_1 - \delta_0)}_{\text{Observed}} + \underbrace{\epsilon_{ij}^1 - \epsilon_{ij}^0}_{\text{Unobserved}}$$

- Whether to use the robot can be modeled in a latent variable framework

$$Robot = 1[R_{ij}^* \geq 0]$$

$$R_{ij}^* = \rho X_{ij} + \mu Skills_j - V_{ij}$$

- Identification problem:

correlation of ϵ_{ij}^{Δ} with V_{ij} or selection into treatment based on unobservables

Identification: set up

- Treatment effect of using the robot

$$Y_{ij} = Robot\, Y_{ij}^1 + (1 - Robot)\, Y_{ij}^0 = Y_{ij}^0 + \underbrace{(Y_{ij}^1 - Y_{ij}^0)}_{\Delta} Robot$$

$$\Delta = \underbrace{X_{ij}(\beta_1 - \beta_0) + Skills_j(\delta_1 - \delta_0)}_{\text{Observed}} + \underbrace{\epsilon_{ij}^1 - \epsilon_{ij}^0}_{\text{Unobserved}}$$

- Whether to use the robot can be modeled in a latent variable framework

$$Robot = 1[R_{ij}^* \geq 0]$$

$$R_{ij}^* = \rho X_{ij} + \mu Skills_j + \gamma Z_{ij} - V_{ij}$$

- Identification problem:

  correlation of ϵ_{ij}^{Δ} with V_{ij} or selection into treatment based on unobservables

- Solution to this problem:
  1. $(\epsilon_{ij}^0, \epsilon_{ij}^1, V_{ij})$ are statistically independent of Z_{ij} conditional on covariates
  2. $\gamma \neq 0$ (*Rank condition*)
  3. ϵ_{ij}^{Δ} does not depend on covariates conditional on quantile of distribution of V_{ij}

Instrumental variable: definition

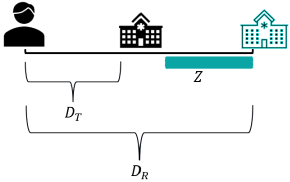
Patient's residence relative distance to robotic hospital

$$Z = D_R - D_T$$

graph

D_R := km to closest hospital for robotic surgery

D_T := km to closest hospital for traditional surgery



- *Validity*: should affect outcomes only through probability of treatment
  - control for km to closest hospital
  - postal area fixed effects
  - falsification test using heart attack patients ✓ show it
- *Relevance & Monotonicity* ✓ show it
- *Common Support* ✓ show it

Instrumental variable: source of exogenous variation

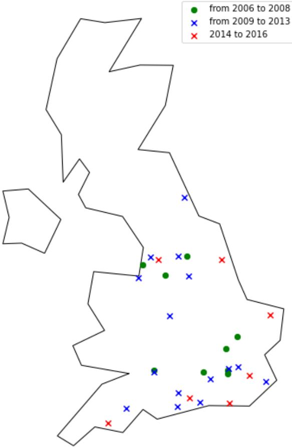
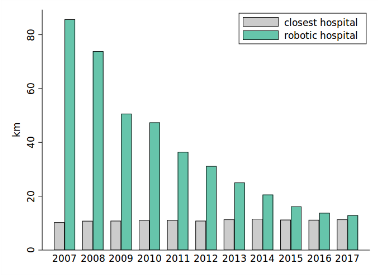


Figure 3: Distance to robot



Staggered adoption $\Longrightarrow$ variation in treatment probability across areas and over time

The Marginal Treatment Effect

- Allows estimating the distribution of treatment effects in the population *Heckman, Urzua, and Vytlacil (2006) and Carneiro, Heckman, and Vytlacil (2011)*
  - goes beyond LATE as it does not depend on the instrument
  - imposes no restrictions on the relationship between ϵ_0, ϵ_1 and V
  - actually allows unobserved returns to be associated with the unobserved resistance V

- The **MTE** is the treatment effect for i with observed characteristics x at u-th quantile of the distribution of V

$$\begin{aligned} MTE(s,x,u) = \quad & E(Y_{1ij} - Y_{0ij} | X_{ij} = x, Skills_j = s, U_{ij} = u) \\ = \quad & x\beta^{\Delta} + s\delta^{\Delta} + \underbrace{E(\epsilon^{\Delta} | U_{ij} = u)}_{\text{K(p)}} \end{aligned}$$

- **The derivative of the outcome Y w.r.t. p identifies the MTE**

$$\frac{\partial E[Y_{ij} | X_{ij} = x, Skills_j = s, P(Z_i) = p]}{\partial p} = x\beta^{\Delta} + s\delta^{\Delta} + \frac{\partial K(p)}{\partial p}$$

  - estimated using Local IV *Heckman (1999)*
  - with parametric or semi-parametric assumptions on $\partial K(p)/\partial p$

Results

Skills and technological gains

Table 1: MTE - Baseline specification (Coefficients) - Normal Model

	Selection equation	Morbidity	Length of stay
δ_0			
Z_{dist}	-0.013***		
	(0.001)		
High Skilled Dummy	0.487***	-0.089***	-0.145***
	(0.044)	(0.011)	(0.019)
$\delta_1 - \delta_0$			
High Skilled Dummy * P(Z)		0.119***	0.287***
		(0.016)	(0.026)
Patient Controls	Yes	Yes	Yes
Patient Postal Area	Yes	Yes	Yes
Year*Month	Yes	Yes	Yes
N	19698	19698	19453
ymean	0.538	0.12	0.7
mean Z_{dist}	20	20	20

Standard errors bootstrapped with 100 repetitions $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Column(1) dependent variable binary indicator of robotic surgery. Estimated using Probit regression model. Patient controls include age, age squared, indicator for white ethnic profile, ten comorbidity variables, and distance to the closest hospitals. Estimation of coefficients under the assumption of normality of unobserved components. Model estimated using postal area and year month fixed effects, not interacted with the propensity score.

linear experience

Skills and technological gains

Figure 4: Selection on skills

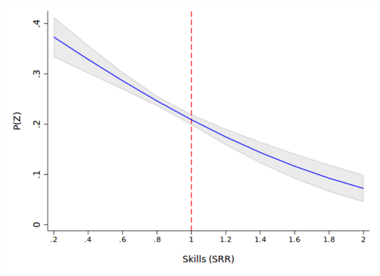
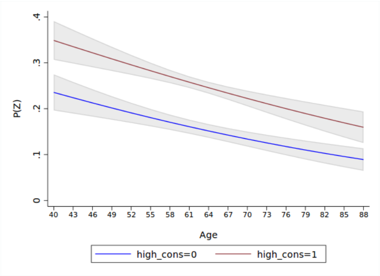


Figure 5: Age and skills



* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bootstrapped standard errors with 100 repetitions. Controls for age, indicator for white ethnic profile, ten comorbidity variables, distance to the closest hospital, postal area, and year-month fixed effects. Joint normality assumption. Fixed effects are not interacted with propensity score. Mean length of stay 2.5, with 13 percent of patients experiencing complications.

Selection on unobservables

Figure 6: Morbidity

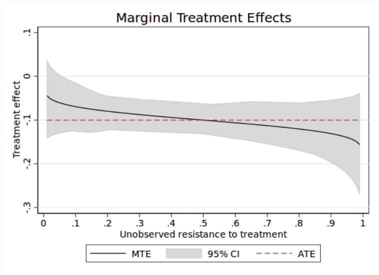
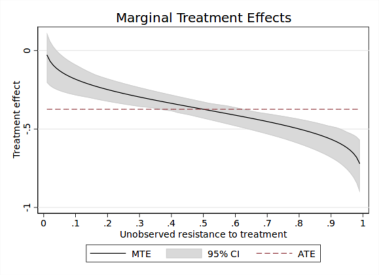


Figure 7: Length of stay



* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bootstrapped standard errors with 100 repetitions. Controls for age, indicator for white ethnic profile, ten comorbidity variables, distance to the closest hospital, postal area, and year-month fixed effects. Joint normality assumption. Fixed effects are not interacted with propensity score. Mean length of stay 2.5, with 13 percent of patients experiencing complications.

polynomial back

Treatment Effects

Table 2: Estimated Treatment Effects

	(1) Morbidity	(2) Length of stay
ATE	-0.100***	-0.374***
	(0.021)	(0.028)
ATT	-0.085***	-0.310***
	(0.021)	(0.030)
ATUT	-0.118***	-0.449***
	(0.025)	(0.033)
LATE	-0.124***	-0.383***
	(0.021)	(0.032)
N	19698	19453

Standard errors bootstrapped with 100 repetitions $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Patient controls include age, age squared, indicator for white ethnic profile, ten comorbidity variables, and distance to the closest hospitals. Model estimated using postal area and year month fixed effects, not interacted with the propensity score.

Robotic surgery: policy simulation

- The structure of the model can be used to simulate policies
  - Mean effect of changing to an alternative policy that provides different incentives to participate in treatment
    (Heckman and Vytlacil, 2001, 2005)
  - The policy relevant treatment effect (PRTE) measures the average effect of switching from a status-quo policy to a counterfactual policy under consideration.

- How much are we losing from surgeons' behavior?
  - Negative selection suggests we could do better
  - We can test a policy that mandates low-skilled to use the robot as much as the high-skilled

Robotic surgery: policy simulation

Figure 8: Morbidity

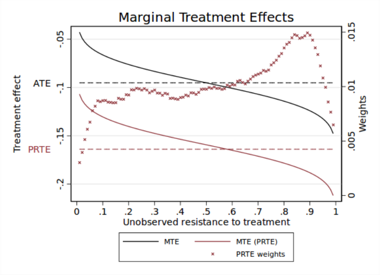
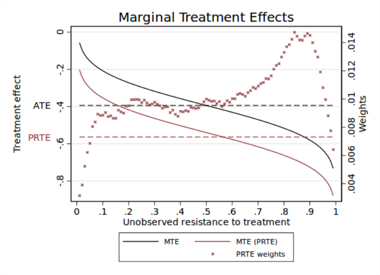


Figure 9: Length of stay



* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bootstrapped standard errors with 100 repetitions. Controls for age, indicator for white ethnic profile, ten comorbidity variables, distance to the closest hospital, postal area, and year-month fixed effects. Joint normality assumption. Fixed effects are not interacted with propensity score. Mean length of stay 2.5, with 13 percent of patients experiencing complications.

Conclusive Remarks

- Good News
  - the technology improves outcomes of patients
  - the robot has an equalizing effect, mostly due to low-skilled surgeons having more significant treatment effects
  - potential to reduce variation in performance and unwarranted differences in outcomes

- Bad News
  - negative selection on gains, low-skilled surgeons use the robot at a rate that impedes significant part of the gains
  - points to the existence of some barrier (actual or perceived) to adoption
  - potential improvements from increasing access is economically relevant

- Future research
  - welfare considerations
  - a model to explain negative selection
  - estimate allocative inefficiency

Thanks!

Diffusion of Robotic Surgery

Figure 10: Intensity of use

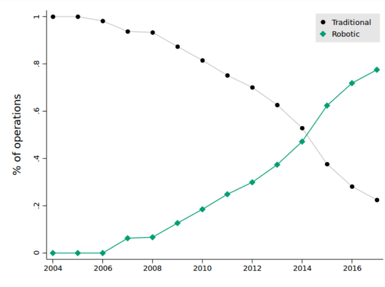
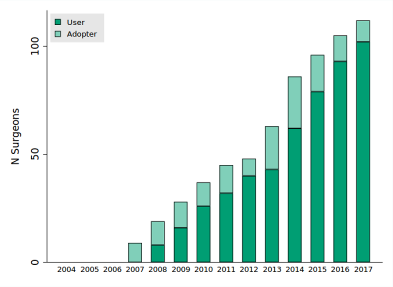


Figure 11: Adoption



back

Instrumental variable exclusion restriction

- Relationship to outcomes of patients unaffected by the robot
- Heart attack patients in England 2005-2009
- High mortality patients
- Outcomes strongly correlated with social and economic indicators

Table 3: AMI death and relative distance

	(1)	(2)	(3)
Z_{dist}	0.000449**	-0.000159	0.000314
	(0.000163)	(0.000185)	(0.000199)
Distance closest hospital		0.00269*	0.00111
		(0.00107)	(0.00130)
Year-month	No	Yes	Yes
Day of the week	No	Yes	Yes
Patient characteristics	No	No	Yes
Mortality rate (%)	19	19	19
Relative distance (km), Z_{dist}	68.64	68.64	68.75
N	68467	68467	67882

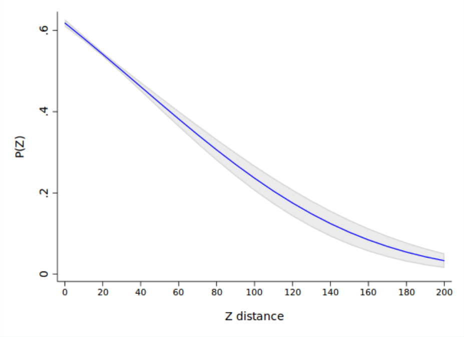
$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$ Standard errors in parentheses. Patient characteristics include age, age squared, ethnicity, rural-urban indicator, ten comorbidity dummies (e.g., malignant neoplasm, diabetes). Sample of AMI patients from 2005 to 2009. Source HES inpatient admissions data for NHS hospital in England.

back

Instrumental variable relevance

- Probit model dependent variable indicator of robotic approach
- Estimated average marginal effects

Figure 12: Relative distance relevance

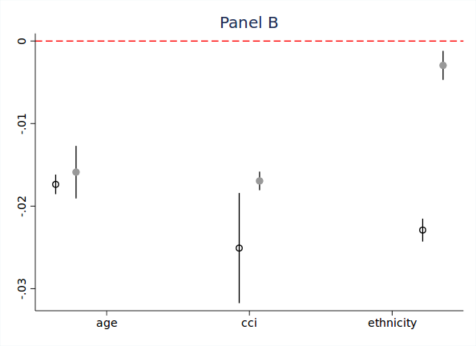


back

Instrumental variable monotonicity

- Assumption of no-defiers
- Run the first stage for different subgroups to provide suggestive evidence
- OLS regression dependent variable is an indicator of treatment

Figure 13: Relative distance monotonicity

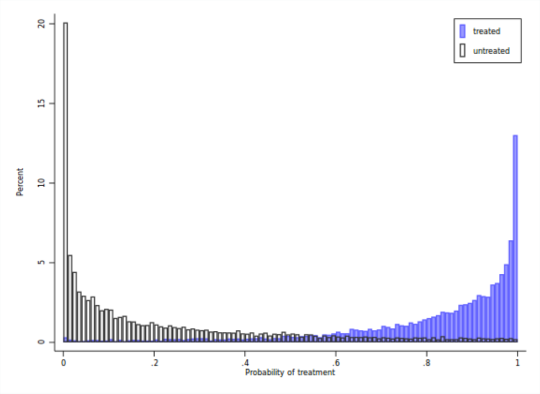


back

Common Support

- Support generated by join variation of instrument and covariates
- Overlap impacts precision and reliability
- Span impacts ability to estimate treatment effect parameters

Figure 14: Unconditional common support of propensity score



back

Polynomial Model

Figure 15: Morbidity

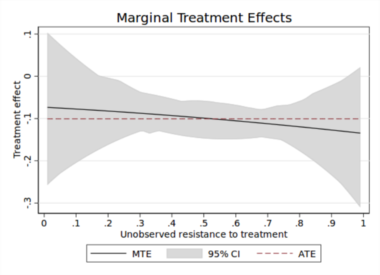
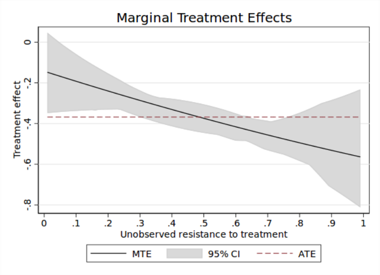


Figure 16: Length of stay



* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bootstrapped standard errors with 100 repetitions. Controls for age, age squared, indicator for white ethnic profile, ten comorbidity variables, distance to the closest hospital, indicator for closest hospital being teaching hospital, urban city indicator, year-month, day of the week controls. Joint normality assumption. All variables interacted with propensity score. Mean length of stay 2.5, with 13 percent of patient experiencing complications.

back

Linear Measure of Skills

Table 4: MTE - Baseline specification (Coefficients) - Normal Model

	(1) Selection equation	(2) Complications	(3) Length of stay
Z_{dist}	-0.013***		
	(0.001)		
Skills (SRR)	-0.597***	-0.114***	-0.175***
	(0.058)	(0.014)	(0.020)
Skills (SRR) * P(Z)		0.149***	0.328***
		(0.021)	(0.025)
Patient Controls	Yes	Yes	Yes
Patient Postal Area	Yes	Yes	Yes
Year*Month	Yes	Yes	Yes
N	19698	19698	19453
ymean	0.538	0.12	0.7
mean Z_{dist}	20	20	20

Standard errors bootstrapped with 100 repetitions $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Column(1) dependent variable binary indicator of robotic surgery. Estimated using Probit regression model. Patient controls include age, age squared, indicator for white ethnic profile, ten comorbidity variables, and distance to the closest hospitals. Skills are measured using the inverse of the post-operative morbidity standardised risk ratio. Estimation of coefficients under the assumption of normality of unobserved components. Model estimated using postal area and year month fixed effects, not interacted with the propensity score.

back

Controls for Surgeon's Experience

Table 5: MTE - Baseline specification (Coefficients) - Normal Model

	(1) Selection equation	(2) Complications	(3) Length of stay
Z_{dist}	-0.016***		
	(0.001)		
High Skilled Dummy	0.244***	-0.107***	-0.191***
	(0.053)	(0.015)	(0.020)
High Skilled Dummy * P(Z)		0.130***	0.329***
		(0.020)	(0.028)
Patient Controls	Yes	Yes	Yes
Patient Postal Area	Yes	Yes	Yes
Year*Month	Yes	Yes	Yes
N	19698	19698	19453

Standard errors bootstrapped with 100 repetitions $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Column(1) dependent variable binary indicator of robotic surgery. Estimated using Probit regression model. Patient controls include age, age squared, indicator for white ethnic profile, ten comorbidity variables, and distance to the closest hospitals. Estimation of coefficients under the assumption of normality of unobserved components. Model estimated using postal area and year month fixed effects, not interacted with the propensity score.

back