Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

ECONOMICS

Ciclo 33

Settore Concorsuale: 13/A1 - ECONOMIA POLITICA

Settore Scientifico Disciplinare: SECS-P/01 - ECONOMIA POLITICA

TIMELINESS, EFFICIENCY AND PUBLIC PROCUREMENT IN HIP SURGERY IN ITALY

Presentata da: Francesca Cassanelli

Coordinatore Dottorato

Maria Bigoni

Supervisore

Daniele Fabbri

Co-supervisore

Maria Bigoni

Esame finale anno 2022

Abstract

This thesis focuses on hip surgeries, which are resource-intensive operations, among the most frequently performed on the elderly, with a trend in volume that has increased in recent years due to the aging population. Taking several perspectives, the scope of the thesis is to investigate how to improve the allocation of healthcare resources and the provision efficiency for this kind of treatment.

Specifically, the effect of Time-To-Surgery (TTS) on health outcomes – i.e. mortality – for patients with hip fracture diagnoses is investigated in the first chapter. In this context, the study attempts to account for TTS endogeneity due to the inability of the researcher to fully control for variables affecting patient delay – e.g. patient severity. The analysis is carried out through an instrumental variable model, where being admitted on Friday or Saturday predicts longer TTS. Results show that exogenous TTS does not have a significant effect on patient mortality. Findings are robust for several model specifications and the inclusion of time spent at the ER. Evidence may suggest surgeons' prioritization scheme effectively neutralizes the adverse effect of longer TTS.

In the second chapter, the volume-outcome relation for total hip replacement surgery is analyzed. Specifically, the research seeks to disentangle the selective referral impact, which may be present when considering elective surgeries, and cause a reverse causality issue in the volume-outcome relation estimation. The analysis exploits a conditional choice model where patient travel distance from all the hospitals in the region is used as the main predictor of hospital choice. The estimated choice probabilities are used to compute a measure of hospital volume cleared from the selective referral component. Results show that the exogenous measure of hospital volume significantly decreases adverse health outcomes probability, especially in the short term, while no association is found for the observed hospital volume.

Finally, in the third chapter, the change in hip prostheses public procurement, enforced in the Romagna LHA (Italy), is exploited to assess the impact on prostheses cost, the effect of the new procurement on surgeons treating choices, and patient health outcomes. Hip prostheses are the major cost-driver of hip replacement surgeries, hence it is crucial to design the public tender such that resulting implant prices are minimized. However, cost-containment policies have to be weighted with patient well-being. Evidence shows that a cost reduction occurred without any significant impact on surgeons' prostheses choices. Also, positive or no effect of surgeons treating style is found on patients outcomes after the new procurement is introduced.

Contents

List of Figures

List of Tables

Chapter 1

Is Faster Better? Treatment delay and patient's outcome in hip fracture surgery[1]

Abstract

In the health care sector, providing timely treatment is particularly valuable. Based on patient conditions, waiting in queue to be treated may be dangerous for health, beyond annoying. In this context, the role of physicians is of primary importance for prioritizing more severe patients over those who can wait longer. This study seeks to identify the effect of in-hospital Time-To-Surgery (TTS) on health outcomes for patients with a hip fracture diagnosis. According to the medical literature, hip fracture patients are considered urgent and need to be treated within two days from hospital admission to significantly reduce mortality. The inability to fully control for patient severity, causes TTS to be endogenous: severe patients need to be stabilized before surgery. An instrumental variable approach is exploited to solve the issue. The time of admission − the day of the week − is used as an instrument for TTS, as patients admitted close to the weekend have a longer delay. Under the exogeneity assumption, TTS is found to have a positive and significant impact on mortality. However, findings indicate that when endogeneity is accounted for, the effect of TTS on mortality is not statistically significant. Results are robust to several different specifications. Evidence may be explained by the fact that even though patients admitted on Friday or Saturday have on average longer TTS, patients who are delayed the most are those who can wait longer, suggesting an effective prioritization scheme implemented by surgeons.

[1] I am grateful to Daniele Fabbri for the essential feedback. I am also thankful to Margherita Fort, Matteo Magnaricotte, Chiara Monfardini, Vincenzo Scrutinio for the useful suggestions. This paper also benefited from participants' comments at the Unibo DSE internal seminar and the 17^{th} Conferência Nacional de Economia da Saúde (Lisbon, Portugal).

1.1 Introduction

The timely treatment of patients is a remarkable aspect of the healthcare sector. Tight budget constraints require efficient use of available resources, and sometimes patient queuing is used as a rationing tool. However, the patient delay may increase pain, discomfort and ultimately lead to adverse health outcomes. Also, a longer time-to-surgery boosts the overall cost of hospital stay. Consequently, the relationship between patient delay and health outcomes has been largely investigated in the literature. However, studies have mainly focused on waiting time for elective surgery as a measure of healthcare efficiency, while less attention has been devoted to emergency and urgent care.

This study investigates the effect of in-hospital Time-To-Surgery (TTS) on health outcomes for hip fracture patients. Hip fracture is a severe trauma frequently experienced by the elderly. According to the medical literature, prompt intervention is required to reduce adverse health outcomes (Casaletto & Gatt, 2004; Doruk, Mas, Yıldız, Sonmez, & Kýrdemir, 2004; Gdalevich, Cohen, Tauber, & Yosef, 2004). Hence hip fractures can be regarded as a typical condition requiring urgent treatment. In particular, this study seeks to account for the possible endogeneity of pre-surgery hospital stay. In this context, endogeneity is due to the inability of the researcher to fully control for unobservable factors influencing the patient delay, such as patient severity. Patient pre-surgery health status may influence TTS since more fragile patients need to be stabilized before surgery. Also, patient severity affects health outcomes since the probability of adverse outcomes increases with severity. The failure to control for unobserved health status causes the estimation of the effect of TTS on health outcomes to be biased. To solve the endogeneity issue, an instrumental variable approach is implemented in this study. The day of the week of hospital admission is used as an instrument for patient delays. The instrument is appropriate if health outcomes are not directly related to the day of the week of patient admission, but they are only indirectly affected through longer delay. Medical staff and operating theaters availability fluctuate along the week: precisely, during weekends, fewer surgeons are on shift, and a lower number of operating theaters is functioning, causing a longer patient time-to-surgery. Like all other types of injuries, hip fracture occurs randomly, therefore the day of the week of hospital admission serves as an exogenous variable to pseudo-randomize patients and enables measurement of the exogenous effect of TTS on health outcomes. This identification strategy has already been presented in the literature (Ho, Hamilton, & Roos, 2000; McGuire, Bernstein, Polsky, & Silber, 2004). This study expands the model in two distinct aspects. Firstly, in order to consider patients as delayed or not, no predetermined cut-off point is established. Instead, TTS is handled as a continuous variable, and its marginal effect on patient mortality at different points in time is investigated. For a subset of observations, TTS is measured in hours instead of days. Also, the time spent by the patient at the ER is included, thus allowing refinement of the estimation. Secondly, the analysis exploits a large dataset of hip fracture surgeries performed in Italy, covering a considerably long period of almost 20 years.

Although under the assumption of exogeneity, a positive impact of time-to-surgery on mortality is found, evidence shows no significant effect exists once allowing for the endogeneity of patient delay. Measuring TTS in hours and including the time spent at the ER does not change the results, and findings are robust to several specifications and further checks. Furthermore, no significant impact is observed, both for the exogenous and endogenous regressions, when considering post-surgery Length of Stay (LoS) – a proxy for patient speed of recovery and absence of complications. Also, findings suggest that mortality is more severely affected by delay for frailer patients. These pieces of evidence hint that surgeons' prioritization scheme effectively compensates for the negative effect caused by longer patient delays during weekends. As long as prioritization works, efforts to treat patients faster need to be weighed against potential savings from shorter pre-surgery hospital stays and patient pain and discomfort reduction. Furthermore, future work should be devoted to assessing the effect of TTS on other sorts of health outcomes.

The rest of the work is structured as follows. The general research background and relevant literature are presented in the next section. Section 1.3 describes the empirical strategy exploited in the analysis, while a description of the data and sample composition is given in section 1.4. Findings and further robustness checks are illustrated in sections 1.5 and 1.6, respectively. Eventually, section 1.7 concludes.

1.2 Background

The fast provision of treatments is a primary concern in the healthcare sector. Tight budget constraints demand efficient use of available resources, particularly when healthcare provision is publicly financed. Hence, patient queuing can sometimes be, more or less effectively, used as a rationing tool (Lindsay & Feigenbaum, 1984; Martin & Smith, 1999). Despite this, waiting in a queue to be treated might be dangerous for health beyond being painful and annoying for the patient. Also, longer in-hospital time-to-surgery increases the cost of hospital stay. For this reason, the role of patient delay on health outcomes has been largely investigated in the literature.

The economic literature has mainly focused on waiting time for elective surgeries and how to tackle this issue (Siciliani & Hurst, 2005; Siciliani, Moran, & Borowitz, 2014). On the contrary, less attention has been devoted to the effect on emergency and urgent conditions, where the patient needs immediate intervention. This relation has been investigated more closely by the medical literature (van Essen, Hans, Hurink, & Oversberg, 2012), instead. While for elective surgeries, the main interest is on waiting time for patients on the list when considering emergency or urgent conditions, the relevant measure of delay is in-hospital time-to-surgery – i.e. the time from hospital admission to surgery.

The purpose of this research is to investigate what is the effect of TTS on health outcomes for hip fracture patients. Hip fracture is one of the most frequent traumas in the elderly. Like all other types of traumas, it occurs unexpectedly, and it is considered an urgent condition that requires timely intervention. These characteristics make hip fracture a good case study to be investigated. There is a general *a-priori* belief that early treatment of this condition decreases the probability of adverse health outcomes, especially patient mortality. Indeed, national guidelines and international standards prescribe quick intervention for patients over 65 with this trauma. For instance, the percentage of hip fractures treated within two days from admission is a national healthcare performance indicator for hospitals in Italy[2] as well as internationally (Mattke et al., 2006).

Despite this strong *a-priori* belief, evidence from the literature is mixed. Some studies found a positive association between patient delay and 30-day mortality (Carretta et al., 2011; Pincus et al., 2017), others claimed small or no significant effect on patient length of stay and in-hospital mortality (B. H. Hamilton & Bramley-Harker, 1999; Ho et al., 2000). Other authors found no impact of TTS on mortality, yet suggesting a positive association with post-surgery complications (Grimes, Gregory, Noveck, Butler, & Carson, 2002) and length of stay (Orosz et al., 2004) exists. In contrast, Bottle and Aylin (2006) found a positive association with in-hospital mortality but little effect on emergency readmission rate within 28 days. These contrasting findings were highlighted in the systematic review by Khan, Kalra, Khanna, Thiruvengada, and Parker (2009) as well. The authors found a significant impact on hospital stay but contrasting evidence for mortality and morbidity, claiming that studies with a more rigorous methodology were more

[2]As stated in Appendix 1, par. 4.6 of Ministerial Decree 70/2015 of the Ministry of Health of 2 April 2015 (https://www.gazzettaufficiale.it/eli/id/2015/06/04/15G00084/sg), at least 60% of cases at the hospital level should be treated within this threshold, while the international standards prescribe a minimum of 80%. According to the AGENAS PNE (*National Program for Outcomes*), in 2016, less than 60% of surgeries were performed within the 2-day threshold in Italy. With large variations occurring both between and within regions. https://www.agenas.gov.it/images/agenas/pne/PNE2018_4_giugno.pdf

likely to find no significant effect.

Patient delay may have several causes, such as availability of the operating theater, low medical staff levels, hospital overcrowding, and patient health conditions. Considering the latter Kenzora, McCarthy, Lowell, and Sledge (1984) argued that despite mortality being significantly correlated with delay, operating before 24 hours from admission may negatively affect survival for hip fracture patients. Accordingly, the authors suggested that at least 24 hours need to be waited to stabilize more severe patients before treating them. Similarly, Weller, Wai, Jaglal, and Kreder (2005) argued that any delay due to non-medical reasons is detrimental to patient health. As a result, several studies in the literature have tried to control for patient health conditions, including a measure of comorbidities or trauma severity in their model. Besides, the availability of medical staff and operating rooms is fundamental for reducing patient delays. However, studies attempting to control for these variables report contrasting results. Evans and Kim (2006) investigated the role of nurse-to-patients ratio for several interventions, finding little evidence of an adverse effect on patient outcomes – namely length of stay, mortality, and readmission – when the staff level experiences a negative shock. In contrast, Bell and Redelmeier (2001) observed a positive association between being admitted during weekends – when most operating theaters are not functioning, since the level of medical staff is incredibly reduced – and patients' mortality for several severe conditions; yet, no association was found for hip fracture patients.

Despite attempts to control for patient pre-surgery conditions, some level of heterogeneity remains unobserved by the researcher. The inability to fully control for patient severity may cause the time-to-surgery to be endogenous. In fact, patients with worst health may have longer TTS due to the need for stabilization before surgery or shorter if surgeons deem the patient as critically requiring care. Treating in-hospital TTS as exogenous, while indeed it is not, may result in biased estimates. The first who tried to address this issue were Ho et al. (2000), proposing an instrumental variable approach where the day of the week of admission is used as a predictor of treatment delay for hip fracture patients. As a matter of fact, the level of medical staff is lower during weekends, and fewer operating theaters are functioning, thus patients are more likely to be delayed during those days. Also, the day of the week of admission does not influence patient outcome, making it a good instrument for the analysis. The authors do not find a significant effect of TTS on in-hospital mortality or length of stay. However, a simple one-way analysis of variance instead of a proper instrumental variable regression model drew the results. Subsequently, McGuire et al. (2004) built on this intuition to implement an instrumental variable approach to investigate the effect of patient delay on 30-day mortality. In contrast to Ho et al. (2000), the authors claimed a 15% increase in the risk of mortality for patients delayed two days or more. Nonetheless, the fact that their analysis considered patients over 65 instead of those over 45 and a stricter selection of fracture types may justify the difference in findings.

To address the possible endogeneity of patient time-to-surgery, a similar approach to that proposed by McGuire et al. (2004) is used in this study. Differently from previous works, which considered a threshold of 3 or 2 days of delay, a predetermined cut-off point for TTS is not established in this analysis, and patients are not *ex-ante* considered as delayed or not. Actual individual time-to-surgery is considered instead. The lack of consensus in the literature regarding the 2-day threshold within which patients should be treated justifies the choice. Even though in the medical field there is general agreement on treating hip fracture patients within two days from hospital admission, Pincus et al. (2017) argued that the relevant threshold should be 24 hours, while Bredahl, Nyholm, Hindsholm, Mortensen, and Olesen (1992) suggested an incredibly tight cut-off point of 12 hours. On top of that, studies typically consider the patient delay in days since more accurate time measures are not available. For most recent data, the analysis exploits the exact time-to-surgery in hours, enabling to refine the measurement of TTS for this subset of observations. This measure of patient TTS also includes the time spent by patients at the ER, whose information was not available in previous studies. Finally, this paper exploits a particularly rich dataset, whose size and period are more extensive than that used in all other studies in our knowledge. Because possible surgery's strategies are highly standardized and leave quite limited discretionary power to the surgeon (Roos, Wennberg, & McPherson, 1988), considering a period of 18 years should not affect the analysis through, for example, changes in technology or surgical techniques. The health outcome considered is patient mortality at different points in time.

1.3 Identification Strategy

Hip fracture is a common and severe trauma typically affecting the elderly, whose bones are more fragile. Like many injuries, it occurs unexpectedly and randomly among the susceptible population. Hence, the incidence frequency and patient case mix do not vary during the week. Furthermore, the pain suffered is such that the injured seeks immediate care at the hospital, where the patient is immobilized at bed until surgery. While it may be argued that the time from accident to surgery should be considered when dealing with critical traumas, this information is usually not available or sufficiently reliable. Moreover, Vidal et al. (2009) found that for hip fracture patients, time-to-surgery from admission to operation is a good proxy for the total interval from fracture to surgery. Consequently, we can regard the in-hospital time-to-surgery as a good proxy for the overall time spent from injury to surgery.

Once at the hospital, surgeons prioritize patients based on severity. Variations in patient pre-surgery health status affect TTS, which may be longer if the patient needs stabilization before surgery, or shorter if prompt care is critical for health. Also, patient condition influences the health outcomes since fragile patients have a lower probability of positive recovery, hence the need to thoroughly control for patient heterogeneity. The impossibility for the researcher to fully account for patient severity, causes TTS to be endogenous: this is due to unobservable factors included in the model error term, which cause it to be correlated with patient delay. Accounting for the endogeneity of patient delay is crucial to avoid biased estimates of the effect of TTS on health outcomes. The direction of the bias depends on the relation between TTS and patient conditions: the probability of survival may be lower for more severe patients even when treated promptly, thus a downward bias. Alternatively, an upward bias may be present if survival improves when treating frailer patients faster. A possible solution to the endogeneity issue is to estimate an instrumental variable model, where the instrument needs to be an exogenous variable predicting TTS and affecting patient outcomes only indirectly through its effect on patient delay. The day of the week of hospital admission may serve as a good instrumental variable. In fact, the day of the week is exogenous to the model since injury occurrence is random, while during weekends, a lower number of medical staff is on shift, and fewer operating theaters are functioning. These changes in healthcare resources availability result in patients admitted closer to the weekend having a longer TTS. The use of the instrumental variable model permits the estimation of the average exogenous effect of TTS on health outcomes by pseudo-randomizing patients. The main model assumption is that unobserved patient severity is homogeneous along the week. This assumption is credible as the injury occurs randomly.

The model specification for patient health outcomes is the following:

$$y_i = \beta_0 + \beta_1 TTS_i + \mathbf{X}_i'\beta_2 + \mathbf{D}_i'\beta_3 + \mathbf{H}_i'\beta_4 + \mathbf{T}_i'\beta_5 + \mathbf{M}_i'\beta_6 + \varepsilon_i \tag{1.1}$$

where y_i is a binary outcome variable equal to 1 if the patient is dead. Patients mortality is measured at different points in time to capture possible short- and medium-term effects. Therefore, equation 1.1 is estimated for in-hospital, 30-day, 60-day, 6-month, and 1-year mortality. The independent variable of interest is TTS_i, which is patient time-to-surgery measured as the number of days of hospital stay before treatment. $\mathbf{X}_i$ is a vector of patient's characteristics: age, gender, the average yearly gross income at the municipality of residence level (in logarithm), and travel distance from patient residency to hospital in minutes. This latter is computed using the ISTAT[3] commuting matrix exploiting information about the patient municipality of residence and that in which the hospital is located. In addition, $\mathbf{D}_i$, $\mathbf{H}_i$, $\mathbf{T}_i$, and $\mathbf{M}_i$ are fixed effects for the type of fracture diagnosed, the hospital of admission, and the year and month of surgery, respectively. These variables are included to control for state-specific effects. Finally, ε_i is the error term.

Specifically, for the year 2017, additional information is included in the analysis. More precisely, TTS is measured as the number of hours from ER access to surgery, therefore considering the extra time spent at the hospital before ward admission. Also, measuring the time in hours, instead of days, gives the possibility to improve the precision of TTS measurement and thus estimates accuracy. Further, the overall patient's health condition is controlled by including dummy variables for the number of comorbidities and cancer presence.

Firstly, equation 1.1 is estimated through a simple Linear Probability Model (LPM) under the assumption of exogeneity of TTS. Then, an Instrumental Variable (IV) model is implemented through two-stage least squares estimation, where the endogeneity of patient delay is accounted for. The first stage equation is as follows:

$$TTS_i = \gamma_0 + \gamma_1 FriSat_i + \mathbf{X}_i'\gamma_2 + \mathbf{D}_i'\gamma_3 + \mathbf{H}_i'\gamma_4 + \mathbf{T}_i'\gamma_5 + \mathbf{M}_i'\gamma_6 + \nu_i \tag{1.2}$$

The instrument is the dummy variable $FriSat_i$ equal to 1 if the patient is admitted to the hospital on Friday or Saturday, given that TTS is longer for patients admitted on those days, as shown in Section 1.4. The other regressors are the same as for equation 1.1. The estimated TTS is then included in 1.1 to determine the average effect of patient delay on health outcomes. Further, to better capture the binary nature of the dependent variables, the analysis is also performed through a probit and an IV probit estimation.

1.4 Data and Sample Composition

The primary source of data is the administrative hospital discharge records database (SDO[4] database), from which information about hospital stays of patients resident in Emilia-Romagna[5] (Italy) admitted to the hospital of the Region are collected. The dataset inclusion criteria follow

[3] Italian National Statistical Institute
[4] Schede di Dimissione Ospedaliera
[5] Emilia-Romagna is an Italian region located in the northern part of the country. With almost 4.5 million resident citizens in 2016, it is one of the most populated Italian regions (about 13% of the overall population).

the guidelines used to compute the national healthcare quality indicator measuring the volume of hip fracture patients treated within two days from admission[6]. Hence, the dataset includes patients aged between 65 and 100 with a diagnosis of fracture of neck of femur[7], treated with total or partial hip prostheses or fracture reduction[8]. Information about the patient's date of death is retrieved from the mortality registry (ReM[9] database). The data span from 2000 to 2017, therefore covering 18 years. The dataset includes surgeries performed in all 84 public and private hospitals of the Region. However, in Italy, the public healthcare sector is predominant, and only 0.52% of surgeries are performed in the 22 private clinics. Patients transferred from one hospital to the other before undergoing surgery are excluded, although amounting to less than 1% of observations in the dataset. Due to the relevant period considered, patients may beak their hip more than once. Since information about what side of the hip experiences the trauma is not available, only the first occurrence per patient is considered, thus excluding subsequent admissions.

For the subset of hip fractures experienced in 2017, additional data are available. Specifically, information about the exact time of ER access – retrieved linking the SDO database with the PS[10] database – ward admission and surgery are collected, while for the rest of the dataset, only the dates are available, and no data on ER access are included. This information allows measuring TTS in hours instead of days, thus improving the precision of variable measurement. Also, the possibility to capture the extra time spent by the patient at the ER gives a better proxy for the total time-to-surgery. Finally, information about the number of relevant comorbidities and whether the patient has cancer are gathered. These variables are included in the model to control for the general patient's health condition.

The final dataset contains 99,130 observations of hip fracture patients, corresponding to an average of about 5,500 surgeries per year, though the trend increases in years. Variables' summary statistics can be found in Table 1.5. The first column reports the mean and standard deviation values for the overall dataset, while statistics for patients admitted on other days or Friday-Saturday are listed in the second and third columns, respectively. Female patients represent individuals majority (76%), and the average age is 83 years. Patient travel distance

[6] https://pne.agenas.it/risultati/protocolli/pro_42.pdf

[7] Defined according to the international classification of diagnosis ICD9-CM (2002 version). Refer to Section 1.8.1 for the detailed list of codes considered. The listed conditions are generally referred to as hip fractures. Therefore, this denomination is used throughout the paper for simplicity.

[8] Defined according to the international classification of surgical procedures ICD9-CM (2002 version). Refer to Section 1.8.2 for the detailed list of codes considered.

[9] Rilevazione Mortalità

[10] Pronto Soccorso

from the admitting hospital is 9 minutes on average. Patients' characteristics and fracture type do not vary significantly between the two groups, apart from one minor exception. This evidence supports the assumption that patients' case-mix does not vary among the two groups. Even if it is not possible to fully control for patient severity with the available data, it seems reasonable to assume that average severity is not affected by the day of the week of patient admission. However, as expected, the average TTS significantly differs between the two groups: it is more than three days and a half for patients admitted on Friday or Saturday, while it is just three days for patients admitted the other days of the week. Similar statistics are found for the subset of observation in 2017 (Table 1.6). The overall average TTS measured in hours is just above two days (50 hours), but the difference between the two groups is still similar (10 hours).

One of the main assumptions of the model is that the inflow of patients admitted to the hospital is constant during the week. As reported in Table 1.1, patients' admissions have a considerably uniform distribution along the week. However, from the third and fourth columns of the table, it is clear that the number of patients undergoing surgery experiences a significant drop during the weekend, while there is a slight increase on Monday. This pattern may be explained by the fact that operating theaters majority are not functioning on Saturday and Sunday, given that the level of medical staff is lower on those days. Accordingly, a longer TTS for patients admitted on Friday and Saturday is presumed since they need to wait all weekend to be treated. This speculation is supported by the evidence shown in Figure 1.1. Indeed, for patients admitted on Friday or Saturday, the TTS distribution is centered around three days of delay, while more than 60% of patients admitted to the hospital on other days is treated within two days from ward admission. Only a small portion of patients is treated on the same day of admission, while almost all patients undergo surgery within a week, apart from rare cases in both groups. The Kolmogorov-Smirnov test for equality of distribution functions is strongly rejected (p-value=0.000), confirming that the two groups have different TTS distributions. Also, Pearson's chi-squared test rejects the independence of patient delay distribution from the day of the week of hospital admission (p-value=0.000). Considering just the 5,517 observations for 2017, Figure 1.2 illustrates the average TTS in hours per day of the week and time slot of ER admission. The bell-shaped pattern is evident around Friday, Saturday, and Sunday. The patient delay starts increasing steadily on Friday morning, it reaches the peak on Saturday afternoon, and then it decreases all along Sunday. On the contrary, a clear pattern for the other days of the week is not observed, and apart from a few exceptions, patients are generally treated within the 48-hour threshold. Furthermore, irrespective of time slot or day of the week of ER admission, patients spend no more than 5 hours at the ER before being admitted to the orthopedic ward. The only exception is for patients admitted on Sunday from 00.00 to 05.59 am. Looking at the evolution of TTS over the 18-year period, there is evidence that the average delay went from 4 days at the beginning of the period to 2 days at the end of it (Figure 1.3). However, this reduction did not alter the difference in TTS between patients admitted on Friday and Saturday and those admitted on the other days, which remained considerably stable.

Considering the dependent variable of interest, the probability of dying is increasing with time-to-surgery (Figure 1.4). This raw relation is in line with findings from the medical literature. However, the patients' mortality rate is almost flat up to 4 days of delay: the period in which most patients are treated. Finally, as reported in Table 1.2, the probability of death at different points in time is reasonably stable along the week – the Pearson chi-squared test is accepted for all outcome variables considered. This confirms the assumption that patient mortality is not related to the day of the week of ward admission. Hence, making it a suitable instrumental variable for the study.

With TTS measured in days from ward admission to surgery, the analysis for the overall dataset is presented in the next section. While, results for the restricted and more rich subset of observations, including information on the ER stay, are presented in Section 1.6 as a further robustness check.

1.5 Results

Firstly, the model under the assumption of exogenous time-to-surgery is estimated. Table 1.7 reports results for the complete specification of the linear probability and probit models for all dependent variables of interest. The coefficients for the average effect of time-to-surgery are positive and statistically significant for all regressions. The effect magnitude is not negligible. All else equal, one extra day of delay increases the probability of death from 0.27 for in-hospital mortality to 0.92 percentage points for 1-year mortality. The findings highlight that the magnitude of the effect is increasing with time to mortality. This result is somehow puzzling, given that in this context, expectations would predict the treatment to have a more substantial impact on events occurring closer to it. However, the impact magnitude expressed in percentage terms results in being larger for short-term mortality outcomes than long-term ones. A probit model is also used to capture better the binary nature of the dependent variable of interest. Results are remarkably similar to those found through LPM estimation. The average partial effects of patient delay are positive and statistically significant for all dependent variables considered.

Also, the magnitude of TTS is close to that found in the LPM estimation, although slightly smaller.

When accounting for the endogeneity of time-to-surgery, regressions' results significantly change. Table 1.8 reports the estimations for the full specification of the IV and IV probit models for all mortality outcomes. The sign of TTS becomes negative, and the coefficients are all statistically insignificant, with the only exception of in-hospital mortality, for which the estimated coefficient is significant only at 10% level. The IV probit estimations report a larger magnitude in absolute values, yet the significance remains unchanged compared to that obtained through the simple IV estimation. From the first stage results, evidence shows that being admitted on Friday or Saturday significantly increases patient TTS by more than half a day (Table 1.9). The Kleibergen-Paap rk Wald test for instrument weakness reports an F-statistic larger than 10, so the instrument is relevant. Also, the exogeneity of TTS is always rejected. Findings are consistent for all mortality outcomes through several different specifications for all the models presented, as reported in Figures 1.5-1.8. The graphs outline point estimates and 95% confidence intervals for the average effect of TTS on patient mortality at different points in time for various specifications, *ceteris paribus*. The models were also estimated clustering standard errors at the individual, hospital, and day of the week level. However, results remain unchanged. For this reason, all the standard errors reported in the Tables and Figures are just robust to heteroskedasticity.

As expected, the impact of TTS on patient mortality is found to be upward biased when exogeneity is assumed. Once controlling for endogeneity, the marginal effect of TTS on mortality becomes negative and insignificant, thus indicating that the omitted variable is positively related to the patient delay. Despite patients admitted on Friday and Saturday experiencing longer TTS, this delay is not significantly affecting their mortality. Evidence suggests severe patients may be delayed to be stabilized before surgery. These findings may also be explained by the fact that surgeons, through their prioritization system, can neutralize the adverse effect of longer TTS, and patients who wait longer are those who can better bear the delay without negative consequences on survival. Furthermore, results hold not only for short-term mortality but also for long-term outcomes.

Findings are limited by the fact that with the available data it is not possible to thoroughly account for patient health condition before surgery, given that the only characteristics are gender, age, and fracture type, and the model estimates an average effect. Also, the not-significant TTS effect on mortality does not rule out the possibility of a positive impact on other health outcomes exists, such as the probability of readmission or complications occurrence. Results presented in the next Section attempt to address these issues, restricting the analysis on frail patients and estimating the model for post-surgery length of stay.

1.6 Robustness Checks

1.6.1 TTS measured in hours from ER admission to surgery

For hip fracture patients in 2017, additional information on the exact time of ER admission and surgery is available, thus enabling refinement of the analysis by measuring TTS in hours instead of days. Moreover, it is possible to account for the extra time spent by the patient at the ER, thus obtaining a better proxy of the overall time from trauma to surgery. The same model presented in 1.1 is estimated to investigate whether these more accurate data change the results. For this subset of observations, patient time-to-surgery is measured in hours from ER access instead of ward admission to surgery. Also, the number of comorbidities and cancer presence variables are included in the regression to better control for patient frailty.

Results do not differ much from the primary evidence. A positive and statistically significant relation is found – though with a smaller magnitude – under the assumption of exogenous TTS for all mortality outcomes (Table 1.10). Estimates for the LPM and probit model are once again strikingly alike in magnitude and significance. As for the main results, the instrumental variable models estimate a negative and insignificant effect of TTS on mortality, with the only exception being in-hospital and 30-day mortality, for which the impact of delay is negative and significant at 10% and 5% level in the IV probit estimation (Table 1.11). The negative sign may support the hypothesis that patients needs to wait for stabilization before undergoing surgery. Results from the first stage estimation of the IV model (Table 1.12) highlight that being admitted on Friday or Saturday significantly increases patient TTS by more than 9 hours. This magnitude is comparable to that found for the main first-stage regression, though somewhat smaller, also due to the fact that overall TTS is diminishing in years. Findings are robust to several model specifications as shown in Figures 1.9-1.12, which report point estimates and 95% confidence intervals for the effect of patient TTS on all outcome variables of interest. Evidence suggests that measuring TTS in hours does not alter the results. Moreover, accounting for the additional time spent by the patient at the ER does not change the findings.

In light of these pieces of evidence, the differential effect of TTS spent at the ER or the orthopedic ward existence is also examined. Quality of care may differ in the two environments, and spending more time in one context may have a different impact on mortality. Table 1.13

reports LPM and probit estimations for all mortality outcomes for the most complete model specification. The magnitude and significance are similar across LPM and probit estimations, tough slightly higher for the latter. The time spent from ER access to ward admission does not significantly affect patient mortality in the short run, while significance is generally found for 6-month and for 1-year mortality. Results may be driven by the amount of time spent at the ER, which may be too short to significantly impact patient mortality. Findings support the idea that measuring patient delay from ward admission to surgery is a sufficiently reliable proxy of overall patient waiting time. Unfortunately, the lack of additional instruments prevents accounting for the endogeneity of TTS through the estimation of the IV model. Therefore, it is not possible to assess whether a causal relation exists.

1.6.2 Frail Patients

Results presented in Section 1.5 are just the average impact among all patients, for which waiting a long time may have a differential effect. According to the literature, older patients have a lower probability of successfully recovering from a hip fracture, given trauma severity and the invasiveness of the surgical procedure. There may be the chance that longer TTS has a more substantial impact on frailer patients. Given the limited available data on patient health conditions before surgery, age is used as a proxy for frailty to check for this possibility. Patients over 85 constitute 40.65% of observations in the dataset. A dummy variable equal to 1 if the patient is over 85 years old is introduced in the model. This threshold is chosen based on evidence from the literature, according to which patients above that age are particularly likely to experience adverse health outcomes after hip fracture surgery. The dummy variable is let interact with TTS to account for the differential effect of delay between the two groups. The regression is estimated under the assumption of exogeneity through a linear probability model since an additional instrument would be necessary to perform an IV estimation.

Findings show that TTS has a positive and significant impact on mortality (Table 1.14). The magnitude of the effect is greater for patients over 85, though significance varies with the outcome considered. In particular, the interaction term is found to be insignificant for in-hospital mortality, while only a 10%-level significance is found for 30-day and 1-year mortality, and a 1%-level one for 60-day and 6-month mortality. Evidence suggests that under the assumption of exogeneity, longer TTS has a more substantial adverse impact for very old patients, although results are not conclusive. Unfortunately, the inability to estimate the model controlling for the endogeneity of TTS considerably limits the ability to comment on these results. It may be the case that effects turn to zero for both groups once endogeneity is considered. To attempt to partially check for this possibility, the instrumental variable model, for the most complete specification, is estimated for the subset of patients over 85. Results, reported in Table 1.15, show no significant effect of TTS is found for all dependent variables of interest, except for in-hospital mortality, whose effect is significant at 10%-level only, and with a negative sign. The negative impact of TTS on in-hospital mortality may be in favor of the practice of stabilizing more severe patients before treating them. However, the generally not significant effect may support the hypothesis that no differential effect exists for patients over 85. Nevertheless, age is just a proxy for severity, and a more thorough assessment of patient health conditions may be necessary.

1.6.3 TTS effect evolution through years

As pointed out in Figure 1.3, average TTS is decreasing in years, though the difference between Friday and Saturday, and the other days of the week remains fairly stable around half a day. The results obtained through the main model assume the effect of TTS is homogeneous across years and may not explain well the effect on mortality for a single year. To understand how the TTS effect evolves across the 18-year period considered in the analysis, the model in 1.1 is estimated per single year, for each outcome variable of interest, through LPM and IV model. The regression includes all fixed effects from 1.1, apart from year fixed effects, which are excluded. Standard errors are robust to heteroskedasticity.

Figure 1.13 and 1.14 report respectively TTS point estimates and 95% confidence intervals for the LPM and IV model estimated for each year separately. While a significant and fairly stable impact is found for LPM estimations – with some exceptions for the first 5 years –, no significant effect of TTS on any mortality outcome is detected across years once accounting for endogeneity. These findings suggest the overall reduction of TTS across years has not affected the estimates obtained through the main model.

1.6.4 Post-Surgery Length of Stay

It is licit to argue that patient mortality may be an extreme health outcome to look at, which may be hardly affected by TTS. Notwithstanding that no significant effect of TTS on mortality is found, a longer patient delay may influence other relevant health outcomes, such as the insurgence of post-surgery complications. The effect of patient TTS on post-surgery Length of Stay (LoS) is examined to investigate if this is the case. Post-surgery LoS can be regarded as a proxy for quick recovery ability. Patients treated faster may have a lower incidence of post-surgery complications, regain physical mobility sooner, and thus be discharged earlier. LoS is also a relevant economic outcome since a shorter hospital stay decreases healthcare expenditure.

The model presented in 1.1 is estimated both through OLS and IV model, where the instrument is the usual dummy variable equal to 1 for patients admitted on Friday or Saturday. The dependent variable of interest is patient post-surgery LoS measured in days from surgery to hospital discharge. Patients have an average post-surgery LoS of 16 days, which is reasonably constant over the day of the week of ward admission (Table 1.3). Hence, being admitted on Friday or Saturday can still be used as a good instrument. Also, studies have assessed that patient post-surgery length of stay is affected by the discharge destination. Statistics reported in Table 1.4 supports this concept: patients transferred to long-term care facilities or rehabilitation centers have on average a shorter LoS, while a longer LoS is experienced by those discharged at home with special care, probably due to particular fragilities. Nevertheless, the majority of patients – more than 56% – are simply discharged at home. To control for the relation between LoS and discharge destination, a set of dummy variables is included in the estimated model.

Results are illustrated in Figure 1.15, which presents point estimates and 95% confidence intervals for several model specifications. The left-hand side plot shows coefficients for the OLS estimation, under exogeneity assumption, while results for the IV model are on the right-hand side plot. Assuming the exogeneity of TTS, findings are mixed. While results are not significant for almost all specifications, a negative impact of patient TTS on LoS is found when hospital and month fixed effects are included. The relation goes in the opposite direction to the initial expectations and may reflect the practice of early discharge to long-term facilities to maintain the overall hospital stay within the standard. However, after accounting for endogeneity, no significant effect is found for all model specifications, even if patients admitted on Friday and Saturday experience a longer delay. Results hold also for the 2017 subset (Figure 1.16). Consequently, treating patients faster does not have a statistically significant impact on the length of post-surgery hospitalization. Regrettably, results can not completely rule out the possibility that TTS significantly affects post-surgery complications, given that this analysis considers only a proxy for them. Still, the lack of significant effect on patient LoS may be interpreted as the incidence of complications during the hospitalization remaining stable. Further research should be devoted to investigating whether TTS impacts patients' health outcomes other than mortality, such as hospital readmission.

1.7 Conclusions

This study aims to assess the effect of in-hospital time-to-surgery on the health outcomes of patients undergoing hip fracture surgery. The analysis attempts to account for the endogeneity of TTS caused by the inability of the researcher to control for unobservable patient severity, which may affect LTT. Being admitted on Friday or Saturday is found to be a valid and relevant instrument for TTS: medical staff levels decrease during the weekend, increasing in-hospital waiting time, while patients' outcomes are directly unaffected by the day of the week of admission. The day of the week of hospital admission is exogenous to the model, as the injury occurs unexpectedly. Assuming unobserved patient severity is homogeneously distributed along the week, the instrumental variable model permits us to estimate the average effect of TTS on patient mortality, controlling for endogeneity. Evidence suggests that under the assumption of exogenous TTS, patient delay positively and significantly affects short- and long-term mortality, while results become negative and insignificant once controlling for endogeneity. Findings are robust to several specifications. Measuring TTS in hours and including the time spent at the ER in the overall measure of patient delay does not change the results. Findings remain stable across the 18-year period considered in the analysis. The evidence fails to support the belief that shorter TTS improves patient survival. Even if patients admitted on Friday or Saturday experience longer – more than half a day – TTS, surgeons' prioritization scheme can effectively neutralize the adverse effect of longer delay on patient survival.

Because mortality can be considered an extremely adverse outcome, the effect of TTS on the patient post-surgery LoS is investigated as well. LoS can be regarded as a measure of fast recovery, also related to the insurgence of post-surgery complications. Contrary to expectations, a significant impact of TTS on patient LoS is not detected once controlling for endogeneity. Nevertheless, it is not reasonable to completely rule out the presence of TTS effect on post-surgery complications, even if the not significant effect found for LoS supports the belief that an impact should not be present, since this would have also affected the length of post-surgery hospital stay. Lastly, patient TTS has a more substantial impact on frailer patients – those over 85 – mortality, although results are not conclusive. Unfortunately, the lack of additional instruments precluded the possibility to evaluate if a differential effect exists once controlling for the endogeneity of TTS. Nevertheless, no significant effect is found when estimating the IV model on the subset of patients over 85.

The analysis is limited by the inability to thoroughly account for patients' frailty and pre-surgery conditions due to the lack of available data. In this respect, the instrumental variable model captures only the average effect across patients. The not significant effect of TTS on patient health outcomes may also be due to the fact that the current differential delay between the two groups is sufficiently small to avoid poor health outcomes for those waiting a longer period. Surgeons may play a crucial role in this respect, by wisely prioritizing patients according to their needs. Another possible explanation may be that the extra time waited by the patient is necessary for stabilization before surgery, which is supported by the upward biased uncovered through the IV estimation. Another limitation of the estimation is that the model assumes patient severity is homogeneous along the week and the estimated effect is an average among all patients.

In conclusion, no substantial evidence supporting a policy aiming at treating patients faster is ascertained, apart from the fact that delay needs to be reduced to limit patient suffering and discomfort. Moreover, this remark holds as long as surgeons can effectively prioritize patients so those who wait longer can bear it without detrimental effect on health outcomes or need time for stabilization before safely undergoing surgery. Also, decisions on where to allocate more resources need to be weighed against the potential benefits for the hospital organization as a whole or at least for the ward. Future work should be devoted to investigating more closely surgeons' decisions regarding patient scheduling, as well as investigating whether a causal effect exists between TTS and other health outcomes other than mortality.

1.8 ICD9-CM Codes for Inclusion Criteria

1.8.1 Diagnosis codes

Codes	Definition
820.00	Closed fracture of intracapsular section of neck of femur, unspecified
820.01	Closed fracture of epiphysis (separation) (upper) of neck of femur
820.02	Closed fracture of midcervical section of neck of femur
820.03	Closed fracture of base of neck of femur
820.09	Other closed transcervical fracture of neck of femur
820.10	Open fracture of intracapsular section of neck of femur, unspecified
820.11	Open fracture of epiphysis (separation) (upper) of neck of femur
820.12	Open fracture of midcervical section of neck of femur
820.13	Open fracture of base of neck of femur
820.19	Other open transcervical fracture of neck of femur
820.20	Closed fracture of trochanteric section of neck of femur
820.21	Closed fracture of intertrochanteric section of neck of femur
820.22	Closed fracture of subtrochanteric section of neck of femur
820.30	Open fracture of trochanteric section of neck of femur, unspecified
820.31	Open fracture of intertrochanteric section of neck of femur
820.32	Open fracture of subtrochanteric section of neck of femur
820.8	Closed fracture of unspecified part of neck of femur
820.9	Open fracture of unspecified part of neck of femur

1.8.2 Surgical Procedures Codes

Codes	Definition
79.00	Closed reduction of fracture without internal fixation, unspecified site
79.05	Closed reduction of fracture without internal fixation, femur
79.10	Closed reduction of fracture with internal fixation, unspecified site
79.15	Closed reduction of fracture with internal fixation, femur
79.20	Open reduction of fracture without internal fixation, unspecified site
79.25	Open reduction of fracture without internal fixation, femur
79.30	Open reduction of fracture with internal fixation, unspecified site
79.35	Open reduction of fracture with internal fixation, femur
79.40	Closed reduction of separated epiphysis, unspecified site
79.45	Closed reduction of separated epiphysis, femur
79.50	Open reduction of separated epiphysis, unspecified site
79.55	Open reduction of separated epiphysis, femur
81.51	Total hip replacement
81.52	Partial hip replacement

1.9 Tables and Figures

1.9.1 Tables

Table 1.1: Average number of admitted to the ward and treated patients per day of the week

	Ward Admission		Surgery	
	No.	%	No.	%
Monday	15.73	14.89	24.36	23.06
Tuesday	15.44	14.62	19.77	18.71
Wednesday	15.35	14.53	18.12	17.16
Thursday	15.43	14.61	16.88	15.98
Friday	15.38	14.56	20.26	19.18
Saturday	14.94	14.14	5.18	4.91
Sunday	13.35	12.64	1.06	1.00
Total	105.62	100.00	105.62	100.00

Table 1.2: Mortality per day of the week of ward admission

	Mortality				
	In-hospital mean	30-day mean	60-day mean	6-month mean	1-year mean
Monday	0.035	0.047	0.081	0.163	0.228
Tuesday	0.033	0.045	0.080	0.159	0.220
Wednesday	0.034	0.044	0.081	0.159	0.219
Thursday	0.038	0.050	0.087	0.164	0.221
Friday	0.034	0.047	0.082	0.161	0.218
Saturday	0.033	0.044	0.081	0.162	0.221
Sunday	0.038	0.052	0.087	0.164	0.223
Total	0.035	0.047	0.083	0.162	0.221
Observations	99,130	99,130	99,130	99,130	99,130

Table 1.3: Patients average post-surgery LoS per day of the week of ward admission

	LoS mean	LoS sd
Monday	15.99	13.88
Tuesday	15.90	13.91
Wednesday	16.26	14.21
Thursday	16.10	13.96
Friday	16.09	14.36
Saturday	16.00	13.97
Sunday	15.70	13.81
Total	16.01	14.02
Observations	99,127	

Table 1.4: Frequency of patients' discharge destination and average post-surgery LoS

	Discharge type No.	Discharge type %	LoS mean
Discharged dead	3,458	3.49	16.95
Discharged at home	56,036	56.53	16.73
Discharged at nursing home	18,021	18.18	16.33
Discharged at patient's home w/ nursing care	455	0.46	25.62
Voluntary	239	0.24	15.26
Transfer to acute care facility	3,060	3.09	12.73
Transfer to a different ward	2,007	2.02	10.65
Transfer to a rehabilitation centre	11,418	11.52	11.23
Discharged w/ home-integrated care	4,432	4.47	20.93
Displacement	1	0.00	7.00
Total	99,127	100.00	16.01

Table 1.5: Sample composition

	Total mean (sd)	Other Days mean (sd)	Friday/Saturday mean (sd)	t-stat for mean difference
Age	83.12	83.11	83.14	-0.528
	(7.304)	(7.308)	(7.293)	
Female	0.762	0.761	0.764	-1.168
	(0.426)	(0.426)	(0.424)	
Log income (thousands)	3.037	3.038	3.037	0.951
	(0.180)	(0.180)	(0.180)	
Hospital distance (minutes)	9.017	9.054	8.925	1.449
	(12.67)	(12.69)	(12.64)	
TTS (days)	3.176	3.000	3.614	-32.52***
	(2.704)	(2.700)	(2.665)	
Fracture type:				
Closed fracture of unspecified part of neck of femur	0.0353	0.0356	0.0345	0.793
	(0.184)	(0.185)	(0.183)	
Open fracture of unspecified part of neck of femur	0.000454	0.000495	0.000351	0.962
	(0.0213)	(0.0222)	(0.0187)	
Closed fracture of intracapsular section of neck of femur, unspecified	0.0510	0.0512	0.0505	0.472
	(0.220)	(0.220)	(0.219)	
Closed fracture of epiphysis (separation) (upper) of neck of femur	0.107	0.107	0.107	-0.0115
	(0.309)	(0.309)	(0.309)	
Closed fracture of midcervical section of neck of femur	0.139	0.140	0.139	0.469
	(0.346)	(0.347)	(0.346)	
Closed fracture of base of neck of femur	0.102	0.102	0.101	0.680
	(0.302)	(0.303)	(0.301)	
Other closed transcervical fracture of neck of femur	0.0682	0.0695	0.0650	2.493**
	(0.252)	(0.254)	(0.247)	
Open fracture of intracapsular section of neck of femur, unspecified	0.00227	0.00235	0.00207	0.824
	(0.0476)	(0.0484)	(0.0455)	
Open fracture of epiphysis (separation) (upper) of neck of femur	0.00154	0.00156	0.00151	0.165
	(0.0393)	(0.0394)	(0.0388)	
Open fracture of midcervical section of neck of femur	0.00674	0.00666	0.00692	-0.450
	(0.0818)	(0.0814)	(0.0829)	
Open fracture of base of neck of femur	0.00294	0.00289	0.00306	-0.450
	(0.0541)	(0.0536)	(0.0552)	
Other open transcervical fracture of neck of femur	0.000293	0.000311	0.000246	0.544
	(0.0171)	(0.0176)	(0.0157)	
Closed fracture of trochanteric section of neck of femur	0.276	0.275	0.279	-1.453
	(0.447)	(0.446)	(0.449)	
Closed fracture of intertrochanteric section of neck of femur	0.155	0.154	0.156	-0.887
	(0.362)	(0.361)	(0.363)	
Closed fracture of subtrochanteric section of neck of femur	0.0451	0.0446	0.0463	-1.136
	(0.208)	(0.206)	(0.210)	
Open fracture of trochanteric section of neck of femur, unspecified	0.00166	0.00166	0.00169	-0.110
	(0.0408)	(0.0407)	(0.0410)	
Open fracture of intertrochanteric section of neck of femur	0.00375	0.00375	0.00376	-0.0246
	(0.0611)	(0.0611)	(0.0612)	
Open fracture of subtrochanteric section of neck of femur	0.00199	0.00190	0.00221	-1.0168
	(0.0445)	(0.0435)	(0.0470)	
Observations	99,130	70,674	28,456	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The logarithm of the average yearly gross income at municipality of residence level is considered. Hospital distance in minutes is retrieved from ISTAT distances' matrix and it is computed from the centroid of the municipality of patient residence to the centroid of the municipality in which the hospital is located.

Table 1.6: Sample composition - 2017 subset

	Total mean (sd)	Other Days mean (sd)	Friday/Saturday mean (sd)	t-stat for mean difference
Age	83.77	83.87	83.52	1.555
	(7.500)	(7.474)	(7.560)	
Female	0.746	0.748	0.739	0.760
	(0.436)	(0.434)	(0.440)	
Log income (thousands)	3.084	3.085	3.082	0.675
	(0.131)	(0.130)	(0.134)	
Hospital distance (minutes)	9.449	9.261	9.916	-1.728*
	(12.75)	(12.51)	(13.30)	
TTS ER-surgery (hours)	50.02	47.17	57.10	-9.173***
	(36.66)	(35.16)	(39.27)	
No. of comorbidities = 0	0.684	0.680	0.694	-1.030
	(0.465)	(0.466)	(0.461)	
No. of comorbidities = 1	0.271	0.275	0.263	0.844
	(0.445)	(0.446)	(0.441)	
No. of comorbidities = 2	0.0393	0.0412	0.0347	1.135
	(0.194)	(0.199)	(0.183)	
No. of comorbidities = 3	0.00471	0.00382	0.00693	-1.529
	(0.0685)	(0.0617)	(0.0830)	
No. of comorbidities = 4	0.000363	0.000254	0.000630	-0.663
	(0.0190)	(0.0160)	(0.0251)	
Cancer	0.00544	0.00356	0.0101	-2.983***
	(0.0735)	(0.0596)	(0.0999)	
Fracture type:				
Closed fracture of unspecified part of neck of femur	0.0234	0.0221	0.0265	-0.963
	(0.151)	(0.147)	(0.161)	
Open fracture of unspecified part of neck of femur	0.00254	0.00204	0.00378	-1.166
	(0.0503)	(0.0451)	(0.0614)	
Closed fracture of intracapsular section of neck of femur, unspecified	0.0301	0.0298	0.0309	-0.217
	(0.171)	(0.170)	(0.173)	
Closed fracture of epiphysis (separation) (upper) of neck of femur	0.142	0.140	0.146	-0.626
	(0.349)	(0.347)	(0.353)	
Closed fracture of midcervical section of neck of femur	0.153	0.149	0.162	-1.175
	(0.360)	(0.356)	(0.369)	
Closed fracture of base of neck of femur	0.0814	0.0847	0.0731	1.431
	(0.273)	(0.279)	(0.260)	
Other closed transcervical fracture of neck of femur	0.0506	0.0550	0.0397	2.343**
	(0.219)	(0.228)	(0.195)	
Open fracture of intracapsular section of neck of femur, unspecified	0.00453	0.00560	0.00189	1.856*
	(0.0672)	(0.0746)	(0.0435)	
Open fracture of epiphysis (separation) (upper) of neck of femur	0.00199	0.00204	0.00189	0.110
	(0.0446)	(0.0451)	(0.0435)	
Open fracture of midcervical section of neck of femur	0.0156	0.0155	0.0158	-0.063
	(0.124)	(0.124)	(0.125)	
Open fracture of base of neck of femur	0.00326	0.00229	0.00567	-1.994*
	(0.0570)	(0.0478)	(0.0751)	
Other open transcervical fracture of neck of femur	0.000181	0	0.000630	-1.574
	(0.0135)	(0)	(0.0251)	
Closed fracture of trochanteric section of neck of femur	0.230	0.234	0.221	1.032
	(0.421)	(0.423)	(0.415)	
Closed fracture of intertrochanteric section of neck of femur	0.194	0.191	0.200	-0.714
	(0.395)	(0.393)	(0.400)	
Closed fracture of subtrochanteric section of neck of femur	0.0569	0.0575	0.0555	0.298
	(0.232)	(0.233)	(0.229)	
Open fracture of trochanteric section of neck of femur, unspecified	0.00471	0.00407	0.00630	-1.095
	(0.0685)	(0.0637)	(0.0792)	
Open fracture of intertrochanteric section of neck of femur	0.00471	0.00407	0.00630	-1.095
	(0.0685)	(0.0637)	(0.0792)	
Open fracture of subtrochanteric section of neck of femur	0.00145	0.000763	0.00315	-2.110**
	(0.0381)	(0.0276)	(0.0561)	
Observations	5,517	3,930	1,587	

*** p<0.01, ** p<0.05, * p<0.1

Note: The logarithm of the average yearly gross income at municipality of residence level is considered. Hospital distance in minutes is retrieved from ISTAT distances' matrix and it is computed from the centroid of the municipality of patient residence to the centroid of the municipality in which the hospital is located.

Table 1.7: LPM and probit estiamtes under exogeneity assumption

	Mortality									
	In-hospital		30-day		60-day		6-month		1-year	
	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TTS (days)	0.00274***	0.00228***	0.00332***	0.00290***	0.00540***	0.00482***	0.00830***	0.00773***	0.00926***	0.00884***
	(0.000296)	(0.000188)	(0.000321)	(0.000218)	(0.000404)	(0.000319)	(0.000505)	(0.000431)	(0.000552)	(0.000493)
Age	-0.00712***	0.00286*	-0.0127***	0.00229	-0.0241***	-2.95e-05	-0.0366***	-0.00293	-0.0397***	-0.00381
	(0.00142)	(0.00161)	(0.00166)	(0.00183)	(0.00217)	(0.00237)	(0.00278)	(0.00309)	(0.00304)	(0.00343)
Age^2	5.76e-05***	-1.99e-06	9.79e-05***	7.99e-06	0.000183***	3.75e-05***	0.000286***	8.25e-05***	0.000324***	0.000106***
	(8.91e-06)	(9.57e-06)	(1.04e-05)	(1.09e-05)	(1.36e-05)	(1.41e-05)	(1.73e-05)	(1.85e-05)	(1.89e-05)	(2.06e-05)
Female	-0.0378***	-0.0323***	-0.0499***	-0.0431***	-0.0784***	-0.0688***	-0.122***	-0.111***	-0.150***	-0.139***
	(0.00168)	(0.00126)	(0.00191)	(0.00144)	(0.00240)	(0.00185)	(0.00299)	(0.00245)	(0.00326)	(0.00275)
Log income (thousands)	-0.0167**	-0.0161**	-0.0190**	-0.0165**	-0.0271**	-0.0249**	-0.0422***	-0.0384***	-0.0572***	-0.0544***
	(0.00725)	(0.00333)	(0.00833)	(0.00382)	(0.0112)	(0.0110)	(0.0148)	(0.0147)	(0.0164)	(0.0165)
Hospital distance (minutes)	-0.000145***	-0.000148**	6.88e-05	6.99e-05	0.000117	0.000109	0.000133	0.000135	2.89e-05	2.23e-05
	(5.36e-05)	(5.04e-05)	(6.11e-05)	(5.36e-05)	(7.91e-05)	(8.06e-05)	(0.000106)	(0.000108)	(0.000117)	(0.000121)
Observations	99,127	98,861	99,127	98,600	99,127	99,058	99,127	99,080	99,127	99,099
R-squared	0.024		0.029		0.045		0.068		0.083	
Pseudo R2		0.0780		0.0731		0.0764		0.0776		0.0801
Fracture type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0349	0.0349	0.0467	0.0467	0.0829	0.0829	0.162	0.162	0.221	0.221

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Columns 1, 3, 5, 7, and 9 report LPM estimated coefficients, while estimated average partial effects for the probit model are reported in columns 2, 4, 6, 8, and 10.

Table 1.8: Instrumental Variable and IV probit estimates

	Mortality									
	In-hospital		30-day		60-day		6-month		1-year	
	IV	IV Probit	IV	IV Probit	IV	IV Probit	IV	IV Probit	IV	IV Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TTS (days)	-0.00351*	-0.0526*	-0.00328	-0.0410	-0.00236	-0.0179	-0.000278	-0.000257	-0.00497	-0.0177
	(0.00206)	(0.0280)	(0.00237)	(0.0255)	(0.00313)	(0.0217)	(0.00414)	(0.0182)	(0.00463)	(0.0167)
Age	-0.00634***	0.0493**	-0.0119***	0.0336*	-0.0232***	0.00621	-0.0355***	-0.00867	-0.0380***	-0.00771
	(0.00146)	(0.0220)	(0.00170)	(0.0199)	(0.00221)	(0.0168)	(0.00283)	(0.0138)	(0.00311)	(0.0127)
Age^2	5.26e-05***	-9.53e-05	9.26e-05***	2.65e-05	0.000177***	0.000220**	0.000279***	0.000336***	0.000312***	0.000344***
	(9.15e-06)	(0.000131)	(1.07e-05)	(0.000119)	(1.38e-05)	(0.000101)	(1.76e-05)	(8.31e-05)	(1.93e-05)	(7.68e-05)
Female	-0.0391***	-0.457***	-0.0513***	-0.478***	-0.0801***	-0.492***	-0.124***	-0.496***	-0.153***	-0.517***
	(0.00173)	(0.0165)	(0.00195)	(0.0151)	(0.00250)	(0.0129)	(0.00313)	(0.0112)	(0.00342)	(0.0104)
Log income (thousands)	-0.0195***	-0.257**	-0.0219***	-0.209**	-0.0306***	-0.197**	-0.0460***	-0.184***	-0.0636***	-0.220***
	(0.00423)	(0.105)	(0.00479)	(0.0901)	(0.0113)	(0.0775)	(0.0149)	(0.0653)	(0.0166)	(0.0607)
Hospital distance (minutes)	-0.000152***	-0.00210***	6.21e-05	0.000681	0.000109	0.000709	0.000125	0.000559	1.44e-05	3.01e-05
	(4.82e-05)	(0.000815)	(5.69e-05)	(0.000676)	(7.94e-05)	(0.000566)	(0.000106)	(0.000479)	(0.000118)	(0.000444)
Observations	99,127	98,861	99,127	98,600	99,127	99,058	99,127	99,080	99,127	99,099
R-squared	0.010		0.019		0.036		0.061		0.072	
Fracture type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0349	0.0349	0.0467	0.0467	0.0829	0.0829	0.162	0.162	0.221	0.221

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Columns 1, 3, 5, 7, and 9 report IV estimated coefficientss, while estimated average partial effects for the IV probit model are reported in columns 2, 4, 6, 8, and 10.

Table 1.9: Instrumental Variable model - first stage estiamtes

	TTS (days)			
	(1)	(2)	(3)	(4)
Friday or Saturday ward admission	0.614***	0.608***	0.605***	0.604***
	(0.0186)	(0.0178)	(0.0171)	(0.0171)
Age	0.145***	0.126***	0.123***	0.123***
	(0.0209)	(0.0201)	(0.0193)	(0.0193)
Age^2	-0.000977***	-0.000823***	-0.000804***	-0.000804***
	(0.000127)	(0.000122)	(0.000117)	(0.000117)
Female	-0.172***	-0.224***	-0.218***	-0.219***
	(0.0203)	(0.0197)	(0.0189)	(0.0189)
Log income (thousands)	-1.161***	0.376***	-0.451***	-0.448***
	(0.0497)	(0.0662)	(0.0994)	(0.0993)
Hospital distance (minutes)	-0.00597***	0.00182***	-0.000929	-0.000926
	(0.000715)	(0.000708)	(0.000711)	(0.000711)
Observations	99,129	99,129	99,127	99,127
Fracture type FE	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes
Hospital FE	No	No	Yes	Yes
Month FE	No	No	No	Yes
Mean dependent	3.176	3.176	3.176	3.176
Weak identification test (F-stat)	1089	1161	1254	1254

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.10: LPM and probit estimates under exogeneity assumption - 2017 subset

	Mortality									
	In-hospital		30-day		60-day		6-month		1-year	
	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TTS ER-surgery (hours)	0.000171*	0.000131***	0.000244**	0.000210***	0.000480***	0.000395***	0.000615***	0.000536***	0.000729***	0.000664***
	(9.24e-05)	(4.79e-05)	(9.90e-05)	(6.03e-05)	(0.000129)	(8.46e-05)	(0.000152)	(0.000117)	(0.000164)	(0.000138)
Age	-0.00146	0.00985	-0.00600	0.0103	-0.0264***	0.0102	-0.0480***	-0.00315	-0.0512***	-0.00514
	(0.00426)	(0.00602)	(0.00595)	(0.00868)	(0.00841)	(0.0110)	(0.0111)	(0.0134)	(0.0124)	(0.0145)
Age^2	1.93e-05	-4.50e-05	5.47e-05	-4.04e-05	0.000199***	-1.69e-05	0.000358***	8.97e-05	0.000396***	0.000119
	(2.67e-05)	(3.52e-05)	(3.70e-05)	(5.07e-05)	(5.26e-05)	(6.45e-05)	(6.94e-05)	(7.96e-05)	(7.69e-05)	(8.63e-05)
Female	-0.0259***	-0.0241***	-0.0468***	-0.0405***	-0.0724***	-0.0647***	-0.103***	-0.0933***	-0.128***	-0.118***
	(0.00559)	(0.00428)	(0.00743)	(0.00544)	(0.00960)	(0.00736)	(0.0120)	(0.00982)	(0.0133)	(0.0112)
Log income (thousands)	0.0556*	0.0494**	0.0436	0.0393	-0.00711	-0.0103	-0.00723	-0.00885	0.0366	0.0336
	(0.0311)	(0.0250)	(0.0383)	(0.0312)	(0.0473)	(0.0444)	(0.0673)	(0.0631)	(0.0737)	(0.0705)
Hospital distance (minutes)	0.000327	0.000253*	0.000457	0.000387*	0.000256	0.000258	0.000617	0.000643	0.000767	0.000766*
	(0.000214)	(0.000153)	(0.000283)	(0.000215)	(0.000331)	(0.000317)	(0.000440)	(0.000415)	(0.000481)	(0.000465)
Cancer	0.0753	0.0429***	0.125*	0.0700***	0.233***	0.151***	0.261***	0.203***	0.372***	0.302***
	(0.0539)	(0.0162)	(0.0679)	(0.0224)	(0.0818)	(0.0340)	(0.0883)	(0.0538)	(0.0899)	(0.0666)
No. of comorbidities = 1	-0.00787*	-0.00857**	-0.00269	-0.000622	0.00103	0.00431	0.0207*	0.0231**	0.0261*	0.0268**
	(0.00474)	(0.00436)	(0.00667)	(0.00614)	(0.00917)	(0.00875)	(0.0122)	(0.0117)	(0.0137)	(0.0132)
No. of comorbidities = 2	0.0106	0.0102	0.0220	0.0221	0.0568**	0.0575**	0.111***	0.109***	0.178***	0.176***
	(0.0138)	(0.0132)	(0.0184)	(0.0172)	(0.0250)	(0.0242)	(0.0312)	(0.0304)	(0.0340)	(0.0339)
No. of comorbidities = 3	-0.0335***	-	-0.0503***	-	-0.00297	0.000540	0.00376	0.00495	0.0242	0.0260
	(0.0105)		(0.0131)		(0.0513)	(0.0490)	(0.0645)	(0.0595)	(0.0760)	(0.0727)
No. of comorbidities = 4	-0.0337	-	-0.0508	-	-0.122**	-	-0.176**	-	0.819***	-
	(0.0241)		(0.0485)		(0.0606)		(0.0767)		(0.105)	
Observations	5,509	5,022	5,509	5,397	5,509	5,422	5,509	5,491	5,509	5,503
R-squared	0.043		0.051		0.075		0.097		0.118	
Pseudo R2		0.168		0.149		0.137		0.116		0.115
Fracture type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0212	0.0212	0.0390	0.0390	0.0767	0.0767	0.151	0.151	0.216	0.216

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Columns 1, 3, 5, 7, and 9 report LPM estimated coefficients, while estimated average partial effects for the probit model are reported in columns 2, 4, 6, 8, and 10.

Table 1.11: Instrumental Variable and IV probit estimates - 2017 subset

	Mortality									
	In-hospital		30-day		60-day		6-month		1-year	
	IV	IV Probit	IV	IV Probit	IV	IV Probit	IV	IV Probit	IV	IV Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TTS ER-surgery (hours)	-0.000567	-0.0122*	-0.000964	-0.0119**	-0.000995	-0.00734	-0.000459	-0.00196	-0.00142	-0.00508
	(0.000434)	(0.00729)	(0.000590)	(0.00598)	(0.000829)	(0.00580)	(0.00111)	(0.00515)	(0.00127)	(0.00448)
Age	-0.000183	0.214*	-0.00391	0.147	-0.0238***	0.0936	-0.0462***	-0.00773	-0.0475***	-0.00682
	(0.00430)	(0.115)	(0.00615)	(0.105)	(0.00867)	(0.0832)	(0.0114)	(0.0655)	(0.0128)	(0.0556)
Age^2	1.12e-05	-0.00103	4.15e-05	-0.000639	0.000182***	-0.000239	0.000347***	0.000382	0.000372***	0.000364
	(2.69e-05)	(0.000666)	(3.83e-05)	(0.000616)	(5.43e-05)	(0.000493)	(7.09e-05)	(0.000391)	(7.95e-05)	(0.000335)
Female	-0.0293***	-0.513***	-0.0524***	-0.537***	-0.0792***	-0.523***	-0.108***	-0.467***	-0.138***	-0.475***
	(0.00614)	(0.0855)	(0.00827)	(0.0707)	(0.0105)	(0.0566)	(0.0131)	(0.0491)	(0.0147)	(0.0439)
Log income (thousands)	0.0429	0.653	0.0228	0.211	-0.0324	-0.259	-0.0256	-0.120	-0.000254	-0.00475
	(0.0322)	(0.561)	(0.0409)	(0.416)	(0.0506)	(0.351)	(0.0706)	(0.318)	(0.0788)	(0.283)
Hospital distance (minutes)	0.000290	0.00378	0.000394	0.00387	0.000179	0.00144	0.000560	0.00286	0.000654	0.00248
	(0.000217)	(0.00308)	(0.000289)	(0.00274)	(0.000337)	(0.00242)	(0.000446)	(0.00203)	(0.000496)	(0.00181)
Cancer	0.0961*	1.187***	0.159**	1.245***	0.275***	1.422***	0.291***	1.096***	0.432***	1.342***
	(0.0569)	(0.349)	(0.0714)	(0.310)	(0.0926)	(0.312)	(0.0974)	(0.295)	(0.103)	(0.287)
No. of comorbidities = 1	-0.00774	-0.172	-0.00246	-0.00800	0.00131	0.0329	0.0209*	0.111**	0.0265*	0.102**
	(0.00482)	(0.105)	(0.00680)	(0.0755)	(0.00938)	(0.0665)	(0.0123)	(0.0548)	(0.0140)	(0.0497)
No. of comorbidities = 2	0.0192	0.334*	0.0361*	0.392**	0.0740***	0.474***	0.124***	0.501***	0.204***	0.658***
	(0.0148)	(0.194)	(0.0201)	(0.158)	(0.0271)	(0.133)	(0.0337)	(0.118)	(0.0378)	(0.105)
No. of comorbidities = 3	-0.0275**	-	-0.0405**	-	0.00895	0.0962	0.0125	0.0572	0.0416	0.157
	(0.0125)		(0.0165)		(0.0517)	(0.380)	(0.0646)	(0.290)	(0.0762)	(0.263)
No. of comorbidities = 4	-0.0367*	-	-0.0557	-	-0.128**	-	-0.180***	-	0.810***	-
	(0.0216)		(0.0404)		(0.0517)		(0.0696)		(0.0900)	
Observations	5,509	5,022	5,509	5,397	5,509	5,422	5,509	5,491	5,509	5,503
Fracture type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0212	0.0212	0.0390	0.0390	0.0767	0.0767	0.151	0.151	0.216	0.216

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Columns 1, 3, 5, 7, and 9 report IV estimated coefficients, while estimated average partial effects for the IV probit model are reported in columns 2, 4, 6, 8, and 10.

Table 1.12: Instrumental Variable Model first stage estimates - 2017 subset

	TTS ER-surgery (hours)		
	(1)	(2)	(3)
Friday or Saturday ER admission	9.738***	9.384***	9.375***
	(1.131)	(1.082)	(1.091)
Age	1.866	1.825	1.742
	(1.179)	(1.128)	(1.129)
Age^2	-0.0116	-0.0115*	-0.0110
	(0.00715)	(0.00684)	(0.00685)
Female	-4.984***	-4.602***	-4.604***
	(1.241)	(1.185)	(1.187)
Log income (thousands)	4.492	-15.81***	-16.25***
	(4.035)	(5.858)	(5.864)
Hospital distance (minutes)	0.0446	-0.0526	-0.0558
	(0.0396)	(0.0386)	(0.0387)
Cancer	24.29**	25.37**	25.60**
	(9.966)	(10.20)	(10.16)
No. of comorbidities = 1	-2.271**	0.322	0.333
	(1.091)	(1.128)	(1.128)
No. of comorbidities = 2	10.62***	12.06***	12.09***
	(3.304)	(3.330)	(3.344)
No. of comorbidities = 3	5.516	6.959	6.930
	(7.385)	(7.521)	(7.443)
No. of comorbidities = 4	-2.647	-4.825	-5.697
	(8.077)	(10.70)	(11.16)
Observations	5,513	5,509	5,509
Fracture type FE	Yes	Yes	Yes
Hospital FE	No	Yes	Yes
Month FE	No	No	Yes
Mean dependent	50.02	50.02	50.02
Weak identification test (F-stat)	74.08	75.22	73.83

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.13: LPM and probit estimates, differential effect of TTS - 2017 subset

	Mortality									
	In-hospital		30-day		60-day		6-month		1-year	
	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit	LPM	Probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TTS ER-ward (hours)	-6.21e-05	0.000136	0.000280	0.000435	0.00140	0.00144**	0.00210*	0.00218**	0.00333***	0.00335***
	(0.000444)	(0.000377)	(0.000682)	(0.000539)	(0.000855)	(0.000695)	(0.00108)	(0.000959)	(0.00124)	(0.00110)
TTS ward-surgery (hours)	0.000175*	0.000131***	0.000243**	0.000206***	0.000465***	0.000379***	0.000591***	0.000509***	0.000687***	0.000619***
	(9.35e-05)	(4.84e-05)	(9.95e-05)	(6.07e-05)	(0.000129)	(8.54e-05)	(0.000153)	(0.000118)	(0.000165)	(0.000139)
Age	-0.00141	0.00985	-0.00601	0.0103	-0.0265***	0.0101	-0.0483***	-0.00341	-0.0518***	-0.00568
	(0.00427)	(0.00600)	(0.00595)	(0.00869)	(0.00842)	(0.0110)	(0.0112)	(0.0134)	(0.0124)	(0.0144)
Age^2	1.90e-05	-4.50e-05	5.48e-05	-4.05e-05	0.000200***	-1.63e-05	0.000360***	9.13e-05	0.000399***	0.000122
	(2.68e-05)	(3.51e-05)	(3.71e-05)	(5.08e-05)	(5.27e-05)	(6.46e-05)	(6.94e-05)	(7.95e-05)	(7.69e-05)	(8.62e-05)
Female	-0.0260***	-0.0241***	-0.0468***	-0.0405***	-0.0720***	-0.0646***	-0.102***	-0.0930***	-0.127***	-0.117***
	(0.00561)	(0.00427)	(0.00746)	(0.00544)	(0.00963)	(0.00736)	(0.0120)	(0.00982)	(0.0133)	(0.0112)
Log Income (thousands)	0.0553*	0.0494**	0.0436	0.0395	-0.00595	-0.00886	-0.00535	-0.00687	0.0399	0.0364
	(0.0311)	(0.0250)	(0.0384)	(0.0312)	(0.0473)	(0.0445)	(0.0673)	(0.0631)	(0.0737)	(0.0705)
Hospital distance (minutes)	0.000329	0.000253*	0.000457	0.000385*	0.000250	0.000247	0.000608	0.000629	0.000750	0.000743
	(0.000214)	(0.000153)	(0.000283)	(0.000215)	(0.000331)	(0.000317)	(0.000440)	(0.000415)	(0.000481)	(0.000465)
Cancer	0.0751	0.0429***	0.125*	0.0702***	0.234***	0.152***	0.262***	0.204***	0.374***	0.304***
	(0.0539)	(0.0162)	(0.0680)	(0.0224)	(0.0819)	(0.0340)	(0.0884)	(0.0537)	(0.0899)	(0.0665)
No. of comorbidities = 1	-0.00781*	-0.00857**	-0.00270	-0.000698	0.000798	0.00401	0.0203*	0.0227*	0.0255*	0.0262**
	(0.00474)	(0.00437)	(0.00668)	(0.00615)	(0.00918)	(0.00876)	(0.0122)	(0.0117)	(0.0137)	(0.0133)
No. of comorbidities = 2	0.0107	0.0102	0.0220	0.0218	0.0563**	0.0565**	0.111***	0.108***	0.177***	0.174***
	(0.0138)	(0.0132)	(0.0184)	(0.0171)	(0.0250)	(0.0240)	(0.0311)	(0.0303)	(0.0340)	(0.0338)
No. of comorbidities = 3	-0.0333***	-	-0.0503***	-	-0.00360	-0.000492	0.00274	0.00353	0.0224	0.0235
	(0.0105)		(0.0131)		(0.0513)	(0.0484)	(0.0648)	(0.0594)	(0.0763)	(0.0725)
No. of comorbidities = 4	-0.0340	-	-0.0508	-	-0.121**	-	-0.174**	-	0.822***	-
	(0.0241)		(0.0486)		(0.0606)		(0.0769)		(0.106)	
Observations	5,509	5,022	5,509	5,397	5,509	5,422	5,509	5,491	5,509	5,503
R-squared	0.043		0.051		0.075		0.098		0.118	
Pseudo-R2		0.168		0.149		0.138		0.117		0.116
Fracture type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0212	0.0212	0.0390	0.0390	0.0767	0.0767	0.151	0.151	0.216	0.216

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Columns 1, 3, 5, 7, and 9 report LPM estimated coefficients, while estimated average partial effects for the probit model are reported in columns 2, 4, 6, 8, and 10.

Table 1.14: Linear Probability Model estimates, differential impact of TTS for patients over 85

	Mortality				
	In-hospital	30-day	60-day	6-month	1-year
	(1)	(2)	(3)	(4)	(5)
TTS (days)	0.00256***	0.00285***	0.00452***	0.00700***	0.00845***
	(0.000329)	(0.000349)	(0.000441)	(0.000566)	(0.000636)
Over 85	0.00177	-0.00264	-0.00673	-0.00645	0.00706
	(0.00279)	(0.00320)	(0.00410)	(0.00530)	(0.00588)
Over 85 * TTS (days)	0.000439	0.00120*	0.00227***	0.00336***	0.00205*
	(0.000567)	(0.000632)	(0.000791)	(0.000983)	(0.00107)
Age	-0.00630***	-0.0125***	-0.0242***	-0.0357***	-0.0362***
	(0.00169)	(0.00199)	(0.00258)	(0.00325)	(0.00352)
Age^2	5.16e-05***	9.64e-05***	0.000183***	0.000280***	0.000298***
	(1.10e-05)	(1.29e-05)	(1.67e-05)	(2.09e-05)	(2.26e-05)
Female	-0.0378***	-0.0499***	-0.0784***	-0.122***	-0.150***
	(0.00168)	(0.00191)	(0.00240)	(0.00299)	(0.00326)
Log Income (thousands)	-0.0167**	-0.0190**	-0.0271**	-0.0422***	-0.0571***
	(0.00725)	(0.00833)	(0.0112)	(0.0148)	(0.0164)
Hospital distance (minutes)	-0.000146***	6.87e-05	0.000117	0.000133	2.90e-05
	(5.36e-05)	(6.11e-05)	(7.91e-05)	(0.000106)	(0.000117)
Observations	99,127	99,127	99,127	99,127	99,127
R-squared	0.024	0.029	0.045	0.069	0.083
Fracture type FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0349	0.0467	0.0829	0.162	0.221

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 1.15: Instrumental Variable Model estimates, effect of TTS on patients over 85

	Mortality				
	In-hospital	30-day	60-day	6-month	1-year
	(1)	(2)	(3)	(4)	(5)
TTS (days)	-0.00700*	-0.00428	-0.00278	-0.00296	-0.00789
	(0.00406)	(0.00474)	(0.00611)	(0.00783)	(0.00858)
Age	0.0266	0.0223	-0.00284	-0.000149	0.00888
	(0.0174)	(0.0206)	(0.0266)	(0.0329)	(0.0353)
Age^2	-0.000130	-9.36e-05	6.74e-05	8.65e-05	5.05e-05
	(9.55e-05)	(0.000113)	(0.000146)	(0.000181)	(0.000194)
Female	-0.0519***	-0.0722***	-0.112***	-0.162***	-0.189***
	(0.00336)	(0.00387)	(0.00481)	(0.00580)	(0.00613)
Log Income (thousands)	-0.0330**	-0.0449***	-0.0388*	-0.0668**	-0.0684**
	(0.0144)	(0.0166)	(0.0219)	(0.0280)	(0.0305)
Hospital distance (minutes)	-0.000283***	-6.98e-06	7.92e-05	9.73e-05	6.19e-05
	(0.000105)	(0.000120)	(0.000154)	(0.000199)	(0.000218)
Observations	40,293	40,293	40,293	40,293	40,293
R-squared	0.000	0.013	0.025	0.036	0.037
Fracture type FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0513	0.0706	0.125	0.236	0.318

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

1.9.2 Figures

Figure 1.1: Distribution of TTS per day of the week of ward admission

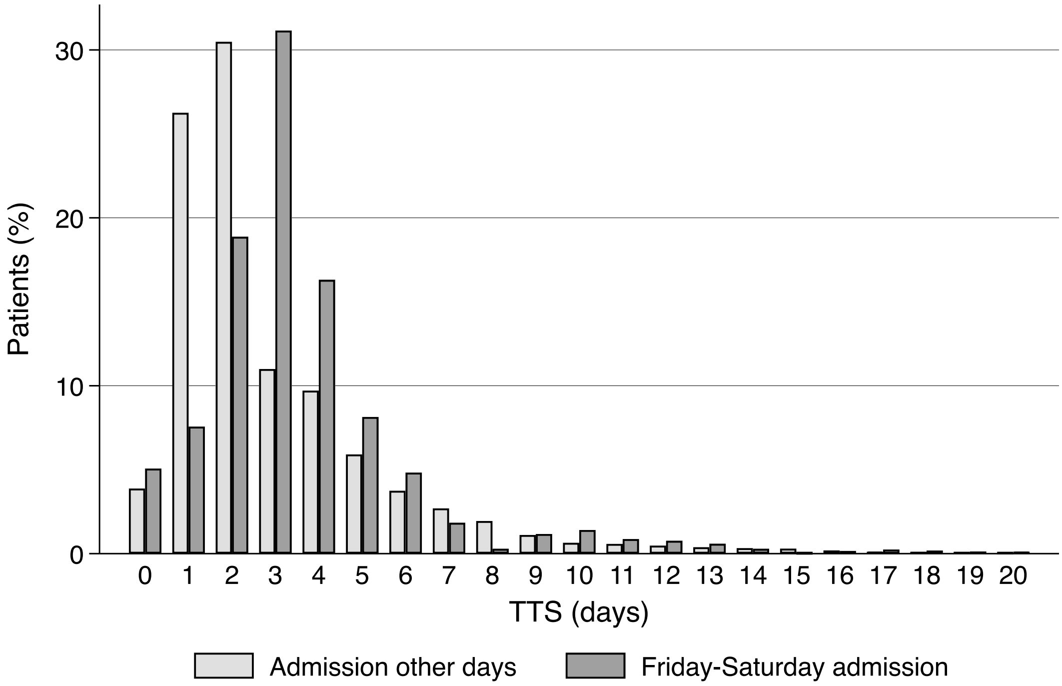


Figure 1.2: Average TTS per day of the week and time slot of ER admission

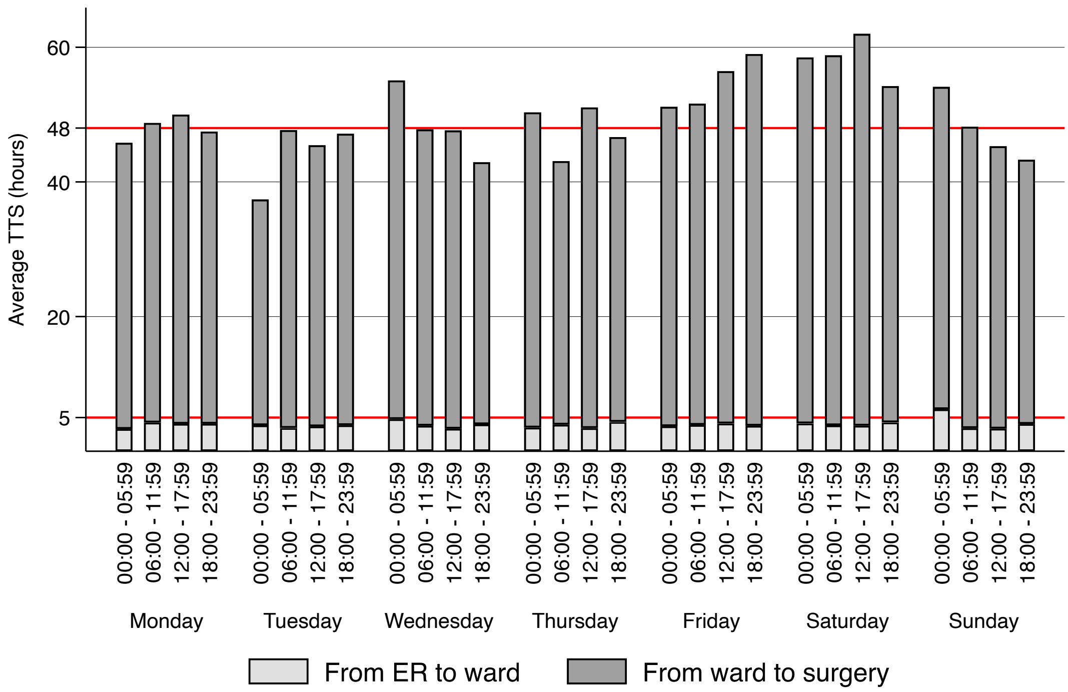


Figure 1.3: Evolution of average TTS in years

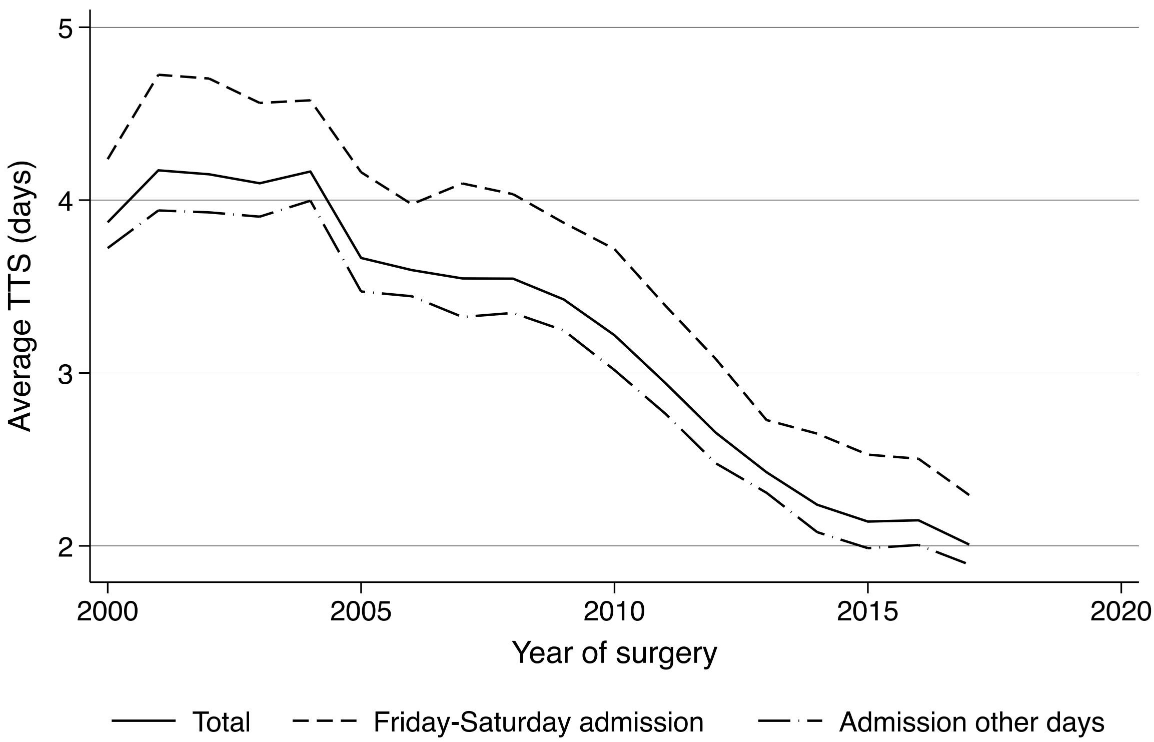


Figure 1.4: Patient mortality per TTS

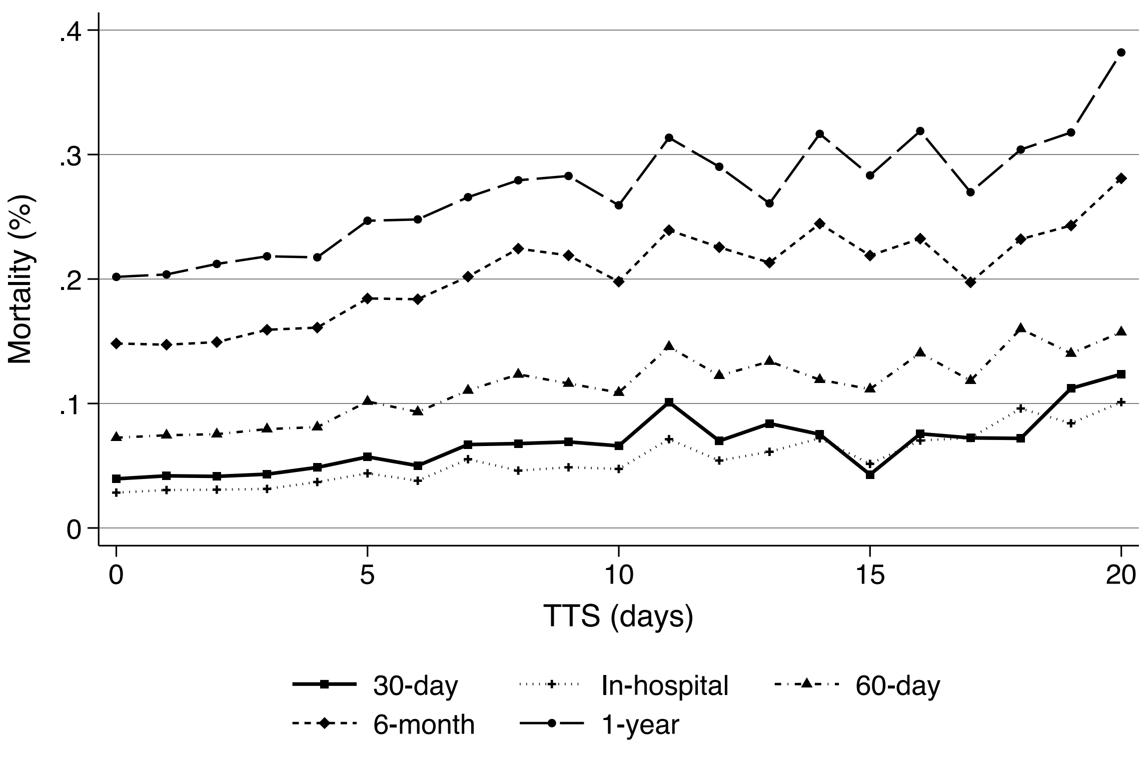


Figure 1.5: TTS estimates - LPM model, various specifications

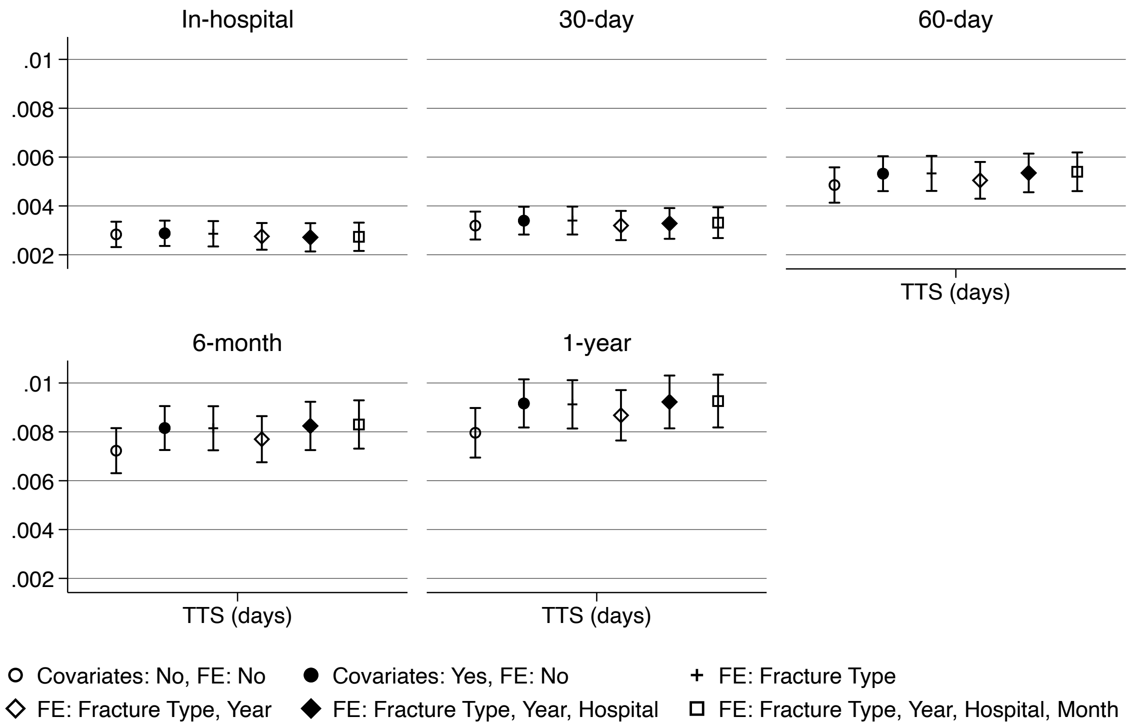


Figure 1.6: TTS estimates - Probit model, various specifications

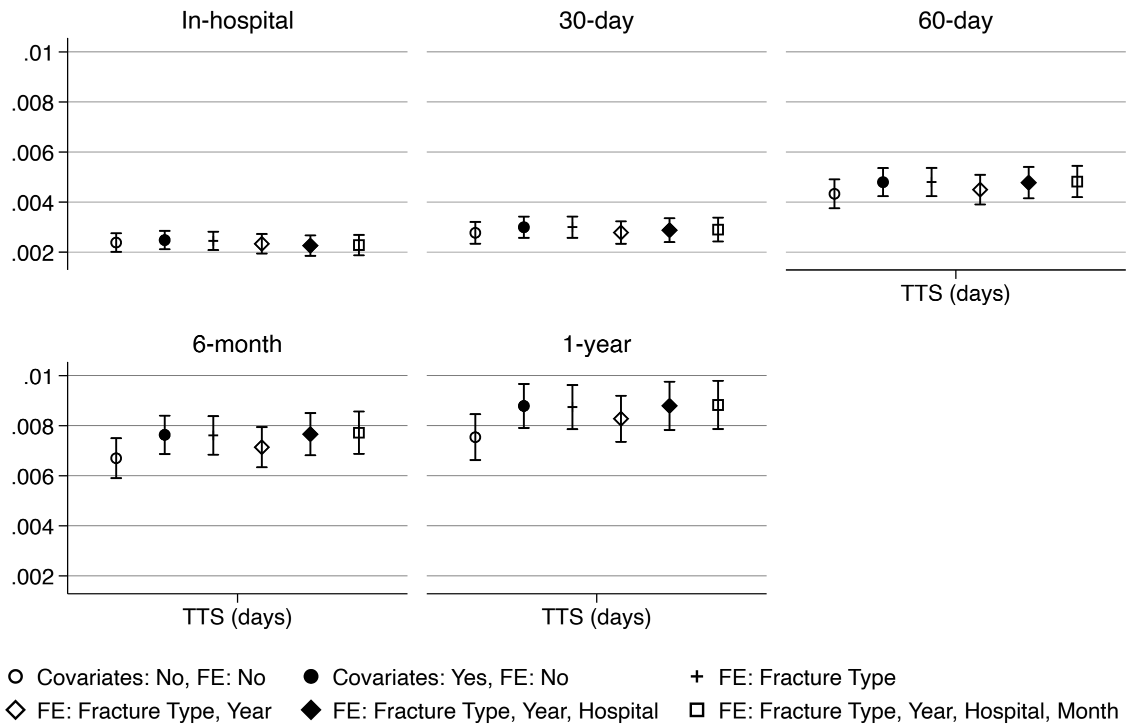


Figure 1.7: TTS estimates - IV model, various specifications

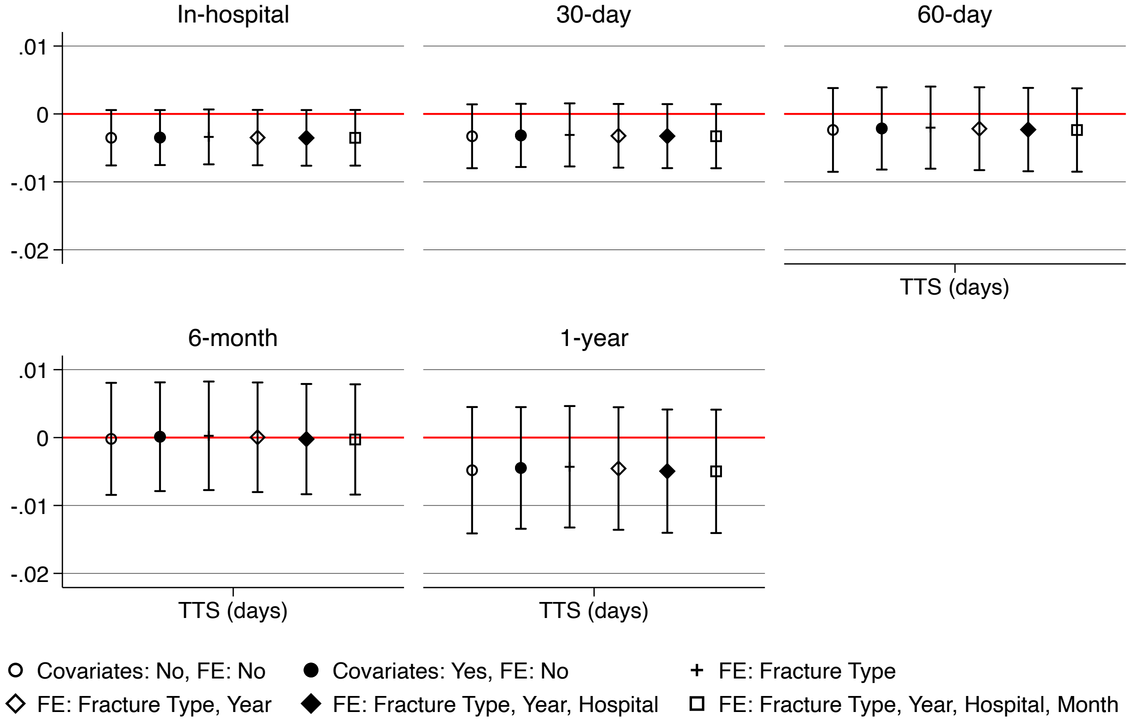


Figure 1.8: TTS estimates - IV probit model, various specifications

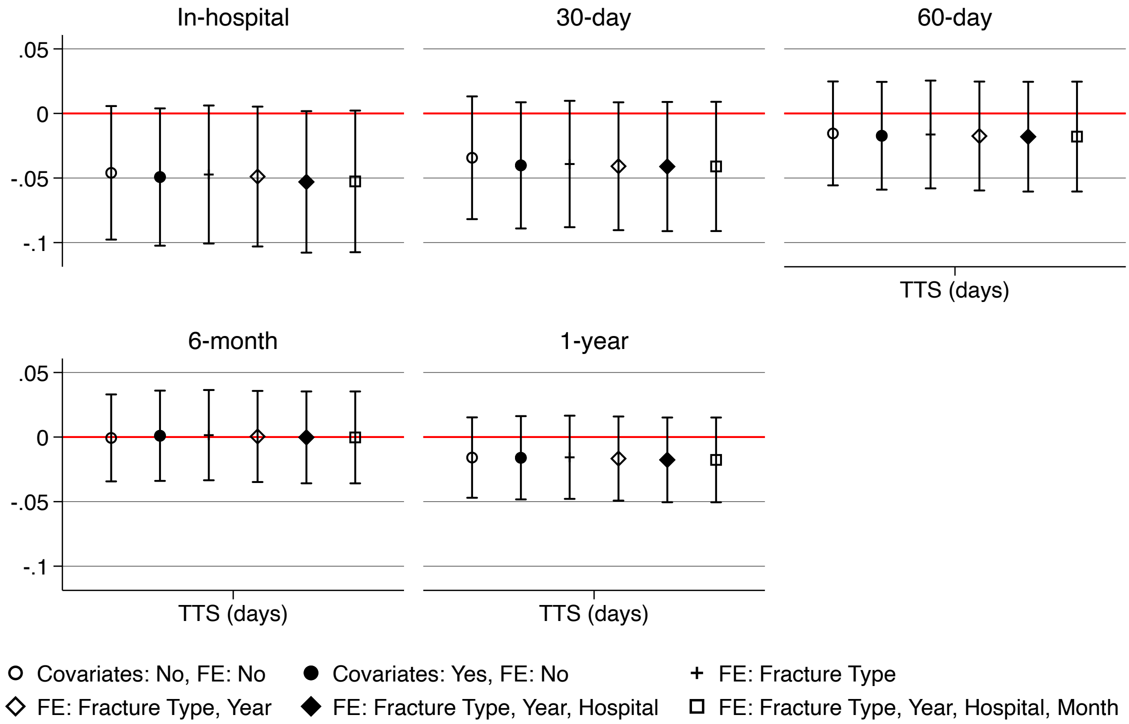


Figure 1.9: TTS (hours) estimates - LPM model, various specifications

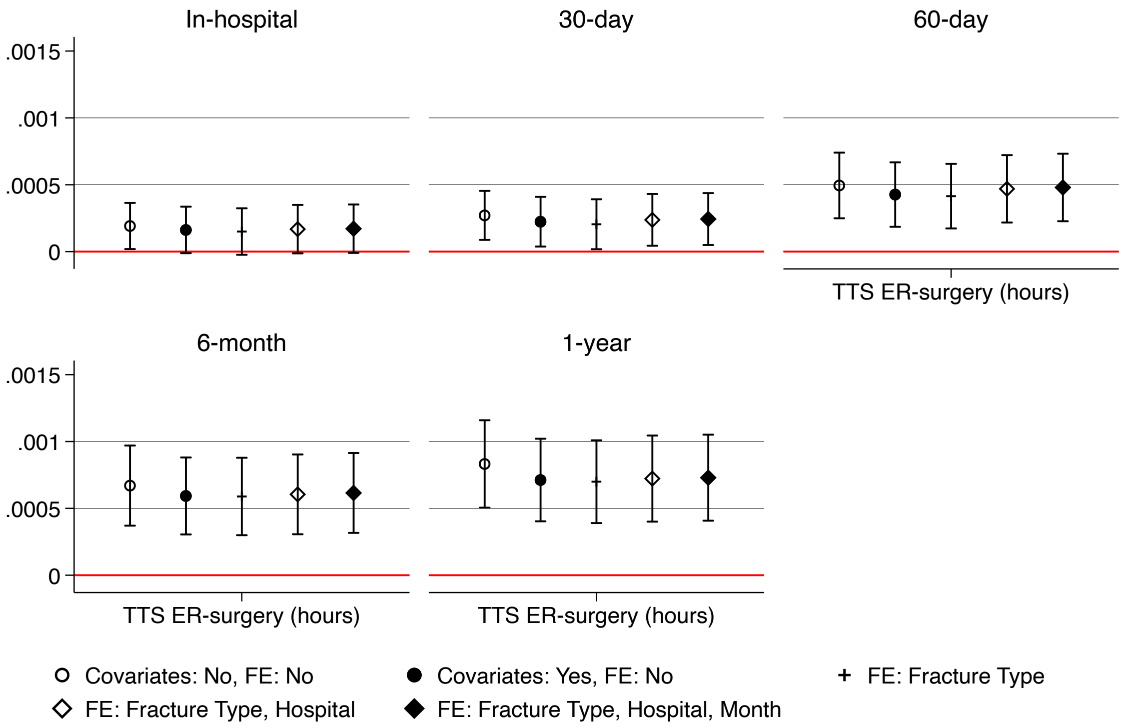


Figure 1.10: TTS (hours) estimates - Probit model, various specifications

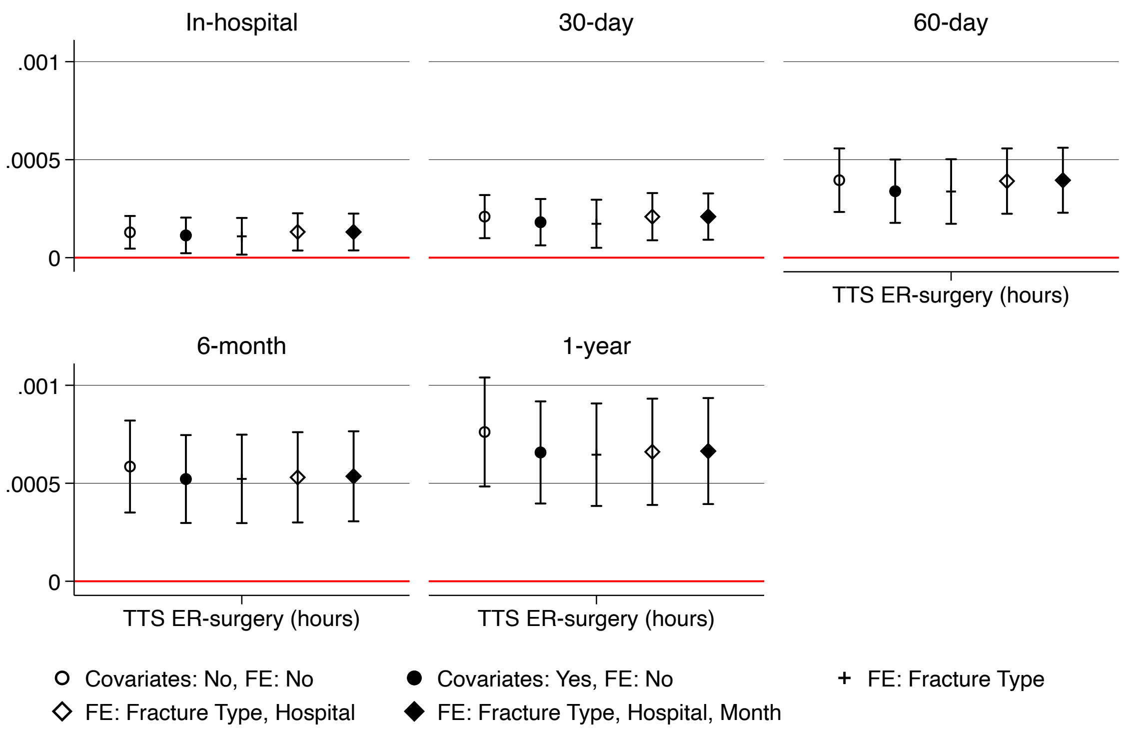


Figure 1.11: TTS (hours) estimates - IV model, various specifications

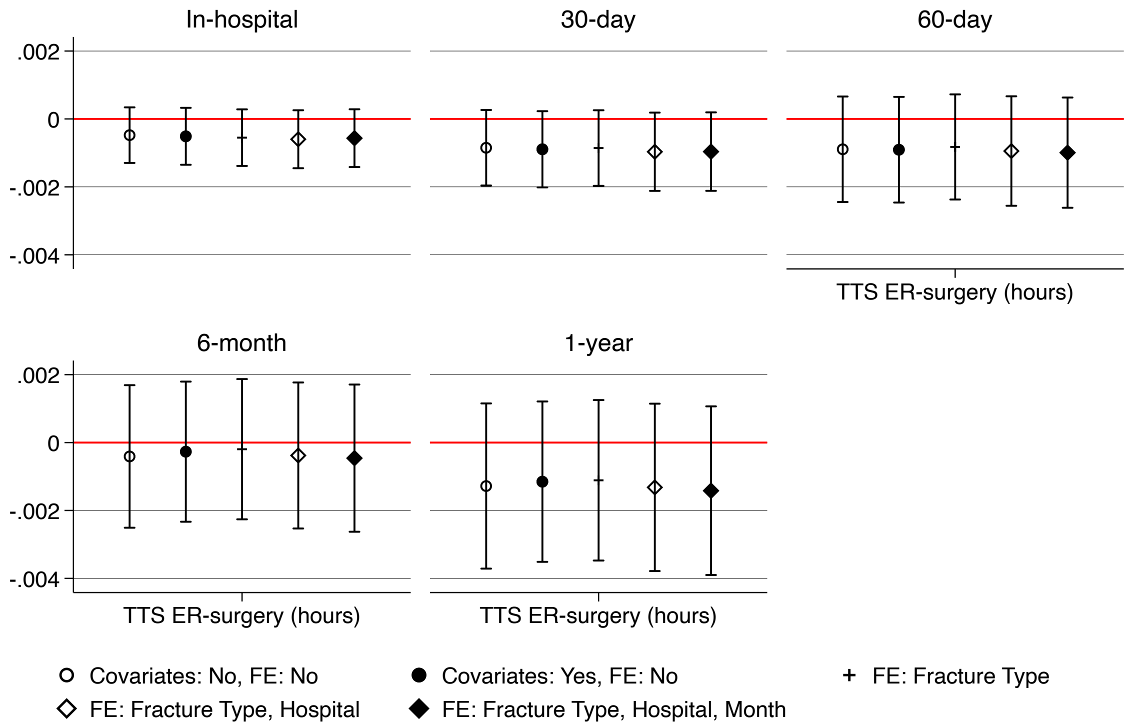


Figure 1.12: TTS (hours) estimates - IV probit model, various specifications

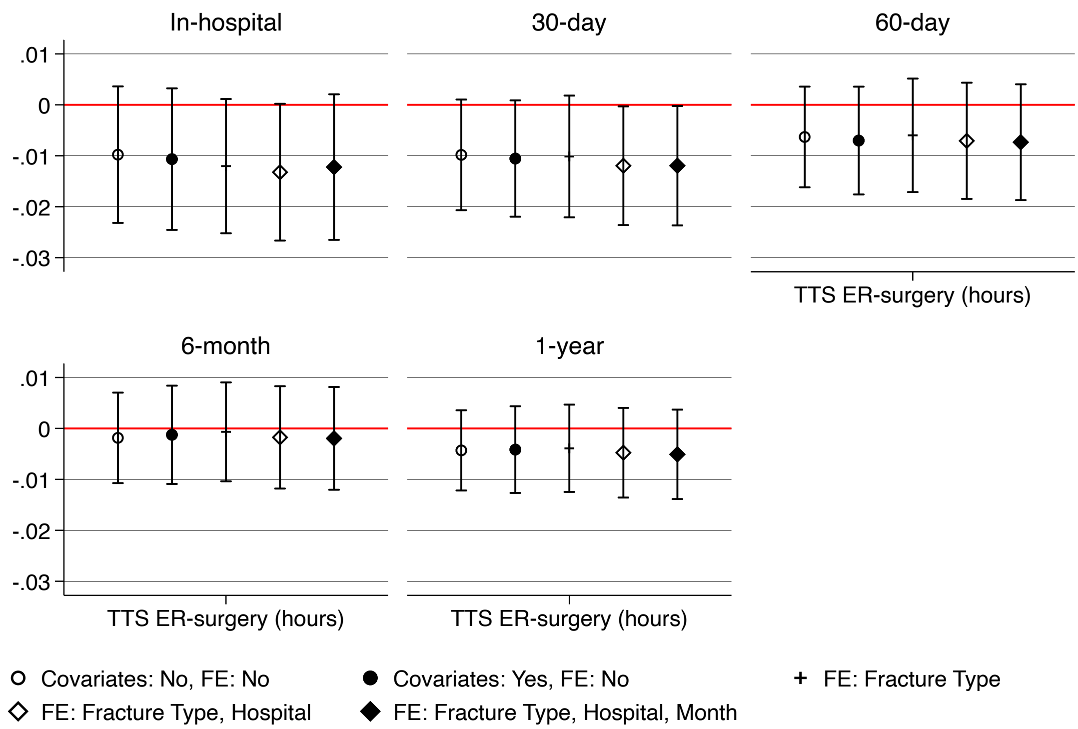


Figure 1.13: Evolution of TTS impact on mortality across years - LPM model

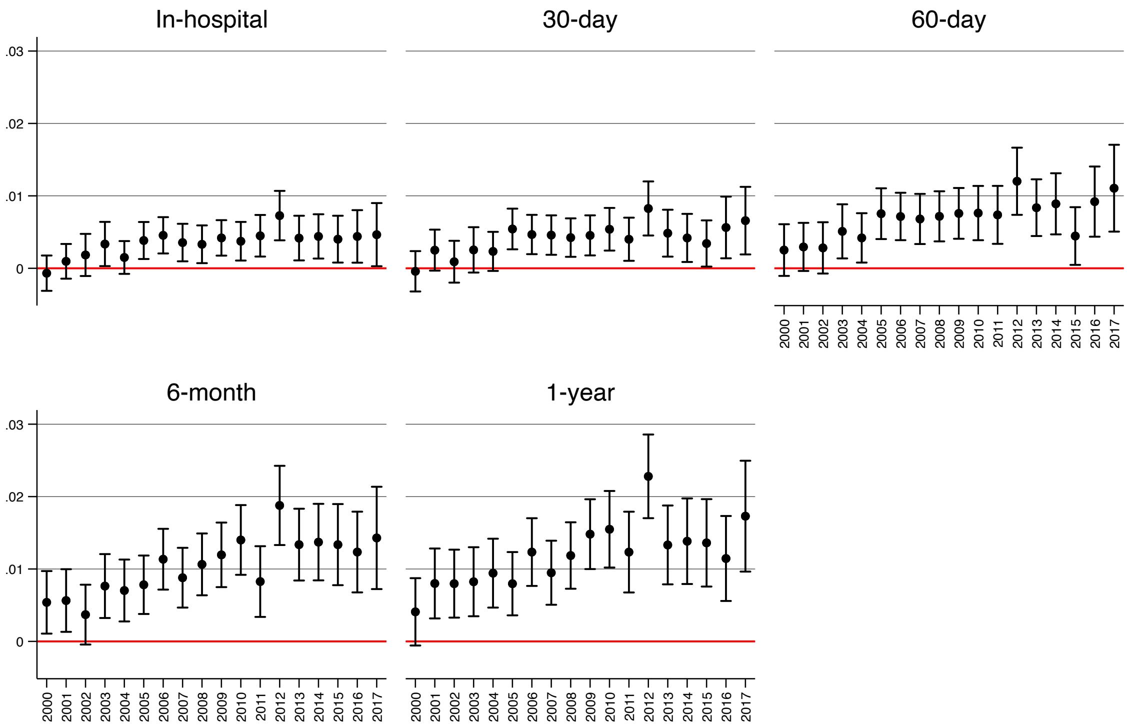


Figure 1.14: Evolution of TTS impact on mortality across years - IV model

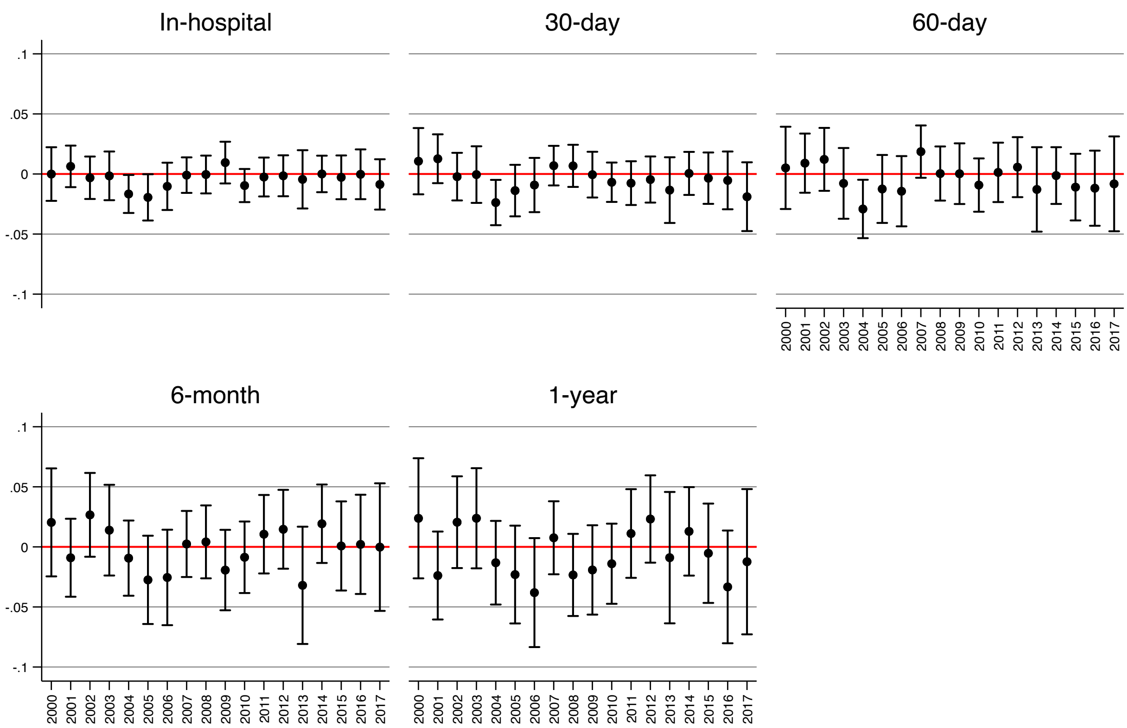


Figure 1.15: TTS impact on LoS - various specifications

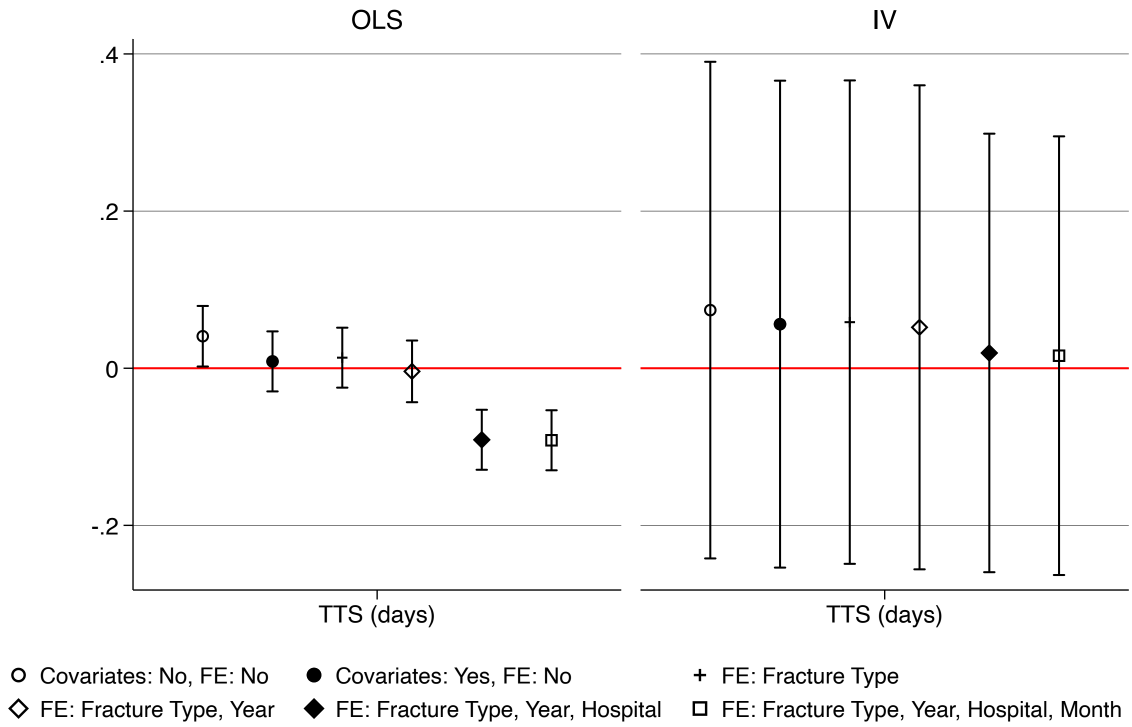
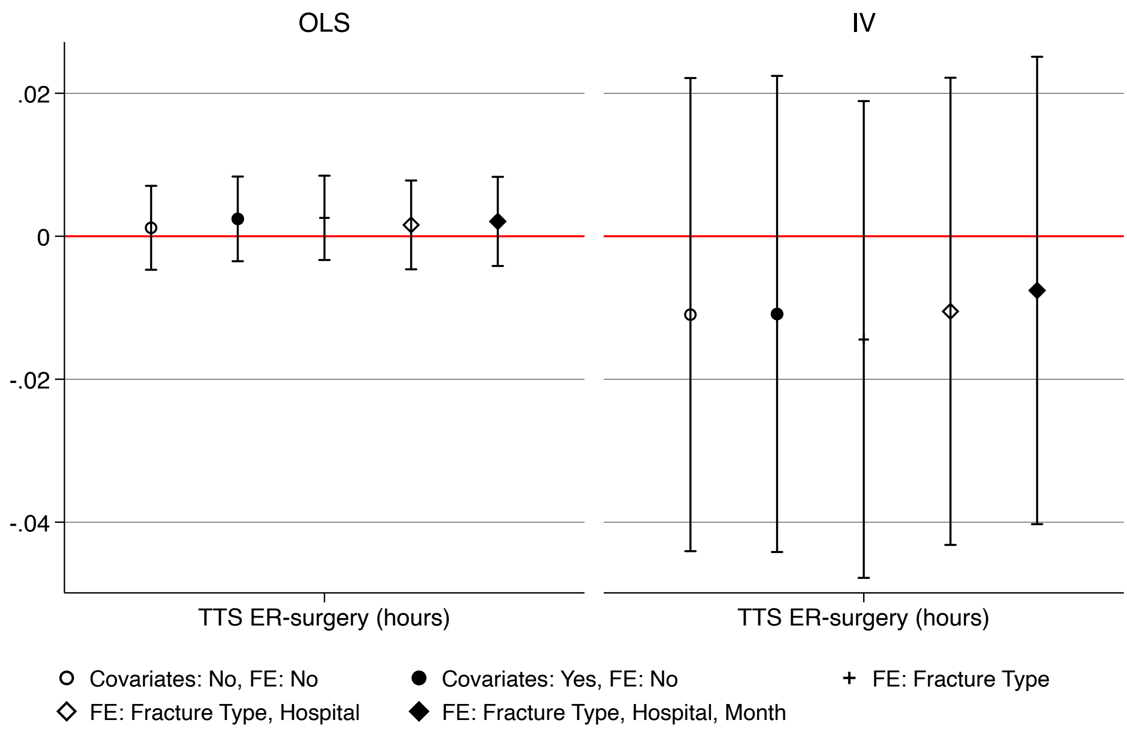


Figure 1.16: TTS (hours) impact on LoS - various specifications



Chapter 2

Volume-Outcome Relation in Elective Surgery An application to hip replacement in Italy[1]

Abstract

In the healthcare sector, the outcome-volume relationship can be explained by two dynamics: the scale economy mechanism, and the selective referral hypothesis. The former implies, the more a task is performed, the better the achieved outcomes, while the latter suggests hospitals with better performances attract more patients, thus leading to higher volumes of treatments. This study seeks to investigate the relation between the volume of elective surgeries and health outcomes for patients undergoing hip replacement surgery. In this context, patients can select in which hospital to be treated, making hospital choice nonrandom. This work attempts to disentangle the volume effect from the selective referral mechanism. To address patient nonrandom selection a conditional choice model is implemented where the probability of hospital choice is estimated based on patient and hospital characteristics and travel distance from residency to hospital. Patients are expected to choose the closest hospital only if the unobserved quality is homogeneous in the region or if the additional disutility of traveling further exceeds the utility of higher hospital quality, and observing longer traveled distance signals higher perceived hospital quality. The estimated probability of choosing each hospital is used to predict a measure of hospital volume purged of the selective referral component, which then is included in the health outcome equation. Findings suggest the presence of both volume effect and selective referral. Policies aiming to aggregate provision may be effective, but caution needs to be posed to the dynamics of these two contrasting mechanisms.

[1]I am grateful to Daniele Fabbri, Annalisa Loviglio and Chiara Monfardini for the essential feedback. This paper also benefited from participants' comments at the EuHEA Seminar Series. In particular, I would like to thank Nils Gutacker for the precious remarks.

2.1 Introduction

In the healthcare sector, analyzing the volume-outcome relation is particularly relevant. The presence of scale economies would justify policy implementation aiming at aggregating the provision of treatments and the introduction of minimum volume standards. Also, if higher volume favors positive health outcomes, this would promote the patients' well-being and reduce the costs of care due to, for example, earlier discharges or lower probability of readmission for complications. Two contrasting dynamics may explain the volume-outcome relation: the scale economy effect and the selective referral hypothesis. The former implies that the more a task is performed, the better the achieved outcomes, implying a volume effect. While the latter reverses the direction of the effect, suggesting hospitals with better performances attract more patients, leading to higher volumes of treatments. Hence, reputation may have a role in increasing hospital volume. The coexistence of these two contrasting mechanisms generates an issue of reverse causality when attempting to estimate the effect of hospital volume on health outcomes. This study aims to investigate the relationship between the volume of elective surgeries and health outcomes for patients undergoing hip replacement surgery. The focus on this type of operation is justified by the fact that hip replacement is one of the most common surgical operations performed, especially in the elderly (OECD, 2019). Furthermore, elective surgeries are operations that are scheduled in advance. Hence, patients have the opportunity to choose the admitting hospital, making the selection nonrandom. This work attempts to disentangle the two mechanisms to estimate the unbiased effect of surgical volume on patient health outcomes and recommend adequate policies for the organization of healthcare provision.

Similarly to Rachet-Jacquet, Gutacker, and Siciliani (2021), the analysis exploits a conditional choice model, where the probability of hospital selection is estimated based on patient and hospital characteristics and travel distance from patient residency to each hospital to address nonrandom patient selection. Patients are expected to choose the closest hospital only if the unobserved quality is homogeneous in the region or if the disutility of traveling further to a higher quality hospital exceeds the utility of the extra quality. Thus, traveling longer distances signals higher net utility with respect to choosing the closest hospital. The estimated probability of choosing each hospital is used to compute a hospital volume measure that accounts for the selective-referral component. The predicted, exogenous hospital volume is then included in the health outcome equation to obtain an unbiased estimation of the volume-outcome relation. This study exploits a hospital discharge records dataset of elective hip replacement surgeries performed in the Emilia-Romagna region (Italy) in 17 years. The dataset is complemented with additional information obtained by combining other data sources. Complications during hospitalization, post-surgery Length of Stay (LoS), 1-year mortality, 3-year failure of the prosthesis, and 30-day adverse outcomes (either mortality, failure, or readmission due to complications) are the outcome variables considered. Although studies suggest patients may be less responsive to raw hospital quality measures, several outcomes are considered to mitigate the possible null effects. Also, estimated effects may be regarded as a lower-bound.

Results suggest the presence of both volume and selective referral effects. After controlling for reverse causality presence, the surgical volume is found to have a significant and positive relation on all outcomes considered apart from 30-day adverse outcomes. Also, the magnitude is relevant. No effect is detected for the observed endogenous hospital volume instead. Estimates for different functional forms of hospital volume generally support the main results. The available data prevents a thorough control of patient health status, although findings are robust for several subsets of individuals, such as patients over 50 or the exclusion of those undergoing the operation due to congenital pathologies. However, the robustness checks show that the most significant effects are found for short-term outcome variables, while only a 10%-level significance is generally found for 1-year mortality and 3-year prosthesis failure. Findings contrast with previous studies and suggest both volume and selective referral effects are present. The positive volume effect may justify the introduction of provision aggregation policies. However, these decisions can not avoid considering the selective referral component. Further, evidence is likely to be type-of-surgery specific, and additional analyses are needed to assess whether results hold in other contexts.

The rest of the work is organized as follows. The general framework and relevant literature are presented in the next section. Section 2.3 illustrates the identification strategy exploited in the analysis, while the description of the dataset and the related descriptive statistics are exposed in Section 2.4. Results and robustness checks are reported in Section 2.5 and 2.6 respectively. Section 2.7 concludes the paper.

2.2 Background

The volume-outcome relationship spreads from the learning-by-doing hypothesis, which implies knowledge increases the more a task is performed, thus leading to outcome improvement. Formerly, this idea was investigated in the industrial sector (Arrow, 1962), in particular with application to aircraft production (Benkard, 2000), and a generally positive association has been found. Later, the volume-outcome relation has also been researched in other fields, sometimes with a more scale economy approach. The learning-by-doing hypothesis is based on the accumulation of knowledge that increases dynamically across time periods, while scale economies imply a more static approach to the volume-outcome relation.

In the healthcare sector, whether performed procedures volume is positively associated with patient health outcomes was investigated. In this setting, analyzing the existence of this relation is relevant because of the associated implications. The presence of a positive relationship would justify the introduction of policies aiming at regionalizing treatments provision and minimum volume standards introduction – according to which the hospital has to perform a specified number of surgeries in a year to continue to do so. Also, if higher surgical volume favors positive health outcomes, this relation improves patients' well-being and reduces healthcare costs through earlier discharges or lower probability of readmission due to complications. However, the association may work in the opposite direction, as stated by the selective referral hypothesis. This mechanism suggests that hospitals with better performances – e.g. lower mortality rates – attract more patients, thus leading to higher treatments volume. Consequently, hospital reputation may play a role in increasing hospital volume. The presence of these conflicting effects causes an endogeneity issue due to reverse causality when estimating the econometric model, which may yield biased measures of volume effect on health outcomes. Accordingly, the policies implemented in light of the volume effect may not be particularly effective in achieving the desired result, being selective referral in place. For this reason, it is crucial to investigate the presence of both effects and understand which dominates the other to adopt conforming policies.

Luft, Bunker, and Enthoven (1979) were among the firsts to examine the volume-outcome relation in the healthcare realm, finding a positive association between hospital volumes and health outcomes for several surgical procedures, thus suggesting aggregation of healthcare providers should be encouraged. The authors acknowledged either the volume effect or the selective referral effect might drive the results. However, they were not able to disentangle the two. Subsequently, Luft, Hunt, and Maerki (1987) tried to capture these effects to uncover whether one dominates the other, estimating a simultaneous equation model. The authors found that both selective referral and volume effect are present, and the magnitude of the two varies with the type of surgery considered. Additionally, they determined emergency care is less likely to be subject to selective referral since the patient may not have the time or possibility to choose the hospital of admission to and typically is taken to the closest one.

Later, several other studies have attempted to investigate the volume-outcome relation, using different estimation techniques in several settings and for various surgical operations. However, results are not conclusive and also depend on the rigorousness of the statistical methodology. In their systematic review of the literature, Halm, Lee, and Chassin (2002) highlighted a significant positive relationship between surgical volumes and health outcomes exists for several treatment procedures, even if the effect is less pronounced after adjusting for patient risk, and magnitude is varying according to the treatment considered. In addition, the systematic review by Mesman, Westert, Berden, and Faber (2015), attempting to assess the factors influencing the volume-outcome relation, concluded that the majority of studies have poor methodological rigor, which may bias the findings. Ho (2002) found a limited influence of hospital surgical volume effect on health outcomes for PTCA[2]. For the same procedure, Kansagra, Curtis, and Schulman (2004) found that regionalization of provision does not increase the travel distance for most patients, though not considering other cost dimensions unrelated to travel. B. H. Hamilton and Hamilton (1997) analyzed the volume-outcome relationship for hip fracture surgeries, considered as an urgent condition, exploiting within-hospital variation in the volume of surgeries performed and introducing hospital fixed effects to account for time-invariant quality. As the volume impact is found insignificant, the authors concluded that the volume-outcome relation reflects differences between high and low-quality hospitals instead of volume effects. Kim, Wolff, and Ho (2016) compared different estimation methods for the volume-outcome relationship for several cancer procedures, finding that logistic regression is always rejected in favor of models controlling for unobserved time-invariant heterogeneity, namely fixed and random effects models. However, the authors were not able to establish the direction of the association.

Many studies attempting to assess this relationship have failed to account for patient hospital choice nonrandomness. The central issue arises from the fact that hospital selection is correlated with unobserved patient severity and perceived hospital quality making hospital volume endogenous. These factors have to be considered for estimated effects to be unbiased. Gaynor, Seider, and Vogt (2005) tried to account for the endogeneity of hospital volume, using an instrumental variable approach where the number of competing hospitals within a certain radius and the number of resident patients is used as an instrument for hospital volume. Although the instruments are found to be relevant, the authors are unable to reject the exogeneity of surgical volume, thus concluding that the direction of the causal relation goes from volume to outcomes. Also, they determined that the current volume, namely the volume of surgery of the previous

[2]Percutaneous Transluminal Coronary Angioplasty

year, has a more consistent impact compared to lagged volume. Therefore, the authors affirmed that the relationship reflects a static scale economy mechanism instead of dynamic learning-by-doing. Other studies have used a similar approach to address the endogeneity concern. A possible solution was proposed by Gowrisankaran and Town (1999), exploiting the exogenous variation generated by the distance from patient residency to each hospital as an instrumental variable for hospital choice. Indeed, Luft et al. (1990) found that distance is a relevant predictor of hospital choice. Assuming unobserved patient severity is distributed homogeneously in the population, patient distance to a given hospital is correlated only with the endogenous hospital choice variables. Gowrisankaran and Town (1999) found the IV model corrects for hospital selection bias, and hospital quality estimates – i.e. mortality rate for patients with pneumonia – are significantly different from those obtained with a GLS estimation. Later, hospital distance has been used in other studies to attempt controlling for the endogeneity of hospital volume. Tsai, Votruba, Bridges, and Cebul (2006) used the linear distance between the patient and hospitals as an instrumental variable for hospital volume. After controlling for the endogeneity, results support the selective referral hypothesis. A similar approach was implemented by Goldstein, Babikian, Rana, Mackenzie, and Millar (2016) for total hip arthroplasty. However, findings do not allow to reject the exogeneity of surgical volume. Subsequently, Kessler and McClellan (2000) proposed the implementation of a hospital choice model based on travel distance – specifically, the differential distance from the closest hospital – as the exogenous determinant of choice to measure the effect of hospital competition on patient health outcomes. Similarly, Geweke, Gowrisankaran, and Town (2003) used hospital distance as an exogenous predictor in a multinomial choice model.

This study seeks to investigate the relation between the volume of elective surgeries and patient health outcomes for patients undergoing hip replacement surgery. The research focuses on this kind of operation precisely because elective surgeries are programmed in advance. Therefore, patients have the time and possibility of selecting the admitting hospital, making hospital choice nonrandom. Additionally, total hip replacement surgery is one of the most performed treatments among the elderly, thus policy-relevant. Following the model presented by Kessler and McClellan (2000) on hospital competition, as further developed by Rachet-Jacquet et al. (2021), this analysis attempts to implement a model of patient hospital choice. The patient chooses which hospital to be admitted to based on observable individual characteristics, hospital features, and the travel distance to each hospital in the region. The estimated probability of choosing the hospital is used to compute a predicted volume measure for each hospital in the dataset. This measure of hospital volume is purged of endogenous factors, such as hospital quality and unobserved patient severity. Then, the outcome regression is estimated through the expected hospital volume of surgeries. This work differs from that presented by Rachet-Jacquet et al. (2021) in two main aspects. First, the health outcomes that are investigated come from administrative data and are not based on patients' surveys that may be subject to self-perception bias. Although studies have found measures of health gain to be more relevant for patient hospital choice, compared to raw hospital quality measures (Gutacker, Siciliani, Moscelli, & Gravelle, 2016), the study attempts to mitigate the possible limitations coming from focusing on administrative data by considering a series of health outcomes other than mortality and revision rate. Further, the possible estimated effect that is found may be interpreted as a lower bound. The health outcomes considered are patient mortality, prosthesis failure, complications during the hospitalization, post-surgery Length of Stay (LoS), and 30-day adverse health outcomes. Second, a larger period of 17 years is considered, which gives the possibility of controlling for fluctuations in hospital volume and aggregation of providers which may have occurred during the years.

2.3 Model Framework and Identification Strategy

This study aims to estimate the impact of hospital surgical volume on health outcomes for patients undergoing elective surgery. In particular, the analysis focuses on hip replacement surgery, which is one of the most common surgical operations performed in the elderly. It is a non-emergency, programmable treatment that leaves the patient the faculty to choose the admitting hospital in advance. The linear health outcome model to be estimated is specified as follows:

$$y_{iht} = \alpha + \delta_1 Q_{ht} + \mathbf{X}'_{ih}\delta_2 + \mathbf{H}'_{ht}\delta_3 + \mathbf{T}'_{ht}\delta_4 + \upsilon_{iht} \tag{2.1}$$

where y_{iht} a variable for patient health outcome, Q_{ht} is hospital volume of hip replacement surgeries performed in the 12 months before patient admission, $\mathbf{X}_{ih}$ is a vector of patient's characteristics (age, gender, type of diagnosis, hip side operated, and average gross yearly income at municipality of residence level), $\mathbf{H}_{ht}$ is a vector of hospital characteristics (ownership, whether it is a teaching hospital, and the number of beds in orthopedic ward), while $\mathbf{T}_{ht}$ are years fixed effects. The model is estimated as a Linear Probability Model for several binary outcome variables: complications during hospitalization, 1-year mortality, 3-year prosthesis failure, and 30-day adverse outcome (either mortality, failure, or readmission). The model is also estimated through OLS for the outcome variable of post-surgery LoS in days.

When estimating the health outcome model, the main concern is the endogeneity of hospital volume due to the reverse causality. Precisely, higher surgical volume can improve hospital health outcomes, yet patients may choose to be treated in the hospital with higher perceived quality, thus increasing the volume of surgery performed in hospitals with better performance. As a result, patient choice is nonrandom but depends, instead, on exogenous observable variables and some not fully measurable variables, such as patient severity and hospital quality. The patient hospital choice model is implemented similarly to Rachet-Jacquet et al. (2021). Patient i utility from choosing a hospital h can be stated as:

$$u_{iht} = V_{iht} + \varepsilon_{iht} \tag{2.2}$$

where V_{iht} is patient i observed utility from choosing hospital h at time t, and ε_{iht} is the unobserved utility. Assuming unobserved patient severity is distributed homogeneously across the population, patients are expected to choose the closest hospital only if hospital quality is also homogeneous on the territory. Alternatively, if the patient is observed traveling a longer distance, this reflects a higher perceived quality of the chosen hospital. This unobserved factor has to be considered when estimating the model to obtain unbiased results of the volume-outcome relation. Patients are free to choose the admitting hospital, and no established guidelines are present for GPs to refer patients to specialized hospitals. While patient priority may differ according to diagnosis and health conditions, the corresponding waiting time is not expected to vary significantly across hospitals.

The observed component of patient i utility from choosing hospital h is specified as:

$$V_{iht} = \beta_1 D_{iht} + \beta_2 D^2_{iht} + \mathbf{Z}'_{ht}\beta_z + \sum_{j=1}^{J} \mathbf{X}_{ijt}\left(\gamma_{1j} + \gamma_{2j} D_{iht} + \gamma_{3j} D^2_{iht}\right) + \beta_3 T_{ht} + \beta_4 T^2_{ht} + \beta_5 T^3_{ht} \tag{2.3}$$

where D_{ih} is the travel distance in minutes from patient municipality of residence to each hospital municipality of the region, a quadratic term is also included to capture possible non-linearities, $\mathbf{Z}_{ht}$ is a vector of hospital characteristics, specifically the number of beds in the orthopedic ward, whether the hospital is the closest to the patient, and the number of competing hospital in the same municipality, $\mathbf{X}_{it}$ is a vector of patient characteristics (age, gender, type of diagnosis, and average gross yearly income at municipality of residence level), which is also interacted with the travel distance to control for differences in preferences related to individual characteristics (Kansagra et al., 2004). Finally, T_{ht} is the year time trend, included as a cubic polynomial to account for non-linear relation in a parsimonious way. The model is estimated as a conditional logit (McFadden, 1973), through maximum likelihood, where the estimated probability that patient i chooses hospital h is given by:

$$\hat{p}_{iht} = \frac{exp(\hat{V}_{iht})}{\sum_{h \in M_{it}} exp(\hat{V}_{iht'})} \tag{2.4}$$

where M_{it} is patient choice set, that varies with years, since some hospitals shut down, while others starts performing this surgery. The estimated probability of choosing hospital h is used to compute predicted volume for hospital h as follows:

$$\hat{Q}_{ht} = \sum_{i=1}^{N} \hat{p}_{iht} \tag{2.5}$$

This estimated measure of hospital volume of surgery accounts for patient selective referral, and it is used in 2.1 to estimate the exogenous volume-outcome relation. Section 2.5 reports both results for the model 2.1 estimation of the endogenous volume-outcome relation, with standard errors clustered at hospital level, and the estimation of the exogenous model 2.1 with the introduction of $\hat{Q}_{ht}$, with standard errors bootstrapped (100 replications) to account for the fact that hospital volume is an estimated measure. The model is also estimate with the inclusion of hospital fixed-effects to control for time-invariant characteristics that may correlate with distance.

2.4 Data and Sample Composition

The analysis is performed exploiting administrative data from the RIPO[3] database. The dataset contains detailed information about Total Hip Arthroplasties (THA)[4] performed from 2000 to 2016 in the Emilia-Romagna region, in the northern part of Italy[5]. In particular, resident patients between 18 and 100 treated in public or private hospitals of the Region are considered. Because of the long panel dimension of the dataset, patients may undergo surgery more than once. For those patients, only the first operation is considered, thus excluding subsequent treatments. Additionally, patients with a diagnosis of hip fracture are excluded from the dataset. The reason for this choice is that hip fracture is an urgent condition that requires quick intervention. Consequently, in this setting, the patient does not have the time to choose the admitting hospital

[3]Registro dell'Implantologia Protesica Ortopedica (Register of the Orthopaedic Prosthetic Implants). http://ripo.cineca.it/authzssl/index.htm

[4]Also known as Total Hip Replacement (THR)

[5]Emilia-Romagna is one of the most populous Italian Regions, with almost 4.5 million resident citizens in 2016, amounting to about 13% of the Italian population.

and she is taken to the closest one instead. Accordingly, the selective referral argument does not hold. Data are complemented with information about patient residency at municipality level, retrieved from hospital discharge records - SDO[6] database. Using this information, coupled with the municipality in which the hospital is located, the travel distance in minutes from the patient municipality of residence to each possible admitting hospital of the Region is computed exploiting the ISTAT[7] commuting matrix. The travel distance expressed in minutes gives a more accurate measure compared to that in meters since, for instance, patients located in remote territories such as mountainous areas need more time to cover the same distance in meters due to roads' slopes and curves. The number of hospital beds in the orthopedic ward is retrieved from regional open data and used as a proxy for hospital capacity. Information about the monthly volume of hip surgeries performed in each hospital from 2000 to 2016 is obtained from the RIPO database. Only THA surgeries are considered when measuring hospital volume. Results for a more comprehensive measure of surgical volume, including also partial hip replacement, hip resurfacing, and revision surgery, are presented as a robustness check in Section 2.6.1. In this sense, even if total hip replacement constitutes the primary type of surgery performed (Table 2.1), the measure of volume considered in the main estimation captures only the effect of replicating and practicing the very same surgery, and the effects may be regarded as lower-bound.

The heat plot depicted in Figure 2.2 shows the evolution of annual surgical volume per hospital in the considered period. The blank rectangles indicate that the hospital is not treating any patient in that year. The number of available hospitals varies across years, as some ceased activity and others started in the study period. Consequently, patient choice set size increases from a minimum of 56 hospitals in 2000 to a maximum of 60 hospitals in the 2007-2009 period and decreases again in subsequent years (Figure 2.1). Two main details are evident from Figure 2.2. First, there are no wide fluctuations in the number of surgeries performed by hospitals across years. Being hospital capacity somehow fixed may be a possible explanation. Instead, a slight and general increase in surgical volume across almost all hospitals through the years is evident. This phenomenon is consistent with the increasing national trend in the number of THA surgeries performed (AGENAS, 2020). Second, one hospital ($ID = 61$) is an obvious outlier: its observed surgical volume is disproportionately higher than that of all other hospitals in the dataset. More precisely, this is a specialized orthopedic hospital with an average volume of 1,106 surgeries performed per year against 129 surgeries on average for all other hospitals in the

[6]Schede di Dimissione Ospedaliera
[7]Italian National Statistical Institute

dataset. On one side, being a specialized hospital may attract more severe patients with a higher probability of adverse health outcomes. On the other side, experiencing such a large volume of surgeries may have significantly improved hospital quality and expertise. In either case, to avoid results being driven by the presence of the outlier, the analysis is performed excluding it. The final dataset consists of 48,383 observations, corresponding to an average of more than 2,840 patients treated per year, though the trend is increasing.

The health outcome variables considered are complications during hospitalization – either intra- or post-surgery –, post-surgery LoS measured in days, 1-year patient mortality, and 3-year failure of the prosthesis. Additionally, only for the years 2012 to 2016, the 30-day adverse outcome – either mortality, prosthesis failure, or readmission – is considered. Table 2.2 reports the unconditional probability of patient health outcomes and the probability per quartile of hospital volume in the 12 months before patient treatment. The probability of incurring complications during hospitalization is higher for hospitals in the low-volume quartile and decreases as the volume quartile increases. The same evidence holds for post-surgery LoS and 3-year failure of the prosthesis, and the difference is statistically significant. Instead, no clear pattern is evident for 30-day adverse outcomes and 1-year mortality. Also, the probabilities are not jointly statistically different across the volume quartiles. Although reverse causality is not addressed, and the relationship may be simply spurious, this preliminary evidence supports the presence of a volume effect.

Average patient characteristics for the overall dataset and per quartile of travel distance from the chosen hospital are presented in Table 2.3. There are statistically significant differences among the groups for all variables considered. In particular, the older the patient, the closer the chosen hospital. Wealthier individuals are also admitted to closer hospitals, even though this pattern may indicate that cities in which hospitals are located have a higher average income per capita. Differences in distance traveled are also present according to gender, hip side operated, and the type of diagnosis leading to hip replacement surgery. Evidence of patient selection into hospitals is clear. Considering hospital characteristics per quartile of distance traveled by patients, there are statistically significant differences based on hospital affiliation, the number of competing hospitals in the same municipality, and the number of beds in orthopedic wards. However, for this latter, the relation is not linear (Table 2.4). A possible explanation may be that hospitals are generally located in bigger cities, thus a large number of patients can avoid traveling long distances. Table 2.5 highlights that patients choosing a hospital that is not the closest go on average to hospitals with a higher volume of surgeries for those facilities belonging to the center of the volume distribution. Also, more than 42% of patients are admitted to a hospital that is not the closest, hence patients' mobility is non-negligible. This fact is evident also from Figure 2.3, which plots the evolution of patient mobility across provinces. Though data are aggregated at the province level for presentation sake, and several hospitals are located within each province, this plot displays how patient medium-distance mobility has changed in the years. Provinces are ordered geographically from west to east on both the x and y axes. While in the first years considered, patients tend to generally move up to neighboring provinces, in more recent years, they tend to travel also to more distant locations.

Summarizing this preliminary evidence, although the endogeneity of hospital volume is unaccounted for, not only patients choosing closer hospitals are different from those traveling further, the admitting hospitals' types are also diverse. Moreover, patient mobility towards hospitals beyond the closest to residency is a significant phenomenon.

2.5 Results

The predicted surgical volume estimated through the conditional logit model appears to be highly correlated with the observed (endogenous) volume: the correlation coefficient is 0.806. Results for the volume-outcome relation are reported in Table 2.6 for all dependent variables of interest, both for observed and predicted hospital volume. The surgical volume is measured as the total number of surgeries (in hundreds) performed in the hospital h 12 months before the patient's treatment. The impact of observed (endogenous) volume is generally not significant. The only exception is for patient complications during hospitalization, whose effect is negative and statistically significant. When controlling for patients' selective referral, the predicted hospital volume effect on the health outcomes is negative and strongly statistically significant for complications and post-surgery LoS. Also, coefficients' magnitude is larger in absolute value and non-negligible. In particular, a hundred increase in hospital volume decreases by almost 40% the probability of incurring complications, by 17.41% the length of post-surgery hospital stay (2 days less), and by 10% the probability of prosthesis failure 3-year after surgery. Results for post-surgery LoS need to be interpreted with caution since they may capture the tendency of high-volume hospitals to discharge patients earlier to rehabilitation or long-term care facilities. The effect of predicted hospital volume is only significant at 10% level, though with large magnitude, for 1-year patient mortality. No effect is found for 30-day adverse outcomes in either specification. These findings suggest the presence of a relevant volume effect after accounting for the reverse causality issue. The presence of some selective referral is supported by the no significant impact found for observed surgical volume.

The presence of a possible quadratic relation between hospital volume and patient outcome is also investigated. However, strong significance is found only for the effect of predicted hospital volume on post-surgery LoS (Table 2.7). The positive coefficient for the squared volume indicates that as the number of surgeries increases, the negative effect on post-surgery LoS becomes smaller. The same result is obtained for the observed hospital volume, though the magnitude of the coefficients is lower. The estimated coefficients for the other health outcomes regressions are not significant.

To check for the presence of additional nonlinearities in the volume-outcome relationship, the analysis is also performed including a set of dummy variables for four different volume bands: $Q < 50$, $50 \leq Q < 100$, $100 \leq Q < 150$, $Q \geq 150$. The volume bands are constructed over the total number of surgeries performed 12 months before patient treatment. Estimates are reported in Table 2.8. In general, the effect of predicted hospital volume increases, in absolute values, with volume band. However, significance varies depending on the outcome variable considered. In particular, a 1%-level significance is found for patients' complications for each volume band, while for 3-year prosthesis failure and 1-year mortality, only the highest band is statistically significant, though only at 10% level for the latter. Results for post-surgery LoS are mixed: the observed hospital volume bands are significant, with increasing magnitude, while the effect size of predicted volume has an inverse u-shaped pattern, with the middle band having the highest impact. Still, no effect is found for 30-days adverse outcomes for either specification. These findings generally support the scale economy hypothesis.

Also, estimates are computed for the measure of monthly hospital volume. As expected, the magnitude of the coefficients is reduced compared to the main findings, due also to the different variable scaling (Table 2.9). Nevertheless, results are significant for the effect of hospital volume on complications and post-surgery LoS, with a relevant magnitude. In fact, one extra surgery performed in the current month decreases the probability of complications by 3.6% and the post-surgery hospital stay by 1.5%.

To check the results' consistency the model is estimated with the inclusion of the outlier hospital. Results are reported in Table 2.10. The estimated coefficient magnitude is generally lower, while significance is almost unchanged. Therefore, results are not largely affected by the exclusion of the outlier. The lower magnitude may suggest that the outlier hospital attracts more severe patients with a higher probability of adverse health outcomes.

Further, the analysis is performed with the inclusion of hospital fixed effects. This enables to control for unobserved time-invariant characteristics specific to the hospital. Results are reported in Table 2.11. No significant effect of observed hospital volume is found on all patient health outcomes. Estimates for predicted hospital volume effect are significant only for complications and post-surgery LoS. In particular, the former has changed the sign and is found to have a positive impact on in-hospital complications. These findings suggest that once accounting for hospital unobserved time-invariant characteristics, such as time-invariant quality, no volume effect is generally found, both for observed and predicted one.

These findings support the idea that selective referral presence is curbing the effect of hospital volume on patients' health outcomes, which emerges after accounting for patient selection. This may be explained by the fact that more severe patients seek care in better hospitals. These findings contrast with evidence from Rachet-Jacquet et al. (2021). Also, evidence suggests that the effect of exogenous hospital volume is more relevant for short-term outcomes, such as intra- and post-surgery complications, while generally a 10%-level significance or no effect is found for 3-year prosthesis failure and 1-year mortality. However, these findings may be in line with the fact that patients are less responsive to raw quality measures.

2.6 Robustness Checks and Extensions

2.6.1 Including other hip surgery types

The main results assess the presence of a volume effect of total hip replacement surgeries on patients outcomes. However, knowledge can also be acquired by performing other closely related types of surgery (Schilling, Vidal, Ployhart, & Marangoni, 2003). Hence, practicing different types of hip surgery may contribute to improving the learning process, and estimations including only the volume of total hip replacement surgeries may underestimate the effect on patient health outcomes. Therefore, the analysis is replicated with the inclusion of partial hip replacement, hip revision, hip prosthesis removal, and hip resurfacing surgeries in the measure of observed hospital volume. These are also elective surgeries that account for 34% of all the operations involving hip prosthesis performed in Emilia-Romagna in the period of the study (Table 2.1). Results do not vary significantly from the main findings (Table 2.12). The magnitude of the observed hospital volume is lower, while significance is generally unchanged. The effect of predicted hospital volume is similar to that of the main findings for all outcome variables considered. Evidence suggests these surgeries may be less relevant in expertise acquisition for performing total hip replacement surgery.

2.6.2 Different Patient Selections

One limitation of the proposed analysis is the inability to fully control for patient heterogeneity in pre-surgery health status that may affect hospital choice. For instance, patients with particular conditions may be referred to specialized hospitals. In order to mitigate this issue and check the robustness of the main findings, the analysis is performed for two subsets of individuals.

The data used in the main analysis includes observations for all patients between 18 and 100 years old. However, young patients may have different health conditions compared to the elderly. Also, younger people can better bear surgery and generally have a faster recovering ability. To estimate hospital volume effect on the elderly the analysis is performed on the subset of patients over 50 years old. This subgroup constitutes 93.86% of observations in the overall dataset, although not surprisingly since hip replacement is a surgical operation typically undergone by the elderly. Evidence from Table 2.13a shows that the significance of the volume effect is identical to that found in the main analysis, though the magnitude of the effect is slightly higher. This finding means that the volume effect has a larger impact on older patients.

Further, the dataset is restricted, excluding patients undergoing total hip replacement due to congenital pathologies. This subgroup of patients constitutes 10% of the main data and partly overlaps the subset of younger individuals under 50 (more than 20%). In fact, patients with this diagnosis are typically treated earlier than the average patient with primary coxarthrosis. Also, the inability to fully control for patient health status may result in the estimation being biased. To check the robustness of the results, the analysis with the exclusion of this subgroup of patients is performed. Estimated coefficients are again larger in absolute value than that of the main analysis (Table 2.13b). Also, the effect of predicted hospital volume on 1-year mortality is significant at 5% level.

Evidence suggests that including patients under 50 and with congenital pathologies leading to hip replacement mitigates the volume effect, which is largely significant for the rest of the sample.

2.6.3 Causal Effect Evolution in Years

The availability of a particular large longitudinal dimension allows assessing whether and how the causal effect of surgical volume varies over the years. For instance, it may be that bigger hospitals uptake new technologies and procedures more quickly. However, after an initial period of comparative advantage, the technology may spread to smaller hospitals that may catch up with larger ones in terms of outcomes. To check for this possibility the causal relationship between volume and health outcomes is estimated over several subsets of a 3-year period each. Results for patient health outcome of interest are shown in Figure 2.4, which plots point estimates and 95% confidence intervals for the observed and predicted volume of surgeries. No results are reported for 30-day adverse health outcomes as coefficients are not significant for all previous estimations. The impact of observed surgical volume is stable and generally not statistically significant across years. On the other hand, the predicted hospital volume effect has different patterns depending on the outcome variable considered. The volume effect on complications during hospitalization is negative and significant starting from the period 2003-2005. The effect magnitude is increasing in absolute values, though the difference does not seem statistically significant. The predicted hospital volume effect on post-surgery LoS is found significantly decreasing with years, although a positive jump is estimated for the last period (2012-2014). Finally, no significant effect is found for 1-year mortality and 3-year failure of the prosthesis. These pieces of evidence do not support the initial hypothesis of diminishing volume effect in years. Instead, it shows that for the most responsive outcome variables the volume effect is improving.

2.6.4 Learning Effect

Finally, the relation between the cumulative hospital volume and patient health outcome is estimated. The measure of cumulative volume is the number of surgical operations performed in all the periods before the current month. Therefore, it is updated at a monthly rate. This relation is just a raw estimate of the hospital learning effect across years. As this estimation is still preliminary, no forgetting is introduced in the model, and caution should be made in interpreting the results, as accumulated knowledge is assumed to have the same impact across several years without any discounting factors. Results are reported in Table 2.14. Though the effect of predicted cumulative hospital volume is very limited in magnitude, it has maintained the significance found for the main results. This preliminary evidence points toward the presence of some learning effect, though more thorough modeling has to be carried out to assess its presence.

2.7 Conclusions

This study analyses the effect of hospital volume of elective surgery on patient health outcomes. In particular, the aim is to disentangle the two opposite mechanisms that may influence this relation: the scale economy mechanism and the selective referral effect. The former implies knowledge increases the more a task is performed, thus leading to outcome improvement, while, according to the latter, hospitals with better health outcomes attract more patients, thus increasing the volume of surgeries. The coexistence of these two contrasting effects is a peculiarity of programmed surgical operations since patients can decide in advance the admitting hospital. Patients' nonrandom hospital choice needs to be considered when estimating the volume-outcome relation to avoiding biased effect estimations. The analysis exploits a hospital choice model where patient travel distance from residency to the hospital is the main proxy for hospital quality. The key assumption is that if hospital quality were distributed homogeneously in the territory, patients would choose the closest facility to avoid higher disutility from traveling further. Consequently, observing a patient admitted to a hospital beyond the closest one reflects the higher quality attached to it. The estimated probability of choosing each hospital is used to predict hospital volume purged of the selective referral component. Finally, the effect of the exogenous surgical volume on patient health outcomes is estimated.

While no association between observed hospital volume and health outcome is found, the predicted volume of surgeries has generally a large and significant effect on the patient outcomes. These findings support the hypothesis that both volume effect and selective referral are present. The non-significant effect of observed hospital volume may be explained by the fact that more severe patients seek care in hospitals with higher quality. Additionally, they contrast with results from previous literature, suggesting no volume effect on health outcomes exists. Results are tested against different specifications, volume functional forms, and for several sub-samples of patients. While the volume effect on post-surgery LoS may be biased by the impossibility to control for patient discharge destination, the large and significant estimated coefficients on complications signal the presence of a volume effect after controlling for the selective referral mechanism. Moreover, the causal relation of hospital volume is generally determined for short-term outcomes, while small or no effect on 1-year mortality and 3-year prosthesis failure is found. This evidence may support previous literature results suggesting patients are less subject to raw measures of hospital quality. In this respect, estimates may be regarded as a lower-bound. The positive volume effect may justify the introduction of healthcare provision aggregation policies. However, these decisions can not avoid considering the selective referral component.

The analysis is limited by the fact that with available data is not possible to account for patient severity thoroughly, and more accurate data on hospital distance may be necessary to refine the estimation. Also, results may be type-of-surgery specific, and further assessment is necessary to understand the volume-outcome relation in other contexts.

2.8 Tables and Figures

2.8.1 Tables

Table 2.1: Total number of surgeries per type (period 2000-2016)

Surgery type	No.	%
Total Hip Replacement	103,789	63,49%
Partial Hip Replacement	39,197	23,98%
Hip Revision Surgery	15,728	9,62%
Hip Prosthesis Removal	2,761	1,69%
Hip Resurfacing	1,314	0,80%
Other	695	0,43%
Total	163,484	100%

Table 2.2: Outcome variables per quartiles of yearly hospital volume

	Total mean (sd)	Low mean (sd)	Low-Mid mean (sd)	Mid-High mean (sd)	High mean (sd)	F-stat for joint mean difference
Complications	0.0359	0.0427	0.0372	0.0375	0.0260	16.41***
	(0.186)	(0.202)	(0.189)	(0.190)	(0.159)	
Post-surgery LoS (days)	12.59	14.70	12.51	11.62	11.53	334.82***
	(8.723)	(9.674)	(8.584)	(8.069)	(8.078)	
30-day adverse outcome	0.0272	0.0294	0.0263	0.0300	0.0237	1.27
	(0.163)	(0.169)	(0.160)	(0.171)	(0.152)	
1-year mortality	0.0132	0.0142	0.0119	0.0142	0.0127	0.33
	(0.114)	(0.118)	(0.108)	(0.118)	(0.112)	
3-year failure	0.0610	0.0663	0.0604	0.0593	0.0577	2.84**
	(0.239)	(0.249)	(0.238)	(0.236)	(0.233)	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Complications during hospitalization are considered. The variable 30-day adverse outcome is available only for the period 2012-2016, with 14,895 observations. Volume quartiles are computed over the hospital volume of surgeries performed in the 12 months before patient treatment.

Table 2.3: Patients characteristics per quartile of distance traveled

	Total mean (sd)	Close mean (sd)	Mid-Close mean (sd)	Mid-Far mean (sd)	Far mean (sd)	F-stat for joint mean difference
Age	69.01	70.10	69.32	68.68	67.59	72.04***
	(10.16)	(9.816)	(9.738)	(10.16)	(10.68)	
Female	0.599	0.617	0.578	0.590	0.593	19.96***
	(0.490)	(0.486)	(0.494)	(0.492)	(0.491)	
Right side	0.578	0.584	0.563	0.579	0.574	4.52**
	(0.494)	(0.493)	(0.496)	(0.494)	(0.494)	
Hospital distance (minutes)	16.58	0	10.31	18.48	42.27	15131.49***
	(18.81)	(0)	(2.171)	(3.343)	(17.76)	
Extra time (minutes)	9.378	0	2.102	6.386	30.00	1198.58***
	(16.44)	(0)	(4.211)	(7.485)	(20.61)	
Log income (thousands)	9.929	9.992	9.915	9.899	9.875	1231.97***
	(0.176)	(0.160)	(0.166)	(0.169)	(0.185)	
Diagnosis:						
Primary coxarthrosis	0.820	0.828	0.842	0.818	0.801	8.83***
	(0.384)	(0.378)	(0.365)	(0.386)	(0.399)	
Congenital pathologies	0.0922	0.0846	0.0800	0.0933	0.109	5.40***
	(0.289)	(0.278)	(0.271)	(0.291)	(0.311)	
Femoral head necrosis	0.0645	0.0627	0.0581	0.0661	0.0689	2.29*
	(0.246)	(0.243)	(0.234)	(0.248)	(0.253)	
Rheumatoid arthritis	0.00769	0.00896	0.00517	0.00761	0.00728	4.57**
	(0.0873)	(0.0943)	(0.0717)	(0.0869)	(0.0850)	
Other	0.00940	0.00964	0.00836	0.00976	0.00926	0.52
	(0.0965)	(0.0977)	(0.0911)	(0.0983)	(0.0958)	
Observations	48,383	17,626	6,577	12,087	12,093	

*** p<0.01, ** p<0.05, * p<0.1

Note: The logarithm of the average yearly gross income at the municipality of residence level is considered. Extra time is the difference between the total traveled time from patient residency to hospital municipality and the distance between patient residency and the municipality of the closes hospital, in minutes.

Table 2.4: Hospital characteristics per quartile of distance traveled

	Total mean (sd)	Close mean (sd)	Mid-Close mean (sd)	Mid-Far mean (sd)	Far mean (sd)	F-stat for join mean difference
No. beds	33.57	37.22	30.46	32.50	31.01	366.42***
	(19.96)	(20.71)	(15.81)	(20.03)	(20.00)	
Affiliation:						
Public	0.588	0.531	0.758	0.586	0.582	520.39***
	(0.492)	(0.499)	(0.428)	(0.493)	(0.493)	
Teaching	0.133	0.185	0.0833	0.124	0.0931	257.32***
	(0.339)	(0.388)	(0.276)	(0.329)	(0.291)	
Private	0.279	0.284	0.158	0.291	0.325	223.51***
	(0.448)	(0.451)	(0.365)	(0.454)	(0.468)	
No. hospitals:						
1	0.426	0.299	0.662	0.435	0.474	1393.82***
	(0.495)	(0.458)	(0.473)	(0.496)	(0.499)	
2	0.0986	0.111	0.0680	0.107	0.0892	52.42***
	(0.298)	(0.314)	(0.252)	(0.309)	(0.285)	
3	0.261	0.336	0.104	0.246	0.251	699.43***
	(0.439)	(0.472)	(0.306)	(0.431)	(0.434)	
4	0.0955	0.108	0.0753	0.104	0.0809	30.07***
	(0.294)	(0.310)	(0.264)	(0.305)	(0.273)	
5+	0.119	0.147	0.0900	0.109	0.105	92.42***
	(0.324)	(0.354)	(0.286)	(0.312)	(0.306)	
Observations	48,383	17,626	6,577	12,087	12,093	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The number of hospitals located in the same municipality is considered. The number of beds in orthopedic ward is reported.

Table 2.5: Average yearly volume per hospital distance & volume quartile

Volume Quartile	Beyond mean (sd)	Closest mean (sd)	Total mean (sd)
Low	47.73	49.07	48.55
	(17.24)	(16.27)	(16.66)
Medium-Low	95.47	94.09	94.72
	(14.23)	(14.40)	(14.34)
Medium-High	144.40	143.78	144.06
	(15.80)	(15.79)	(15.80)
High	231.16	231.93	231.63
	(54.14)	(50.46)	(51.93)
Total	128.92	129.49	129.25
	(71.70)	(75.20)	(73.74)
Observations	20,457	27,926	48,383

Note: Volume quartiles are computed over the distribution of hospital volume of surgeries performed in the 12 months before patient treatment.

Table 2.6: Estimates of observed and predicted hospital volume

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Yearly volume (hundreds)	-0.00899**	-0.0144***	-1.215	-2.188***	-0.000166	-0.00164	-0.000992	-0.00254*	-0.00257	-0.00624**
	(0.00356)	(0.00217)	(0.857)	(0.115)	(0.00186)	(0.00354)	(0.000896)	(0.00139)	(0.00233)	(0.00302)
Female	-0.000339	-0.000289	0.842***	0.852***	-0.00105	-0.00103	-0.00575***	-0.00573***	-0.0200***	-0.0200***
	(0.00280)	(0.00168)	(0.143)	(0.0838)	(0.00245)	(0.00283)	(0.00121)	(0.00114)	(0.00256)	(0.00228)
Age	-0.00154*	-0.00153*	-0.252***	-0.252***	-0.00440***	-0.00440***	-0.00316***	-0.00316***	-0.00827***	-0.00824***
	(0.000805)	(0.000835)	(0.0526)	(0.0343)	(0.00126)	(0.00154)	(0.000478)	(0.000498)	(0.00103)	(0.000974)
Age^2	1.53e-05**	1.52e-05**	0.00285***	0.00284***	3.81e-05***	3.81e-05***	3.04e-05***	3.04e-05***	8.03e-05***	8.01e-05***
	(6.20e-06)	(6.18e-06)	(0.000397)	(0.000263)	(9.91e-06)	(1.18e-05)	(3.86e-06)	(4.00e-06)	(8.32e-06)	(7.61e-06)
Log income (thousands)	-0.00799	-0.00431	1.275	1.929***	0.00318	0.00343	0.00114	0.00187	-0.00857	-0.00631
	(0.0108)	(0.00746)	(2.022)	(0.360)	(0.0102)	(0.0104)	(0.00554)	(0.00398)	(0.0116)	(0.00879)
Right side	-0.00373*	-0.00376**	0.0262	0.0270	0.00475	0.00475	0.000648	0.000644	0.000264	0.000204
	(0.00194)	(0.00167)	(0.0902)	(0.0779)	(0.00295)	(0.00298)	(0.000897)	(0.00112)	(0.00193)	(0.00201)
Hospital type, Teaching	-0.00488	-0.00204	-1.762	-1.342***	0.00275	0.00299	0.00480	0.00534**	0.00508	0.00622
	(0.0125)	(0.00281)	(1.473)	(0.188)	(0.00525)	(0.00560)	(0.00383)	(0.00235)	(0.00871)	(0.00462)
Hospital type, Private	-0.00551	-0.00703***	-0.974	-1.237***	0.00369	0.00340	-0.00265**	-0.00292***	0.00119	0.000261
	(0.00623)	(0.00212)	(1.107)	(0.0783)	(0.00326)	(0.00315)	(0.00117)	(0.00112)	(0.00370)	(0.00251)
No. beds	9.69e-05	4.35e-05	0.0267	0.0223***	-0.000151**	-0.000137*	-8.16e-06	-4.61e-06	1.40e-05	2.34e-05
	(0.000192)	(5.84e-05)	(0.0286)	(0.00253)	(7.20e-05)	(7.96e-05)	(3.80e-05)	(3.80e-05)	(9.69e-05)	(7.25e-05)
Constant	0.151	0.121	4.041	-1.502	0.115	0.113	0.0805	0.0746*	0.342***	0.321***
	(0.108)	(0.0763)	(19.91)	(3.647)	(0.111)	(0.107)	(0.0534)	(0.0406)	(0.115)	(0.0847)
Observations	45,585	45,532	45,584	45,531	14,791	14,791	45,585	45,532	45,585	45,532
R-squared	0.003	0.003	0.046	0.049	0.005	0.005	0.018	0.018	0.021	0.021
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0361	0.0361	12.57	12.57	0.0272	0.0272	0.0134	0.0134	0.0612	0.0612

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Observed and predicted yearly volume is the number of surgeries performed by hospital h in the previous 12 months, in hundreds. The logarithm of the average yearly gross income (in thousands) at the municipality of residence level is considered. Standard errors in columns 1, 3, 5, 7, and 9 are clustered at the hospital level, while they are bootstrapped (100 replications) in columns 2, 4, 6, 8, and 10.

Table 2.7: Estimates of observed and predicted hospital volume - Quadratic function

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Yearly volume (hundreds)	-0.0121	-0.0142*	-6.761***	-8.902***	-0.000938	0.00737	-0.00616	-0.00185	-0.00913	-0.00266
	(0.00825)	(0.00751)	(1.128)	(0.288)	(0.00553)	(0.0139)	(0.00380)	(0.00580)	(0.00654)	(0.00938)
Yearly volume (hundreds)^2	0.00101	-8.13e-05	1.770***	3.371***	0.000212	-0.00452	0.00162*	-0.000629	0.00209	-0.00179
	(0.00242)	(0.00333)	(0.280)	(0.145)	(0.00113)	(0.00638)	(0.000951)	(0.00273)	(0.00157)	(0.00415)
Observations	45,585	45,532	45,584	45,531	14,791	14,791	45,585	45,532	45,585	45,532
R-squared	0.003	0.003	0.069	0.059	0.005	0.005	0.014	0.014	0.021	0.021
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0361	0.0361	12.57	12.57	0.0272	0.0272	0.0252	0.0252	0.0612	0.0612

*** p<0.01, ** p<0.05, * p<0.1

Note: Observed and predicted yearly volume is the number of surgeries performed by hospital h in the previous 12 months, in hundreds. Standard errors in columns 1, 3, 5, 7, and 9 are clustered at the hospital level, while they are bootstrapped (100 replications) in columns 2, 4, 6, 8, and 10. Covariates include age, age^2, gender, average yearly gross income at the municipality of residence level (in log), side of the hip treated, hospital type, number of beds in the orthopedic ward.

Table 2.8: Estimates of observed and predicted hospital volume - Different volume bands

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$50 \leq$ Volume < 100	-0.00391	-0.0104***	-1.002*	-2.507***	-0.00741	0.000439	-0.00269	-0.00113	-0.000922	-0.00320
	(0.00507)	(0.00223)	(0.551)	(0.102)	(0.00599)	(0.00379)	(0.00345)	(0.00183)	(0.00501)	(0.00302)
$100 \leq$ Volume < 150	-0.00382	-0.0184***	-3.241***	-3.564***	-0.00712	0.00248	-0.00455	-0.00199	-0.00550	-0.00270
	(0.00809)	(0.00281)	(0.687)	(0.109)	(0.00643)	(0.00450)	(0.00351)	(0.00216)	(0.00546)	(0.00328)
Volume ≥ 150	-0.0167**	-0.0249***	-3.636***	-1.680***	-0.00644	-0.00559	-0.00232	-0.00524*	-0.00302	-0.0108**
	(0.00670)	(0.00316)	(0.914)	(0.184)	(0.00643)	(0.00643)	(0.00350)	(0.00297)	(0.00544)	(0.00509)
Observations	45,585	45,532	45,584	45,531	14,791	14,791	45,585	45,532	45,585	45,532
R-squared	0.003	0.004	0.060	0.061	0.005	0.005	0.014	0.014	0.021	0.021
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0363	0.0363	12.16	12.16	0.0273	0.0273	0.0240	0.0240	0.0590	0.0590

*** p<0.01, ** p<0.05, * p<0.1

Note: Observed and predicted volume is the number of surgeries performed by hospital h in the previous 12 months. The baseline volume band is Volume < 50. Standard errors in columns 1, 3, 5, 7, and 9 are clustered at the hospital level, while they are bootstrapped (100 replications) in columns 2, 4, 6, 8, and 10. Covariates include age, age^2, gender, average yearly gross income at the municipality of residence level (in log), side of the hip treated, hospital type, number of beds in the orthopedic ward.

Table 2.9: Estimates of observed and predicted monthly hospital volume

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Monthly volume	-0.000691**	-0.00132***	-0.101	-0.192***	-0.000241	-0.000244	-3.79e-05	-3.14e-05	-0.000146	-0.000416
	(0.000281)	(0.000203)	(0.0617)	(0.0111)	(0.000176)	(0.000319)	(6.64e-05)	(0.000137)	(0.000191)	(0.000277)
Observations	48,099	48,099	48,097	48,097	14,791	14,791	48,099	48,099	48,099	48,099
R-squared	0.007	0.007	0.050	0.052	0.009	0.009	0.022	0.022	0.025	0.025
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0363	0.0363	12.16	12.16	0.0273	0.0273	0.0126	0.0126	0.0590	0.0590

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Observed and predicted monthly volume is the number of surgeries performed by hospital h in the current month. Standard errors in columns 1, 3, 5, 7, and 9 are clustered at the hospital level, while they are bootstrapped (100 replications) in columns 2, 4, 6, 8, and 10. Covariates include age, age^2, gender, average yearly gross income at the municipality of residence level (in log), side of the hip treated, hospital type, number of beds in the orthopedic ward.

Table 2.10: Estimates of observed and predicted yearly hospital volume with outlier inclusion

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Yearly volume (hundreds)	-0.00213	-0.00546***	-0.848**	-2.297***	0.00239	0.00288	-0.000796	-0.00330*	-0.00212	-0.00560*
	(0.00368)	(0.00184)	(0.412)	(0.0980)	(0.00160)	(0.00455)	(0.000895)	(0.00191)	(0.00143)	(0.00302)
Observations	51,017	50,964	51,016	50,963	16,813	16,813	51,017	50,964	51,017	50,964
R-squared	0.002	0.002	0.063	0.067	0.005	0.005	0.012	0.012	0.021	0.021
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0362	0.0362	12.15	12.15	0.0273	0.0273	0.0240	0.0240	0.0590	0.0590

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Observed and predicted yearly volume is the number of all hip surgeries types performed by hospitals h in the previous 12 months, in hundreds. Standard errors in columns 1, 3, 5, 7, and 9 are clustered at the hospital level, while they are bootstrapped (100 replications) in columns 2, 4, 6, 8, and 10. Covariates include age, age^2, gender, average yearly gross income (in thousands) at the municipality of residence level (in log), side of the hip treated, hospital type, number of beds in the orthopedic ward.

Table 2.11: Estimates of observed and predicted yearly hospital volume with hospital fixed effects

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Yearly volume (hundreds)	0.00567	0.0106***	-0.612	-1.255***	0.00272	0.0213	-0.00298	-0.00190	-0.00339	-0.00479
	(0.00577)	(0.00383)	(0.585)	(0.184)	(0.00904)	(0.0137)	(0.00212)	(0.00411)	(0.00354)	(0.00508)
Female	0.000744	0.000741	0.768***	0.771***	-0.00119	-0.00120	-0.00591***	-0.00588***	-0.0201***	-0.0201***
	(0.00264)	(0.00203)	(0.120)	(0.0760)	(0.00246)	(0.00296)	(0.00164)	(0.00153)	(0.00263)	(0.00220)
Age	-0.00121*	-0.00120	-0.295***	-0.296***	-0.00423***	-0.00424***	-0.00360***	-0.00357***	-0.00807***	-0.00805***
	(0.000721)	(0.000875)	(0.0508)	(0.0381)	(0.00125)	(0.00147)	(0.000689)	(0.000596)	(0.00104)	(0.000976)
Age^2	1.26e-05**	1.25e-05*	0.00316***	0.00316***	3.69e-05***	3.69e-05***	3.46e-05***	3.44e-05***	7.86e-05***	7.84e-05***
	(5.52e-06)	(6.68e-06)	(0.000390)	(0.000295)	(9.83e-06)	(1.12e-05)	(5.50e-06)	(4.76e-06)	(8.48e-06)	(7.55e-06)
Log income (thousands)	0.00335	0.00306	-1.351	-1.315***	0.00476	0.00470	-0.000387	-0.000112	-0.0112	-0.0111
	(0.00816)	(0.00919)	(1.106)	(0.399)	(0.0132)	(0.0141)	(0.00776)	(0.00715)	(0.0122)	(0.00964)
Right side	-0.00391**	-0.00392**	0.00963	0.00639	0.00466	0.00470*	0.00122	0.00119	0.000432	0.000359
	(0.00193)	(0.00177)	(0.0897)	(0.0833)	(0.00298)	(0.00244)	(0.00134)	(0.00134)	(0.00195)	(0.00202)
No. beds	-0.000614	-0.000589***	0.0245	0.0227***	-0.000194	-0.000287	-0.000104	-0.000157	-2.19e-05	-5.34e-05
	(0.000396)	(0.000138)	(0.0243)	(0.00455)	(0.000215)	(0.000322)	(8.72e-05)	(0.000109)	(0.000154)	(0.000175)
Constant	0.0392	0.0400	36.14***	36.02***	0.122	0.120	0.115	0.113	0.362***	0.361***
	(0.0822)	(0.0960)	(10.25)	(4.012)	(0.141)	(0.148)	(0.0773)	(0.0741)	(0.125)	(0.0974)
Observations	45,585	45,532	45,584	45,531	14,791	14,791	45,585	45,532	45,585	45,532
R-squared	0.017	0.017	0.146	0.147	0.009	0.010	0.016	0.016	0.024	0.024
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0361	0.0361	12.57	12.57	0.0272	0.0272	0.0252	0.0252	0.0612	0.0612

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Observed and predicted yearly volume is the number of surgeries performed by hospital h in the previous 12 months, in hundreds. Standard errors in columns 1, 3, 5, 7, and 9 are clustered at the hospital level, while they are bootstrapped (100 replications) in columns 2, 4, 6, 8, and 10.

Table 2.12: Estimates of observed and predicted yearly hospital volume of total hip surgeries

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Yearly volume (hundreds)	-0.00499*	-0.0143***	-1.110**	-2.187***	0.000769	-0.00163	-0.000458	-0.00254*	-1.81e-05	-0.00620**
	(0.00257)	(0.00220)	(0.518)	(0.109)	(0.00161)	(0.00376)	(0.000643)	(0.00140)	(0.00167)	(0.00285)
Observations	45,583	45,530	45,582	45,529	14,789	14,789	45,583	45,530	45,583	45,530
R-squared	0.003	0.003	0.049	0.049	0.005	0.005	0.018	0.018	0.021	0.021
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0361	0.0361	12.57	12.57	0.0272	0.0272	0.0134	0.0134	0.0612	0.0612

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Observed and predicted yearly volume is the number of all hip surgeries types performed by hospitals h in the previous 12 months, in hundreds. The hip surgery types included in the measure of hospital volume are total and partial hip prostheses, hip revision surgery, hip resurfacing, and hip prosthesis removal. Standard errors in columns 1, 3, 5, 7, and 9 are clustered at the hospital level, while they are bootstrapped (100 replications) in columns 2, 4, 6, 8, and 10. Covariates include age, age^2, gender, average yearly gross income (in thousands) at the municipality of residence level (in log), side of the hip treated, hospital type, number of beds in the orthopedic ward.

Table 2.13: Estimates of observed and predicted hospital volume - Several patient selection

(a) Patient over 50

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Yearly volume (hundreds)	-0.00930**	-0.0158***	-1.244	-2.323***	-0.000860	-0.00258	-0.000936	-0.00291*	-0.00245	-0.00657**
	(0.00364)	(0.00226)	(0.860)	(0.120)	(0.00200)	(0.00397)	(0.000953)	(0.00159)	(0.00240)	(0.00280)
Observations	43,428	43,377	43,428	43,377	14,002	14,002	43,428	43,377	43,428	43,377
R-squared	0.003	0.004	0.045	0.048	0.005	0.005	0.017	0.018	0.022	0.022
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0362	0.0362	12.65	12.65	0.0273	0.0273	0.0137	0.0137	0.0621	0.0621

*** p<0.01, ** p<0.05, * p<0.1

(b) Excluding patients with congenital pathologies

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Yearly volume (hundreds)	-0.00996***	-0.0150***	-1.317	-2.810***	-0.00193	-0.00488	-0.00145	-0.00377**	-0.00347	-0.00805**
	(0.00361)	(0.00254)	(0.863)	(0.134)	(0.00198)	(0.00433)	(0.000993)	(0.00177)	(0.00235)	(0.00330)
Observations	41,489	41,439	41,488	41,438	13,737	13,737	41,489	41,439	41,489	41,439
R-squared	0.003	0.003	0.047	0.052	0.005	0.005	0.018	0.018	0.021	0.021
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0356	0.0356	12.57	12.57	0.0264	0.0264	0.0141	0.0141	0.0631	0.0631

*** p<0.01, ** p<0.05, * p<0.1

Note: Observed and predicted yearly volume is the number of surgeries performed by hospital h in the previous 12 months, in hundreds. Standard errors in columns 1, 3, 5, 7, and 9 are clustered at the hospital level, while they are bootstrapped (100 replications) in columns 2, 4, 6, 8, and 10. Covariates include age, age^2, gender, average yearly gross income (in thousands) at the municipality of residence level (in log), side of the hip treated, hospital type, number of beds in the orthopedic ward.

Table 2.14: Estimates of observed and predicted cumulative hospital volume

	Complications		Post-surgery LoS		30-day adverse outcome		1-year mortality		3-year failure	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Cumulative volume	-8.69e-06*	-1.07e-05***	-0.00145**	-0.00140***	-6.48e-07	-1.65e-06	-1.72e-06	-1.61e-06	-1.89e-06	-3.81e-06*
	(4.73e-06)	(1.45e-06)	(0.000686)	(6.92e-05)	(2.08e-06)	(2.41e-06)	(1.51e-06)	(1.23e-06)	(2.12e-06)	(2.04e-06)
Observations	48,099	48,099	48,097	48,097	14,791	14,791	48,099	48,099	48,099	48,099
R-squared	0.003	0.003	0.047	0.048	0.005	0.005	0.014	0.014	0.021	0.021
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dependent	0.0361	0.0361	12.57	12.57	0.0272	0.0272	0.0252	0.0252	0.0612	0.0612

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The cumulative volume at monthly rate is considered. Standard errors in columns 1, 3, 5, 7, and 9 are clustered at the hospital level, while they are bootstrapped (100 replications) in columns 2, 4, 6, 8, and 10. Covariates include age, age^2, gender, average yearly gross income at the municipality of residence level (in log), side of the hip treated, hospital type, number of beds in the orthopedic ward.

2.8.2 Figures

Figure 2.1: Number of available hospitals in patient choice set per years

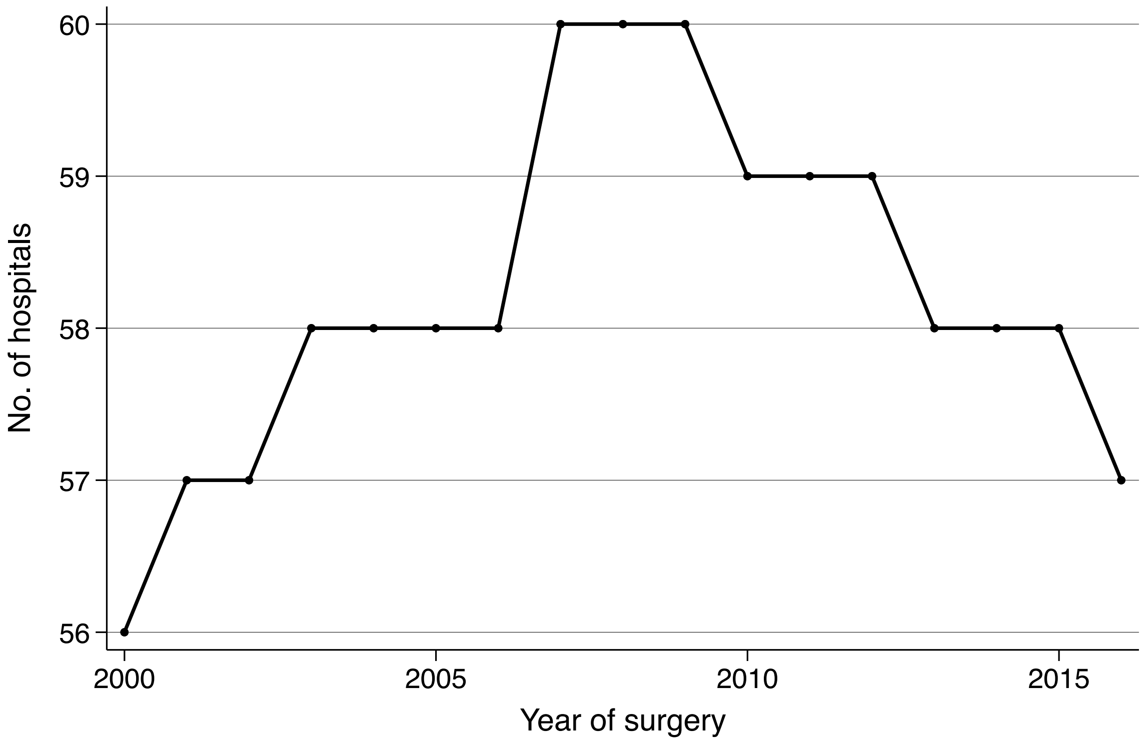


Figure 2.2: Observed yearly hospital volume per year

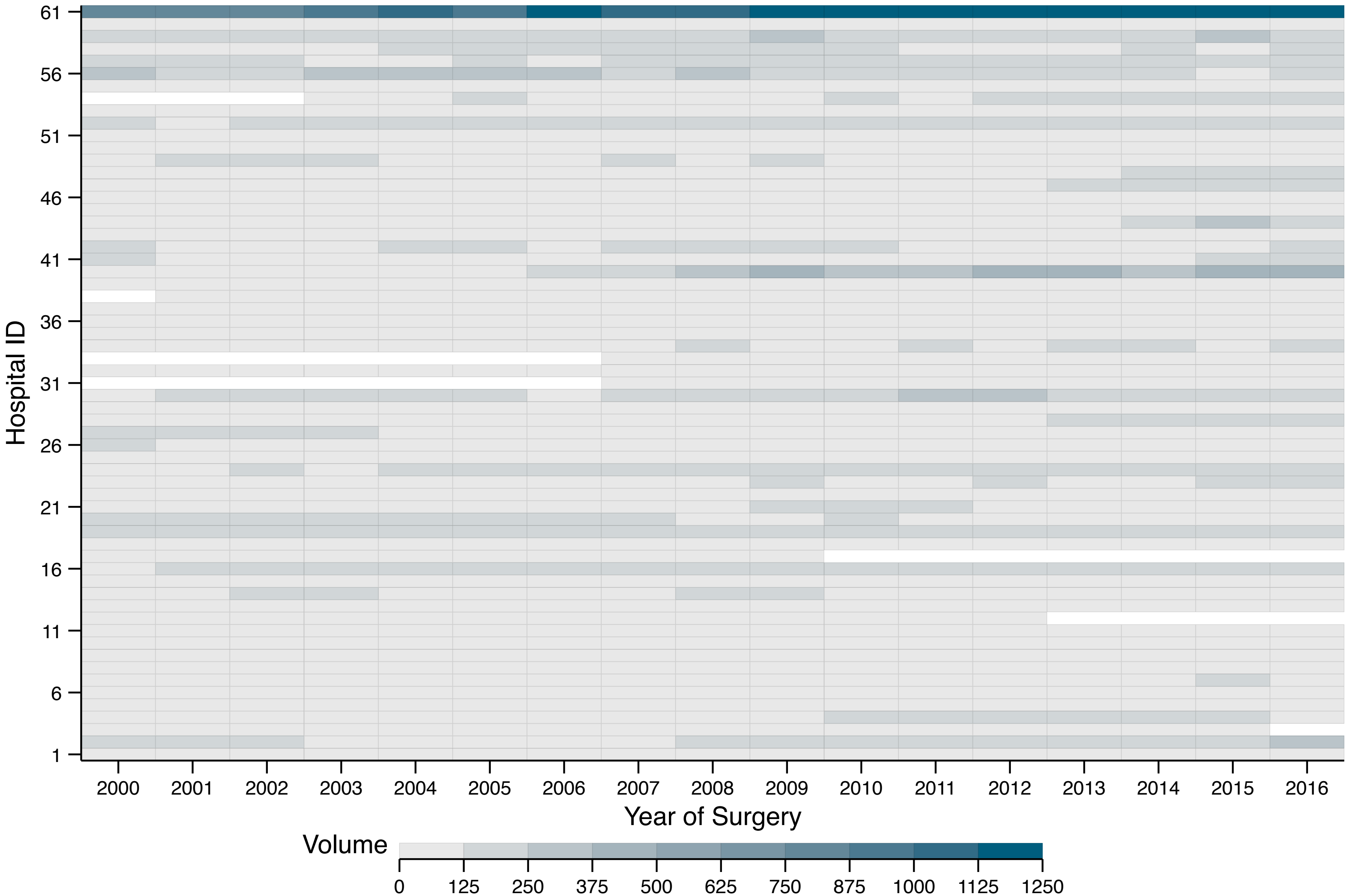


Figure 2.3: Evolution of inter-province patient mobility across years

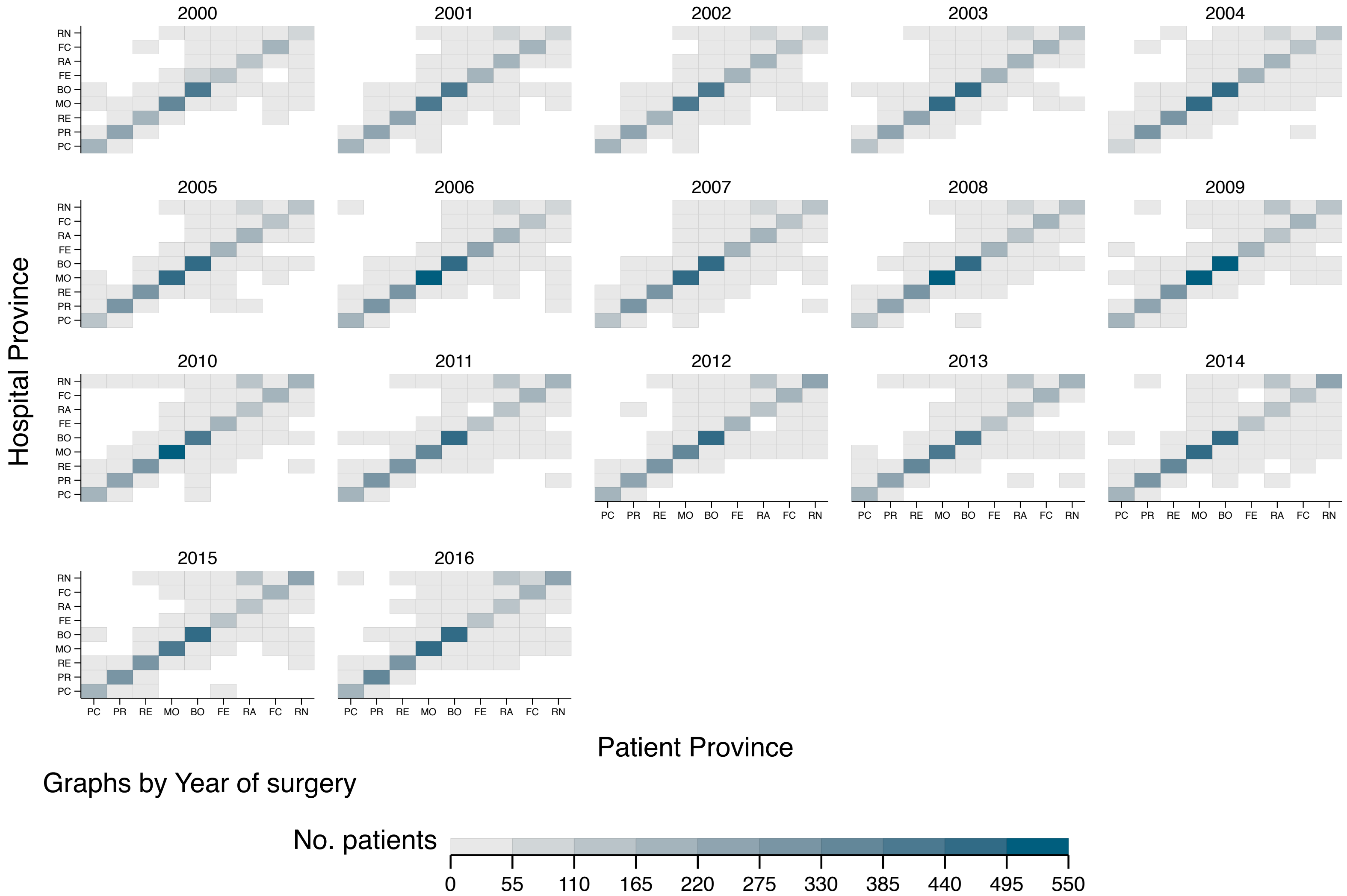
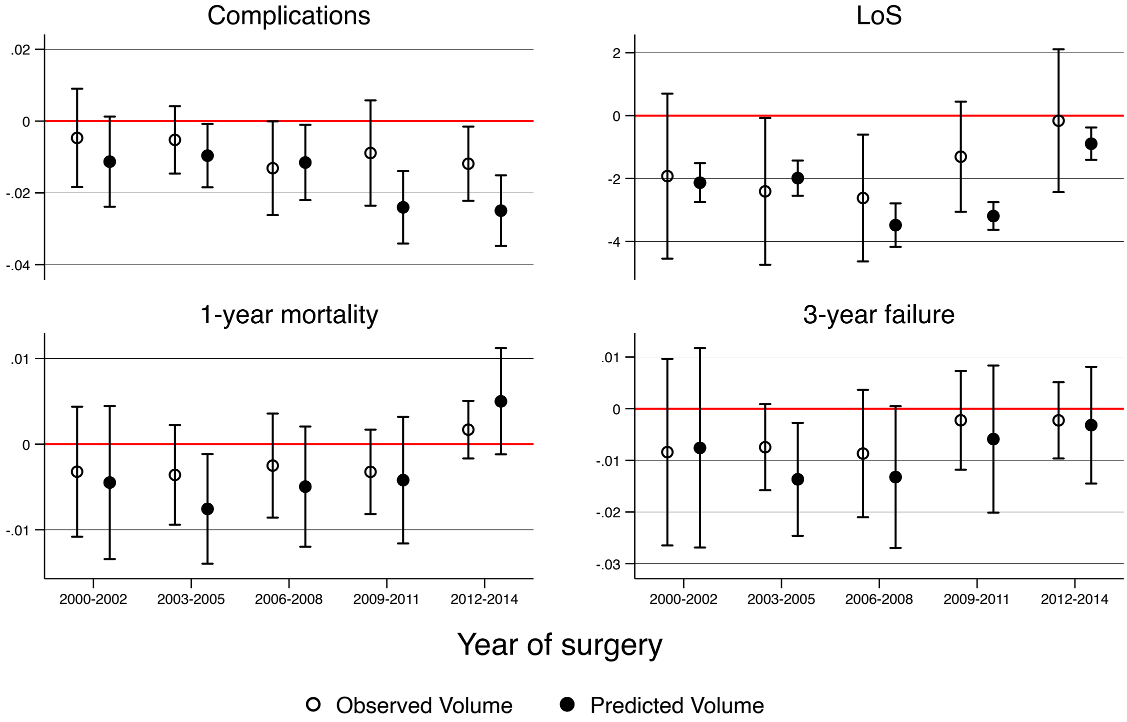


Figure 2.4: Estimated coefficients for observed and predicted hospital volume across years



Chapter 3

Searching for Efficiency Public procurement and healthcare providers' behavior in the Italian NHS[1]

Abstract

This study exploits the change in hip prostheses public procurement that occurred in the Romagna LHA (Italy) to assess its impact on prostheses cost, the effect on surgeons specialization, and patient health outcomes. Hip prostheses are the major cost-driver of hip replacement surgeries. The cost-containment pressure, especially in the public healthcare sector, makes it crucial to design the public tender such that resulting implant prices are minimized. However, cost-containment policies have to be weighed against physicians' preferences and patient well-being. Evidence shows that a cost reduction occurred without any significant impact on surgeons' prostheses choices. Also, positive or no effect of surgeons treating style is found on patients outcomes after the new procurement is introduced.

[1]I am grateful to Chiara Cassanelli for helping in developing the sketch of the hip prosthesis presented in this chapter.

3.1 Introduction

The pressure to contain healthcare expenditure has grown incredibly in recent years, especially in contexts of public provision. Therefore, efforts to boost health provision efficiency are particularly salient. However, the introduction of efficiency-enhancing policies has to be weighted with patient health outcomes. Typically, public institutions' cost-savings attempts generate concern among the public opinion on the sufficient level of services' quality retention.

One of the most substantial contributions to healthcare costs comes from medical devices expenditure (Schreyögg, Bäumler, & Busse, 2009). Consequently, attempts have been devoted to containing expenses by improving provision efficiency through standardization, rationalization and market competition, and transparency promotion. Specifically, policies aiming to standardize medical devices trigger concerns among physicians since they may be specialized in using a limited number of devices, thus restricting their choice may impact their performance. Among the wide range of efficiency measures healthcare institutions might undertake, this study focuses on those affecting the discretionary power of healthcare providers. The reason for this choice is that physician treatment decisions have an impact on patient health outcomes. Therefore, limiting physicians' discretion to increase healthcare efficiency might cause an unwanted deterioration of patient health. However, this aspect has not been fully explored by the existing literature.

In particular, this research exploits the change in hip prostheses public procurement agreement, carried out by the Romagna Local Health Authority (LHA), Italy, to assess its impact on hip prostheses cost, surgeons' implant choice, and patient health outcomes. Hip replacement surgery is one of the most performed operations (OECD, 2020), with an increasing trend due to the aging population. This volume increase boosted healthcare costs in recent years. The new procurement agreement is designed to reduce the hip prosthesis price through a demanded implant types rationalization, resulting in a shrinkage of surgeons' choice set. Surgeons' preferred type of implant may not be available after the rationalization, or they may be required to choose among fewer types of implants. Whether the price reduction materialized is examined, as well as the possible variation in surgeons treating behavior – i.e. prosthesis choice – captured by the Herfindahl-Hirschman Index (HHI). The ultimate impact of the new procurement agreement on patient health outcomes is also assessed through a multivariate linear regression model. Nevertheless, it is not possible to identify the impact channel since the analysis does not allow to disentangle the impact of the new procurement from that of surgeons' treatment choice on patient health.

Although a choice-set reduction materialized, the results highlight that the proportion of used implants in the choice set increased in the post-period. The analysis confirmed that the price of implants and single prostheses' components significantly decreased, both for the choice set and the implanted devices. However, in the post-period, the average implanted prostheses price is found to be higher than the average price in the available choice set, meaning that surgeons are selecting more expensive implants. Therefore, mitigating the cost-saving effect of the new procurement. No variation is found in surgeon level of brand utilization concentration, as measured by the HHI. The evidence may be due to the same brands of implants being available in both periods, allowing surgeons to maintain their preferred choices. Finally, regression results for patient health outcomes suggest that the lower the number of prostheses brands used by the surgeon, the lower the probability of incurring complications. However, in the post-period, the association is significant only for the probability of incurring complications overall or just intra-surgery. Findings suggest the procurement successfully decreased prosthesis cost while not modifying surgeons' implant choices nor worsening patient health outcomes. Nevertheless, the analysis is conducted at the brand level due to the inability to compare homogeneously defined tender lots across periods, and a higher level of detail is left for further research.

The rest of the work is organized as follows. The relevant literature is illustrated in the next section. The general framework and the tender characteristics are presented in section 3.3, while the methods are outlined in section 3.4. A description of the dataset and descriptive statistics are provided in section 3.5. Section 3.6 illustrates the results. Finally, section 3.7 concludes the paper.

3.2 Literature Review

The public sector typically faces tight budget constraints, which require a rationalization of available resources and efficiency-enhancing policies introduction. Particularly in the public healthcare sector, the pressure to improve system efficiency has increased in recent years. Nevertheless, patient health protection remains the ultimate priority. Indeed, physicians' professional norms pose the attention on patient well-being while lowering concerns about the cost of healthcare provision (Kesternich, Schumacher, & Winter, 2015). A possible way of reducing healthcare costs is through procedures standardization or medical devices and equipment rationalization. Generally, physicians tend to specialize in a relatively limited number of surgical devices and techniques utilization. Therefore, their discretionary power may be restricted by efficiency-enhancing policies. In this context, the limitation of treatment decisions could work through two different channels. First, physicians' preferred devices may be no longer available after the standardization. Second, physicians may experience a decline in the number of available options among which to choose. This limitation may compromise patient health outcomes since physicians would be forced to adopt a technique or medical device for which their level of expertise is suboptimal. The choice set restriction may also trigger adverse and unwanted effects on patients' health, given that physicians may be less trained.

The variation in physicians' medical practice has been investigated in the literature. Epstein and Nicholson (2009) found that variation in treatment behaviors exists and residency programs or peer effects can not fully explain it. Thus, physician practice styles tend not to converge over time. How physicians develop their practice style can be described by the learning-by-doing model proposed by Jovanovic and Nyarko (1996), according to which, the more expert the physician is in a procedure utilization, the higher the drop in expertise once switching to new technology. Ramdas, Saleh, Stern, and Liu (2018) investigated the effect of device-specific learning and forgetting on surgeons' productivity for hip replacement surgery, claiming that first-time usage of a given type of device increases the duration of surgery significantly. Also, the authors detected some forgetting when the time between repeated uses of a specific implant increases. Accordingly, physicians may resist the up-taking of new techniques, given that they feel more confident in using those with which they already have familiarity. However, Corallo et al. (2014), in their systematic review of medical practice variation, highlighted the lack of sufficient research on the causes of differences and their impact on patients' health outcomes. Additionally, they advised focusing more on resource-intensive treatments and those relevant from a clinical and policy perspective.

The relation between physicians' practice and patient health outcomes has been found in the literature. Surgeons specialization is a relevant predictor of patient mortality for several surgical techniques, independently of the volume of surgeries performed (Sahni, Dalton, Cutler, Birkmeyer, & Chandra, 2016). There is evidence that surgeons' procedure specialization is inversely related to patient mortality, even if findings are sensitive to the definition of specialization (Hall, Hsiao, Majercik, Hirbe, & Hamilton, 2009). The effect of practice style on health outcomes for patients with AMI[2] was documented by Currie, MacLeod, and Van Parys (2016): the more aggressive the treatment, the better the health outcome. Also, treatment aggressiveness is found to be positively related to its costs. Likewise, Grytten and Sørensen (2003) found that variation

[2] Acute Myocardial Infarction

in clinical practice accounts for a large portion of expenditure for primary physician services in Norway. The mortality-cost of care trade-off has been further investigated by Schreyögg and Stargardt (2010), suggesting caution should be posed when introducing cost-containment policies, though not explicitly considering variation in medical practice. Also, the authors looked just at AMI patients, and the relationship needs further analysis to establish consistency for other specialties.

Apart from procedure standardization, cost-saving policies may also regard the rationalization of available medical devices. While procedures may often be decided internally, through the introduction of guidelines, medical devices uniformity is achieved by acting on the procurement process. Public procurement of medical devices can have different designs, which may accomplish different outcomes depending on their features. Kjerstad (2005) found that auctions create more competitive markets compared to negotiations. Nevertheless, auctions fail to significantly lower prices. Similar results, robust for quality differences, have also been identified by Vellez (2011). Interestingly, Sorenson and Kanavos (2011), in their comparative analysis of major EU countries, observed that price is the predominant feature considered when evaluating procurement offers, while other aspects, such as quality, physician acceptance, innovation, and ancillary services, are largely ignored. This approach may affect physicians' treatment choices as well as patient health outcomes. A higher degree of centralization of public procurement plays a role in reducing administrative costs for hospitals and increasing their market power (Cappellaro, Fattore, & Torbica, 2009), especially in low-quality areas (Ferraresi, Gucciardi, & Rizzo, 2021). Also, the size of the public buyer is positively related to competence in the public procurement of medical devices, while for low-skilled buyers, the introduction of mandatory reference prices could lead to efficiency gains (Bucciol, Camboni, & Valbonesi, 2020). Analyzing the features of orthopedic implants tenders in Italy, Atella and Decarolis (2019) detect competition is incentivized by open auctions, with a higher reserve price. Also, the lowest price awarding criterion leads to higher rebates compared to scoring rules like Most Economically Advantageous Tender (MEAT), though quality may be lower. However, there is limited evidence regarding the relationship between restriction of physician discretionary power over the type of medical devices they can use and patient health outcomes. Ricketts and Sherry (2018) analyzed the impact of a cost-containment policy, which forced surgeons to use a single femoral component in hip replacement surgeries. They compared patient outcomes based on surgeons' prior familiarity with the implant component, finding no difference in clinical outcomes.

In the general research framework of analyzing the impact of public policies aiming at increasing efficiency by targeting the healthcare providers' discretion on treatment choices, this study exploits the variation in public procurement enforced in the Romagna Local Health Authority (LHA) in Italy. The new procurement aimed to reduce implant cost by aggregating prosthesis types, thus affecting surgeons' prostheses choice set. The variation in public procurement design is exploited to assess its impact on implant cost, surgeons' specialization, and patient health outcomes. Hip replacement surgery is one of the most performed healthcare treatments, and hip implants are the main cost driver of surgery total cost. The market for hip prostheses has reached a mature stage, with very little innovation. Also, the level of concentration is significant, with few firms holding a large market share (Davies & Davies, 2021). Differently from Ricketts and Sherry (2018), this work exploits a considerably larger dataset of patients and surgeons and considers a change in the procurement structure that affected all implant components instead of just the femoral stem, thus possibly expanding the procurement policy effect. Also, this analysis focuses on patient complications during hospitalization, while Ricketts and Sherry (2018) analyzed outcomes 12-months after surgery. A detailed description of the hip prostheses context and the public procurement agreement is provided in the next section.

3.3 General Framework

3.3.1 Total Hip Arthroplasty

Total Hip Arthroplasty (THA)[3] is one of the most frequent surgeries among the elderly. In Italy, more than 100,000 individuals underwent this operation in 2016, and the trend has been increasing in years as a result of the aging population (AGENAS, 2020). In Emilia-Romagna, THA is the fourth most performed surgery, with almost 8,000 patients treated in 2016. THA is a well-consolidated surgical procedure, and the level of technology has reached a mature stage. The principal source of variation comes from the type of implants used. Specifically, the prosthesis can vary in material and shape mainly to accommodate anatomical specificities of treated patients, although producers have developed their distinct designs. The implant is composed of 5 different elements: a femoral stem, a femoral neck – which most of the time is fixed to the stem, but can also be modular –, a femoral head, a liner or insert, and an acetabular cup or shell (Figure 3.2). The femoral head and the liner are quite alike between producers, and they only vary in diameter to adjust patient anatomy and material – typically ceramic, metal, or polyethylene –, this latter being the main cost driver. Instead, the stem and the cup are highly diverse, and they may require specific surgical techniques depending on their shape. In 2016, 110

[3]Also known as Total Hip Replacement (THR)

different types of cups and 140 types of stems were implanted in Emilia-Romagna. Among these, 23 cups and 20 stems were used for the first time (R.I.P.O., 2018). Typically, all components of the implanted prosthesis belong to the same producer, and surgeons are advised against mixing parts from different brands due to no guarantees on the compatibility of the components across brands. However, sometimes mixed prostheses are implanted. Implants represent the main cost driver of hip replacement surgery price (Stargardt, 2008). Coupled with the volume of surgeries performed, this makes the total healthcare expenditure for this treatment a particularly relevant issue for cost-containment policies.

3.3.2 Public Procurement in the Romagna LHA

This study analyses the change in public procurement tender for hip prostheses occurring in the Romagna LHA[4]. Figure 3.1 outlines the timing of the two public procurement tenders examined in this study. The public procurement pre-period starts in 2009. The agreement endures three years with the chance of renewal – indeed obtained – for an additional 3-year period. In 2015, contracts are extended for one year to allow the LHA to conduct a new public procurement tender without the interruption of implants supply. The following year, an additional extension is granted to meet the start of the new public procurement contracts expected for May 2017. The new (post-period) procurement lasts two years without renewal possibility.

The pre-period tender comprises ten separate lots, each corresponding to different hip prostheses types. Each lot is composed of sub-lots for several implant components. Firms compete by submitting sealed offers for each lot, with the technical and economic proposals placed in separate envelopes. Firms are not required to tender for all lots. The awarding criterion is based on a scoring rule. A maximum of 60 points is granted for the technical offer, where quality is assessed based on pre-determined multi-dimensional features. The evaluation of the technical proposal needs to obtain a minimum of 35 points for the economic bid to be evaluated. Subsequently, 40 points are awarded to the firm offering the cheapest bid k, while the economic offers of firm i receive points proportional to the most economical one, according to the following formula:

$$\text{points}_i = \frac{\text{value cheapest bid}_k * 40}{\text{value economic bid}_i} \tag{3.1}$$

Firms are then ranked based on the total score of the technical and economical offer. The bid with the highest total score is awarded the lot so that just a single bidder is designated to sign

[4] The Romagna LHA belongs to the Emilia-Romagna Region, located in the northern part of Italy. This public institution provides healthcare services to a population of more than a million citizens, which is roughly 24% of the total Emilia-Romagna population in 2016.

the procurement contract with the LHA per lot. The competition was almost null: on average less than two firms competed per lot, and the majority of lots had a single bidder (Table 3.1). The fact is in line with Atella and Decarolis (2019) findings, and it is not surprising, given the market of hip implants is quite concentrated (Davies & Davies, 2021). Eventually, five different firms won at least one lot, with one firm obtaining two lots and two firms obtaining three lots each. Nevertheless, the LHA could purchase implants in the free market, not exceeding 20% of the total value of contracted procurement, and on justified grounds only.

The post-period tender's purpose is to decrease the price of hip prostheses, maintaining the current level of surgeons' discretionary power over which type of implant to use. To this aim, the tender scheme has significant differences compared to that in the pre-period. The LHA decided to group prostheses in a smaller number of lots – just five –, thus increasing the quantity demanded per lot. Additionally, seven lots per spare component are tendered. These components are tendered alone because they are typically employed in revision surgery, where only some prosthesis elements are substituted. To prevent an excessive restriction of surgeon's choice-set each lot – either per total implant or single component – is awarded to the first four bidding firms in the final ranking, drawn up with the same 40-60 points scheme as in the pre-period tender. However, no minimum quantity is guaranteed to any of the four winning firms per lot. Differently from the previous tender[5], a maximum reserve price is set per lot, below which economic offers had to remain. The procurement agreement lasts two years without the possibility of renewal. In this new setting, the competition was fiercer, with 8.25 bidders on average competing per lot. A total of 12 different firms are awarded at least one lot. Specifically, all winning firms in the pre-period are also awarded at least one lot in the post-period, and seven new firms enter the procurement agreement in the post-period. A comparison of tenders features across periods is reported in Table 3.1.

3.4 Methods

The investigation exploits the change in the design of the public procurement of hip prostheses, which occurred in the Romagna LHA. In particular, it is examined whether the new procurement has successfully reduced the cost of hip implants without limiting surgeons' treatment choices, and patient health outcomes. By aggregating different categories of prostheses, the new procurement setup aimed to reduce the cost of prosthetic implants to increase the quantity demanded.

[5] No explicit price threshold is set in the pre-period tender. However, the Ministerial Decree 296/2007 of the Ministry of Health of 11 October 2007 introduced tenders' maximum prices per hip prosthesis components (https://www.gazzettaufficiale.it/eli/id/2007/11/13/07A09617/sg).

The possible envisaged shortcoming of this design is a restriction of the physician prosthesis choice set. Surgeons unable to use their preferred prosthesis implant are forced to switch to a less familiar alternative, which may lead to a lower surgery success rate and a worsening of patient health outcomes.

In the post-period, surgeons are all subject to the new procurement agreement being effective in all the Romagna LHA territory. Therefore, the variation in practice style is tested by comparing surgeons' implant choices before and after the new procurement introduction, without the possibility to exploit the presence of a control group. Variation in prostheses cost is assessed with t-tests for differences in average values across periods. Similarly to Hall et al. (2009), to measure treatment variation the Herfindahl-Hirschman Index (HHI) is computed for all surgeons across periods. The HHI is generally used in economics to measure the level of market competition. In this context, it is used to indicate the concentration level of implant brand usage by surgeons. The HHI for surgeon s is computed as follows:

$$HHI_{st} = \sum_{j=1}^{N} s_{jt}^2 \tag{3.2}$$

where s_{jt} is the share of implants of firm j used, over the total surgeries performed by surgeon s in period t. In this analysis $t = 0$ in the pre-period and $t = 1$ in the post-period. By construction $HHI \in [0, 1]$, where 0 indicates a highly competitive setting, while a monopolistic situation gets values close to 1. Therefore, the higher the index, the more concentrated the prosthesis utilization, thus the lower the number of different brands implanted by the surgeon. A t-test is computed to assess whether a statistically significant difference exists in average surgeon HHI across periods. The index is also computed at the stem-cup combination level to assess the utilization concentration at a more disaggregated level.

Finally, the effect on patient i health outcomes is investigated through a simple linear probability model, where the independent variable of interest is the surgeon's HHI. The relation is modeled as follows:

$$\begin{aligned} y_i = \alpha + \beta_1 log(HHI_{st}) + \beta_2 post_i + \beta_3 log(HHI_{st}) * post_t + \\ + \mathbf{X}_i'\beta_4 + \mathbf{L}_s'\beta_5 + \mathbf{S}_i'\beta_6 + \mathbf{B}_i'\beta_7 + \mathbf{H}_i'\beta_8 + \mathbf{T}_i'\beta_9 + \varepsilon_i \end{aligned} \tag{3.3}$$

where the dependent variable is a dummy equal to 1 if the patient had complications during the hospitalization (intra-, post-surgery, or total), HHI_{st} is surgeon s Herfindahl-Hirschman Index in period t expressed in logarithm, $post_i$ is a dummy equal to 1 if the patient is treated in the post-period, $\mathbf{X}_i$ is a vector of patient characteristics (age, gender, side of the hip treated, a dummy equal to 1 if the patient is resident in the region, and a set of dummies for the type of diagnosis), $\mathbf{L}_s$ is a set of dummies for the years since the surgeons obtained the license to practice, which is included as a proxy for surgeons' expertise. Finally, $\mathbf{S}_i$, $\mathbf{B}_i$, $\mathbf{H}_i$ and $\mathbf{T}_i$ are surgeon, prosthesis brand, hospital and year-quarter fixed effects to control for fixed state-specific features. Both hospital and surgeon fixed effect are included in the model as some practitioners work across more than one hospital. The coefficients of interest are β_1 and β_3. Specifically, the impact of surgeons practice style in the post period is captured by $\beta_1 + \beta_3$. The interaction term captures differential effect of surgeon HHI after the new procurement is in place. The HHI is included in the regression in logarithm, as it is sensitive to small fluctuations. Thus, expressing it in log form makes the interpretation of the estimated coefficient in percentage terms, which are easier to understand.

3.5 Data and Sample Composition

The principal source of data is the RIPO[6] database. The dataset contains information on patient characteristics (age, gender, province of residence, side of the hip treated, and diagnosis) and details on implanted prosthesis (brand, material, reference code, and commercial name of each component). Patients between 18 and 100 years old, admitted to public hospitals in the Romagna LHA undergoing total hip replacement surgery are considered for the analysis. Four patients whose implanted prosthesis is custom-made are excluded from the dataset. The RIPO database also provides data about the gender and date of birth of the treating surgeon. The latter was not available for all orthopedic surgeons in the dataset. Hence, additional data are obtained from the online registry of medical practitioners, where information regarding the year of orthopedic specialization completion is also collected. Surgeons performing less than ten surgeries in the overall period of the analysis are excluded from the dataset. Information about the price of each prosthesis component is retrieved from the Romagna LHA procurement legal documents both in the pre- and post-period. This data is used to compute implanted prosthesis total price. The analysis considers the period January 1st, 2015 - December 31st, 2018, as highlighted in blue in Figure 3.1. The post-period begins May 1st, 2017. The patients' health outcomes are intra- and post-surgery complications.

The final dataset includes 3,682 total hip replacement surgeries performed by 29 male sur-

[6] Registro dell'Implantologia Protesica Ortopedica (Register of the Orthopaedic Prosthetic Implants). http://ripo.cineca.it/authzssl/index.htm

geons in 8 public hospitals. The number of observations amounts to somewhat less than 1,000 total hip replacement surgeries per year. Four surgeries are excluded as the implanted prosthesis is custom made. Average patient characteristics are reported in Table 3.3 both for the overall sample and for the pre- and post-period. Patients are on average 69 years old. The proportion of females is slightly higher than males (57%), and less than 15% of patients are residents outside the Region. No statistically significant differences in average patient characteristics are found between the two periods apart from minor exceptions regarding patient diagnosis. Finally, the unconditional probability of post-surgery complications is found to be significantly lower in the post-period, though no difference is detected for intra-surgery complications.

3.6 Results

The number of firms winning at least one lot increases from 5 to 12 in the post-period (Table 3.1). Consequently, the number of different brands implanted also increases. Interestingly, all suppliers in the pre-period result being suppliers also in the post-period. However, the size of surgeons' available choice set shrinks considerably with the introduction of the new procurement agreement, from 746 to 167 prostheses (Table 3.2). The reduction is not due to the lower availability of different stems and cups – which is higher in the post-period instead – but to a lower number of possible stem-cup combinations provided by the same firm. Also, the number of different combinations implanted is lower, in absolute value in the post-period while corresponding to a higher portion of the available choice set. This evidence suggests that even if the choice set is smaller in the post-period, the percentage of utilization increases.

The price of the awarded prostheses significantly decreased after the introduction of the new procurement agreement (Table 3.4a). The largest contribution to the price reduction – in absolute values – comes from the stem and cup components, while no difference is found for neck price. A statistically significant difference across periods is also found for the average implanted prostheses price, though the magnitude is more modest with respect to that of the choice set (Table 3.4b). While the overall reduction in choice set prices is 24% – corresponding to 666.39 euros –, the price reduction for the implanted prosthesis is only 17%, which amounts to 447.07 euros. It is also worth noting that while, in the pre-period, the average price of the implanted prostheses is in line with the average price of implants in the choice set, this is not the case in the post-period. In fact, the average implanted prostheses price is above the average price of implants in the choice set, thus mitigating the cost-saving effect of the new procurement. The evidence suggests that, in the post-period, surgeons choose more expensive prostheses within their newly available cheaper choice set. Also, with a back of the envelope calculation, it is possible to suggest that the overall saving would have been more than 927 thousand euros if the choice set average price were the same as that for the implanted devices, while the actual savings amount to slightly more than 622 thousand euros, thus canceling more than 300 thousand euros of potential savings.

Surgeons' treatment behavior per year since they obtained the license to practice is summarized in Table 3.5. On average surgeons use prostheses belonging to only three different brands. Also, the HHI is higher than 0.5, thus indicating a high level of concentration. The index computed at the stem-cup combination level is lower but still of significant magnitude. There are no significant differences between years of seniority. Furthermore, as reported in Table 3.6, surgeon practice style is not found to vary across periods: both the Wilcoxon rank-sum and the nonparametric K-sample test are not statistically significant. Although the number of available firms significantly increases in the post-period (Table 3.1), this does not induce a greater diversification of brand usage by surgeons. A possible explanation may be that since all the winning firms in the pre-period are also awarded at least one lot in the post-period, surgeons can maintain their treatment style. Indeed, 34% of surgeons use prostheses from the same brands across periods. This idea is further reinforced by the HHI value computed at the brand level, which is higher than 0.5, thus suggesting a high concentration level in both periods. The same result is obtained by computing the HHI at the stem-cup combination level: the degree of utilization concentration is lower than that computed at the brand level but still of relevant magnitude. These findings are limited since it was not possible to group prostheses in homogeneously defined lots across periods. This prevented performing a more detailed analysis on the impact of the implant availability variation in lot-specific choice sets.

Estimates for the impact of surgeon practice style on patient health outcomes are reported in Table 3.7. The probability of incurring complications – either total, intra-, or post-surgery – decreases as the concentration of prostheses brands increases. The magnitude of the impact is large and statistically significant. This result may be justified by surgeons using a lower number of different prostheses, being more specialized, thus improving patient outcomes. The impact of surgeons' prosthesis utilization concentration in the post-period is obtained by the linear combination of the estimated coefficients of $log(HHI)$ and $post{*}log(HHI)$. The interaction term is positive with a smaller magnitude than the HHI estimated coefficient. Also, the significance is generally only at 10% level. The resulting impact of HHI in the post-period is negative for all dependent variables, although it is significant only for total and intra-surgery complications.

Specifically, a 1% increase in the HHI leads to a 4% decrease in the probability of intra-surgery complications and a 9.7% decrease in total complications. This finding means that after the new procurement introduction, the higher the HHI, the lower the complication rate. However, the impact is lower than that in the pre-period. Also, in the post-period, surgeons' brand utilization concentration is not found to have a significant impact on patients' post-surgery complications. The robustness of findings is tested by estimating the same linear probability model in 3.3 for the subset of patients over 50. In fact, although patients below 50 years old represent less than 8% of the dataset, they may have a disproportionately lower complication rate due to their ability to bear the operation and recover faster. Results reported in Table 3.8 show that the impact magnitude and significance of HHI are strikingly similar to that of the main regression, while the interaction term magnitude is lower and not significant. The resulting impact of HHI in the post-period is slightly increased and still significant for total and intra-surgery complications. A 1% change in surgeons' HHI in the post-period is associated with a decrease of 4.3% in the probability of incurring intra-surgery complications and an 11.5% decrease in the probability of total complications, though the significance is still just at 10% level. The impact magnitude is again lower in the post period. Still, no impact is found for post-surgery complications. Hence, the main results are confirmed.

3.7 Conclusions

This study aimed to analyze the impact of the change in public procurement design on medical devices cost, surgeons treating decisions, and patient health outcomes. The analysis exploited the variation in the hip prostheses public procurement agreement, occurred in the Romagna LHA, Italy. The new procurement aimed to reduce implant cost through prosthesis types aggregation, thus increasing the quantity demanded per implant type. The new public procurement agreement may lead to a limitation of surgeons' implant choice set. This restriction of surgeon autonomy may have an impact on patient health outcomes since the surgeon's preferred implant could be no longer available after the rationalization, or the surgeon may have to choose from a smaller choice set. To safeguard surgeons' treatment autonomy, a higher number of firms were awarded per single lot. Being all surgeons subject to the new public procurement, a pre-post analysis was conducted to assess the impact of different procurement designs on surgeons treating behavior.

A tenders structure and resulting public procurements comparison highlighted that although a choice set restriction occurred in the post-period, the percentage of choice set utilization increased. All suppliers providing implants in the pre-period were among the suppliers also in the post-period. This evidence could justify the lack of variation in practice style among the two periods. Further, the level of concentration of surgeons' brand utilization, as measured by the HHI, was found to be high and stable. The prosthesis cost significantly decreased after the new procurement design implementation, both at the choice set and implanted level. However, in the post-period, the average price of the implanted prostheses appeared to be higher than that in the choice set, meaning that surgeons chose more expensive implants, thus mitigating the cost-saving effect of the new procurement. Finally, surgeon practice style was found to be significantly associated with a lower probability of incurring complications. However, in the post-period, the impact of surgeons' brand utilization concentration is significantly associated with lower total and intra-surgery complications, while no impact is found for post-surgery complications. Evidence suggests that the new procurement successfully reduced prostheses costs while allowing surgeons sufficient autonomy in treatment choices and without worsening patients' health outcomes. Medical device public procurement should be designed to balance the need for improving healthcare efficiency and surgeons treating decisions. In the analyzed context, this was achieved by the same prostheses brands being available in both periods but at a lower overall price in the post period. Also, the fact that the average price of implanted prostheses was higher than the average price of the available choice set in the post-period may signal that surgeons are looking for higher quality products among those available.

This study provides a first explorative analysis of the impact of procurement structure on physicians' practice style and the effect on patient health outcomes. Findings are limited since it was not possible to aggregate implants in homogeneously defined lots across periods. This prevented investigating the impact of the prostheses availability variation in lot-specific choice sets. Thus, the comparison presented remains at a higher (brand) level of aggregation, and a more detailed analysis remains for future work. Also, the analysis does not enable the comparison of the quality of prostheses in the choice sets. Finally, the effect of surgeons' implant choice on other relevant health outcomes needs to be considered.

3.8 Tables and Figures

3.8.1 Tables

Table 3.1: Tender features

	Period	
	Pre	Post
Total lots tendered	10	5 (tot. implants) + 7 (spare components)
Winners per lot	1	4
Potential maximum winners	10	48
Average competing firms per lot	1.7	8.25
Average excluded offers per lot	0.2	1.58
Actual winning firms	5	12
Implanted brands	7	9
Non-winning implanted brands	2	1

Table 3.2: Choice set composition and implanted prostheses

	Period	
	Pre	Post
Choice set size	746	167
Same brand stem-cup combinations in choice set	194	107
Stems in choice set	39	58
Cups in choice set	35	44
Same brand stem-cup implanted combinations	100	78
As % of stem-cup combinations in choice set	0.515	0.729
Stem implanted	48	42
As % of stems in choice set	1.23	0.724
Cups implanted	25	26
As % of cups in choice set	0.714	0.591

Note: The choice set size is given by the total number of possible combinations among same-brand awarded components.

Table 3.3: Patients characteristics per period

	Total mean (sd)	Pre mean (sd)	Post mean (sd)	t-stat for mean diff
Age	68.70	68.48	69.06	-1.54
	(11.00)	(10.91)	(11.15)	
Female	0.573	0.576	0.569	0.40
	(0.495)	(0.494)	(0.495)	
Resident	0.871	0.865	0.881	-1.41
	(0.336)	(0.342)	(0.324)	
Hip Side: Right	0.544	0.538	0.555	-0.99
	(0.498)	(0.499)	(0.497)	
Diagnosis:				
Primary coxarthrosis	0.681	0.678	0.685	-0.40
	(0.466)	(0.467)	(0.465)	
Outcome of congenital pathologies	0.0559	0.0685	0.0352	4.22***
	(0.230)	(0.253)	(0.184)	
Femoral head necrosis	0.0500	0.0493	0.0510	-0.23
	(0.218)	(0.217)	(0.220)	
Rheumatoid arthritis	0.00380	0.00262	0.00575	-1.50
	(0.0615)	(0.0511)	(0.0756)	
Bone fractures	0.159	0.155	0.164	-0.68
	(0.365)	(0.362)	(0.370)	
Other	0.0510	0.0458	0.0596	-1.84*
	(0.220)	(0.209)	(0.237)	
Health outcomes:				
Complications (total)	0.235	0.257	0.199	4.01***
	(0.424)	(0.437)	(0.399)	
Intra-surgery complications	0.0109	0.0109	0.0108	0.04
	(0.104)	(0.104)	(0.103)	
Post-surgery complications	0.227	0.247	0.193	3.87***
	(0.419)	(0.432)	(0.394)	
Observations	3,682	2,291	1,392	

*** p<0.01, ** p<0.05, * p<0.1

Table 3.4: Price of total prosthesis and single components in the choice set and implanted per period

(a) Choice set

Price in euros	Total mean (sd)	Pre mean (sd)	Post mean (sd)	Δ pre-post	t-stat for mean diff
Total	2,690	2,794	2,128	666.39	9.48***
	(769.5)	(767.6)	(484.6)		
Stem	1,056	1,098	846.8	251.05	6.27***
	(442.2)	(459.3)	(259.1)		
Cup	962.9	1,006	772.4	233.20	5.49***
	(486.5)	(512.0)	(281.4)		
Liner	377.9	400.2	283.3	116.93	7.74***
	(171.6)	(176.9)	(103.3)		
Head	279.4	292.3	216.4	75.91	11.14***
	(79.03)	(74.45)	(70.18)		
Neck	372.6	383.6	335.0	48.57	1.30
	(194.4)	(184.7)	(223.3)		

*** p<0.01, ** p<0.05, * p<0.1

(b) Implanted

Price in euros	Total mean (sd)	Pre mean (sd)	Post mean (sd)	Δ pre-post	t-stat for mean diff
Total	2,577	2,742	2,295	447.07	25.09***
	(539.4)	(475.9)	(525.0)		
Stem	1,009	1,060	920.8	139.25	15.14***
	(263.9)	(236.9)	(283.8)		
Cup	835.5	887.4	746.6	140.85	17.52***
	(233.1)	(245.1)	(178.9)		
Liner	362.8	397.6	303.0	94.64	18.38***
	(148.8)	(157.9)	(108.1)		
Head	321.5	348.9	274.5	74.43	37.97***
	(65.17)	(56.87)	(49.87)		
Neck	212.9	218.3	204.4	13.92	2.78***
	(71.58)	(79.65)	(55.71)		

*** p<0.01, ** p<0.05, * p<0.1

Table 3.5: Surgeons' implant utilization per years since license to practice

	Years since license to practice			
	Total	<20	[20; 30)	$\geq$ 30
	mean	mean	mean	mean
	(sd)	(sd)	(sd)	(sd)
No. brands implanted	3.038	3	2.789	3.333
	(1.192)	(1.211)	(0.976)	(1.372)
HHI (brand)	0.595	0.554	0.621	0.606
	(0.215)	(0.221)	(0.225)	(0.207)
HHI (stem-cup combination)	0.381	0.424	0.400	0.322
	(0.207)	(0.235)	(0.200)	(0.184)

Table 3.6: Surgeons' implants utilization per period

	Total median mean (sd)	Pre median mean (sd)	Post median mean (sd)	Wilcoxon rank-sum test z-stat	Nonparam. K-sample test Pearson χ^2
No. brands implanted	1.192	1.241	1.143	1.02	1.13
	3.038	3.185	2.885		
	(1.192)	(1.241)	(1.143)		
HHI (brand)	0.215	0.216	0.219	0.11	0.18
	0.595	0.601	0.589		
	(0.215)	(0.216)	(0.219)		
HHI (stem-cup combination)	0.207	0.199	0.217	0.69	0.17
	0.381	0.393	0.369		
	(0.207)	(0.199)	(0.217)		

*** p<0.01, ** p<0.05, * p<0.1

Table 3.7: Estimated effect of surgeon practice style (brand level) on patient health outcomes

	Complications		
	Total	Intra-surgery	Post-surgery
Log(HHI)	-0.183***	-0.0551**	-0.143**
	(0.0640)	(0.0234)	(0.0623)
Post * Log(HHI)	0.0802*	0.0146	0.0763*
	(0.0423)	(0.00972)	(0.0419)
Post	0.0341	0.0268**	0.0254
	(0.0562)	(0.0126)	(0.0560)
Age	-0.0117**	-0.00158	-0.0111**
	(0.00568)	(0.00168)	(0.00558)
Age^2	0.000113**	9.69e-06	0.000110**
	(4.37e-05)	(1.26e-05)	(4.30e-05)
Female	0.0732***	0.00651*	0.0685***
	(0.0136)	(0.00348)	(0.0135)
Resident	0.0380*	0.000640	0.0355*
	(0.0199)	(0.00530)	(0.0196)
Hip Side: Right	-0.0160	0.000208	-0.0141
	(0.0131)	(0.00344)	(0.0130)
Diagnosis:			
Outcome of congenital pathologies	0.0764**	-0.00574	0.0831***
	(0.0322)	(0.00722)	(0.0320)
Femoral head necrosis	0.0297	-0.00229	0.0280
	(0.0331)	(0.00804)	(0.0326)
Rheumatoid arthritis	0.0440	-0.00913*	0.0489
	(0.107)	(0.00480)	(0.108)
Bone fractures	0.0606***	-0.00828	0.0674***
	(0.0222)	(0.00511)	(0.0220)
Other	0.106***	0.00493	0.104***
	(0.0328)	(0.00928)	(0.0324)
Years since license:			
$20 \leq$ Year license < 30	0.0226	-0.0151	0.0223
	(0.0698)	(0.0246)	(0.0692)
Year license ≥ 30	0.159	-0.0244	0.161*
	(0.0973)	(0.0289)	(0.0966)
Constant	0.187	0.0425	0.185
	(0.201)	(0.0664)	(0.198)
Log(HHI) + [Post * Log(HHI)]	-0.0972*	-0.0405*	-0.0670
Observations	3,682	3,682	3,682
R-squared	0.158	0.054	0.158
Year-quarter FE	Yes	Yes	Yes
Surgeon FE	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes
Mean dependent	0.235	0.0109	0.227

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Log(HHI) is the logarithm of surgeon HHI computed at brand level. Primary coxarthrosis is considered as baseline diagnosis. Year license indicate the years since surgeon obtained the license to practice, the baseline is Year license < 20.

Table 3.8: Estimated effect of surgeon practice style (brand level) on health outcomes for patients over 50

	Complications		
	Total	Intra-surgery	Post-surgery
Log(HHI)	-0.183***	-0.0489**	-0.139**
	(0.0662)	(0.0233)	(0.0643)
Post * Log(HHI)	0.0679	0.00620	0.0650
	(0.0449)	(0.00925)	(0.0445)
	(0.392)	(0.101)	(0.389)
Post	0.0162	0.0174	0.00691
	(0.0594)	(0.0114)	(0.0591)
Log(HHI) + [Post * Log(HHI)]	-0.115*	-0.0427*	-0.0743
Observations	3,402	3,402	3,402
R-squared	0.160	0.054	0.160
Covariates	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes
Surgeon FE	Yes	Yes	Yes
Brand FE	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes
Mean dependent	0.235	0.0109	0.227

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Log(HHI) is the logarithm of surgeon HHI computed at brand level. Covariates include: age, age^2, gender, a dummy variable equal to 1 if the patient is resident in the Emilia-Romagna Region, hip side treated, a set of dummies for patient diagnosis, and a set of dummies for the years since surgeon obtained the license to practice.

3.8.2 Figures

Figure 3.1: Public Procurement Timeline

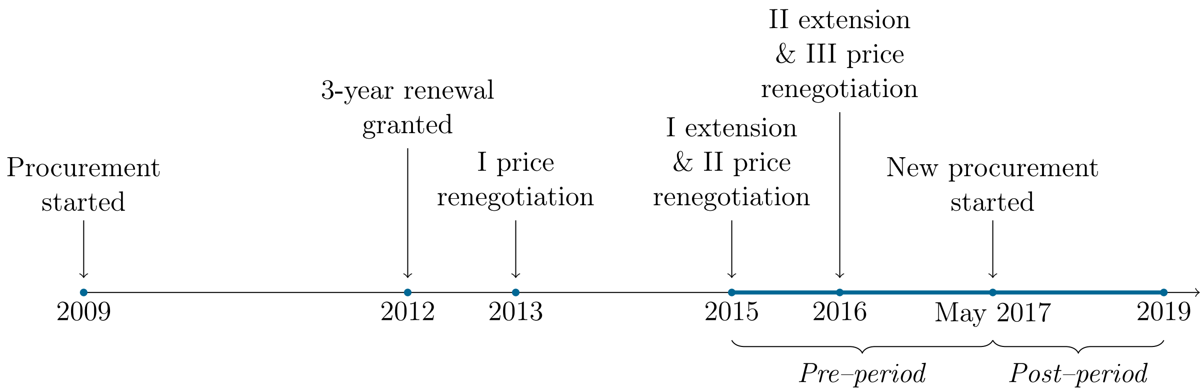
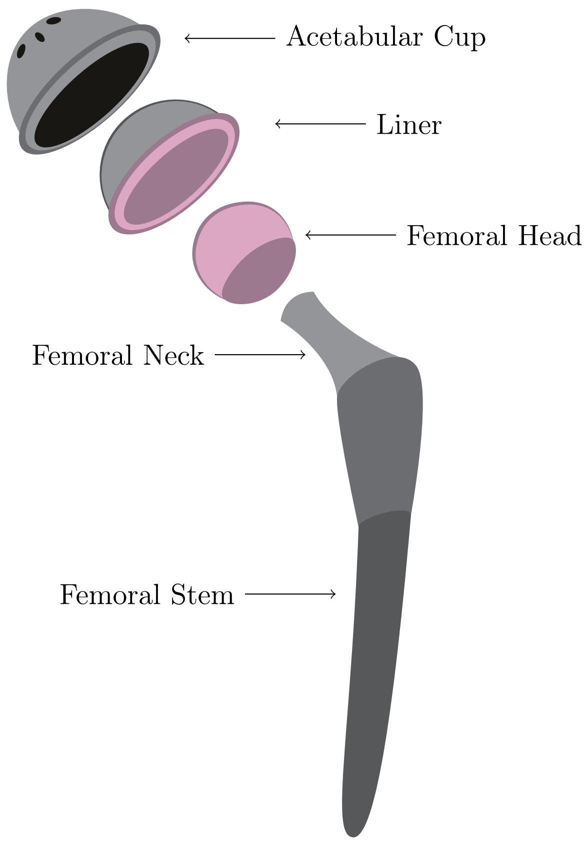


Figure 3.2: Sketch of hip prosthesis components



Bibliography

AGENAS. (2020). *Programma Nazionale Esiti*. Retrieved from https://pne.agenas.it/main/doc/Report_PNE_2020.pdf

Arrow, K. J. (1962). The Economic Implications of Learning by Doing. *The Review of Economic Studies*, *29*(3), 155. doi:10.2307/2295952

Atella, V., & Decarolis, F. (2019). Procuring Medical Devices: Evidence from Italian Public Tenders. *SSRN Electronic Journal*. doi:10.2139/ssrn.3467466

Bell, C. M., & Redelmeier, D. A. (2001). Mortality among patients admitted to hospitals on weekends as compared with weekdays. *The New England journal of medicine*, *345*(9), 663–8. doi:10.1056/NEJMsa003376

Benkard, C. L. (2000). Learning and Forgetting: The Dynamics of Aircraft Production. *American Economic Review*, *90*(4), 1034–1054. doi:10.1257/aer.90.4.1034

Bottle, A., & Aylin, P. (2006). Mortality associated with delay in operation after hip fracture: Observational study. *British Medical Journal*, *332*(7547), 947–950. doi:10.1136/bmj.38790.468519.55

Bredahl, C., Nyholm, B., Hindsholm, K., Mortensen, J., & Olesen, A. (1992). Mortality after hip fracture: results of operation within 12 h of admission. *Injury*, *23*(2), 83–86. doi:10.1016/0020-1383(92)90037-S

Bucciol, A., Camboni, R., & Valbonesi, P. (2020). Purchasing medical devices: The role of buyer competence and discretion. *Journal of Health Economics*, *74*, 102370. doi:10.1016/j.jhealeco.2020.102370

Cappellaro, G., Fattore, G., & Torbica, A. (2009). Funding health technologies in decentralized systems: A comparison between Italy and Spain. *Health Policy*, *92*(2-3), 313–321. doi:10.1016/j.healthpol.2009.05.004

Carretta, E., Bochicchio, V., Rucci, P., Fabbri, G., Laus, M., & Fantini, M. P. (2011). Hip fracture: effectiveness of early surgery to prevent 30-day mortality. *International Orthopaedics*, *35*(3), 419–424. doi:10.1007/s00264-010-1004-x

Casaletto, J. A., & Gatt, R. (2004). Post-operative mortality related to waiting time for hip fracture surgery. *Injury*, *35*(2), 114–120. doi:10.1016/S0020-1383(03)00210-9

Corallo, A. N., Croxford, R., Goodman, D. C., Bryan, E. L., Srivastava, D., & Stukel, T. A. (2014). A systematic review of medical practice variation in OECD countries. *Health Policy*, *114*(1), 5–14. doi:10.1016/j.healthpol.2013.08.002

Currie, J., MacLeod, W. B., & Van Parys, J. (2016). Provider practice style and patient health outcomes: The case of heart attacks. *Journal of Health Economics*, *47*, 64–80. doi:10.1016/j.jhealeco.2016.01.013

Davies, C., & Davies, S. (2021). Assessing competition in the hip implant industry in the light of recent policy guidance. *Social Science & Medicine*, *287*, 114055. doi:10.1016/j.socscimed.2021.114055

Doruk, H., Mas, M., Yıldız, C., Sonmez, A., & Kýrdemir, V. (2004). The effect of the timing of hip fracture surgery on the activity of daily living and mortality in elderly. *Archives of Gerontology and Geriatrics*, *39*(2), 179–185. doi:10.1016/j.archger.2004.03.004

Epstein, A. J., & Nicholson, S. (2009). The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics*, *28*(6), 1126–1140. doi:10.1016/j.jhealeco.2009.08.003

Evans, W. N., & Kim, B. (2006). Patient outcomes when hospitals experience a surge in admissions. *Journal of Health Economics*, *25*(2), 365–388. doi:10.1016/j.jhealeco.2005.10.003

Ferraresi, M., Gucciardi, G., & Rizzo, L. (2021). Savings from public procurement centralization in the healthcare system. *European Journal of Political Economy*, *66*, 101963. doi:10.1016/j.ejpoleco.2020.101963

Gaynor, M., Seider, H., & Vogt, W. B. (2005). The Volume–Outcome Effect, Scale Economies, and Learning-by-Doing. *American Economic Review*, *95*(2), 243–247. doi:10.1257/000282805774670329

Gdalevich, M., Cohen, D., Tauber, C., & Yosef, D. (2004). Morbidity and mortality after hip fracture: the impact of operative delay. *Archives of Orthopaedic and Trauma Surgery*, *124*(5), 334–340. doi:10.1007/s00402-004-0662-9

Geweke, J., Gowrisankaran, G., & Town, R. J. (2003). Bayesian inference for hospital quality in a selection model. *Econometrica*, *71*(4), 1215–1238. doi:10.1111/1468-0262.00444

Goldstein, J. P., Babikian, G. M., Rana, A. J., Mackenzie, J. A., & Millar, A. (2016). The Cost and Outcome Effectiveness of Total Hip Replacement: Technique Choice and Volume–Output Effects Matter. *Applied Health Economics and Health Policy*, *14*(6), 703–718. doi:10.1007/s40258-016-0260-3

Gowrisankaran, G., & Town, R. J. (1999). Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics*, *18*(6), 747–767. doi:10.1016/S0167-6296(99)00022-3

Grimes, J. P., Gregory, P. M., Noveck, H., Butler, M. S., & Carson, J. L. (2002). The effects of time-to-surgery on mortality and morbidity in patients following hip fracture. *American Journal of Medicine*, *112*(9), 702–709. doi:10.1016/S0002-9343(02)01119-1

Grytten, J., & Sørensen, R. (2003). Practice variation and physician-specific effects. *Journal of Health Economics*, *22*(3), 403–418. doi:10.1016/S0167-6296(02)00105-4

Gutacker, N., Siciliani, L., Moscelli, G., & Gravelle, H. (2016). Choice of hospital: Which type of quality matters? *Journal of Health Economics*, *50*, 230–246. doi:10.1016/j.jhealeco.2016.08.001

Hall, B. L., Hsiao, E. Y., Majercik, S., Hirbe, M., & Hamilton, B. H. (2009). The Impact of Surgeon Specialization on Patient Mortality. *Annals of Surgery*, *249*(5), 708–716. doi:10.1097/SLA.0b013e3181a335f8

Halm, E. A., Lee, C., & Chassin, M. R. (2002). Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature. *Annals of Internal Medicine*, *137*(6), 511. doi:10.7326/0003-4819-137-6-200209170-00012

Hamilton, B. H., & Bramley-Harker, R. E. (1999). The Impact of the NHS Reforms on Queues and Surgical Outcomes in England: Evidence From Hip Fracture Patients. *The Economic Journal*, *109*(457), 437–462. doi:10.1111/1468-0297.00456

Hamilton, B. H., & Hamilton, V. H. (1997). Estimating surgical volume—outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Health Economics*, *6*(4), 383–395. doi:10.1002/(SICI)1099-1050(199707)6:4<383::AID-HEC278>3.0.CO;2-L

Ho, V. (2002). Learning and the evolution of medical technologies: The diffusion of coronary angioplasty. *Journal of Health Economics*, *21*(5), 873–885. doi:10.1016/S0167-6296(02)00057-7

Ho, V., Hamilton, B. H., & Roos, L. L. (2000). Multiple approaches to assessing the effects of delays for hip fracture patients in the United States and Canada. *Health services research*, *34*(7), 1499–518. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10737450

Jovanovic, B., & Nyarko, Y. (1996). Learning by Doing and the Choice of Technology. *Econometrica*, *64*(6), 1299. doi:10.2307/2171832

Kansagra, S. M., Curtis, L. H., & Schulman, K. A. (2004). Regionalization of percutaneous transluminal coronary angioplasty and implications for patient travel distance. *JAMA*, *292*(14), 1717–23. doi:10.1001/jama.292.14.1717

Kenzora, J. E., McCarthy, R. E., Lowell, J. D., & Sledge, C. B. (1984). Hip fracture mortality. Relation to age, treatment, preoperative illness, time of surgery, and complications. *Clinical Orthopaedics and Related Research*, (186), 45–56. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6723159

Kessler, D. P., & McClellan, M. B. (2000). Is Hospital Competition Socially Wasteful? *The Quarterly Journal of Economics*, *115*(2), 577–615. doi:10.1162/003355300554863

Kesternich, I., Schumacher, H., & Winter, J. (2015). Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus ? *Journal of Public Economics*, *131*, 1–11. doi:10.1016/j.jpubeco.2015.08.009

Khan, S. K., Kalra, S., Khanna, A., Thiruvengada, M. M., & Parker, M. J. (2009). Timing of surgery for hip fractures: A systematic review of 52 published studies involving 291,413 patients. *Injury*, *40*(7), 692–697. doi:10.1016/j.injury.2009.01.010

Kim, W., Wolff, S., & Ho, V. (2016). Measuring the Volume-Outcome Relation for Complex Hospital Surgery. *Applied Health Economics and Health Policy*, *14*(4), 453–464. doi:10.1007/s40258-016-0241-6

Kjerstad, E. (2005). Auctions vs negotiations: a study of price differentials. *Health Economics*, *14*(12), 1239–1251. doi:10.1002/hec.1004

Lindsay, C. M., & Feigenbaum, B. (1984). Rationing by waiting lists. *The American economic review*, *74*(3), 404–17. Retrieved from https://www.jstor.org/stable/1804016

Luft, H. S., Bunker, J. P., & Enthoven, A. C. (1979). Should Operations Be Regionalized? *New England Journal of Medicine*, *301*(25), 1364–1369. doi:10.1056/nejm197912203012503

Luft, H. S., Garnick, D. W., Mark, D. H., Peltzman, D. J., Phibbs, C. S., Lichtenberg, E., & Mcphee, S. J. (1990). Does Quality Influence Choice of Hospital? *JAMA: The Journal of the American Medical Association*, *263*(21), 2899–2906. doi:10.1001/jama.1990.03440210049031

Luft, H. S., Hunt, S. S., & Maerki, S. C. (1987). The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health services research*, *22*(2), 157–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3112042

Martin, S., & Smith, P. C. (1999). Rationing by waiting lists: an empirical investigation. *Journal of Public Economics*, *71*(1), 141–164. doi:10.1016/S0047-2727(98)00067-X

Mattke, S., Kelley, E., Scherer, P., Hurst, J., Luisa, M., & Lapetra, G. (2006). *HEALTH CARE QUALITY INDICATORS PROJECT INITIAL INDICATORS REPORT*. Retrieved from https://www.oecd.org/els/health-systems/36262514.pdf

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In *Zarembka, p. (ed.), frontiers in econometrics* (Chap. 4, pp. 105–142). Academic Press, New York. Retrieved from https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf

McGuire, K. J., Bernstein, J., Polsky, D., & Silber, J. H. (2004). The 2004 Marshall Urist Award: Delays until Surgery after Hip Fracture Increases Mortality. *Clinical Orthopaedics and Related Research*, *428*(428), 294–301. doi:10.1097/01.blo.0000146743.28925.1c

Mesman, R., Westert, G. P., Berden, B. J., & Faber, M. J. (2015). Why do high-volume hospitals achieve better outcomes? A systematic review about intermediate factors in volume–outcome relationships. *Health Policy*, *119*(8), 1055–1067. doi:10.1016/j.healthpol.2015.04.005

OECD. (2019). *Health at a Glance 2019*. Health at a Glance. doi:10.1787/4dd50c09-en

OECD. (2020). *Health at a Glance: Europe 2020*. Health at a Glance: Europe. doi:10.1787/82129230-en

Orosz, G. M., Magaziner, J., Hannan, E. L., Morrison, R. S., Koval, K., Gilbert, M., . . . Siu, A. L. (2004). Association of Timing of Surgery for Hip Fracture and Patient Outcomes. *JAMA*, *291*(14), 1738–1743. doi:10.1001/jama.291.14.1738

Pincus, D., Ravi, B., Wasserstein, D., Huang, A., Paterson, J. M., Nathens, A. B., . . . Wodchis, W. P. (2017). Association Between Wait Time and 30-Day Mortality in Adults Undergoing Hip Fracture Surgery. *JAMA*, *318*(20), 1994. doi:10.1001/jama.2017.17606

R.I.P.O. (2018). *REPORT of R.I.P.O. Regional Register of Orthopaedic Prosthetic Implantology OVERALL DATA HIP, KNEE AND SHOULDER ARTHROPLASTY IN EMILIA-ROMAGNA REGION (ITALY)*. Retrieved from http://ripo.cineca.it/authzssl/pdf/report_eng_2016.pdf

Rachet-Jacquet, L., Gutacker, N., & Siciliani, L. (2021). Scale economies in the health sector: The effect of hospital volume on health gains from hip replacement surgery. *Journal of Economic Behavior & Organization*, *190*, 704–729. doi:10.1016/j.jebo.2021.08.014

Ramdas, K., Saleh, K., Stern, S., & Liu, H. (2018). Variety and Experience: Learning and Forgetting in the Use of Surgical Devices. *Management Science*, *64*(6), 2590–2608. doi:10.1287/mnsc.2016.2721

Ricketts, J., & Sherry, P. (2018). A Single Femoral Component for All Total Hip Replacements Performed by a Trust? Does This Affect Early Clinical and Radiological Outcomes? *The archives of bone and joint surgery*, *6*(1), 27–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/29430492

Roos, N. P., Wennberg, J. E., & McPherson, K. (1988). Using diagnosis-related groups for studying variations in hospital admissions. *Health care financing review*, *9*(4), 53–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10312632

Sahni, N. R., Dalton, M., Cutler, D. M., Birkmeyer, J. D., & Chandra, A. (2016). Surgeon specialization and operative mortality in United States: retrospective analysis. *BMJ*, *354*, i3571. doi:10.1136/bmj.i3571

Schilling, M. A., Vidal, P., Ployhart, R. E., & Marangoni, A. (2003). Learning by Doing Something Else: Variation, Relatedness, and the Learning Curve. *Management Science*, *49*(1), 39–56. doi:10.1287/mnsc.49.1.39.12750

Schreyögg, J., Bäumler, M., & Busse, R. (2009). Balancing adoption and affordability of medical devices in Europe. *Health Policy*, *92*(2-3), 218–224. doi:10.1016/j.healthpol.2009.03.016

Schreyögg, J., & Stargardt, T. (2010). The trade-off between costs and outcomes: the case of acute myocardial infarction. *Health services research*, *45*(6 Pt 1), 1585–601. doi:10.1111/j.1475-6773.2010.01161.x

Siciliani, L., & Hurst, J. (2005). Tackling excessive waiting times for elective surgery: A comparative analysis of policies in 12 OECD countries. *Health Policy*, *72*(2), 201–215. doi:10.1016/j.healthpol.2004.07.003

Siciliani, L., Moran, V., & Borowitz, M. (2014). Measuring and comparing health care waiting times in OECD countries. *Health Policy*, *118*(3), 292–303. doi:10.1016/j.healthpol.2014.08.011

Sorenson, C., & Kanavos, P. (2011). Medical technology procurement in Europe: A cross-country comparison of current practice and policy. *Health Policy*, *100*(1), 43–50. doi:10.1016/j.healthpol.2010.08.001

Stargardt, T. (2008). Health service costs in Europe: cost and reimbursement of primary hip replacement in nine countries. *Health Economics*, *17*(S1), S9–S20. doi:10.1002/hec.1328

Tsai, A. C., Votruba, M., Bridges, J. F. P., & Cebul, R. D. (2006). Overcoming Bias in Estimating the Volume-Outcome Relationship. *Health Services Research*, *41*(1), 252–264. doi:10.1111/j.1475-6773.2005.00461.x

van Essen, J. T., Hans, E. W., Hurink, J. L., & Oversberg, A. (2012). Minimizing the waiting time for emergency surgery. *Operations Research for Health Care*, *1*(2-3), 34–44. doi:10.1016/j.orhc.2012.05.002

Vellez, M. (2011). Auctions Versus Negotiations: Evidence from Public Procurement in the Italian Healthcare Sector. *SSRN Electronic Journal*, *9*(4). doi:10.2139/ssrn.1803297

Vidal, E. I. O., Moreira-Filho, D. C., Coeli, C. M., Camargo Jr., K. R., Fukushima, F. B., & Blais, R. (2009). Hip fracture in the elderly: does counting time from fracture to surgery or from hospital admission to surgery matter when studying in-hospital mortality? *Osteoporosis International*, *20*(5), 723–729. doi:10.1007/s00198-008-0757-1

Weller, I., Wai, E. K., Jaglal, S., & Kreder, H. J. (2005). The effect of hospital type and surgical delay on mortality after surgery for hip fracture. *Journal of Bone and Joint Surgery - Series B*, *87*(3), 361–366. doi:10.1302/0301-620X.87B3.15300