Improving Performance Through Allocation and Competition: Evidence from a Patient Choice Reform*

Mika Kortelainen[a] ⓡ Liisa T. Laine[b] ⓡ Konsta Lavaste[c] ⓡ Tanja Saxell[d] ⓡ Luigi Siciliani[e]

[a] *University of Turku, Finnish Institute for Health and Welfare, VATT Institute for Economic Research*
[b] *University of Missouri*
[c] *Finnish Institute for Health and Welfare, University of Jyvaskyla*
[d] *Aalto University, VATT Institute for Economic Research, Helsinki GSE*
[e] *University of York*

Abstract

We study the allocative effects of enhancing consumer choice and non-price competition in markets with heterogeneous producers. We use comprehensive administrative data and a difference-in-differences design based on the introduction of a regional patient choice reform for planned surgeries in Finland. We find that large teaching hospitals attracted more patients and concentration increased in their markets. Waiting times decreased in hospitals exposed to the reform and more patients were treated, with little effect on clinical quality or average surgical expenditure after the reform. Our results suggest that increased choice can reallocate patients towards large producers and improve public hospital performance.

Keywords: Reallocation, Heterogeneous Producers, Performance, Competition, Concentration, Patient Choice, Market Structure

JEL: I11, L11, I18, L32, L38

*Acknowledgements:** We thank Saku Aura, Cristina Bratu, Petri Böckerman, Heather Brown, Partha Deb, Michele Fioretti, Geir Godager, Martin Hackmann, Unto Häkkinen, Tor Iversen, David Kaplan, Laura Lasio, Teemu Lyytikäinen, Jennifer Mayo, Jaakko Pehkonen, Tuomas Pekkarinen, Michael Pesko, Carol Propper, Adam Sacarny, Henri Salokangas, Stephan Seiler, Markku Siikanen, Otto Toivanen, Janne Tukiainen, Benjamin Ukert, and Jouko Verho for their comments and suggestions. We also thank all the conference and seminar participants who provided comments on this study at ASHEcon 2023, CEPR Health Economics 2023, EARIE 2023, IHEA 2023, NHESG meeting 2022, University of Missouri, University of Turku, and VATT Institute for Economic Research. For further guidance on the institutional details we appreciate the views of Göran Honga, Tapani Hämäläinen, Alpo Karila, Petra Kokko, Riitta Luosujärvi, Heikki Miettinen, Veli Penttilä, Jaakko Pihlajamäki, and Eva Salomaa. This research is supported by the Academy of Finland (Decision No. 325110), the Yrjö Jahnsson Foundation, the Instrumentarium Science Foundation, and Vakuutustiedon kehittämissäätiö. Author order randomized using Ray and Robson ⓡ (2018).

Email addresses: mika.kortelainen@utu.fi (Mika Kortelainen), lainel@missouri.edu (Liisa T. Laine), konsta.lavaste@thl.fi (Konsta Lavaste), tanja.saxell@vatt.fi (Tanja Saxell), luigi.siciliani@york.ac.uk (Luigi Siciliani)

I. Introduction

Market frictions related to consumer choice can distort the allocation of economic activity among heterogeneous producers and hamper industry performance (de Loecker and Syverson, 2021; Syverson, 2004). A central feature of well-functioning markets is that consumers can easily choose and substitute among producers. It enables markets to allocate more consumers towards larger, better-performing producers and can provide producers with greater competitive pressures to improve performance (Syverson, 2004; Chandra et al., 2016; Van Reenen, 2018). Even though the potential benefits of enhanced consumer choice are recognized in various industries, their empirical implications on market structure and producer performance are less well-understood due to lack of quasi-experimental variation in the extent of choice.

Market frictions are particularly salient in the health care sector, where patient choice is restricted by regulation, incomplete information about available alternatives, or narrow health insurance networks. Financial incentives related to patient choice are also diminished by health insurance, and prices are set administratively by regulators, rather than being market determined (Chandra et al., 2016; Gaynor, Ho and Town, 2015). Furthermore, in response to the rising health care expenditures, the increasing needs of aging populations, and the challenges of long waiting times for health care systems, governments are seeking effective policies to improve performance and use of scarce resources in the sector (OECD, 2020; National Audit Office, 2021; Gødøy et al., 2023). Yet, there is limited evidence on the effectiveness of friction-reducing policies to mitigate these challenges.

We study the effects of enhanced patient choice in a decentralized, universal health care system, where choice for public hospitals is highly restricted by regulation and waiting times rather than prices are used to allocate scarce resources as in other types of public services such as elderly care and daycare. To obtain quasi-experimental variation in the extent of choice, we use a regional patient choice reform for planned surgeries in Finland. Prior to this reform, patients were typically referred to the closest hospital within their own health care district. After the reform, patients in the reform area could choose a hospital in any of the health care districts in the reform area, leaving patients and hospitals outside the reform area unaffected. We use a difference-in-differences (DiD) design and administrative hospital discharge data to study whether the enhanced patient choice led to a reallocation of patients towards large hospitals, and what were the implications on hospital market structure and performance.

Patient choice reforms have been adopted at the national level in many countries such as the United Kingdom, the Netherlands, Sweden, Denmark, and Norway to facilitate choice and stimulate non-price (waiting time or quality) competition among public hospitals at administratively set prices. Estimating the effects of reforms is challenging, however, when reforms are implemented simultaneously nationwide and thus no suitable control group exists. Due to these challenges, the existing literature has studied changes in hospital performance after the choice reform in more versus less concentrated areas, where the intensity of competition induced by the reform might be higher due to a higher density of hospitals (Gaynor, Moreno-Serra and Propper, 2013; Cooper et al., 2011; Moscelli et al., 2018; Brekke et al., 2021; Moscelli, Gravelle and Siciliani, 2021).[1] On the other hand, by enabling patient choice among all reform area hospitals, the reform expanded markets and exposed all of them to competition. Our research design based on the regional reform and a DiD methodology provides us with a unique opportunity to estimate the effects of reform over all markets, rather than changes in the marginal effects of concentration post-reform.

We find that the reform had substantial effects on patients' hospital choices and allocation across several commonly performed planned surgeries: hip replacements, knee replacements, and all musculoskeletal surgeries. Specifically, the reform led to a reallocation of patients towards large teaching hospitals, which have better a reputation and resources compared to non-teaching hospitals.[2] Given that teaching hospitals attracted more patients, concentration, as measured by the Herfindahl-Hirschman Index, increased by up to 9 percent in their markets.

We then study the consequences of the reform for hospital performance in the health care system, which is characterized by long waiting times, approximately 5–6 months for hip and knee replacements. We find evidence that all hospital types treated more patients with shorter waiting times post-reform, consistent with hospital performance improving in response to greater choice and competition. Moreover, hospitals shortened the length of stays for all musculoskeletal surgeries by 8 percent, with little impact on hospitals' clinical quality (such as emergency readmissions), patient mix, or their average total surgical expenditures.[3] Based on our results, hospitals used resources more efficiently to increase health care production.

By showing how a friction-reducing policy can improve allocation and performance for public hospitals, we contribute to the large literature on the effects of competition and related reforms in the health care sector, as reviewed by Gaynor, Ho and Town (2015) and Handel and Ho (2021). Our paper is most closely related to the literature on patient choice reforms estimating changes in the marginal effects of concentration post-reform (Gaynor, Moreno-Serra and Propper, 2013; Cooper et al., 2011; Moscelli, Gravelle and Siciliani, 2021; Brekke et al., 2021). This literature has primarily focused on the effects on hospital performance, rather than the effects on choices or market concentration. An important exception is Gaynor, Propper and Seiler (2016), who estimated a structural model of demand to study changes in the quality elasticity of demand faced by hospitals and how this is linked to changes in mortality rates post-reform.

We differ from this literature in two ways. First, as mentioned earlier, our paper focuses on the effects of enhanced choice using a design-based approach based on DiD methodology. Second, we document comprehensive evidence on the effects along various dimensions related to choice and hospital performance, with a specific focus on the roles of producer size and resources employed. Our DiD design with variation created by the regional patient choice reform also allows us to use market concentration as an outcome instead of

[1] This quasi-DiD approach is commonly used in evaluating the marginal effects of continuous treatments in settings where nationwide reforms or shocks apply at the same time to all individuals or the population of interest (Duflo, 2001; Acemoglu, Autor and Lyle, 2004; Finkelstein, 2007).

[2] Despite their better reputation and higher quality expectations (Newsweek and Statista, 2021), we find that teaching hospitals do not provide higher clinical quality or have higher (risk-adjusted) costs of care for surgeries such as hip and knee replacements, consistent with the evidence from other settings (Silber et al., 2020; Burke et al., 2019). Teaching hospitals, however, have waiting times below the average of all hospitals.

[3] Moreover, there ware no hospital entries or exits post-reform.

a variable defining the intensity of competition induced by the reform as used in prior work estimating the marginal effects. We show how improvements in competition conditions due to greater choice can promote market concentration towards large producers.

Our paper thus contributes to the debates on market concentration and competition. On one hand, higher concentration has been seen as resulting from lower competition and the market dominance of large producers, in extreme, a monopoly. On the other hand, higher concentration can signal a well-functioning market, where large producers are better and thus, gain more consumers. (de Loecker and Syverson, 2021; Syverson, 2011; Berry, Gaynor and Scott Morton, 2019; Van Reenen, 2018) Our results support the latter explanation and show the positive effects of the pro-competitive choice reform on both market concentration and performance for public hospitals.

Finally, we link the literature analyzing patient choice reforms with an extensive literature studying the relationship between consumer allocation and producer performance across various industries (e.g., Syverson, 2011; de Loecker and Syverson, 2021). Our paper is inspired by the previous research by Chandra et al. (2016), who find that higher-quality hospitals tend to attract more patients. This finding suggests that patient demand plays an important role in market allocation despite various frictions in health care markets (Chandra et al., 2016). Unlike this paper, we present quasi-experimental evidence on the allocative effects of enhanced patient choice, with implications for the performance of public hospitals. We show that even though the resulting reallocation towards large, better-resourced teaching hospitals did not enhance clinical quality, it shortened waiting times and helped the health care system to meet patient needs with the existing resources.

The rest of the paper proceeds as follows. Section II describes the institutional setting. Section III presents the data and descriptive statistics. Section IV describes our baseline DiD approach. In Section V, we present our baseline results for the average effects of the choice reform using patient-level data. Section VI shows our results from hospital-level analyses, as well our results regarding the allocative effects between teaching and non-teaching hospitals. In Section VII, we compare these results to those obtained for the marginal effects of concentration using a quasi-DiD approach. The last section concludes.

II. Institutional Setting

II.A. The Finnish Health Care System

Finland has a decentralized, universal health care system that is financed primarily through taxation. All permanent residents are entitled to public health care services through universal, public health insurance, which are characterized by moderate co-payments and rationing, and long waiting times for non-emergency conditions (Keskimäki et al. (2019), Section II.B below). Public primary care is organized and financed by the municipalities ($N = 326$ in 2010) for their residents by law and it is provided in municipality-owned health centers by primary care physicians. Primary care physicians act as gatekeepers for planned non-emergency hospital-based specialized care in the public sector.

Specialized health care such as surgeries is provided by public hospitals governed by health care districts known as hospital districts ($N = 20$ in 2010). Each municipality is a member of one of the hospital districts and is also responsible of the governance and financing of that district together with other member municipalities. The hospital districts are responsible for organizing specialized health care services in their region. The market share for private outpatient surgeries is small, approximately 5 percent (Keskimäki et al., 2019), and is not covered by public health insurance.

The Finnish hospital industry consists of heterogeneous producers in terms of size, resources, location, and the services produced. There are large university-based teaching hospitals, medium-sized central hospitals, and small regional hospitals. Every hospital district has either a teaching hospital or central hospital typically located close to the center of the district. In addition, hospital districts may have one or more regional hospitals that only provide services for most common medical conditions. In contrast, teaching hospitals have a much greater range of services compared to other types of hospitals such as central and regional hospitals, which we refer to as non-teaching hospitals hereafter. This includes services for common medical conditions as well as specialist services for more serious and rare diseases. Moreover, teaching hospitals partner with local university medical schools to provide medical education and conduct medical research. There are five such hospitals in Finland, but they have much higher patient volumes, better resources and capacity compared to non-teaching hospitals ($N = 41$) (Karhunen, 2020).

II.B. Regional Patient Choice Reform and Incentives

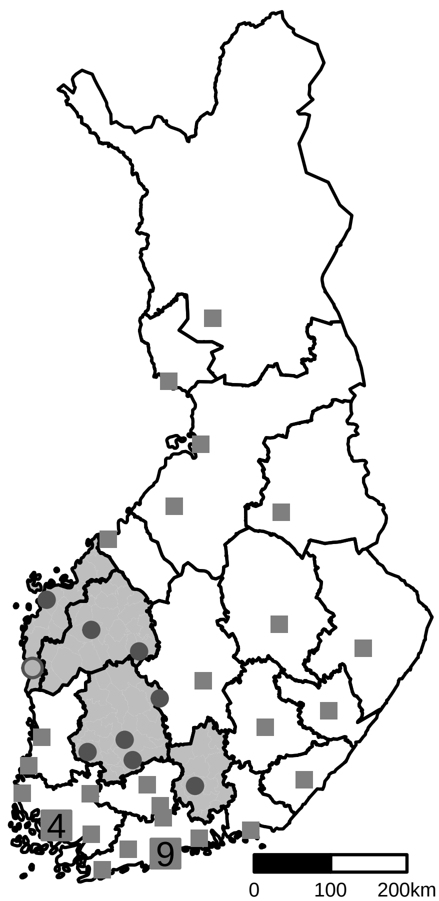
We study a regional patient choice reform that was introduced in four hospital districts in South-West Finland in October 2007, comprising approximately one-fifth of the Finnish population. Prior to the reform, patients were typically referred to the closest hospital within their own hospital district (Government Proposal 90/2010).[4] The reform allowed planned surgical care patients in the reform area to choose any public hospital within and across hospital districts in the reform area. The reform was effective until patient choice was allowed nationwide in April 2011. Figure 1 shows the shaded reform area, hospital districts, and geographical distribution of hospitals in Finland in 2004–2010. Online Appendix Figure A1 shows the locations of different hospital types: teaching, central, and regional.

The central policy goals of the reform were to enhance patient choice and timely access to care, improve hospital performance such as quality of care and waiting times, and increase non-price competition between hospitals (Pirkanmaa Hospital District, 2007). According to our hospital expert interviews, policymakers in hospital districts in the reform area also hoped that patients would substitute to hospitals with shorter waiting times, consequently shifting demand to ease the pressure in overly crowded hospitals. In fact, hospital

[4]In a few hospital districts in and outside the reform area physicians could refer patients to hospitals located outside patients' hospital districts in specific circumstances, such as long travel distance or substantial waiting time. Our results are robust to excluding such cases from the econometric analyses (Section V.C).

waiting times are long in Finland: for example, in 2007, 13 percent of patients had to wait more than six months for hospital care, which is the national waiting time target set by law in March 2005 (THL, 2012). The reform was inspired by earlier market-oriented patient choice reforms conducted in many other Nordic countries and the United Kingdom (Pirkanmaa Hospital District, 2007), but in contrast to these national reforms the Finnish reform was regional.

Figure 1: The 2007 Reform Area and Hospital Locations



Notes: Borders indicate hospital districts in 2007 and the shaded area constitutes the 2007 reform area. The dots indicate reform area hospitals (the dot with an empty middle marks a hospital which closed down in the pre-reform period) and the squares control area hospitals. The large squares mark the capital region, which had 9 hospitals, and the Turku region, which had 4 hospitals. In total, there were $N = 9$ hospitals in the reform area and $N = 37$ hospitals in the control area. The figure includes all public hospitals that performed planned surgeries (excluding municipal-owned hospitals), although some of them did not perform hip and/or knee replacement surgeries.

Patients access planned surgical care by a referral from their primary care physician, and patients' hospital choices are guided by these physicians. The financial incentives associated with hospital referral decisions are minimal as public primary care physicians are salaried employees of municipalities rather than hospital districts. Private primary care physicians are also able to make referrals to public hospitals, but the receiving hospitals' specialists assess whether it is necessary for the patient to undergo the procedure or not.

In the Finnish health care system, public hospitals are reimbursed for the services produced from their patients' municipalities of residence. Each hospital district sets the reimbursement rates of their own hospitals administratively, and many base them on nationally fixed diagnosis-related groups (DRGs) (Kautiainen, Häkkinen and Lauharanta, 2011). For example, all of the reform area hospital districts reimbursed DRG tariffs either exclusively or combined with a fee-for-service model during our study period 2004—2010.[5] DRG systems incentivize hospitals to control costs while increasing activity levels by providing them with a predetermined, flat reimbursement rate for treating patients within a single DRG category based on their average, rather than actual, costs. In DRG systems, hospitals, however, have incentives to cream-skim and compete for profitable patients whose reimbursement is expected to be above the actual costs (Ellis, 1998).

Patients are publicly and universally insured, which reduces their financial consequences related to hospital choice. In the publicly administered health care system, patients' co-payments are generally moderate, capped by national legislation, and do not vary much between hospital districts (Hetemaa et al., 2018). For example, in 2008, the maximum co-payment for a surgery was 83.90 euros (Government Decree 464/2008). Hence, we do not expect hospital co-payments to impact patients' hospital choices to a great extent. However, significant monetary costs can result for patients from traveling to a distant hospital because the Finnish population is spread out over a large geographical area, the distances between hospitals are long, and the National Health Insurance Scheme covers travel costs based on the cheapest mode of transport to the nearest hospital, regardless of the actual mode of transport or hospital choice (Paltta, 2008).

Although public hospitals may not be profit maximizers in the same way as private for-profit hospitals are, they face significant pressures to perform well financially due to tight public sector budgets. Thus, attracting patients is important for hospitals because their funding and financial performance depends on it through municipality reimbursements. Given that administered co-payments are almost fixed, the way hospitals and their managers can increase demand is by making effort to improve performance in terms of quality or waiting times. The patient choice reform brought about a substantial shift in hospitals' ability to attract and compete for patients (see Gaynor, Moreno-Serra and Propper (2013) and Gaynor, Ho and Town (2015) for discussions of the role of non-price competition and related reforms for public hospitals with administered prices).

Patients' ability to choose their hospital, and thereby the intensity of non-price competition, depends on the available information (Brown et al., 2023). There is publicly available information on hospital performance outcomes, although no specific patient review system similar to the one maintained by the English National Health Service (NHS) is provided in Finland. Hospital districts publish information on hospital-level waiting times by specialty on their own websites. Nationwide statistics on hospital district-level waiting times are also collected for common procedures such as hip and knee replacements (THL, 2012)—the surg-

[5] We are not aware of any major changes in the hospital reimbursement systems during the study period. However, one of the four districts in the reform area switched its pricing from its own grouping to nationally set DRG grouping in 2005. We have confirmed that our main results remain intact if we exclude this hospital district from the econometric analyses.

eries we study. In terms of clinical quality information, the Finnish Institute for Health and Welfare publishes information on the outcomes of hip and knee replacements, acute myocardial infarction (AMI), and stroke patients at hospital or hospital district level at regular time intervals, for example in 2007 (THL, 2021). In addition to this public information, patients can receive information informally from their referring physicians and unofficial sources such as friends, family, and peers.

III. Data

We use a nationwide patient-level hospital discharge dataset that contains the universe of public hospital admissions and discharges in Finland in 2004–2010. We create three samples for our analyses using the information on hospital admission and discharge-related procedures and diagnoses. The first two samples include patients who had planned primary hip and knee replacement surgery. We choose hip and knee replacement surgeries as they have been analyzed in prior work (Moscelli, Gravelle and Siciliani, 2021; Feng et al., 2015; Goude et al., 2022), were amongst the most common planned surgeries, were available in all types of hospitals, and we expect scope for producer competition as a result of the choice reform. The third sample includes all planned musculoskeletal surgeries, which gives a more comprehensive picture of the potential effects of the choice reform. This sample also includes the primary hip and knee replacement surgeries that account for approximately 16 percent of the observations in the sample.

In total, our samples contain 45 hospitals during the observation period, including in total 29,625 observations for the samples of planned hip replacement surgeries, 35,884 observations for knee replacement surgeries, and 418,109 observations for all musculoskeletal surgeries.

We focus on patients aged 18—74 years at the time of hospital admission.[6] We match each observation with administrative data from Statistics Finland on the patient's date of death, demographics, and residence location at the end of each year. Next we provide the relevant information from our main variables and sample construction, while leaving the more detailed description to online Appendix Section A1. In the remainder of this section, we first describe our variables and then proceed to the descriptive statistics.

III.A. Measures of Hospital Choice

We construct four outcome variables that relate to the extent to which planned surgical patients exercise hospital choice. First, we construct our main choice variable of interest: a binary indicator equal to one if the patient chose a teaching hospital. Compared with non-teaching hospitals, teaching hospitals are generally large, have better resources, and are perceived to be higher quality compared to non-teaching hospitals (Section II.A, Silber et al. 2020; Newsweek and Statista 2021).

[6]We do not have detailed information on the residence location of patients over 74 years of age, and our data contain only a few patients under 18 years of age. The previous literature on patient choice reforms also focuses on patients under 75 years of age (Gaynor, Moreno-Serra and Propper, 2013).

Our second measure of hospital choice is the distance traveled, which is the straight-line distance between each patient's residence location and the location of the hospital where the patient was treated. For the patient, we use information on the residence location according to a one-kilometer-by-one-kilometer grid cell from Statistics Finland. For the hospital, we use the coordinates of the center of the municipality in which the hospital was located. The third measure is a binary indicator equal to one if the patient chose the nearest hospital. The fourth measure is a binary indicator equal to one if the patient was treated outside their hospital district of residence.

III.B. Measures of Hospital Performance

Clinical quality measures. We measure hospital performance based on a clinical quality metric that is a binary indicator equal to one if the patient had an emergency readmission within 30 days of discharge from last hospital in the treatment spell.[7] Emergency readmissions are a commonly used metric of hospital quality in health economics and clinical studies (Benbassat and Taragin, 2000; Varkevisser, van der Geest and Schut, 2012; Moscelli, Gravelle and Siciliani, 2021), and their incidence is also used as a quality indicator in pay for performance schemes (Gupta, 2021). Because it is possible that emergency readmissions are not sensitive enough to capture changes in all quality attributes, we also conduct additional analyses using more detailed measures of clinical quality such as mechanical complications and infections in the prosthesis (Section V.C).

Waiting time. In addition to clinical quality, we examine other commonly studied aspects of hospital performance. We measure waiting time, which is the number of days from a patient being placed on the waiting list (after a specialist's final assessment of the need for surgery) to being admitted to hospital for surgery.[8] From an economics perspective, waiting times act as a non-price rationing device in publicly-funded health systems, which combine moderate to zero co-payments with the presence of capacity constraints (Sá, Siciliani and Straume, 2019). Waiting times for planned surgery can, however, be long despite pre-specified policy targets (OECD, 2020; Siciliani, Moran and Borowitz, 2014). Thus, waiting times can reflect congestion and performance in public hospitals.

Measures of efficiency and resource use. Measuring performance based on hospital efficiency (how well resources are utilized to achieve the output) is a long-standing challenge because of absence of high-quality data on costs (Cooper, Gibbons and Skellern, 2018). Even our detailed discharge data do not include comprehensive information on costs or resources used and, thus we follow the previous literature and use length of stay as a proxy for hospital efficiency (Robinson et al., 1988; Gaynor, Moreno-Serra and Propper, 2013; Cooper, Gibbons and Skellern, 2018; Moscelli, Gravelle and Siciliani, 2021); to the extent that clinical

[7] We calculate the 30-day follow-up period of emergency readmission from discharge from the last, rather than the initial, hospital to account for transfers to another hospital after surgery (Torkki, 2012). The practice is similar to transferring patients to post-acute care facilities in the U.S. In our data, 4–19 percent of patients (depending on the sample) were transferred to another hospital for post-acute care.

[8] Some hospitals or hospital districts, however, have reported their waiting times less consistently than others, although in most cases waiting times are observed (online Appendix Section A1). In Section V.C, we show that the results regarding waiting times remained similar when we excluded such hospital districts from our samples.

quality does not change, a shorter stay indicates faster discharge, and thereby lower costs and less resources used for the same patient outcomes. Moreover, we separately analyze a coarse measure of annual hospital operating expenditure (e.g., purchases of labor and material inputs) for all surgeries collected from all individual hospitals by the Finnish Institute for Health and Welfare (Section VI). In the discharge data, planned musculoskeletal surgeries represent more than 20 percent of all surgeries.

III.C. Measure of Hospital Market Structure

We measure market structure at the hospital level using the Herfindahl-Hirschman Index (HHI) based on observed hospital choices (Gaynor, Moreno-Serra and Propper, 2013). In the first step, we calculate an HHI value for each municipality by taking the sum of the squared patient market shares of hospitals using data on patients from that municipality only. In the second step, we calculate the hospital-level HHI values by taking a weighted average of the values of the municipal-level HHI, where each municipality is weighted according to its share of the hospital's total patient volume. Because we use market shares in this calculation, the HHI varies between 0 and 1, and hospitals located in highly concentrated markets (high hospital-level HHI) also have high market shares (correlation approximately 0.9). The hospital-level HHI measure captures the degree of concentration in each hospital's municipality markets and allows large hospitals to operate in a larger markets compared to smaller hospitals. We refer to this HHI calculated from observed hospital choices as the *actual* HHI and we calculate it separately for each estimation sample.

III.D. Patient and Surgery Covariates

Our main covariates are patient's age and sex at the time of admission in addition to the patient's surgery type, because many possible covariates (such as financial position or staffing) may be endogenous. In our robustness checks (Section V.C), we also estimate models using two additional covariates. The first is an indicator for a weekend admission (equal to one if admitted on Saturday and Sunday), as staff may be more limited in the weekend. The second is the number of past emergency admissions each patient had within one year prior to surgery, as a proxy for severity and morbidity (case mix).

III.E. Descriptive Statistics

Table 1 reports the descriptive statistics of hospital choice outcomes during our observation period for the samples of hip replacements, knee replacements, and all musculoskeletal surgeries (panels A–C). The first three columns refer to all patients in each of these samples, followed by the descriptive statistics split for patients receiving care from hospitals in the reform and control areas. In each sample, approximately one fourth of the patients resided in areas affected by the reform.

Table 1 shows that patients in the reform and controls area differed in the way they exercise choice for hip and knee replacement surgeries, but less so for all musculoskeletal surgeries. The hip or knee replacement patients were more likely to be treated in a teaching instead of non-teaching hospital in the reform than in the control area (55–56 versus 36–40 percent). They also traveled longer distances, approximately 34–37 kilometers in the areas affected by the reform, as opposed to approximately 28–29 kilometers in the control areas. Moreover, the hip or knee replacement patients were less likely to be treated in the nearest hospital (81–84 versus 87 percent) and more likely to be treated outside their own hospital district in the reform than in the control area (6–7 versus 2 percent).

Table 1 also shows that hospital volumes were larger in the reform than in the control area in the samples of hip and knee replacements, whereas the reverse was true in the sample of all musculoskeletal surgeries. Moreover, the hospital-level means of actual HHI indicate a high degree of market concentration at the level of 0.69—0.90, with fairly high variation across hospitals and over time (SD 0.09–0.16).[9]

Table 1: Descriptive Statistics of Choice Outcomes and Market Concentration

	All			*Reform*			*Control*		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Panel A. Hip replacement surgeries									
Teaching hospital	0.44	0.50	29,625	0.55	0.50	6,974	0.40	0.49	22,651
Distance (km)	30.55	39.23	29,625	37.21	43.60	6,974	28.50	37.55	22,651
Nearest hospital	0.85	0.35	29,625	0.81	0.39	6,974	0.87	0.34	22,651
Different hospital district	0.03	0.18	29,625	0.07	0.26	6,974	0.02	0.15	22,651
Hospital volume	36.94	35.14	802	62.27	46.13	112	32.83	31.17	690
Actual HHI	0.87	0.12	802	0.87	0.09	112	0.87	0.13	690
Panel B. Knee replacement surgeries									
Teaching hospital	0.41	0.49	35,884	0.56	0.50	8,276	0.36	0.48	27,608
Distance (km)	29.25	37.61	35,884	34.29	39.57	8,276	27.74	36.87	27,608
Nearest hospital	0.86	0.34	35,884	0.84	0.37	8,276	0.87	0.33	27,608
Different hospital district	0.03	0.17	35,884	0.06	0.23	8,276	0.02	0.15	27,608
Hospital volume	44.30	39.64	810	73.89	59.08	112	39.55	33.23	698
Actual HHI	0.89	0.12	810	0.90	0.10	112	0.88	0.12	698
Panel C. All musculoskeletal surgeries									
Teaching hospital	0.43	0.49	418,109	0.29	0.45	72,532	0.46	0.50	345,577
Distance (km)	27.99	39.66	418,109	29.40	35.25	72,532	27.70	40.52	345,577
Nearest hospital	0.84	0.37	418,109	0.77	0.42	72,532	0.85	0.35	345,577
Different hospital district	0.04	0.20	418,109	0.04	0.21	72,532	0.04	0.19	345,577
Hospital volume	466.64	489.34	896	370.06	263.68	196	493.68	532.78	700
Actual HHI	0.76	0.14	896	0.69	0.16	196	0.78	0.12	700

Notes: The table reports descriptive statistics for 18–74-year-old patients in 2004–2010. Distance is continuous and the other choice outcomes are binary. Hospital volume (number of patients) and actual HHI are calculated at the hospital-quarter level. The HHI is measured on a 0–1 scale, where greater value indicates more concentration.

Table 2 reports the descriptive statistics of our hospital performance outcomes in addition to the characteristics for surgical patients. Hospital performance was similar between the reform and control areas in terms of the clinical quality outcome, the probability of 30-day emergency readmission (6–11 percent), and length of stay (2–7 days). In contrast, waiting times were approximately 5–10 percent longer in the reform area than in the control area (136–194 versus 129–184 days). Despite these differences in waiting times, patient characteristics (age and sex) were similar in the reform and control areas. Only the average number

[9]Based on the distributions of the actual HHI, the concentration decreased in the reform area after the reform, while remaining relatively stable in the control area, especially in the sample of hip and knee replacements (online Appendix Figure A2).

Table 2: Descriptive Statistics of Quality, Length of Stay, Waiting Time and Patient Characteristic Measures

	All			*Reform*			*Control*		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Panel A. Hip replacement surgeries									
Emergency readmission within 30 days	0.08	0.27	29,625	0.07	0.26	6,974	0.08	0.28	22,651
Length of stay (days)	7.14	7.54	29,625	6.89	6.97	6,974	7.21	7.70	22,651
Waiting time (days)	154.74	141.15	23,481	165.74	152.12	6,394	150.62	136.59	17,087
Age	62.29	9.00	29,625	62.53	9.01	6,974	62.22	9.00	22,651
Female	0.52	0.50	29,625	0.51	0.50	6,974	0.53	0.50	22,651
N of pre-surgery emergency admissions	0.40	1.16	29,625	0.32	0.99	6,974	0.42	1.21	22,651
Panel B. Knee replacement surgeries									
Emergency readmission within 30 days	0.11	0.31	35,884	0.11	0.32	8,276	0.11	0.31	27,608
Length of stay (days)	6.77	5.96	35,884	6.58	5.28	8,276	6.82	6.14	27,608
Waiting time (days)	186.42	173.46	28,269	194.32	188.13	7,541	183.54	167.72	20,728
Age	64.06	7.43	35,884	64.57	7.23	8,276	63.91	7.48	27,608
Female	0.65	0.48	35,884	0.66	0.47	8,276	0.65	0.48	27,608
N of pre-surgery emergency admissions	0.38	1.09	35,884	0.32	0.91	8,276	0.40	1.13	27,608
Panel C. All musculoskeletal surgeries									
Emergency readmission within 30 days	0.06	0.24	418,109	0.06	0.23	72,532	0.06	0.25	345,577
Length of stay (days)	2.23	6.11	418,109	2.55	5.90	72,532	2.16	6.15	345,577
Waiting time (days)	130.20	146.29	294,233	135.68	148.24	56,368	128.90	145.80	237,865
Age	51.87	13.70	418,109	52.90	13.80	72,532	51.66	13.67	345,577
Female	0.52	0.50	418,109	0.52	0.50	72,532	0.53	0.50	345,577
N of pre-surgery emergency admissions	0.59	1.40	418,109	0.49	1.24	72,532	0.61	1.43	345,577

Notes: The table reports descriptive statistics for 18–74-year-old surgical patients in 2004–2010. The hip replacement sample includes 8, the knee replacement sample includes 7, and the musculoskeletal sample includes 648 types of planned surgeries. Waiting time is missing for some patients, which is depicted as a smaller number of observations.

of pre-surgery emergency admissions was smaller in the reform than in the control area, suggesting that hospitals in the reform area treated less severe patients. In Section IV, we address the differences between the reform and control areas and the plausibility of the parallel trends assumption in our econometric approach.

Finally, we present the descriptive statistics by hospital type to shed light on heterogeneity in outcomes and resources across different hospital types (online Appendix Tables A1–A3). As expected, large teaching hospitals had much larger volumes than central and regional hospitals on average (92–1,282 versus 30–416 and 16–217 patients per quarter, respectively; Table A1). In terms of hospital performance, teaching hospitals had shorter mean waiting times than medium-sized central hospitals for hip and knee replacements (Table A2). The waiting times of teaching hospitals were also approximately 20 days below the overall average of all hospitals for hip and knee replacements (153 and 197 versus 175 and 216 patients per quarter, respectively) and fairly close to the overall average for all musculoskeletal surgeries. Moreover, even though teaching hospitals are generally considered higher-quality and more expensive (Silber et al., 2020; Burke et al., 2019), their risk-adjusted readmission rates (Table A2) and surgical expenditures per spell (Table A3) did not differ much from those of non-teaching hospitals after accounting for differences across hospitals in patient case mix or the types of surgeries provided. This finding is consistent with the evidence for common medical and surgical conditions in other settings (Silber et al., 2020; Burke et al., 2019). In sum, much of the surgical activity was concentrated in large teaching hospitals without the costs of care or clinical quality being significantly higher than in non-teaching hospitals. Teaching hospitals had relatively short waiting times for hip and knee replacements, which might have incentivized patients to exercises choice.

IV. Econometric Approach for Estimating Choice Reform Effects

The patient choice reform should affect hospital choices, allocation, and performance by increasing substitution and non-price competition across hospitals (Syverson, 2011; Chandra et al., 2016; Moscelli, Gravelle and Siciliani, 2021). Thus, we estimate the average effects of the reform on such outcomes, using nationwide administrative data and a difference-in-differences (DiD) approach. We use the reform area as a treatment group and the remaining areas of the country as a control group. Specifically, we employ the following baseline specification:

$$y_{imht} = \beta_1 \mathbb{1}[\text{Treated}_h] + \beta_2 \mathbb{1}[\text{Treated}_h] \times \mathbb{1}[\text{Post}_t] + \mathbf{X}'_{it}\gamma + \lambda_t + \mu_m + \varepsilon_{imht}, \tag{1}$$

where y_{imht} is the outcome for patient i living in municipality m and treated by hospital h in period (quarter) t. $\mathbb{1}[\text{Treated}_h]$ is a binary indicator for the treatment group equal to one if hospital h was located in the reform area, and equal to zero if located in the control area. $\mathbb{1}[\text{Post}_t]$ is a binary post-reform indicator equal to one after the introduction of the reform in the fourth quarter of 2007 (Q4/07). We include quarter fixed effects λ_t to control for time-varying national-level shocks that may affect the outcome (they also absorb $\mathbb{1}[\text{Post}_t]$) in addition to the patient's municipality of residence fixed effects, μ_m, to control for any time-invariant differences between municipalities (and also between the reform and control areas), for example in their average population size and morbidity. $\mathbf{X}_{it}$ includes patient-specific control variables: type of surgery, sex, and 10-year age bins.

To avoid the bad control problem (Angrist and Pischke, 2009) we use a minimal set of patient characteristics (age and sex) as control variables in our baseline analysis. To address the concern that our results are driven by changes in the patient mix, we show that the reform had little impact on hospitals' patient age and sex mix (Section VI). Moreover, our results are also robust to controlling for hospital fixed effects, which capture time-invariant hospital-level unobserved factors such as average patient mix, and to controlling for the patient's pre-existing health status (Section V.C). We cluster standard errors at the level of the patient's municipality of residence ($N = 326$) to account for within-area correlation in patients' unobservables in hospital choice and to ensure consistency across different outcomes and specifications.[10] We view the municipality-level clustering as a conservative choice because many Finnish municipalities are geographically large and unobservables in hospital choice are even more strongly correlated within smaller regions such as postal codes.

The key coefficient of interest β_2 identifies the average treatment effect of the choice reform on the patient

[10]Unobservables in hospital choice are expected to be primarily correlated between patients from the same region because they have similar distances to hospitals and possibly also similarities in latent health status. The statistical significance of the DiD results regarding hospital performance outcomes does not change much if we use hospital-level clustering instead.

outcome, using variation across regions in the adoption of the patient choice reform and the assignment to separate treatment and control groups. This holds to the extent that in the absence of policy adoption, patient outcomes would have evolved under parallel trends in the reform and control areas. To examine potential pre-existing trends and the dynamic effects of the choice reform graphically, we estimate the following binary treatment event study specification with 6-month time intervals calculated from the adoption of the reform in October 2007:

$$y_{imht} = \delta_1 \mathbb{1}[\text{Treated}_h] + \mathbb{1}[\text{Treated}_h] \times \sum_{l=-7}^{+5} \delta_{2,l} \mathbb{1}[l = t] + \mathbf{X}'_{it}\tau + \lambda_t + \mu_m + \varepsilon_{imht}. \tag{2}$$

The coefficients for the pre-reform period $\delta_{2,l}$, $l < -1$ capture a possible pre-existing trend in the outcome variable, whereas the coefficients $\delta_{2,l}$, $l > -1$ for the post-adoption periods capture the dynamic effect of the choice reform in each of these periods. We use the same set of controls and fixed effects as in specification (1) and follow the standard practice by normalizing the coefficients for the indicators "one period before adoption" to zero, $\delta_{2,-1} = 0$.

V. Results

V.A. Effects on Hospital Allocation and Choice

We investigate whether the reform led to the reallocation of patients towards large, better-resourced teaching hospitals. We present the results from estimating the DiD specification (1) for hip replacements, knee replacements, and all musculoskeletal surgeries in panels A–C of Table 3, respectively. As expected, we find that patients were much more willing to undergo a surgery in a teaching, rather than in non-teaching hospital after the reform (column 1). More specifically, the probability of choosing a teaching hospital increased by 5–6 percentage points (10–14 percent) for hip and knee replacements. Besides having better resources and reputation, teaching hospitals have waiting times much below the overall average for hip and knee replacements (Section III.E), which might have incentivized patients to exercise choice. Compared with hip and knee replacements, we find a smaller effect (an increase of 2 percentage points or 3 percent) for all musculoskeletal surgeries, which also include a large number of less-invasive surgeries and surgeries with short waiting times.[11]

Figure 2 presents the corresponding event study estimates from estimating specification (2). We find that the probability of choosing teaching hospital began to increase half a year after the implementation of the reform in every sample (Figure 2). The lag in the effects may result from waiting times for planned surgery and patients or physicians adjusting to the new choice system. The event study estimates reveal

[11] When the health risks of surgery are small and waiting times are short, patients might be less willing to choose a possibly more distant but better perceived teaching hospital. Consistent with this, we find that the point estimate for the choice of teaching hospital is no longer statistically significant and the probability of choosing the nearest hospital increased when we exclude hip and knee replacements from the sample of all musculoskeletal surgeries (online Appendix Table A4).

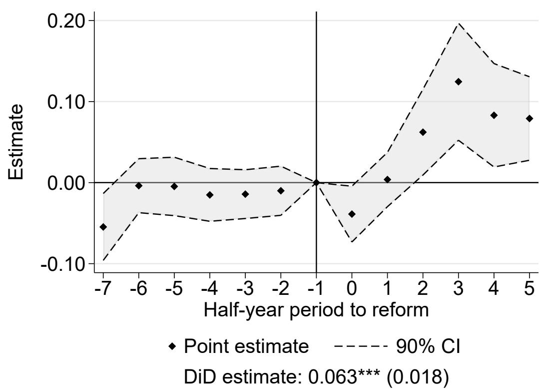
very little evidence of pre-trends in the choice outcome, providing supporting evidence for the credibility of our research design.

We also investigate whether the reform that expanded patient choice of hospitals within and across hospital districts affected the travel distance to the hospital. The reform may have induced patients to choose more distant but preferred hospitals outside their hospital district, thereby increasing the travel distance. On the other hand, the reform allowed some patients to choose nearer hospitals than before the reform, which may have decreased the travel distance.[12] We find that the surgical patients responded to the reform by choosing more distant hospitals (column 2 of Table 3, online Appendix Figure A4). Depending on the sample, the travel distance increased by approximately by 2–5 kilometers or 6–16 percent compared to the pre-reform mean distance. The DiD point estimate is statistically significant at 5 percent level for all musculoskeletal surgeries, and at 10 percent level for hip replacements, while not statistically significant for knee replacements.

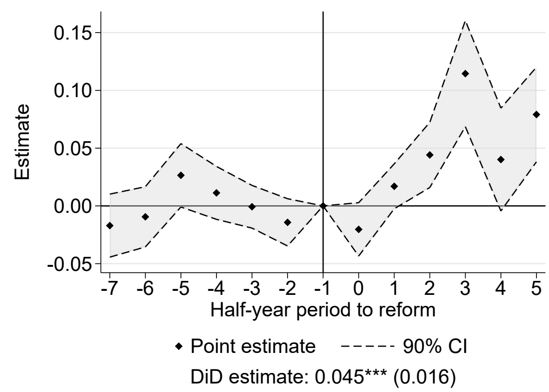
We also tested whether patients' tendency to be treated in the geographically nearest hospital changed after the reform, but the parameter estimates are small in magnitude and statistically significant only for knee replacements (column 3 of Table 3 and online Appendix Figure A5). In turn, we find evidence that the reform substantially increased the probability of a patient being treated outside their own hospital district (column 4 of Table 3 and online Appendix Figure A6). The magnitude of the DiD point estimates (Table 3) ranges from 1 to 3 percentage points (24–83 percent in comparison to the mean) depending on the sample. The statistically and economically significant estimates for hip replacements and all musculoskeletal surgeries mean that a large share of the patients chose a hospital beyond their own hospital district after the reform. In sum, the reform had considerable effects on patients' behavior and allocation to hospitals despite the relatively low densities of patients and hospitals in Finland and the very small financial consequences of publicly-insured patients' hospital choice.

[12] According to our data, the mean distance decreased for patients living in some hospital district border areas.

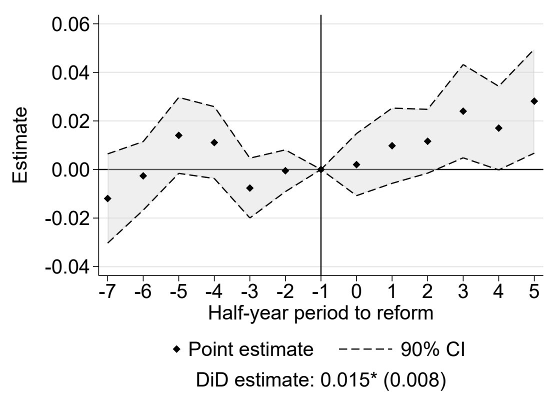
Figure 2: Effect of the Reform on Probability of Surgical Patients Being Treated In Teaching Hospital



Panel A. Hip replacements



Panel B. Knee replacements



Panel C. All musculoskeletal surgeries

Notes: Includes the DiD estimates corresponding to column 4 in Table 3.

Table 3: Effects of the Reform on Hospital Choice and Allocation Outcomes

	Teaching hospital[a] (1)	Distance (km)[b] (2)	Nearest hospital[c] (3)	Different hospital district[d] (4)
Panel A. Hip replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.063***	4.651*	0.008	0.030**
	(0.018)	(2.385)	(0.020)	(0.012)
$\text{mean}(y_{imht} \| \text{Post}_t = 0)$	0.460	29.847	0.864	0.036
N	29,625	29,625	29,625	29,625
Panel B. Knee replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.045***	1.725	0.042**	0.013
	(0.016)	(1.820)	(0.020)	(0.011)
$\text{mean}(y_{imht} \| \text{Post}_t = 0)$	0.436	28.953	0.870	0.033
N	35,884	35,884	35,884	35,884
Panel C. All musculoskeletal surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.015*	2.063**	0.020	0.010*
	(0.008)	(0.993)	(0.015)	(0.005)
$\text{mean}(y_{imht} \| \text{Post}_t = 0)$	0.434	28.046	0.838	0.041
N	418,090	418,090	418,090	418,090
Surgery type FEs	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using equation (1). Standard errors clustered at the level of patient's home municipality ($N = 326$).
[a] Equals one if patient was treated in teaching (university) hospital.
[b] Distance from patient's residence to the hospital in kilometers.
[c] Equals one if patient was treated in the geographically nearest hospital.
[d] Equals one if patient was treated in a hospital that was located in another hospital district than the one the patient lived in.

V.B. Effects on Hospital Performance

Table 4 displays the results on the effects of the choice reform on hospital performance, as measured by the probability of 30-day readmission, waiting time, and length of stay. Column 1 shows that the reform did not affect the emergency readmission probability for hip replacement (panel A), knee replacement (panel B), and all musculoskeletal surgery patients (panel C). The point estimates are small in magnitude and not statistically significant. The corresponding event study estimates are shown in online Appendix Figure A7 and they confirm that almost none of the post-reform point estimates are statistically different from zero.

Column 2 in Table 4 and Figure 3 show the estimated effects on waiting times. We find that in every sample, waiting times decreased dramatically post-reform. The waiting time estimates are precise, and similarly to the results for other outcomes and samples, there is very little evidence of pre-trends for hip replacements and all musculoskeletal surgeries. However, in one sample–knee replacements– waiting times started to decrease more quickly in the reform area than in the control area already in the pre-reform period. Based on raw data patterns (online Appendix Figure A3), this is mostly explained by a law change introducing a national 6-month treatment time guarantee at the beginning of our observation period (Q1/2005). Even though our econometric approach captures national-level shocks such as law changes through time fixed effects, the introduction of the treatment guarantee disproportionately affected long waiting times for knee replacements in the reform area.

Table 4: Effects of the Reform on Surgical Care Quality, Waiting Time, and Length of Stay

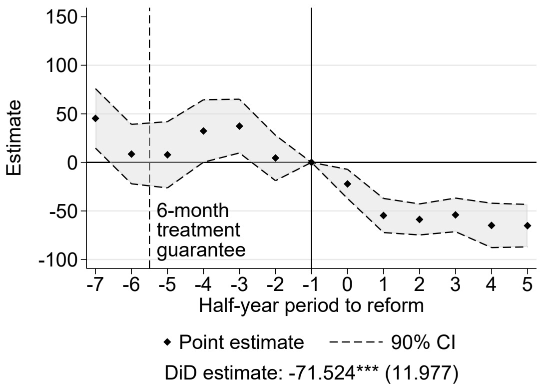
	Readmission[a] (1)	Waiting time[b] (2)	Length of stay[c] (3)
Panel A. Hip replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	-0.004	-71.524***	-0.140
	(0.007)	(11.977)	(0.307)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.080	183.757	8.044
N	29,625	23,481	29,625
Panel B. Knee replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.002	-97.614***	-0.375
	(0.009)	(18.255)	(0.288)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.103	229.671	7.528
N	35,884	28,269	35,884
Panel C. All musculoskeletal surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.001	-18.535**	-0.187***
	(0.003)	(7.844)	(0.047)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.062	150.935	2.340
N	418,090	294,198	418,090
Surgery type FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using equation (1). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality ($N = 326$).
[a] Emergency readmission (to any hospital) within 30 days after discharge.
[b] Number of days. Part of the values are missing, which results in smaller N compared to other columns (see online Appendix Section A1.6 for more details).
[c] Number of days.

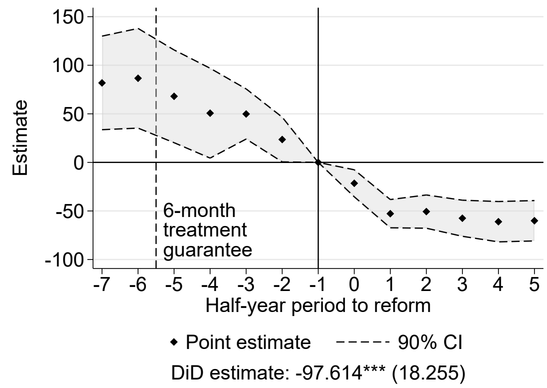
To isolate the effect of the patient choice reform from that of the treatment guarantee, we use data from periods when waiting times were already adjusted to this law change, Q4/2006-Q4/2010 (Table 5). We find that the decrease in waiting times was substantial after the patient choice reform: 36 percent (in comparison to the pre-reform mean) for hip and knee replacements, and 23 percent for all musculoskeletal surgeries. Based on the results and raw data patterns, the patient choice reform was effective in reducing waiting times below the 6-month maximum for waiting times set by national legislation. In Section VI, we study the mechanisms behind the reduction in waiting times and show that the reduction coincides with a large increase in hospital volumes.

Hospitals can reduce the length of stay in order to improve efficiency and to free resources to treat more patients. We find that the estimates of the effects of the reform on length of stay are not statistically significant for hip and knee replacement patients (Column 3 of Table 4, Figure 4). On the other hand, for all musculoskeletal surgeries, including many minor procedures, the length of stay decreased by 8 percent from the pre-reform mean. The reform may have incentivized hospitals to reduce the length of stay for patients needing relatively less invasive musculoskeletal surgery, whose health is less likely to deteriorate due to shorter stays compared with those needing major surgery such as a hip or knee replacement. To the extent that clinical quality did not change, as indicated by our earlier results on readmissions, shorter stays imply improved efficiency post-reform.

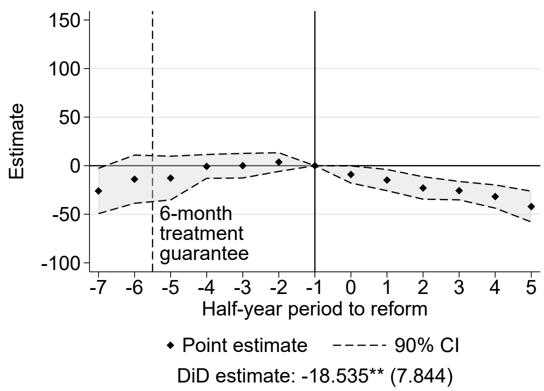
Figure 3: Effect of the Reform on Waiting Times of Planned Surgery Patients Using Binary Treatment



Panel A. Hip replacements



Panel B. Knee replacements



Panel C. All musculoskeletal surgeries

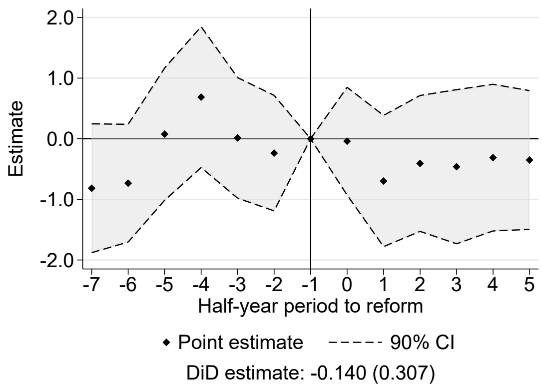
Notes: Includes the DiD estimates corresponding to column 2 in Table 4.

Table 5: Effect of the Choice Reform on Waiting Time Estimates After the 6-month Medical Treatment Guarantee

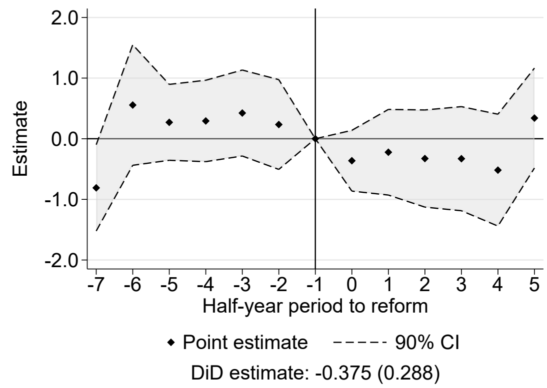
	Hip replacements		Knee replacements		All musculoskeletal surgeries	
	Whole time period	After treatment guarantee	Whole time period	After treatment guarantee	Whole time period	After treatment guarantee
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Treated}_h \times \text{Post}_t$	-71.524***	-51.192***	-97.614***	-58.931***	-18.535**	-27.368***
	(11.977)	(8.580)	(18.255)	(10.895)	(7.844)	(6.092)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	183.757	140.893	229.671	165.659	150.935	120.628
N	23,481	14,858	28,269	18,304	294,198	187,211
Surgery type FEs	✓	✓	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Outcome is waiting time in days. Estimated using equation (1). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality ($N = 326$). Columns (1), (3), and (5) re-display baseline estimates from Table 4 column (2). Columns (2), (4), and (6) display estimates using data from Q4/2006–Q4/2010 only.

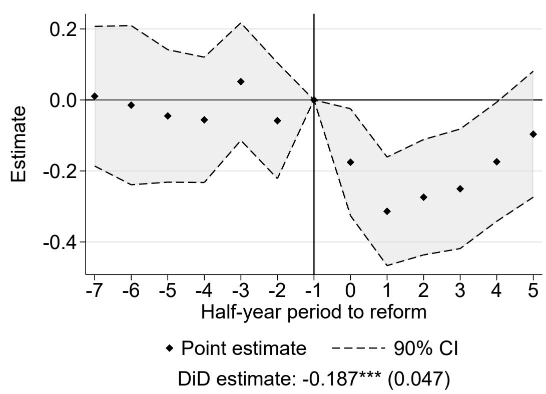
Figure 4: Effect of the Reform on Length of Stay of Planned Surgery Patients Using Binary Treatment



Panel A. Hip replacements



Panel B. Knee replacements



Panel C. All musculoskeletal surgeries

Notes: Includes the DiD estimates corresponding to column 3 in Table 4.

V.C. Robustness Checks and Additional Quality Outcomes

We conduct a number of robustness tests (online Appendix Section A4). We begin by testing whether our results are sensitive to the inclusion of additional controls to the baseline specifications. For choice-related outcomes, we add controls for staffing and morbidity (indicator for weekend admission, and the number of pre-surgery emergency admissions). For the hospital performance outcomes, we also add hospital fixed effects to control for time-invariant hospital-level factors such as their average patient mix. The results remain intact (Tables A5 and A6).

We also study the robustness of our results regarding our baseline samples. First, although there was no hospital entry or exit post-reform, one small hospital in the reform area closed down during the pre-reform period and we excluded it from our baseline estimations.[13] The baseline results are robust for excluding all patients living in municipalities near the closed-down hospital (Tables A7 and A8). Second, in a few hospital districts, patients had the opportunity to obtain referrals to hospitals outside of their own hospital district under specific circumstances, such as long travel distance. Our results are robust for excluding these hospital districts from the econometric analyses, and in some cases this exclusion improves the precision and increases the magnitude of our estimates compared to the baseline estimates (Tables A9–A12). Third, one reform area hospital district implemented DRG pricing in the beginning of 2005. The results are robust for excluding the district from the estimations, although some estimates lose statistical significance (Tables A13–A14). Fourth, some hospitals had a joint hospital identifier in the data and cannot be distinguished from each other. This creates some measurement error in our outcomes related to travel distances and the indicator of choosing the nearest hospital, which are calculated based on the location of the largest hospital under the joint identifier. When we exclude hospitals with joint identifiers from the sample, we find larger effects compared with the baseline estimates (Tables A15 and A16).

To address potential bias from missing values in waiting times, we re-estimated the results regarding waiting times i) without hospital districts in which more than 30 percent of waiting time values were missing, ii) without surgeries for which more than 30 percent of waiting time values were missing, iii) without one reform area hospital which did not report its waiting times in Q1/2008–Q4/2009, and iv) using data from years 2006 and 2010 only, when the share of missing values was generally low. The re-estimated results are similar or even stronger than our baseline results (Table A17).

To address a potential concern of emergency readmissions not being sensitive enough to capture changes in all clinical quality attributes, we also estimated the effects of the reform on more detailed measures for planned hip and knee replacements: indicators for revision surgery, mechanical complication in the prosthesis, and infection or inflammation in the prosthesis (Bayliss et al., 2017; Fleischman et al., 2019; Urquhart et al., 2010; Mäkelä et al., 2011). The results in online Appendix Table A18 show that the point

[13] The exclusion may generate some bias because it removes most of the patients from the nearby area in the pre-closure period, but not in the post-closure period.

estimates are not statistically significant on these quality outcomes, except for the revision probability of knee replacement surgeries, which increased 0.7 percentage points or 23 percent compared to to the mean. We also estimated the effects of the reform on *emergency* care clinical quality and length of stay for acute myocardial infarction (AMI), stroke, and hip fracture patients. The benefit of using AMI, stroke, and hip fracture indicators is that it mitigates patient selection into hospitals (Kessler and McClellan, 2000; Moscelli et al., 2018). Complementing our baseline results, the DiD estimates in online Appendix Table A19 show no other statistically significant effects than a decrease in stroke patients' emergency readmissions (4 percentage points or 22%).

VI. Hospital-Level Analyses and Heterogeneity

VI.A. Hospital Volume, Concentration, and Mean Performance

The results in Section V.A showed that the reform had substantial effects on hospital choices as patients were more likely to choose teaching, instead of non-teaching, hospitals in the post-reform period. In this section, we study whether these effects translated into changes in hospitals' patient volumes, market concentration, and mean performance and whether these changes were different in teaching versus non-teaching hospitals. We estimate the following specification using hospital-quarter-level data:

$$\begin{aligned} y_{ht} = &\beta_1 \mathbb{1}[\text{Treated}_h] \times \mathbb{1}[\text{Post}_t] + \beta_2 \mathbb{1}[\text{Post}_t] \times \mathbb{1}[\text{Teaching}_h] \\ &+ \beta_3 \mathbb{1}[\text{Treated}_h] \times \mathbb{1}[\text{Post}_t] \times \mathbb{1}[\text{Teaching}_h] + \mathbf{X}'_{ht}\gamma + \lambda_t + \eta_h + \varepsilon_{ht}, \end{aligned} \tag{3}$$

where y_{ht} is the outcome for hospital h and period (quarter) t, and $\mathbb{1}[\text{Teaching}_h]$ is the teaching hospital indicator. η_h are hospital fixed effects, which absorb time-invariant hospital-level factors such as $\mathbb{1}[\text{Treated}_h]$, $\mathbb{1}[\text{Teaching}_h]$, their interaction, hospital location, and average patient mix. Covariates $\mathbf{X}_{ht}$ are the same as in the patient-level specification (equation (1)), but transformed to the hospital-level means in quarter t. The coefficient β_1 shows the effect of the reform in non-teaching hospitals and β_3 shows the potentially heterogeneous effect in teaching compared to non-teaching hospitals. For the specification estimating the average hospital-level effects (without heterogeneity), we set $\beta_2, \beta_3 = 0$.[14] We cluster the standard errors at the hospital.

Table 6 shows the effects of the patient choice reform on hospital volumes and market concentrations, as measured by the actual hospital-level HHI. The results for the average effects are shown in columns (1) and (3), while the heterogeneous effects are shown in columns (2) and (4). We find that hospital volumes increased by 9–31 percent on average after the reform (column 1). The increase is statistically significant for knee replacements and all musculoskeletal surgeries and not statistically significant for hip replacements. In

[14] More specifically, we estimate the following specification: $y_{ht} = \beta_1 \mathbb{1}[\text{Treated}_h] \times \mathbb{1}[\text{Post}_t] + \mathbf{X}'_{ht}\gamma + \lambda_t + \eta_h + \varepsilon_{ht}$.

every sample, hospital volumes increased disproportionally in teaching hospitals (column 2). For hip and knee replacements, this increase was 22–30 patients per quarter (the sum of the coefficients on $\text{Treated}_h \times \text{Post}_t$ and $\text{Treated}_h \times \text{Post}_t \times \text{Teaching}_h$), which corresponds to an increase of 56–64 percent compared with the mean volume of all hospitals (35–41 patients per quarter) and 24–29 percent with the mean volume of teaching hospitals (92–105 patients per quarter, see online Appendix Table A1).

The DiD estimates do not display any statistically significant effect on the actual HHI on average (column 3). The average treatment effect, however, masks considerable heterogeneity in the effects across different types of hospitals (column 4): depending on the sample, the measure of market concentration increased by 4–9 percent for teaching hospitals and decreased by 1–4 percent for non-teaching hospitals, although the latter effect was not statistically significant. Thus, reflecting the disproportionate increase in teaching hospitals' volumes, concentration in their markets increased.

Table 6: Effects of the Reform on Hospital Volumes and Market Structure

	Hospital volume[a]		Actual HHI[b]	
	DiD (1)	Heterogeneity (2)	DiD (3)	Heterogeneity (4)
Panel A. Hip replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	4.745	0.084	-0.016	-0.038
	(3.053)	(2.078)	(0.026)	(0.028)
$\text{Treated}_h \times \text{Post}_t \times \text{Teaching}_h$		22.269***		0.114**
		(6.703)		(0.050)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	35.021	35.021	0.883	0.883
N	802	802	802	802
Panel B. Knee replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	12.414**	7.494***	-0.015	-0.032
	(4.620)	(1.838)	(0.034)	(0.041)
$\text{Treated}_h \times \text{Post}_t \times \text{Teaching}_h$		22.860***		0.101*
		(5.709)		(0.056)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	40.640	40.640	0.899	0.899
N	810	810	810	810
Panel C. All musculoskeletal surgeries				
$\text{Treated}_h \times \text{Post}_t$	39.702**	24.110	-0.002	-0.008
	(18.244)	(16.063)	(0.010)	(0.011)
$\text{Treated}_h \times \text{Post}_t \times \text{Teaching}_h$		104.107***		0.035*
		(34.634)		(0.019)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	462.267	462.267	0.766	0.766
N	896	896	896	896
Hospital and time FEs	✓	✓	✓	✓
Age & sex mix[c]	✓	✓	✓	✓
Surgery types[d]	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Includes hospital-quarter-level observations from Q1/2004 to Q4/2010. Standard errors clustered at the hospital level (N=29–32 depending on the sample).
[a] Number of surgeries.
[b] Observed market concentration measured on a 0–1 scale. See A1.9 for more information.
[c] Shares of females, 18–29, 30–39, 40–49, 50–59, and 70–74-year-old patients of hospital's total patient volume. Baseline = share of 60–69-year-old male patients.
[d] Incidence of different procedure codes among hospital's patients.

Table 7 shows the effects of the reform on hospital-level means of performance outcomes. We find that the estimates for emergency readmissions were generally small and not statistically significant (columns 1 and 2). Moreover, in line with the increase in hospital volumes (Table 6), we find evidence that the reform reduced mean waiting times in both teaching and non-teaching hospitals (columns 3 and 4). Even though some of the point estimates for mean waiting time are imprecisely estimated, this finding suggests that hospital competition improved in response to increased choice. For hip and knee replacements, the mean length of stay decreased by 11–12 percent in non-teaching hospitals, whereas for all musculoskeletal surgeries, the statistically insignificant point estimates suggest a larger reduction in teaching hospitals (column 6).

Table 7: Effects of the Reform on Hospital-level Means of Care Quality, Waiting Times and Length of Stay

	Emergency readmission[a]		Waiting time[b]		Length of stay[c]	
	DiD (1)	Hetero-geneity (2)	DiD (3)	Hetero-geneity (4)	DiD (5)	Hetero-geneity (6)
Panel A. Hip replacement surgeries						
$\text{Treated}_h \times \text{Post}_t$	0.005	0.008	-86.808**	-69.829	-0.640	-0.966***
	(0.009)	(0.010)	(34.512)	(43.052)	(0.481)	(0.342)
$\text{Treated}_h \times \text{Post}_t \times \text{Teaching}_h$		-0.019		-62.892		1.048*
		(0.016)		(45.713)		(0.561)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.085	0.085	174.557	174.557	9.006	9.006
N	802	802	668	668	802	802
Panel B. Knee replacement surgeries						
$\text{Treated}_h \times \text{Post}_t$	0.010	0.010	-104.156**	-96.983*	-0.756	-0.988*
	(0.019)	(0.024)	(40.003)	(54.100)	(0.508)	(0.549)
$\text{Treated}_h \times \text{Post}_t \times \text{Teaching}_h$		-0.004		-24.556		0.773
		(0.027)		(58.179)		(0.612)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.111	0.111	215.895	215.895	8.257	8.257
N	810	810	664	664	810	810
Panel C. All musculoskeletal surgeries						
$\text{Treated}_h \times \text{Post}_t$	-0.003	-0.004	-14.575	-8.785	-0.120	-0.087
	(0.004)	(0.005)	(16.787)	(16.685)	(0.124)	(0.145)
$\text{Treated}_h \times \text{Post}_t \times \text{Teaching}_h$		0.008		-33.458*		-0.188
		(0.006)		(19.196)		(0.158)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.064	0.064	135.748	135.748	2.154	2.154
N	896	896	845	845	896	896
Hospital and time FEs	✓	✓	✓	✓	✓	✓
Age & sex mix[d]	✓	✓	✓	✓	✓	✓
Surgery types[e]	✓	✓	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Includes hospital-quarter-level observations from Q1/2004 to Q4/2010. Standard errors clustered at the hospital level (N=29–32 depending on the sample).
[a] 30-day emergency readmission rate.
[b] Mean waiting time in days. Some hospitals did not report waiting times in some quarters, which results in a smaller N compared to the other columns (see online Appendix Section A1.6 for more details).
[c] Mean length of stay in days.
[d] Shares of females, shares of 18–29, 30–39, 40–49, 50–59, and 70–74-year-old patients of hospital's total patient volume. Baseline = share of 60–69-year-old male patients.
[e] Incidence of different procedure codes among a hospital's patients.

VI.B. Resource Use and Case Mix

Earlier, we found that the reform led to shorter lengths of stays for musculoskeletal patients, which may have enabled hospitals to reduce costs and to free up capacity and human capital resources to treat more patients. Next, we study the effects on hospitals' total annual surgical expenditures and a coarse measure of lower productivity: surgical expenditure per treatment spell. We estimate a specification similar to the one in equation (3), except that we use hospital-year (rather than hospital-quarter)-level data and only include the case mix index in $\mathbf{X}_{ht}$.[15] We find statistically insignificant and very small negative effect on hospitals' total surgical expenditure on average (online Appendix Table A20), so total resource use such as purchases of material and labor inputs did not increase for all hospitals on average after the reform despite the increase in their volumes (Table 6). The only exception is teaching hospitals, whose total expenditure increased by 30 percent, but not as much as their volumes (an increase of over 60 percent in Table 6). Moreover, we do not find any economically or statistically significant effect on expenditure per spell in teaching and non-teaching hospitals, so hospitals did not use more resources to produce better patient outcomes. Using information in hospital districts' annual reports, we also confirmed that teaching hospitals did not operate at full capacity or increase capacity in terms of number of beds post-reform.

Because the reform changed hospitals' volumes and patient allocation into different types of hospitals, it could also have affected hospitals' patient composition and case mix. For example, if hospitals treated sicker patients post-reform, it could explain the null effects on emergency readmissions (quality of care) for surgical patients. To explore changes in patient composition and allocation into hospitals based on morbidity, we use the same estimation strategy as in equation (3) while using hospital-wise means of pre-determined patient characteristics as outcomes. The results in online Appendix Table A21 do not reveal any impact on the mean age, sex composition, or morbidity (number of pre-surgery emergency admissions) of a hospital's patients.[16] The only exception is the sample of hip replacements, where non-teaching hospitals treated sicker patients post-reform, whereas teaching hospitals attracted healthier patients.

VII. Comparison to Quasi-DiD Approach

We also use a complementary quasi-DiD approach to estimate differential changes in hospital performance in more versus less concentrated markets after the reform allowed all patients to choose their hospital (within the reform area) to compare our main DiD approach and results to prior research studying nationwide patient choice reforms (Cooper et al., 2011; Gaynor, Moreno-Serra and Propper, 2013; Moscelli et al., 2018; Moscelli, Gravelle and Siciliani, 2021; Roos et al., 2020; Brekke et al., 2021).[17] In general, a quasi-DiD approach with a continuous treatment provide different information about the treatment effects than a binary DiD approach, but requires relatively strong identification assumptions, for example, on treatment effect homogeneity at different levels of treatment (Fricke, 2017; de Chaisemartin and D'Haultfœuille, 2018; Callaway, Goodman-

[15]The case mix index is part of the official benchmarking statistics on the costs and productivity of hospitals from the Finnish Institute for Health and Welfare.

[16]We also confirmed that within each sample of planned surgeries, there is no clear change in hospitals' surgery types or mix, which would have indicated teaching hospitals performing more demanding hip or knee replacement surgeries post-reform.

[17]It would also be possible in our setting to use data from both reform and control areas. It would, however, make the model and its interpretation more complicated. Analyzing the reform area (treatment group) only also makes the approach comparable to the previous choice reform literature and the literature evaluating the continuous treatment approach (Callaway, Goodman-Bacon and Sant'Anna, 2021).

Bacon and Sant'Anna, 2021). Specifically, we estimate the following specification, using patient-level data from the reform area only ($\mathbb{1}[\text{Treated}_h] = 1$):

$$y_{imht} = \alpha_1 \widehat{HHI}_h + \alpha_2 \widehat{HHI}_h \times \mathbb{1}[\text{Post}_t] + \mathbf{X}_{it}'\nu + \lambda_t + \mu_m + \hat{r}_{it}'\theta + \epsilon_{imht}, \tag{4}$$

where $\widehat{HHI}_h$ is the continuous treatment assigned to hospital h, that is, the pre-reform mean of hospital-level predicted HHI.[18] To mitigate the endogeneity of the market structure or continuous treatment, we follow the standard practice (Kessler and McClellan, 2000; Gaynor, Moreno-Serra and Propper, 2013; Moscelli, Gravelle and Siciliani, 2021) and use the predicted, rather than actual, HHI in the pre-reform period. Following the literature, we predict the HHI based on distance and other covariates except for our measures of hospital performance for all patients that we subject to the choice reform. Moreover, to ensure that time-varying patient allocation to hospitals based on unobserved morbidity does not bias the estimates, we follow Moscelli, Gravelle and Siciliani (2021) and control for the first-stage stage residuals ($\hat{r}_{it}$) from a hospital choice model.[19] We bootstrap the clustered standard errors (N=74 municipalities in the reform area) because $\widehat{HHI}_h$ is based on predicted choices. We base our inference on p-values and confidence intervals or their bounds because the bootstrap algorithm, a wild cluster bootstrap, does not produce standard errors. See online Appendix Section A2 for further details on the quasi-DiD approach.

The coefficient of interest, α_2, captures the differential response to the choice reform between hospitals in more versus less concentrated markets, measured by $\widehat{HHI}_h$. The change in the marginal effect of concentration after the choice reform is identified through differences in the treatment intensity, rather than assignment to a separate treatment or control group. The choice reform is the treatment, and its effect or intensity is assumed to vary across hospitals exposed to different market structures.

The sign of α_2 is ambiguous, however, depending on whether concentration is associated with either less or more competition (see, e.g., Syverson, 2019). Hospitals in less concentrated markets can be exposed to the reform and non-price competition more due to a higher density of competing hospitals and more options for patients to choose from post-reform (Gaynor, Moreno-Serra and Propper, 2013; Brekke, Siciliani and Straume, 2011). If this were the case, then we would expect hospital performance to improve more in less concentrated markets post-reform with $\alpha_2 < 0$. On the other hand, by removing constraints on choice across different hospitals and hospital districts, the reform expanded the markets and exposed all reform area hospitals and markets to competition. If hospitals in more versus less concentrated markets were affected similarly by the reform, then $\alpha_2 = 0$, even if the reform itself had a large effect on the outcome. Finally, if the reform predominantly improved choice and competition for hospitals faced with relatively high levels of

[18] Note that the HHI is hospital-specific, municipal fixed effects μ_m are patient-specific, and the hospital treats patients from multiple municipalities. Hence, the time-invariant HHI is not perfectly collinear with the patient municipality fixed effects.

[19] Changes in patient allocation into hospitals based on morbidity do not seem to drive our DiD and quasi-DiD results on hospital performance. This is shown by the results using a control function strategy in the DiD approach (online Appendix Table A22) and the quasi-DiD results without the control function approach (online Appendix Table A23).

market concentration (in our setting, medium-sized central hospitals and large teaching hospitals), then we would expect $\alpha_2 > 0$.[20]

The results for the marginal effects are shown in Table 8. The estimates for the non-emergency quality measure, surgical patients' readmission probability, are not statistically significant (column 1). Following the previous choice reform literature (Gaynor, Moreno-Serra and Propper, 2013; Brekke et al., 2021; Cooper et al., 2011), we also study the effects on emergency care quality and find that concentration was associated with an increase in AMI patients' mortality and stroke patients' emergency readmissions post-reform (online Appendix Table A24).

For hip and knee replacements, the waiting time point estimates are not statistically significant (column 2 of Table 8). For all musculoskeletal surgeries, hospitals in more concentrated markets had a larger decrease in waiting times: the statistically significant point estimate of -217 implies that a one standard deviation (0.11) increase in the predicted HHI (more concentration) is associated with a $217 \times 0.11 \approx 24$ days (15 percent) decrease in waiting time post-reform.

We also find that hospitals in more concentrated markets had a larger reduction in the length of stay (column 3 of Table 8). For example, in the sample of hip replacements, a one standard deviation (0.11) increase in the predicted HHI is associated with a $20 \times 0.11 \approx 2.2$ days (approximately 28 percent) reduction in the length of stay. Provided that the clinical quality (readmission probability) did not change much, shorter stays indicate improved productivity and lower costs after the reform in more concentrated markets.

A comparison of our quasi-DiD and DiD (Section V.B) results show their similarity regarding surgical patients' readmission probability, as all the point estimates regarding this outcome are small and not statistically significant. For hip and knee replacements, waiting times decreased after the reform based on the DiD estimates and the decrease was larger, although not statistically significant, in less concentrated (more competitive) areas. For all musculoskeletal surgeries, waiting times also decreased after the reform, but the decrease was smaller in less concentrated areas. Therefore, the results on waiting times differ between the two empirical approaches. Similarly, length of stay was longer in less concentrated areas. Overall, our results suggest that the quasi-DiD approach commonly used in the literature gives different results on the changes in the marginal effects of concentration post-reform than the DiD approach employed in this study that estimates the effect of the reform over all markets.

[20] Specifically, central hospitals had a mean HHI of 0.85–0.94, followed by teaching hospitals (0.85–0.89), and regional hospitals (0.64–0.80), see online Appendix Table A1.

Table 8: Marginal Effects of Concentration and Reform on Surgical Care Quality, Waiting Time, and Length of Stay

	Readmission[a] (1)	Waiting time[b] (2)	Length of stay[c] (3)
Panel A. Hip replacement surgeries			
PredictedHHI$_h$ × Post$_t$	0.152	170.199	-19.150***
	[-0.042, 0.329]	[-322.737, 639.127]	[-25.362, -12.596]
mean(y_{imht}|Post$_t = 0$)	0.071	212.851	7.756
N	6,974	6,393	6,974
Panel B. Knee replacement surgeries			
PredictedHHI$_h$ × Post$_t$	0.038	411.592	-19.526***
	[-0.299, 0.412]	[-156.686, 981.157]	[-24.287, -14.198]
mean(y_{imht}|Post$_t = 0$)	0.103	265.332	7.485
N	8,276	7,541	8,276
Panel C. All musculoskeletal surgeries			
PredictedHHI$_h$ × Post$_t$	0.015	-216.998**	-2.219***
	[-0.040, 0.076]	[-341.779, -88.250]	[-2.901, -1.511]
mean(y_{imht}|Post$_t = 0$)	0.056	160.512	2.763
N	72,483	56,307	72,483
Surgery type FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓
Control function residuals	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using equation (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=74). HHI measured on a 0–1 scale, where higher value indicates more concentration.
[a] Emergency readmission (to any hospital) within 30 days of departing from the last hospital in the treatment spell.
[b] Number of days. Some of the values are missing, which results in smaller N compared to other columns (see online Appendix Section A1.6 for more details).
[c] Number of days.

VIII. Conclusions

We studied the allocative effects of enhancing patient choice and non-price competition among public hospitals. We used a difference-in-differences (DiD) approach based on a unique quasi-experiment in Finland: a regional patient choice reform for planned surgeries. Using nationwide administrative data, we found that surgical patients responded to the reform by traveling longer distances and across hospital districts to receive care from large teaching hospitals. The effects on patient behavior were considerable despite relatively low densities of population and hospitals in Finland, and very small financial incentives related to hospital choice due to public insurance.

After the reform, hospitals were also able to treat more surgical patients with shorter waiting times. Hospital volumes increased disproportionately in large teaching hospitals post-reform, increasing concentration in their geographical markets. Thus, an improvement in choice and competition conditions can promote concentration by re-allocating patients towards large hospitals.

Our results suggest that market-oriented choice reforms have the potential to enhance hospital performance without raising average total expenditure or expenditure per treatment for surgeries. The reform had direct benefits for patients, as their waiting times (i.e., non-monetary costs) decreased significantly. In addition to increased competition, the decrease in waiting times can reflect improved patient allocation to hospitals. As a result of improved choice, patient demand is reallocated towards large hospitals with better resources or capacity, potentially easing the pressure in crowded hospitals with long waiting lists and large excess demand relative to their capacity. Lastly, the reform had only little impact on clinical quality or hospitals' patient mix, despite the large allocative effects and increased supply.

Our results provide suggestive evidence on the roles of enhancing consumer choice and friction-reducing public policy in improving allocation in the presence of limited resources and long waiting times, which are present in many health care systems (OECD, 2020). Our results are directly relevant to health care systems relying on public production (including those of the United Kingdom, Sweden, and Norway, for example), but also to other systems with administratively set reimbursements to producers (such as the Medicare system in the United States). Our results also have relevance in settings, such as public housing, nursing homes, and daycare centers, where long waiting times are a potential concern.

More broadly, our results based on a choice reform are informative about the role of market-based mechanisms in health care. The results closely tie greater concentration to more competition, as implied by the increased amount of choice post-reform. The more competitive a market is, the more concentrated is the market towards large, better-resourced producers, therefore potentially improving the performance of the market. Our results do not, however, imply that smaller hospitals should be closed, as their volumes did not decrease after the reform (waiting lists just became shorter) and there are likely to be social gains from a larger number of hospitals due to better scope for choice and competition. While our results suggest that the reform improved allocation and performance in terms of waiting times and there has been no hospital entry or exit after the choice reform, further research is needed to disentangle the demand and supply side responses, as well as to study the social welfare implications of enhancing consumer choice.

References

Acemoglu, Daron, David H. Autor, and David Lyle. 2004. "Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury." *Journal of Political Economy*, 112(3): 497–551.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton: Princeton University Press.

Bayliss, Lee E., David Culliford, A. Paul Monk, Sion Glyn-Jones, Daniel Prieto-Alhambra, Andrew Judge, Cyrus Cooper, Andrew J. Carr, Nigel K. Arden, David J. Beard, and Andrew J. Price. 2017. "The Effect of Patient Age at Intervention on Risk of Implant Revision After Total Replacement of the Hip or Knee: a Population-Based Cohort study." *The Lancet*, 389(10077): 1424–1430.

Benbassat, Jochanan, and Mark Taragin. 2000. "Hospital Readmissions as a Measure of Quality of Health Care: Advantages and Limitations." *Archives of Internal Medicine*, 160(8): 1074–1081.

Berry, Steven, Martin Gaynor, and Fiona Scott Morton. 2019. "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization." *Journal of Economic Perspectives*, 33(3): 44–68.

Brekke, Kurt R., Chiara Canta, Luigi Siciliani, and Odd Rune Straume. 2021. "Hospital Competition in a National Health Service: Evidence from a Patient Choice Reform." *Journal of Health Economics*, 79: 102509.

Brekke, Kurt R., Luigi Siciliani, and Odd Rune Straume. 2011. "Hospital Competition and Quality with Regulated Prices." *Scandinavian Journal of Economics*, 113(2): 444–469.

Brown, Zach Y., Christopher Hansman, Jordan Keener, and Andre F. Veiga. 2023. "Information and Disparities in Health Care Quality: Evidence from GP Choice in England." National Bureau of Economic Research Working Paper 31033.

Burke, Laura G., Dhruv Khullar, Jie Zheng, Austin B. Frakt, E. John Orav, and Ashish K. Jha. 2019. "Comparison of Costs of Care for Medicare Patients Hospitalized in Teaching and Nonteaching Hospitals." *JAMA Network Open*, 2(6): e195229–e195229.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with a Continuous Treatment." *Working paper. https://doi.org/10.48550/arXiv.2107.02637.*

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." *American Economic Review*, 106(8): 2110–44.

Cooper, Zack, Stephen Gibbons, and Matthew Skellern. 2018. "Does Competition from Private Surgical Centres Improve Public Hospitals' Performance? Evidence from the English National Health Service." *Journal of Public Economics*, 166: 63–80.

Cooper, Zack, Stephen Gibbons, Simon Jones, and Alistair McGuire. 2011. "Does Hospital Competition Save Lives? Evidence From The English NHS Patient Choice Reforms." *Economic Journal*, 121(554): F228–F260.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2018. "Fuzzy Differences-in-Differences." *Review of Economic Studies*, 85(2): 999 – 1028.

de Loecker, Jan, and Chad Syverson. 2021. "An Industrial Organization Perspective on Productivity." In *Handbook of Industrial Organization, Vol. 4.* , ed. Kate Ho, Ali Hortaçsu and Alessandro Lizzeri, Chapter 3, 141–223. Amsterdam: Elsevier.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*, 91(4): 795–813.

Ellis, Randall P. 1998. “Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins.” *Journal of Health Economics*, 17(5): 537–555.

Feng, Yan, Michele Pistollato, Anita Charlesworth, Nancy Devlin, Carol Propper, and Jon Sussex. 2015. “Association Between Market Concentration of Hospitals and Patient Health Gain Following Hip Replacement Surgery.” *Journal of Health Services Research & Policy*, 20(1): 11–17.

Finkelstein, Amy. 2007. “The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare.” *The Quarterly Journal of Economics*, 122(1): 1–37.

Finnish Intitute for Health and Welfare (THL). 2012. “Trends in Access to Specialised Health Care in 2007–2012.” *https://urn.fi/URN:NBN:fi-fe2012113010110*.

Fleischman, Andrew N., Majd Tarabichi, Zachary Magner, Javad Parvizi, and Richard H. Rothman. 2019. “Mechanical Complications Following Total Hip Arthroplasty Based on Surgical Approach: A Large, Single-Institution Cohort Study.” *The Journal of Arthroplasty*, 34(6): 1255–1260.

Fricke, Hans. 2017. “Identification Based on Difference-in-Differences Approaches with Multiple Treatments.” *Oxford Bulletin of Economics and Statistics*, 79(3): 426–433.

Gaynor, Martin, Carol Propper, and Stephan Seiler. 2016. “Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service.” *American Economic Review*, 106(11): 3521–57.

Gaynor, Martin, Kate Ho, and Robert J. Town. 2015. “The Industrial Organization of Health-Care Markets.” *Journal of Economic Literature*, 53(2): 235–84.

Gaynor, Martin, Rodrigo Moreno-Serra, and Carol Propper. 2013. “Death by Market Power: Reform, Competition, and Patient Outcomes in the National Health Service.” *American Economic Journal: Economic Policy*, 5(4): 134–66.

Gødøy, Anna, Venke F. Haaland, Ingrid Huitfeldt, and Mark Votruba. 2023. “Hospital Queues, Patient Health and Labor Supply.” *American Economic Journal: Economic Policy, Forthcoming.*

Goude, Fanny, Göran Garellick, Sverre Kittelsen, Henrik Malchau, Mikko Peltola, and Clas Rehnberg. 2022. “Effects of Competition and Bundled Payment on the Performance of Hip Replacement Surgery in Stockholm, Sweden: Results from a Quasi-Experimental Study.” *BMJ Open*, 12(7).

Government Decree 464/2008. 2008. *https://www.finlex.fi/fi/laki/alkup/2008/20080464*.

Government Proposal 90/2010. 2010. *https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Documents/he_90+2010.pdf*.

Gupta, Atul. 2021. “Impacts of Performance Pay for Hospitals: The Readmissions Reduction Program.” *American Economic Review*, 111(4): 1241–83.

Handel, Ben, and Kate Ho. 2021. "The Industrial Organization of Health Care Markets." In *Handbook of Industrial Organization, Vol. 5.* , ed. Kate Ho, Ali Hortaçsu and Alessandro Lizzeri, Chapter 16, 521–614. Amsterdam: Elsevier.

Hetemaa, Tiina, Katja Ilmarinen, Satu Kapiainen, Ilmo Keskimäki, Meri Koivusalo, Merja Korajoki, Ismo Linnosmaa, Eeva Liukko, Pasi Moisio, Susanna Mukkila, Kimmo Parhiala, Mikko Peltola, Eeva Reissell, Liina-Kaisa Tynkkynen, and Maria Vaalavuo. 2018. "Social and Health Care Client Fees: Allocation, Effects and Fairness." *Prime Minister's Office Publication Series 30/2018.*

Karhunen, Elina. 2020. "Study on the Number and Regional Distribution of Hospital Beds in Finland – Background Study for Assessing the Need for Construction Investments." *Reports and Memorandums of the Ministry of Social Affairs and Health 2020:36.*

Kautiainen, Kirsi, Unto Häkkinen, and Jorma Lauharanta. 2011. "Finland: DRGs in a Decentralized Health Care System." In *Diagnosis-Related Groups in Europe – Moving towards transparency, efficiency and quality in hospitals.* , ed. Reinhard Busse, Alexander Geissler and Wilm Quentin, Chapter 18, 321–338. New York, USA: McGraw-Hill Open University Press.

Keskimäki, Ilmo, Liina-Kaisa Tynkkynen, Eeva Reissell, Meri Koivusalo, Vesa Syrjä, Lauri Vuorenkoski, Rechel Bernd, and Marina Karanikolos. 2019. *Finland: Health System Review. Health Systems in Transition, vol. 21 (2)*, World Health Organization. Regional Office for Europe.

Kessler, Daniel P., and Mark B. McClellan. 2000. "Is Hospital Competition Socially Wasteful?" *The Quarterly Journal of Economics*, 115(2): 577–615.

Moscelli, Giuseppe, Hugh Gravelle, and Luigi Siciliani. 2021. "Hospital Competition and Quality for Non-Emergency Patients in the English NHS." *The RAND Journal of Economics*, 52(2): 382–414.

Moscelli, Giuseppe, Hugh Gravelle, Luigi Siciliani, and Rita Santos. 2018. "Heterogeneous Effects of Patient Choice and Hospital Competition on Mortality." *Social Science & Medicine*, 216: 50–58.

Mäkelä, Keijo T, Mikko Peltola, Reijo Sund, Antti Malmivaara, Unto Häkkinen, and Ville Remes. 2011. "Regional and Hospital Variance in Performance of Total Hip and Knee Replacements: A National Population-Based Study." *Annals of Medicine*, 43(Suppl. 1): S31–S38.

National Audit Office. 2021. "NHS Backlogs and Waiting Times in England." `https://www.nao.org.uk/reports/nhs-backlogs-and-waiting-times-in-england`.

Newsweek and Statista. 2021. "World's Best Hospitals 2021. `https://www.newsweek.com/best-hospitals-2021` [accessed 12th September 2023]."

OECD. 2020. *Waiting Times for Health Services: Next in Line. OECD Health Policy Studies*, Paris: OECD Publishing.

Paltta, Päivi. 2008. "Eligibility for Reimbursements of Health-Related Travel Costs in Finland and Selected Other Countries." *Studies in Social Security and Health*, 99.

Pirkanmaa Hospital District. 2007. "Miljoona suomalaista voi valita leikkaushoitopaikkansa." *https://web.archive.org/web/20070826180414/http:/www.pshp.fi/default.aspx?contentid=1604&nodeid=9469&contentlan=1 [archived in August 26th 2007].*

Robinson, James C., Harold S. Luft, Stephen J. McPhee, and Sandra S. Hunt. 1988. "Hospital Competition and Surgical Length of Stay." *JAMA*, 259(5): 696–700.

Roos, Anne-Fleur, Owen O'Donnell, Frederik T. Schut, Eddy Van Doorslaer, Raf Van Gestel, and Marco Varkevisser. 2020. "Does Price Deregulation in a Competitive Hospital Market Damage Quality?" *Journal of Health Economics*, 72: 102328.

Siciliani, Luigi, Valerie Moran, and Michael Borowitz. 2014. "Measuring and Comparing Health Care Waiting Times in OECD Countries." *Health Policy*, 118(3): 292–303.

Silber, Jeffrey H., Paul R. Rosenbaum, Bijan A. Niknam, Richard N. Ross, Joseph G. Reiter, Alexander S. Hill, Lauren L. Hochman, Sydney E. Brown, Alexander F. Arriaga, and Lee A. Fleisher. 2020. "Comparing Outcomes and Costs of Medical Patients Treated at Major Teaching and Non-teaching Hospitals: A National Matched Analysis." *Journal of General Internal Medicine*, 35(3): 743–752.

Syverson, Chad. 2004. "Product Substitutability and Productivity Dispersion." *The Review of Economics and Statistics*, 86(2): 534–550.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2): 326–65.

Syverson, Chad. 2019. "Macroeconomics and Market Power: Context, Implications, and Open Questions." *Journal of Economic Perspectives*, 33(3): 23–43.

Sá, Luís, Luigi Siciliani, and Odd Rune Straume. 2019. "Dynamic Hospital Competition Under Rationing by Waiting Times." *Journal of Health Economics*, 66: 260–282.

THL. 2021. "PERFECT – PERFormance, Effectiveness and Cost of Treatment episodes." *https://thl.fi/en/web/thlfi-en/research-and-development/research-and-projects/perfect-performance-effectiveness-and-cost-of-treatment-episodes [accessed in June 23rd 2022].*

Torkki, Paulus. 2012. "Best Practice Processes - What Are the Reasons for Differences in Productivity Between Surgery Units." PhD diss. Aalto University School of Science.

Urquhart, Donna M., Fahad S. Hanna, Sharon L. Brennan, Anita E. Wluka, Karin Leder, Peter A. Cameron, Stephen E. Graves, and Flavia M. Cicuttini. 2010. "Incidence and Risk Factors for Deep Surgical Site Infection After Primary Total Hip Arthroplasty: A Systematic Review." *The Journal of Arthroplasty*, 25(8): 1216–1222.

Van Reenen, John. 2018. "Increasing Difference Between Firms: Market Power and the Macro Economy." *Changing Market Structures and Implications for Monetary Policy, Kansas City Federal Reserve: Jackson Hole Symposium*, 19–65.

Varkevisser, Marco, Stéphanie A. van der Geest, and Frederik T. Schut. 2012. "Do Patients Choose Hospitals With High Quality Ratings? Empirical Evidence from the Market for Angioplasty in the Netherlands." *Journal of Health Economics*, 31(2): 371–378.

ONLINE APPENDIX

Improving Performance Through Allocation and Competition: Evidence from a Patient Choice Reform

Mika Kortelainen[a] ⓡ Liisa T. Laine[b] ⓡ Konsta Lavaste[c] ⓡ Tanja Saxell[d] ⓡ Luigi Siciliani[e]

[a] University of Turku, Finnish Institute for Health and Welfare, VATT Institute for Economic Research

[b] University of Missouri

[c] Finnish Institute for Health and Welfare, University of Jyväskylä

[d] Aalto University, VATT Institute for Economic Research, Helsinki GSE

[e] University of York

A1. Data

Our main data source is the Finnish hospital discharge data, officially known as the Care Register for Health Care, from the Finnish Institute for Health and Welfare. The original data include the universe of inpatient care discharges and admissions, day surgeries, and specialized outpatient care in Finland since 1994, but we consider years 2004–2010 in our analyses.

A1.1. Treatment Spell Construction

The discharge data do not contain identifiers for treatment spells. We use information on hospital admission and discharge dates to construct treatment spells. We made the following assumptions when assigning observations into treatment spells: observations a and b of patient i belong to the same treatment spell if

- admission date in observation $a \leq$ admission date in observation $b \leq$ discharge date in admission a, or
- admission date in observation $a =$ admission date in observation b.

A1.2. Sample Construction

For our baseline estimations, we construct three estimation samples of surgeries from the discharge data: planned primary hip replacement surgeries, planned primary knee replacement surgeries, and all planned musculoskeletal surgeries. We identified the surgeries based on the main procedure code and the care type code. For our analyses on the quality spillover effect on emergency care, we also constructed three emergency care samples based on the main diagnosis code (ICD-10): AMI, stroke, and hip fracture admissions. The inclusion criteria for different samples are shown in the table below. Only admissions which began a treatment spell were included in the estimation samples, meaning that we excluded observations for which the admission date was not equal to the spell beginning date. We also identified all planned surgeries for the purpose of calculating the predicted HHI that we use in our quasi-DiD approach.

We made the following restrictions to all samples. First, treatment spells which began between January 1st 2004 and December 31st 2010 (our observation period) were included. Second, each admission to hospital h in quarter t was excluded if the sample-wise number of admissions to hospital h in quarter t was less than 5. Third, the reform concerned only public hospitals which were owned by reform area hospital districts, hence we excluded private hospitals ($N = 37$) and small municipal-owned public hospitals ($N = 21$) which performed planned musculoskeletal surgeries. We also excluded military, prison, and psychiatric hospitals which had some emergency AMI, stroke, and hip fracture admissions ($N = 6$). Note that these hospitals were, however, included in the treatment spells when calculating length of stay and identifying emergency readmission within 30 days of departure. Fourth, we excluded patients from the Åland islands and admissions to Åland central hospital (0.3–0.7% of observations depending on the sample). Seventh, we excluded patients without a Finnish identification number or home municipality, as well as patients from foreign countries (0.1–0.6%). Eight, we excluded patients under 18 or over 74 years of age (14–31% of observations in the baseline

Inclusion Criteria of Different Estimation Samples

Sample	Procedure codes[a]	Diagnosis codes (ICD-10)[b]
Planned surgeries:		
Primary hip replacement surgeries	NFB*	
Primary knee replacement surgeries	NGB*	
Musculoskeletal surgeries	N* excluding the ones in which third character is a number[c]	
All surgeries	A*–Q* excluding the ones in which third character is a number[c]	
Emergency admissions:		
AMI admissions		I21.0*, I21.1*, I21.2*, I21.3*, I21.4*, I21.9*, I22.0*, I22.1*, I22.8*, I22.9*
Stroke admissions		I60*, I61*, I62*, I63* I16.4*, I16.6*, G46*, I67.2*, I69.8*, R47.0*
Hip fracture admissions		S72.0, S72.1, S72.2

[a] See classification of surgical procedures (in Finnish) in: `https://www.terveysportti.fi/terveysportti/toimenpideluokitus.koti`.
[b] See ICD-10 codes in: `https://icd.who.int/browse10/2010/en#/`.
[c] Identification of surgical procedures taken from: `https://www.julkari.fi/bitstream/handle/10024/80384/Tr41_11.pdf`.

surgical samples and 48–71% in the emergency care samples). Ninth, we excluded patients who lived within (outside) the reform area, but underwent surgery in control (reform) area hospitals (1–2%). Tenth, we excluded patients who underwent surgery in a small hospital which closed down in the pre-reform period (0.3–0.4%). Eleventh, we excluded patients who traveled more than 400 kilometers (0.1–0.2%) or whose length of stay lasted more than 180 days (<0.1%) as outliers. In the samples of emergency AMI, stroke, and hip fracture patients, we also excluded admissions to hospital h in quarter t when h did not perform at least 5 planned surgeries in quarter t. We also excluded emergency patients who were admitted to hospital h, but transferred immediately to hospital k. It is likely that these were more severe cases that required advanced treatment only available in certain hospitals.

A1.3. Additional Data Sources

After constructing the samples from the discharge data, we combined them with information on patients and hospitals from additional administrative data sources. First, patients' dates of death are from Statistics Finland causes of death registry. Second, locations of patients' residences at the end of each year are also from Statistics Finland. The locations are recorded at December 31st each year and are thus not available for those who died or emigrated from Finland mid-year. We assume that their location at the time of their admission was as in the end of the previous year. For the remaining patients whose residence locations are missing ($\approx 0.5\%$), we use the coordinates of the center of their home municipality. Throughout the paper, we use the 2010 municipality classification (there were 62 municipality mergers in 2005–2010). Third, hospitals' locations are determined by the centers of the municipalities where they are located. We link the discharge data to the provider registry (TOPI) from the Finnish Institute for Health and Welfare to obtain information on hospitals' municipalities. We also link municipalities with the coordinates of their centers. Fourth, we linked patients' and hospitals' municipalities to hospital districts by data from the Association of Finnish Municipalities.

A1.4. Measures of Hospital Choice

	Outcome	Source	Description
(1)	Distance	Discharge data, patient location (grid) data, & hospital location[a]	A straight line distance between patient i's residence and hospital h's location in kilometers. Some neighboring hospitals are recorded under a joint identifier, and in these cases we use the location of the largest hospital under the joint identifier.
(2)	Nearest hospital	Discharge data, patient location (grid) data, & hospital location[a]	An indicator equal to one if the patient was treated in the hospital nearest to their residence location. Created based on (1).
(3)	Different hospital district	Discharge data, patient location (grid) data, & hospital location[a]	An indicator equal to one if patient i's hospital district was not the same as hospital h's hospital district.
(4)	Teaching hospital	Discharge data	An indicator equal to one if the patient was treated in a teaching (university-based) hospital.

[a] See (7) in A1.7.

A1.5. Measures of Hospital Volume and Market Structure

	Outcome	Source	Description
(1)	Hospital volume	Discharge data	Sample-wise number of patient admissions in hospital h and quarter t.
(2)	Actual HHI	Discharge data	Concentration of hospital h's market in quarter t. Ranges from 0 (minimal concentration) to 1 (monopoly). For more information, see A1.9.

A1.6. Hospital Performance: Clinical Quality, Waiting Time, and Length of Stay

	Outcome	Source	Description
Panel A. Main clinical quality measure			
(1)	Emergency readmission within 30 days	Discharge data	An indicator equal to one if the patient had an emergency admission to *any* hospital for *any* reason within 30 days of being discharged from the last hospital in the treatment spell.
Panel B. Additional planned clinical care quality measures			
(2)	Revision surgery within 2 years[a]	Discharge data	An indicator equal to one if the patient had planned admission with main procedure code NFC* or NGC* to *any* hospital within 730 days of the initial surgery.
(3)	Mechanical complication of prosthesis within 2 years[a]	Discharge data	An indicator equal to one if the patient had admission with ICD-10 diagnosis code T84.0 to *any* hospital within 730 days of the initial surgery.
(4)	Infection or inflammation in prosthesis within 2 years[a]	Discharge data	An indicator equal to one if the patient had admission with ICD-10 diagnosis code T84.5 to *any* hospital within 730 days of the initial surgery.
Panel C. Additional emergency care quality measure			
(5)	Death within 30 days	Causes of death registry	An indicator equal to one if the patient died within 30 days after the initial admission began (includes deaths in or out of the hospital).
Panel D. Length of stay			
(6)	Length of stay	Discharge data	Number of days between the initial admission and departure from the last hospital in the treatment spell. Includes days in multiple hospitals in case there were hospital transfers. Admissions to nursing homes and other long-term care facilities are excluded before constructing treatment spells. May include days in hospitals that were not part of the reform (especially the municipal-owned health centre hospitals).
Panel E. Waiting time			
(7)	Waiting time	Discharge data	Number of days between the specialist's referral and surgery.

Notes: See classification of surgical procedures (in Finnish) in: `https://www.terveysportti.fi/terveysportti/toimenpideluokitus.koti`. See ICD-10 codes in: `https://icd.who.int/browse10/2010/en#/`.
[a] The discharge data does not allow us to identify whether the revision surgery, mechanical complication, infection or inflammation concerned the same prosthesis as in the initial surgery. It is possible that we identify that, for example, the patient underwent revision surgery within 2 years, even if the initial surgery concerned the left knee and the infection concerned the right knee. We assume that the magnitude of the bias resulting from these false positives is of minor magnitude.

Waiting time is readily available in the discharge data, but its reporting has been inconsistent across hospitals. In the hip replacement sample, 21% of waiting time values are missing, while the same is true for 22% of observations in the knee replacement sample and for more than 30% of the sample of all musculoskeletal surgeries.[21] We implement comprehensive robustness checks in Section VII, and confirm that the missing

[21]Compared to hip and knee replacements, it is more likely that hospitals do not record waiting times for minor procedures

values do not bias the estimates of the effects on waiting times.

A1.7. Other Variables

	Variable	Source	Description
Panel A. Covariates			
(1)	Patient age	Discharge data	Patient age at the the time of admission.
(2)	Patient sex	Discharge data	An indicator equal to one if the patient is female.
(3)	N of emergency admissions within 1 year	Discharge Data	Number of emergency admissions that the patient had to *any* hospital within 365 days before the admission under consideration. May include multiple admissions from one treatment spell.
(4)	Weekend admission	Discharge data	An indicator equal to one if the patient's admission date was Saturday or Sunday.
Panel B. Other variables			
(5)	Patient residence location	Location (grid) data	Finland was divided into 1x1 kilometer squares. The residence location is the easting and northing coordinates (ETRS-TM95FIN) of the square in which the patient's residence was located.
(6)	Patient's municipality of residence	Location (grid) data	The municipality in which the patient was resident at the end of the year.
(7)	Hospital location	Provider registry (TOPI) & municipality center locations (Google Maps)	Easting and northing coordinates of the municipality in which hospital was located (ETRS-TM95FIN). Constructed based on hospital's municipality in TOPI and municipality center locations.
(8)	Hospital districts of patients and hospitals	(6), provider registry (TOPI), & hospital district data from Association of Finnish Municipalities	The hospital district to which the patient's municipality of residence or the hospital's municipality belonged. We use publicly available information on the hospital district of each municipality (patient's or hospital's) from the Association of Finnish Municipalities.
(9)	Reform area	Discharge data & (8)	An indicator equal to one if the hospital in which the patient was treated belonged to Vaasa, Etelä-Pohjanmaa, Pirkanmaa, or Päijät-Häme hospital districts.

A1.8. Hospital expenditure

The discharge data does not include hospital expenditures, hence we use Hospital benchmarking data from the Finnish Institute for Health and Welfare.[22] The data provide hospitals' operating expenditures by specialty,

included in the broad class of all musculoskeletal surgeries.

[22]See `https://thl.fi/en/web/thlfi-en/statistics-and-data/statistics-by-topic/health-care-services/hospital-benchmarking`

allowing us to measure hospitals' surgery-related expenditures, and a case mix index for controlling patients' severity.

	Outcome	Source	Description
Panel A. Expenditure outcomes			
(1)	Total expenditure	Hospital benchmarking data	Hospital h's annual surgery-related operating expenditures (millions of €*) in year t.
(2)	Expenditure per treatment spell	Hospital benchmarking data	Hospital h's annual surgery-related operating expenditures (€*) divided by hospital's DRG-weighted number surgical treatment spells in year t.
Panel B. Covariates			
(3)	Case mix index	Hospital benchmarking data	Number of DRG-weighted treatment spells in surgical ward of hospital h in year t is divided by absolute number of treatment spells in surgical ward of hospital h in year t and then transformed into an index by setting the whole country to equal 1.

* Deflated using Statistics Finland's price index of public health care services (see `https://stat.fi/en/statistics/jmhi`). Base year = 2000.

A1.9. Actual HHI

We constructed the hospital-level Herfindahl-Hirschman Index (HHI) in two steps. First, we calculated a municipality-level HHI for each municipality m in each quarter t as a sum of squared market shares:

$$HHI_{mt}^{MUN} = \sum_{h=1}^{H} \left(\frac{n_{hmt}}{N_{mt}} \right)^2, \tag{A.1}$$

where n_{hmt} is the number of patients from municipality m who underwent surgery in hospital h in quarter t. N_{mt} is the total number of surgical patients from municipality m.

Second, we calculated the hospital-level HHI as a weighted average of the values of municipality-level HHI, using each municipality's share of the hospital's total patient volume as weights:

$$HHI_{ht}^{HOSP} = \sum_{m=1}^{M} \left(\frac{n_{mht}}{N_{ht}} * HHI_{mt}^{MUN} \right), \tag{A.2}$$

where n_{mht} is the number of patients from municipality m who underwent surgery in hospital h in quarter t. N_{ht} is the total number of patients (from any municipality) who underwent surgery in hospital h. We refer to HHI_{ht}^{HOSP} as the *actual* HHI. We calculated it separately for samples of hip replacement surgeries, knee replacement surgeries, and all musculoskeletal surgeries.

A2. Quasi-DiD Approach: Further Details

A2.1. Predicted HHI

We follow Kessler and McClellan (2000) and construct a predicted version of the HHI and use it as the continuous treatment variable in the quasi-DiD specifications (4). We calculate the predicted HHI in four steps. First, we estimate the predicted patient flows of *all* planned surgery patients. We restrict the sample to patients aged 18–74 and allow them to choose any hospital in Finland. We estimate how each observable patient and hospital characteristic (except hospital quality) affects the probability of patient i choosing hospital h in quarter t. Specifically, we estimate the following conditional logit model separately for each quarter during Q1/2004–Q4/2010:

$$
\begin{aligned}
U_{iht} &= V_{iht} + \varepsilon_{iht} \\
&= \alpha_{0t} + \alpha_{1t}\mathbf{X}_{iht} + \alpha_{2t}\mathrm{km}_{iht} + \alpha_{3t}\mathrm{km}^2_{iht} \\
&+ \alpha_{4t}(\mathbf{X}_{iht} * \mathrm{km}_{iht}) + \alpha_{5t}(\mathbf{X}_{iht} * \mathrm{km}^2_{iht}) + \varepsilon_{iht},
\end{aligned}
\tag{B.1}
$$

where $\mathbf{X}_{iht}$ are the hospital characteristics: an indicator for hospital h being located in the same hospital district as patient i, an indicator for teaching hospital, and an indicator for regional hospital. km_{iht} is the distance between patient i's residence and hospital h in kilometers.[23] Patient i chooses hospital h with the probability of:

$$
P_{iht} = \exp(V_{iht})\Big[\sum_{h' \in S_i} \exp(V_{ih't})\Big]^{-1}, \tag{B.2}
$$

Second, we calculate the predicted patient flows of each hospital as a sum of probabilities over all patients.[24] Third, we calculate the predicted municipality-level HHI as the sum of the squares of the predicted market shares (based on predicted patient flows):

$$
\widehat{HHI}^{MUN}_{mt} = \sum_{h=1}^{H} \left(\frac{\widehat{n}_{hmt}}{\widehat{N}_{mt}}\right)^2 \tag{B.3}
$$

$\widehat{n}_{hmt}$ is the predicted number of patients from municipality m who were treated by hospital h in quarter t. $\widehat{N}_{mt}$ is the predicted number of patients from municipality m (treated by any hospital).

Fourth, the predicted hospital-level HHI is calculated as a weighted average of the values of the predicted municipality-level HHI, using each municipality's share of the hospital's predicted total patient volume as

[23] The specification for predicting hospital choices is similar to Moscelli et al. (2018) and Moscelli, Gravelle and Siciliani (2021). An alternative specification used by Gaynor, Moreno-Serra and Propper (2013) includes differential distances and interactions between hospital and patient characteristics, and it yields flows that results in similar HHI values.

[24] An alternative way is to determine the option with the highest probability and regard it as the predicted choice. Tje predicted flows would be counted as sums of these choices.

weights:

$$\widehat{HHI}_{ht}^{HOSP} = \sum_{m=1}^{M} \left(\frac{\widehat{n}_{mht}}{\widehat{N}_{ht}} * \widehat{HHI}_{mt}^{MUN} \right) \tag{B.4}$$

$\widehat{n}_{mht}$ is the predicted number of patients from municipality m who were treated by hospital h in quarter t. $\widehat{N}_{ht}$ is the total predicted number of patients (from any municipality) treated by hospital h.

Our model predicts approximately 85% of patients' choices correctly in 2004Q1–2010Q4. The share is higher than what Gaynor, Moreno-Serra and Propper (2013) predict in a United Kingdom setting (approximately 75%). The correlation between the actual and predicted HHI is 0.87. The correlation is higher than what Moscelli et al. (2018) predict in the United Kingdom setting (0.65–0.80).

Finally, we fix the predicted HHI to its pre-reform mean, as follows:[25]

$$\widehat{HHI}_{h}^{HOSP} = \frac{\sum_{t=2004Q1}^{2007Q3} \widehat{HHI}_{ht}^{HOSP}}{15} \tag{B.5}$$

We refer to $\widehat{HHI}_{h}^{HOSP}$ used in our analyses as the *predicted* HHI.

A2.2. Control Function Strategy

The continuous treatment variable in the quasi-DiD approach (predicted HHI) is affected by patients' choices, which may correlate with unobserved quality. We follow Moscelli, Gravelle and Siciliani (2021) to control for time-varying patient selection to hospitals using a control function strategy. In the first stage, we estimate the following conditional logit model separately for each quarter in Q1/2004–Q4/2010:

$$\begin{aligned} U_{iht} &= V_{iht} + \xi_{iht} \\ &= \lambda_{1t}\text{km}_{ih} + \lambda_{2t}\text{km}_{ih}^2 + \lambda_{3t}\text{km}_{ih}^3 + \lambda_{4t}\text{nearest}_{ih} \\ &+ \lambda_{5t}\text{sameHD}_{ih} + \lambda_{6t}\text{teaching}_h + \lambda_{7t}\text{regional}_h + \xi_{iht}, \end{aligned} \tag{B.6}$$

where km_{ih} is the distance between the patient and the hospital, nearest_{ih} is an indicator for h being the geographically nearest hospital for patient i, sameHD_{ih} is an indicator for patient i and hospital h being located in the same hospital district, and teaching_h and regional_h are indicators for teaching (i.e., the largest) and regional (i.e., the smallest) hospitals.[26] The chosen covariates reflect geographical access (km_{ih}, nearest_{ih}) and factors potentially relevant to hospital choice post-reform (sameHD_{ih}, teaching_h, regional_h).[27]

[25] Note that the conditional logit model (equation B.1) and its parameters are estimated separately for each period. In practice, we only use estimates from the pre-reform period in constructing the predicted HHI.

[26] Baseline = central hospitals. Unlike Moscelli, Gravelle and Siciliani (2021), we do not restrict patients' choice sets, but instead let them choose among all hospitals ($N = 29$ or 32 depending on the sample).

[27] The covariates differ from the choice model in section A2.1, because the predicted HHI concerns only the restricted *pre-reform* choices, for which these factors were much less relevant, while the control function targets also the unrestricted *post-reform* choices in the reform area.

Patient i chooses hospital h with the probability of

$$P_{iht} = \exp(V_{iht})\Big[\sum_{h' \in S_i} \exp(V_{ih't})\Big]^{-1}, \tag{B.7}$$

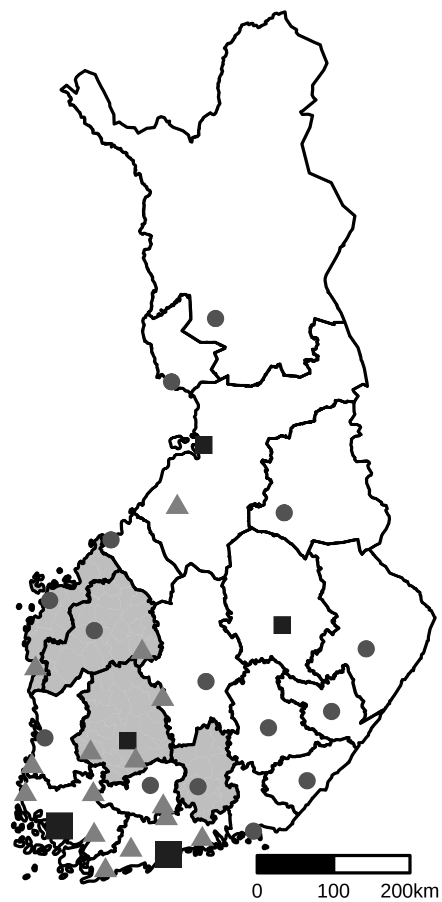
After calculating the choice probabilities, we can derive the set of residuals (one for each hospital h):

$$\widehat{r}'_{it} = [\widehat{r}_{i1t}, \widehat{r}_{i2t}, \ldots, \widehat{r}_{iHt}] = C_{iht} - \widehat{P}_{iht} \tag{B.8}$$

All these residuals ($H = 29$ or 32 depending on the sample) are then added to the second-stage specification (4).

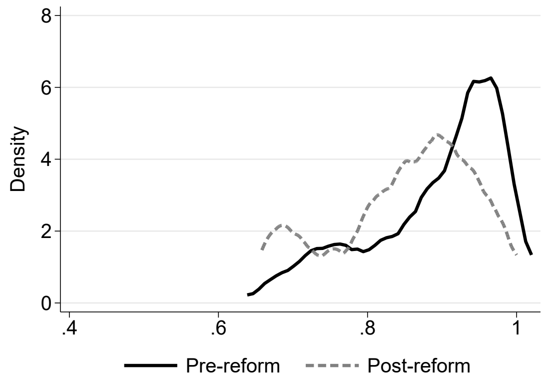
A3. Figures

Figure A1: Teaching, Central, and Regional Hospital Locations

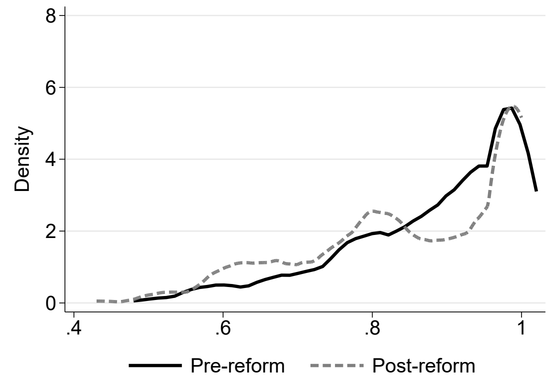


Notes: Borders indicate hospital districts in 2007 and the shaded area constitutes the 2007 reform area. Squares are teaching hospitals, dots central hospitals, and triangles regional hospitals. The large squares mark the capital region (9 hospitals) and the Turku region (4 hospitals), both of which had one teaching hospital and several regional hospitals. In total, there were $N = 9$ hospitals in the reform area and $N = 37$ hospitals in the control area. The figure includes all public hospitals which performed planned surgeries (excluding municipal-owned hospitals), although some of them did not perform hip and/or knee replacement surgeries.

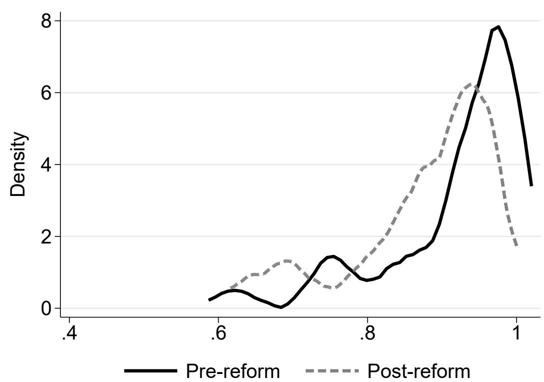
Figure A2: Kernel Density Estimates for the Distribution of the Actual HHI Before and After the Reform



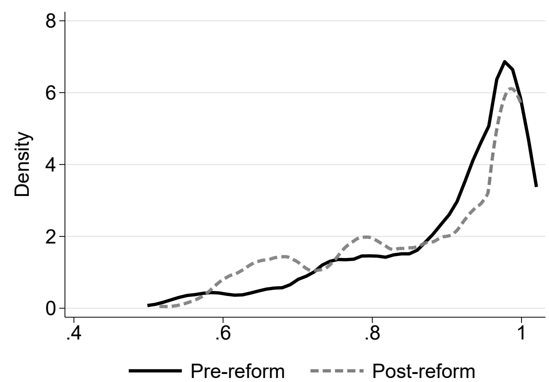
Panel A. Hip replacements in the reform area



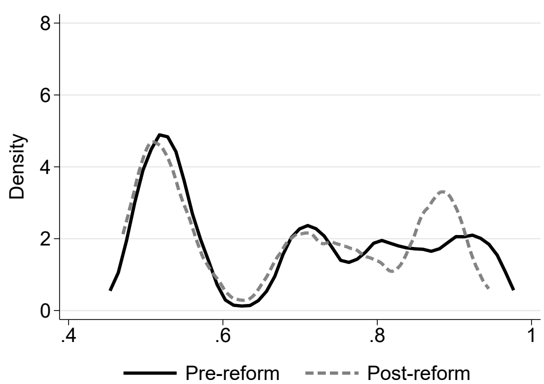
Panel B. Hip replacements in the control area



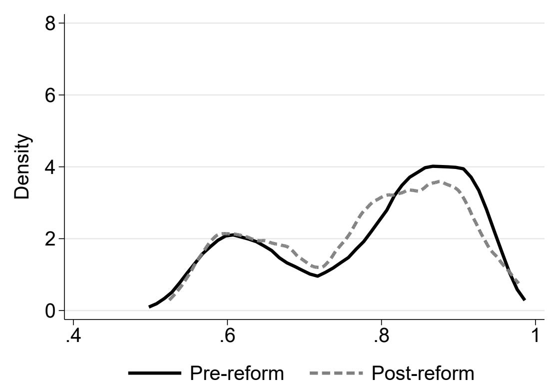
Panel C. Knee replacements in the reform area



Panel D. Knee replacements in the control area



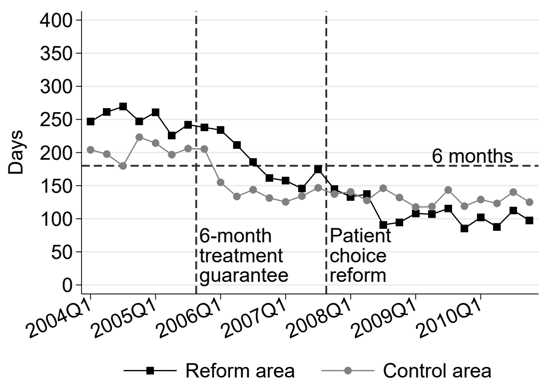
Panel E. Musculoskeletal surgeries in the reform area



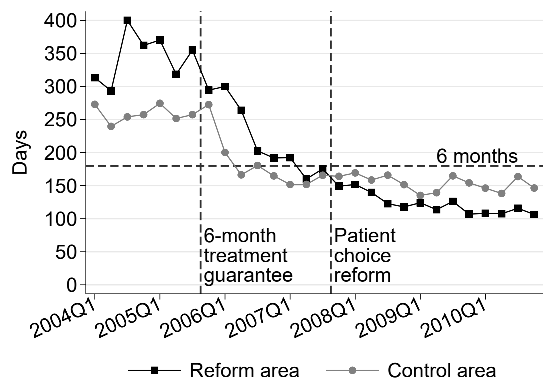
Panel F. Musculoskeletal surgeries in the control area

Notes: The figure plots Epanechnikov kernel density estimates for the distribution of the actual hospital-quarter-level HHI that are calculated based on observed patient flows in reform (Panels A, C, and E) and control areas (Panels B, D, and F). N of hospital IDs = 4–7 (reform area) and 25 (control area). Pre-reform corresponds to Q1/2004–Q3/2007 and post-reform to Q4/2007–Q4/2010.

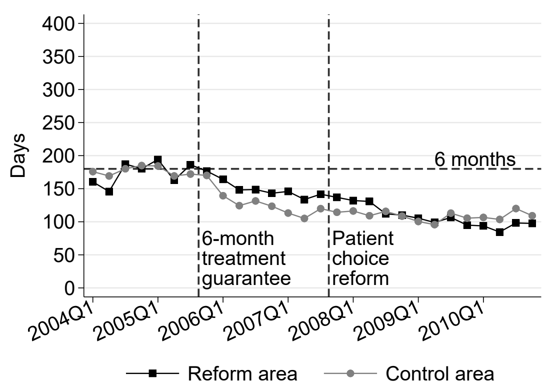
Figure A3: Mean Waiting Time



Panel A. Hip replacements

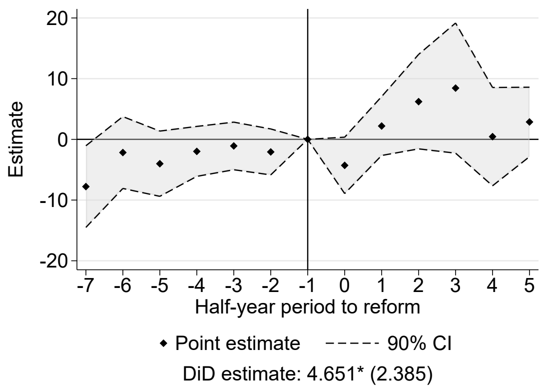


Panel B. Knee replacements

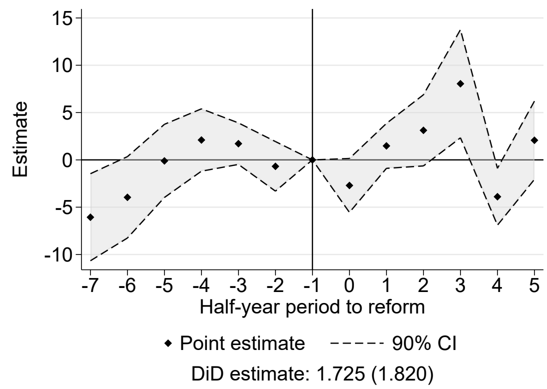


Panel C. All musculoskeletal surgeries

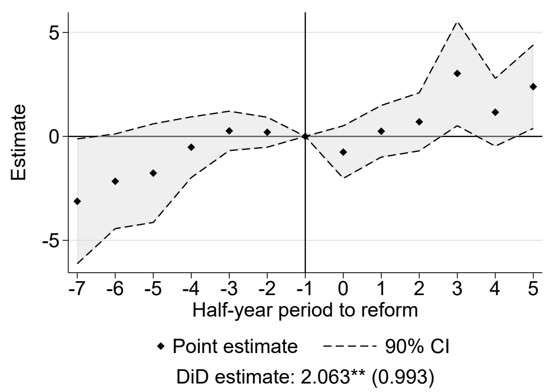
Figure A4: Effect of the Reform on Travel Distance of Surgical Patients



Panel A. Hip replacements



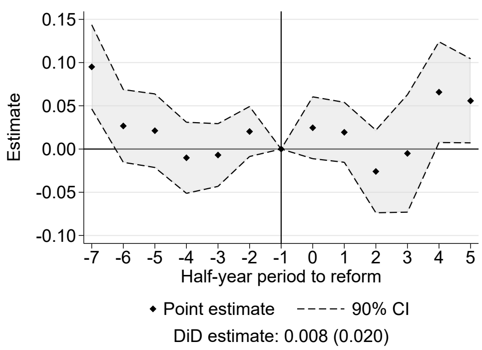
Panel B. Knee replacements



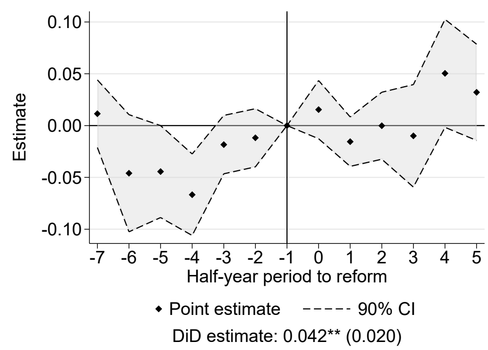
Panel C. All musculoskeletal surgeries

Notes: Includes the DiD estimates corresponding to column 1 in Table 3.

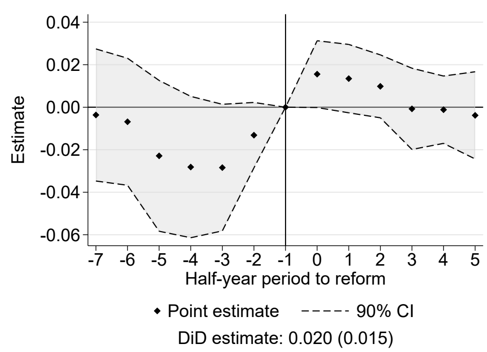
Figure A5: Effect of the Reform on Probability of Surgical Patients Being Treated In Their Nearest Hospital



Panel A. Hip replacements



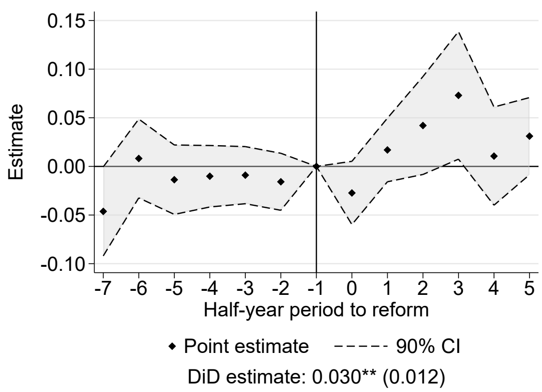
Panel B. Knee replacements



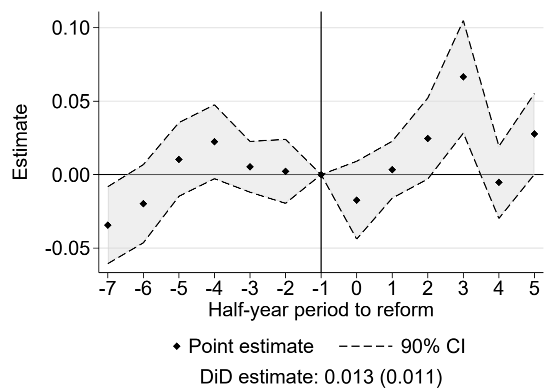
Panel C. All musculoskeletal surgeries

Notes: Includes the DiD estimates corresponding to column 2 in Table 3.

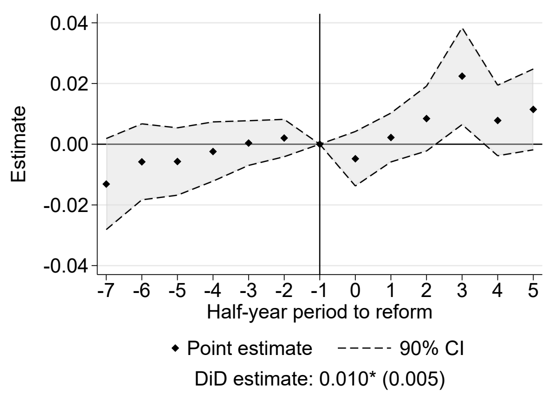
Figure A6: Effect of the Reform on Probability of Surgical Patients Being Treated Outside Their Own Hospital District



Panel A. Hip replacements



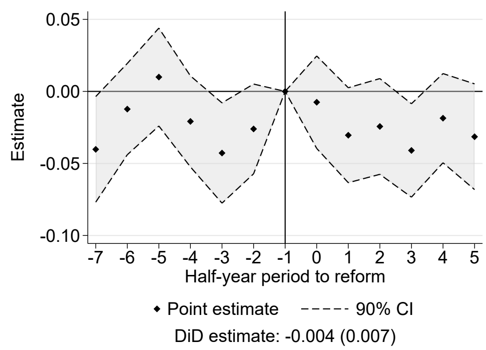
Panel B. Knee replacements



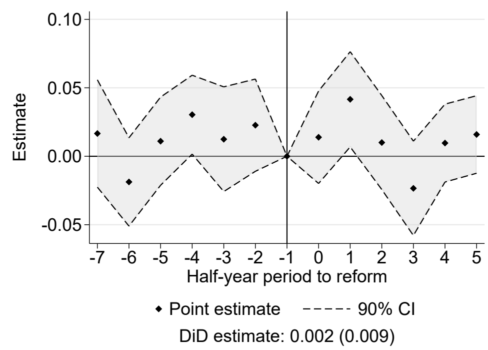
Panel C. All musculoskeletal surgeries

Notes: Includes the DiD estimates corresponding to column 3 in Table 3.

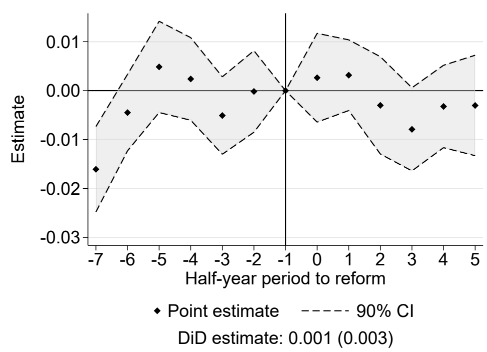
Figure A7: Effect of the Reform on Emergency Readmissions of Surgical Patients Using Binary Treatment



Panel A. Hip replacements



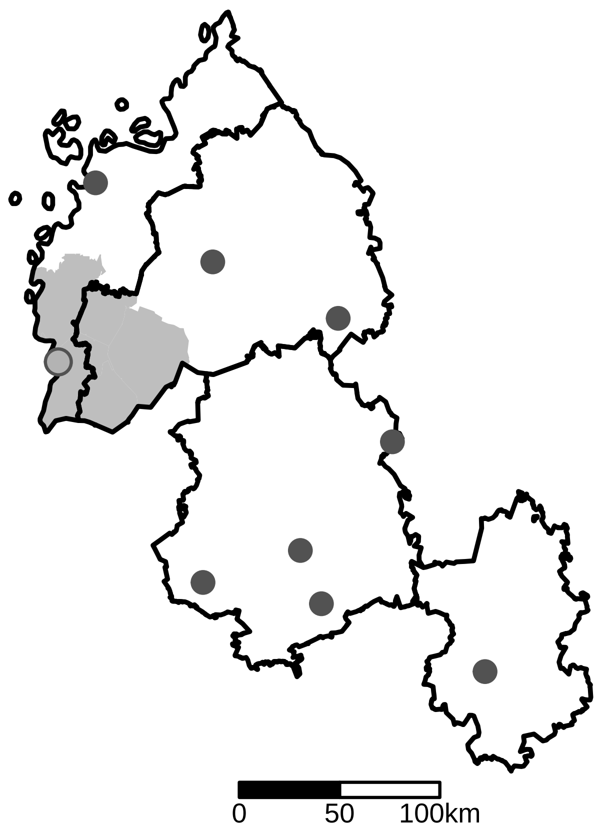
Panel B. Knee replacements



Panel C. All musculoskeletal surgeries

Notes: Includes the DiD estimates corresponding to column 1 in Table 4.

Figure A8: Areas Excluded As a Robustness Test for Being Affected by a Closure of a Reform Area Hospital



Notes: The figure shows the map of the reform area in 2007 with hospital district borders. Dots mark reform area hospitals. The dot with an empty middle marks the hospital which closed down in the pre-reform period. Patients living in the shaded area around the closed hospital were excluded from the robustness test estimation.

A4. Tables

Table A1: Descriptive Statistics: Hospital Volume and Actual HHI by Hospital Type

	Hospital volume[a]			Actual HHI[b]		
	Teaching	Central	Regional	Teaching	Central	Regional
Panel A. Hip replacement surgeries						
Mean	92.29	30.46	16.28	0.89	0.92	0.77
SD	(46.73)	(16.66)	(8.92)	(0.09)	(0.09)	(0.12)
N	140	418	244	140	418	244
Panel B. Knee replacement surgeries						
Mean	104.73	37.73	22.15	0.89	0.94	0.80
SD	(52.88)	(20.89)	(12.86)	(0.11)	(0.08)	(0.12)
N	140	418	242	140	418	242
Panel C. All musculoskeletal surgeries						
Mean	1282.28	415.56	217.25	0.85	0.85	0.64
SD	(743.66)	(184.80)	(103.37)	(0.09)	(0.06)	(0.07)
N	140	418	244	140	418	244

Notes: Values calculated from hospital-quarter level data spanning from Q1/2004 to Q4/2010.
[a] Mean of hospital-quarter level volumes (N of surgeries).
[b] Mean of hospital-quarter level actual HHI (0–1).

Table A2: Descriptive Statistics: Quality, Waiting Time, and Length of Stay, by Hospital Type

	Risk-adjusted readmission rate[a]			Mean waiting time[b]			Mean length of stay[c]		
	Teaching	Central	Regional	Teaching	Central	Regional	Teaching	Central	Regional
Panel A. Hip replacement surgeries									
Mean	0.08	0.08	0.08	153.01	177.30	126.33	6.63	8.26	8.24
SD	(0.01)	(0.01)	(0.01)	(58.85)	(78.19)	(65.27)	(2.03)	(2.87)	(2.86)
N	140	418	244	135	315	218	140	418	244
Panel B. Knee replacement surgeries									
Mean	0.11	0.11	0.11	196.95	210.56	145.12	6.63	7.62	7.30
SD	(0.01)	(0.02)	(0.02)	(89.35)	(103.28)	(68.81)	(1.80)	(2.06)	(1.98)
N	140	418	242	135	312	211	140	418	242
Panel C. All musculoskeletal surgeries									
Mean	0.07	0.06	0.06	138.47	140.21	101.67	2.83	2.15	2.00
SD	(0.01)	(0.01)	(0.01)	(46.98)	(63.90)	(38.90)	(0.97)	(0.60)	(0.67)
N	140	418	244	140	398	237	140	418	244

Notes: Values calculated from hospital-quarter level data spanning from Q1/2004 to Q4/2010.
[a] Hospital-quarter-level mean of patient's probability of emergency readmission to any hospital within 30 days of the surgery. Risk-adjusted by predicting patients' probability of readmission within 30 days with age, sex, number of emergency admissions within one year prior to the surgery, weekend indicator, time fixed effects, and surgery type fixed effect.
[b] Hospital-quarter-level mean of waiting time in days (no risk-adjustment). Some hospitals did not report any waiting times in some quarters, which is depicted as a smaller N.
[c] Hospital-quarter-level mean of length of stay in days (no risk-adjustment).

Table A3: Descriptive Statistics: Hospital Expenditure for All Surgeries by Hospital Type

	No weights			DRG weights[a]		
	Teaching	Central	Regional	Teaching	Central	Regional
Panel A. Total expenditure (millions of €)[b]						
Mean	100.42	24.22	6.77			
SD	(67.96)	(8.64)	(3.09)			
N	35	112	77			
Panel B. Expenditure per treatment spell (€)[c]						
Mean	921.71	619.59	477.22	460.62	425.86	394.43
SD	(111.10)	(87.72)	(76.17)	(38.18)	(44.31)	(39.14)
N	35	112	77	35	112	77

Notes: Values calculated from hospital-year-level data spanning from 2004 to 2010.
[a] Number of treatment spells multiplied by DRG weights, which describe the relative average expenditure for treating patients in a particular DRG category.
[b] Hospital's annual care-related expenditure in the surgical ward (€, deflated using prices in 2000)
[c] Hospital's annual care-related expenditure in the surgical ward (€, deflated using prices in 2000) divided by number of treatment spells in the surgical ward.

Table A4: Robustness Test: All Musculoskeletal Surgeries Except Hip and Knee Replacements (Choice Outcomes)

	Teaching hospital[a]	Distance (km)[b]	Nearest hospital[c]	Different hospital district[d]
	(1)	(2)	(3)	(4)
$\text{Treated}_h \times \text{Post}_t$	-0.002	0.187	0.027*	-0.002
	(0.006)	(0.285)	(0.015)	(0.004)
$\text{mean}(y_{imht} \vert \text{Post}_t = 0)$	0.431	27.820	0.840	0.042
N	352,537	352,537	352,537	352,537
Surgery type / Diagnosis code FEs	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=74–326 depending on the sample).
[a] Equals one if patient was treated in teaching (university) hospital.
[b] Distance from patient's residence to the hospital in kilometers.
[c] Equals one if patient was treated in the geographically nearest hospital.
[d] Equals one if patient was treated in a hospital that was located in another hospital district than the one the patient lived in.

Table A5: Robustness Test: Additional Controls (Choice Outcomes)

	Teaching hospital[a] (1)	Distance (km)[b] (2)	Nearest hospital[c] (3)	Different hospital district[d] (4)
Panel A. Hip replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.064***	4.599*	0.008	0.030**
	(0.017)	(2.376)	(0.020)	(0.012)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.460	29.847	0.864	0.036
N	29,625	29,625	29,625	29,625
Panel B. Knee replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.046***	1.630	0.042**	0.013
	(0.016)	(1.779)	(0.020)	(0.011)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.436	28.953	0.870	0.033
N	35,884	35,884	35,884	35,884
Panel C. All musculoskeletal surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.015*	2.064**	0.020	0.010*
	(0.008)	(0.994)	(0.015)	(0.005)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.434	28.046	0.838	0.041
N	418,090	418,090	418,090	418,090
Surgery type / Diagnosis code FEs	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓
Hospital FEs				
N of emergency admissions	✓	✓	✓	✓
Weekend	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=74–326 depending on the sample).
[a] Equals one if patient was treated in teaching (university) hospital.
[b] Distance from patient's residence to the hospital in kilometers.
[c] Equals one if patient was treated in the geographically nearest hospital.
[d] Equals one if patient was treated in a hospital that was located in another hospital district than the one the patient lived in.

Table A6: Robustness Test: Additional Controls (Hospital Performance Outcomes)

	Readmission[a] (1)	Waiting time[b] (2)	Length of stay[c] (3)
Panel A. Hip replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	-0.005	-67.851***	-0.206
	(0.008)	(12.473)	(0.296)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.080	183.757	8.044
N	29,625	23,481	29,625
Panel B. Knee replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.003	-96.444***	-0.410
	(0.009)	(18.574)	(0.284)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.103	229.671	7.528
N	35,884	28,269	35,884
Panel C. All musculoskeletal surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.001	-18.118**	-0.198***
	(0.002)	(7.875)	(0.047)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.062	150.935	2.340
N	418,090	294,198	418,090
Surgery type / Diagnosis code FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓
Hospital FEs	✓	✓	✓
N of emergency admissions	✓	✓	✓
Weekend	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=69–326 depending on the sample).

[a] Emergency readmission (to any hospital) within 30 days of discharge.

[b] Number of days. Some of the values are missing, which results in smaller N compared to other columns (see online Appendix Section A1.6 for more details).

[c] Number of days.

Table A7: Robustness Test: Excluding Patients Near the Reform Area Hospital That Closed Down in the Pre-Reform Period (Choice Outcomes)

	Teaching hospital[a] (1)	Distance (km)[b] (2)	Nearest hospital[c] (3)	Different hospital district[d] (4)
Panel A. Hip replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.061***	4.545*	-0.002	0.030**
	(0.018)	(2.461)	(0.019)	(0.013)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.464	29.276	0.870	0.035
N	29,293	29,293	29,293	29,293
Panel B. Knee replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.046***	1.841	0.029	0.015
	(0.016)	(1.843)	(0.018)	(0.011)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.439	28.525	0.874	0.032
N	35,544	35,544	35,544	35,544
Panel C. All musculoskeletal surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.014*	2.081**	0.000	0.011**
	(0.008)	(1.002)	(0.008)	(0.005)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.436	27.762	0.842	0.041
N	415,348	415,348	415,348	415,348
Surgery type / Diagnosis code FEs	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=67–319 depending on the sample). All patients from Kristiinankaupunki, Isojoki, Karijoki, Kaskinen, Närpiö, Kauhajoki, and Teuva excluded. See Figure A8 in A3 for a map.

[a] Equals one if patient was treated in teaching (university) hospital.

[b] Distance from patient's residence to the hospital in kilometers.

[c] Equals one if patient was treated in the geographically nearest hospital.

[d] Equals one if patient was treated in a hospital that was located in another hospital district than the one the patient lived in.

Table A8: Robustness Test: Excluding Patients Near the Reform Area Hospital That Closed Down in the Pre-Reform Period (Hospital Performance Outcomes)

	Readmission[a] (1)	Waiting time[b] (2)	Length of stay[c] (3)
Panel A. Hip replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	-0.003	-72.214***	-0.138
	(0.008)	(12.309)	(0.318)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.080	183.154	8.018
N	29,293	23,199	29,293
Panel B. Knee replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.002	-97.832***	-0.379
	(0.009)	(18.688)	(0.297)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.103	229.208	7.510
N	35,544	27,985	35,544
Panel C. All musculoskeletal surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.000	-19.173**	-0.173***
	(0.003)	(7.984)	(0.047)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.062	150.843	2.330
N	415,348	292,413	415,348
Surgery type / Diagnosis code FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations 1 and 4. Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=62–319 depending on the sample). All patients from Kristiinankaupunki, Isojoki, Karijoki, Kaskinen, Närpiö, Kauhajoki, and Teuva excluded. See Figure A8 in A3 for a map.

[a] Emergency readmission (to any hospital) within 30 days of discharge.

[b] Number of days. Some of the values are missing, which results in smaller N compared to other columns (see online Appendix Section A1.6 for more details).

[c] Number of days.

Table A9: Robustness Test: Excluding Three Western Hospital Districts With Possibilities to Obtain Referral Outside Own Hospital District (Choice Outcomes)

	Teaching hospital[a] (1)	Distance (km)[b] (2)	Nearest hospital[c] (3)	Different hospital district[d] (4)
Panel A. Hip replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.057***	2.978**	-0.012	0.034***
	(0.015)	(1.499)	(0.019)	(0.010)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.489	27.495	0.880	0.024
N	27,175	27,175	27,175	27,175
Panel B. Knee replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.052***	2.507*	0.014	0.029***
	(0.015)	(1.486)	(0.017)	(0.010)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.458	27.206	0.880	0.024
N	33,179	33,179	33,179	33,179
Panel C. All musculoskeletal surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.011	1.903*	-0.001	0.015**
	(0.010)	(0.997)	(0.009)	(0.006)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.458	27.163	0.843	0.038
N	391,237	391,237	391,237	391,237
Surgery type / Diagnosis code FEs	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=72–326 depending on the sample). Excludes patients who lived in the three western hospital districts that allowed patient choice to some degree.

[a] Equals one if patient was treated in teaching (university) hospital.

[b] Distance from patient's residence to the hospital in kilometers.

[c] Equals one if patient was treated in the geographically nearest hospital.

[d] Equals one if patient was treated in a hospital that was located in another hospital district than the one the patient lived in.

Table A10: Robustness Test: Excluding Three Western Hospital Districts With Possibilities to Obtain Referral Outside Own Hospital District (Hospital Performance Outcomes)

	Readmission[a]	Waiting time[b]	Length of stay[c]
	(1)	(2)	(3)
Panel A. Hip replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	-0.001	-74.513***	0.347
	(0.009)	(14.778)	(0.299)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.081	180.051	7.881
N	27,175	21,379	27,175
Panel B. Knee replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.010	-110.759***	-0.024
	(0.011)	(20.999)	(0.315)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.103	228.605	7.398
N	33,179	26,041	33,179
Panel C. All musculoskeletal surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.002	-25.012***	-0.108*
	(0.003)	(9.595)	(0.057)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.063	150.245	2.291
N	391,237	276,910	391,237
Surgery type / Diagnosis code FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=38–326 depending on the sample). Excludes patients who lived in the three western hospital districts that allowed patient choice to some degree.

[a] Emergency readmission (to any hospital) within 30 days of discharge.

[b] Number of days. Some of the values are missing, which results in smaller N compared to other columns (see online Appendix Section A1.6 for more details).

[c] Number of days.

Table A11: Robustness Test: Excluding Four Northern Hospital Districts With Possibilities to Obtain Referral Outside Own Hospital District (Choice Outcomes)

	Teaching hospital[a] (1)	Distance (km)[b] (2)	Nearest hospital[c] (3)	Different hospital district[d] (4)
Panel A. Hip replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.066***	4.545*	0.011	0.030**
	(0.018)	(2.438)	(0.021)	(0.012)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.490	28.524	0.859	0.036
N	27,547	27,547	27,547	27,547
Panel B. Knee replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.048***	1.723	0.043**	0.013
	(0.017)	(1.904)	(0.021)	(0.011)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.464	27.883	0.864	0.033
N	33,404	33,404	33,404	33,404
Panel C. All musculoskeletal surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.015*	2.101**	0.021	0.010*
	(0.008)	(1.064)	(0.015)	(0.006)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.462	26.280	0.833	0.037
N	386,224	386,224	386,224	386,224
Surgery type / Diagnosis code FEs	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=74–325 depending on the sample). Excludes patients who lived in the four western hospital districts that allowed patient choice to some degree.

[a] Equals one if patient was treated in teaching (university) hospital.

[b] Distance from patient's residence to the hospital in kilometers.

[c] Equals one if patient was treated in the geographically nearest hospital.

[d] Equals one if patient was treated in a hospital that was located in another hospital district than the one the patient lived in.

Table A12: Robustness Test: Excluding Four Northern Hospital Districts With Possibilities to Obtain Referral Outside Own Hospital District (Hospital Performance Outcomes)

	Readmission[a] (1)	Waiting time[b] (2)	Length of stay[c] (3)
Panel A. Hip replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	-0.008	-69.584***	-0.191
	(0.008)	(12.046)	(0.306)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.077	183.916	7.887
N	27,547	22,301	27,547
Panel B. Knee replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.001	-93.675***	-0.434
	(0.010)	(18.405)	(0.289)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.102	231.280	7.398
N	33,404	26,730	33,404
Panel C. All musculoskeletal surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.000	-15.109*	-0.207***
	(0.003)	(7.799)	(0.048)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.061	152.674	2.321
N	386,224	273,869	386,224
Surgery type / Diagnosis code FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=69–325 depending on the sample). Excludes patients who lived in the three western hospital districts that allowed patient choice to some degree.

[a] Emergency readmission (to any hospital) within 30 days of discharge.

[b] Number of days. Some of the values are missing, which results in smaller N compared to other columns (see online Appendix Section A1.6 for more details).

[c] Number of days.

Table A13: Robustness Test: Excluding the Reform Area Hospital District Which Implemented DRG pricing in 2005 (Choice Outcomes)

	Teaching hospital[a] (1)	Distance (km)[b] (2)	Nearest hospital[c] (3)	Different hospital district[d] (4)
Panel A. Hip replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.053***	3.813	0.023	0.019
	(0.019)	(2.673)	(0.021)	(0.013)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.477	30.336	0.861	0.037
N	28,565	28,565	28,565	28,565
Panel B. Knee replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.037**	0.880	0.058***	0.005
	(0.017)	(1.943)	(0.022)	(0.012)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.449	29.346	0.867	0.034
N	34,642	34,642	34,642	34,642
Panel C. All musculoskeletal surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.013	1.702*	0.032*	0.005
	(0.009)	(1.016)	(0.019)	(0.005)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	0.449	28.595	0.833	0.043
N	402,497	402,497	402,497	402,497
Surgery type / Diagnosis code FEs	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=72–326 depending on the sample). Excludes patients who were treated in the hospital district's hospitals.

[a] Equals one if patient was treated in teaching (university) hospital.

[b] Distance from patient's residence to the hospital in kilometers.

[c] Equals one if patient was treated in the geographically nearest hospital.

[d] Equals one if patient was treated in a hospital that was located in another hospital district than the one the patient lived in.

Table A14: Robustness Test: Excluding the Reform Area Hospital District Which Implemented DRG pricing in 2005 (Hospital Performance Outcomes)

	Readmission[a] (1)	Waiting time[b] (2)	Length of stay[c] (3)
Panel A. Hip replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	-0.007	-85.847***	0.009
	(0.007)	(9.175)	(0.321)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.081	184.920	8.014
N	28,565	22,550	28,565
Panel B. Knee replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	-0.006	-114.849***	-0.181
	(0.010)	(16.103)	(0.284)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.103	230.862	7.496
N	34,642	27,169	34,642
Panel C. All musculoskeletal surgeries			
$\text{Treated}_h \times \text{Post}_t$	-0.000	-17.953*	-0.171***
	(0.003)	(9.487)	(0.056)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.063	149.926	2.352
N	402,497	281,338	402,497
Surgery type / Diagnosis code FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=38–326 depending on the sample). Excludes patients who were treated in the hospital district's hospitals.
[a] Emergency readmission (to any hospital) within 30 days of discharge.
[b] Number of days. Some of the values are missing, which results in smaller N compared to other columns (see online Appendix Section A1.6 for more details).
[c] Number of days.

Table A15: Robustness Test: Excluding Hospitals That Used a Joint Hospital ID (Choice Outcomes)

	Teaching hospital[a] (1)	Distance (km)[b] (2)	Nearest hospital[c] (3)	Different hospital district[d] (4)
Panel A. Hip replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.069***	7.524***	-0.043**	0.056***
	(0.017)	(2.732)	(0.017)	(0.016)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.355	33.089	0.876	0.043
N	19,814	19,814	19,814	19,814
Panel B. Knee replacement surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.048***	4.035***	-0.008	0.033***
	(0.011)	(1.378)	(0.012)	(0.009)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.341	31.085	0.888	0.038
N	24,330	24,330	24,330	24,330
Panel C. All musculoskeletal surgeries				
$\text{Treated}_h \times \text{Post}_t$	0.021**	1.797***	0.002	0.014***
	(0.008)	(0.637)	(0.009)	(0.004)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.274	30.083	0.859	0.047
N	264,905	264,905	264,905	264,905
Surgery type / Diagnosis code FEs	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=73–326 depending on the sample).
[a] Equals one if patient was treated in teaching (university) hospital.
[b] Distance from patient's residence to the hospital in kilometers.
[c] Equals one if patient was treated in the geographically nearest hospital.
[d] Equals one if patient was treated in a hospital that was located in another hospital district than the one the patient lived in.

Table A16: Robustness Test: Excluding Hospitals That Used a Joint Hospital ID (Hospital Performance Outcomes)

	Readmission[a] (1)	Waiting time[b] (2)	Length of stay[c] (3)
Panel A. Hip replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.010	-84.872***	0.130
	(0.007)	(12.937)	(0.345)
$\text{mean}(y_{imht} \| \text{Post}_t = 0)$	0.083	185.398	7.960
N	19,814	14,941	19,814
Panel B. Knee replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.010	-122.254***	-0.312
	(0.010)	(18.574)	(0.355)
$\text{mean}(y_{imht} \| \text{Post}_t = 0)$	0.109	233.832	7.585
N	24,330	18,334	24,330
Panel C. All musculoskeletal surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.001	-32.061***	-0.136**
	(0.003)	(8.841)	(0.058)
$\text{mean}(y_{imht} \| \text{Post}_t = 0)$	0.063	148.667	2.295
N	264,905	183,821	264,905
Surgery type / Diagnosis code FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=51–326 depending on the sample).

[a] Emergency readmission (to any hospital) within 30 days of discharge.

[b] Number of days. Some of the values are missing, which results in smaller N compared to other columns (see online Appendix Section A1.6 for more details).

[c] Number of days.

Table A17: Robustness Test: Tests Regarding Waiting Times

	Baseline	Districts for which <30% missing[a]	Surgeries for which <30% missing[b]	Excluding hospital which did not report waiting times in 2008–2009[c]	2006 vs 2010[d]
	(1)	(2)	(3)	(4)	(5)
Panel A. Hip replacement surgeries					
$\text{Treated}_h \times \text{Post}_t$	-71.524***	-70.601***		-78.705***	-94.832***
	(11.977)	(12.280)		(12.849)	(17.237)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	183.757	188.893		183.231	158.503
N	23,481	19,892		22,475	7,848
Panel B. Knee replacement surgeries					
$\text{Treated}_h \times \text{Post}_t$	-97.614***	-97.301***		-113.376***	-114.880***
	(18.255)	(19.015)		(18.466)	(20.460)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	229.671	234.900		230.647	196.198
N	28,269	23,809		27,045	9,716
Panel c. All musculoskeletal surgeries					
$\text{Treated}_h \times \text{Post}_t$	-18.535**	-23.358***	-27.868***	-23.524***	-42.332***
	(7.844)	(8.754)	(9.573)	(8.414)	(9.461)
$\text{mean}(y_{imht} \mid \text{Post}_t = 0)$	150.935	150.075	158.491	151.307	134.116
N	294,198	245,032	237,948	285,303	92,702
Surgery type / Diagnosis code FEs	✓	✓	✓	✓	✓
Municipal FEs	✓	✓	✓	✓	✓
Age & sex	✓	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using Equations (1) and (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Outcome = waiting time in days.

[a] Including only those hospital districts for which less than 30% of the waiting time values were missing in 2004–2010.

[b] Including only those surgeries for which less than 30% of the waiting time values were missing in 2004–2010.

[c] Excluding one reform area hospital which did not report most of its waiting times in Q1/2008–Q4/2009.

[d] Including only the years 2006 and 2010, when the share of missing waiting time values was generally low across all regions and hospitals. This analysis mitigates the potential bias which may arise when hospitals' shares of missing waiting time values fluctuate over time.

Table A18: Effects of the Reform Additional Quality Measures

	Revision[a] (1)	Infection[b] (2)	Complication[c] (3)
Panel A. Hip replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.007	-0.001	0.004
	(0.005)	(0.004)	(0.006)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.034	0.012	0.056
N	29,625	29,625	29,625
Panel B. Knee replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.007*	0.001	-0.008
	(0.004)	(0.003)	(0.005)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.031	0.019	0.043
N	35,884	35,884	35,884
Surgery type FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using equation (1). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality ($N = 326$).

[a] Revision surgery within 2 years of the initial surgery.

[b] Infection or inflammation in the prosthesis within 2 years of the initial surgery.

[c] Mechanical complication in the prosthesis within 2 years of the initial surgery.

Table A19: Effects of the Reform on Emergency Care Quality and Length of Stay

	Death within 30 days[a] (1)	Readmission[b] (2)	Length of stay[c] (3)
Panel A. AMI			
$\text{Treated}_h \times \text{Post}_t$	0.010	-0.007	-0.200
	(0.008)	(0.010)	(0.591)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.072	0.179	8.724
N	32,107	32,107	32,107
Panel B. Stroke			
$\text{Treated}_h \times \text{Post}_t$	0.004	-0.044***	-0.615
	(0.007)	(0.012)	(0.783)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.082	0.199	16.231
N	48,495	48,495	48,495
Panel C. Hip fracture			
$\text{Treated}_h \times \text{Post}_t$	-0.007	-0.025	-1.697
	(0.007)	(0.019)	(1.480)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.030	0.195	21.805
N	10,747	10,747	10,747
Diagnosis code FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using equation (1). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=296–317 depending on the sample).

[a] Death (before or after discharge) within 30 days of admission.

[b] Emergency readmission (to any hospital) within 30 days of departing from the last hospital of the treatment spell.

[c] Number of days.

Table A20: Effects on Hospitals' Surgical Expenditure

	Total expenditure (millions of €)[a]		Expenditure per treatment spell (€)[b]	
	DiD (1)	Heterogeneity (2)	DiD (3)	Heterogeneity (4)
$\text{Treated}_h \times \text{Post}_t$	-0.521	-1.129	2.213	1.050
	(1.495)	(0.816)	(18.802)	(20.636)
$\text{Treated}_h \times \text{Post}_t \times \text{Teaching}_h$		10.220**		-0.143
		(3.998)		(23.247)
$\text{mean}(y_{ht} \mid \text{Post}_t = 0)$	28.740	28.740	417.671	417.671
N	224	224	224	224
Hospital and time FEs	✓	✓	✓	✓
Case mix index	✓	✓	✓	✓

Notes: Estimated using hospital-year-level data in 2004–2010. t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the hospital level ($N = 32$).

[a] Hospital's annual care-related expenditure in the surgical ward (millions of €, deflated using prices in 2000).

[b] Hospital's annual care-related expenditure in the surgical ward (€, deflated using prices in 2000) divided by DRG-weighted number treatment spells in the surgical ward.

Table A21: Effect of the Reform on Patient Characteristics at the Hospital-Level

	Mean age[a]		Female (%)[b]		Emergency admissions[c]	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Hip replacement surgeries						
Treated$_h$ × Post$_t$	0.834	0.828	-0.009	-0.017	0.119*	0.111*
	(0.726)	(0.945)	(0.017)	(0.016)	(0.059)	(0.063)
Treated$_h$ × Post$_t$ × Teaching$_h$		-0.181		0.021		-0.016
		(1.081)		(0.023)		(0.070)
mean(y_{imht}|Post$_t = 0$)	63.134	63.134	0.522	0.522	0.480	0.480
N	802	802	802	802	802	802
Panel B. Knee replacement surgeries						
Treated$_h$ × Post$_t$	-0.084	-0.036	-0.034*	-0.044**	0.027	0.014
	(0.224)	(0.269)	(0.018)	(0.020)	(0.060)	(0.068)
Treated$_h$ × Post$_t$ × Teaching$_h$		-0.188		0.045**		0.035
		(0.382)		(0.021)		(0.070)
mean(y_{imht}|Post$_t = 0$)	64.494	64.494	0.659	0.659	0.465	0.465
N	810	810	810	810	810	810
Panel C. All musculoskeletal surgeries						
Treated$_h$ × Post$_t$	-0.438	-0.511	0.013*	0.010	0.038	0.030
	(0.280)	(0.324)	(0.006)	(0.007)	(0.052)	(0.057)
Treated$_h$ × Post$_t$ × Teaching$_h$		0.439		0.017**		0.066
		(0.339)		(0.008)		(0.061)
mean(y_{imht}|Post$_t = 0$)	51.907	51.907	0.520	0.520	0.595	0.595
N	896	896	896	896	896	896
Surgery/diagnosis FEs	✓	✓	✓	✓	✓	✓
Hospital FEs	✓	✓	✓	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality ($N = 326$).

[a] Mean age at the time of admission (18–74).

[b] Share of females out of all patients.

[c] Mean number of emergency admissions hospital's patients had within 1 year before their surgery.

Table A22: Robustness Test: Effect of the Reform when Controlling First-Stage Residuals (Hospital Performance Outcomes)

	Readmission[a] (1)	Waiting time[b] (2)	Length of stay[c] (3)
Panel A. Hip replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	-0.003	-100.662***	-1.692***
	(0.005)	(11.834)	(0.264)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.080	183.757	8.044
N	29,625	23,481	29,625
Panel B. Knee replacement surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.016**	-140.146***	-1.584***
	(0.007)	(15.667)	(0.275)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.103	229.671	7.528
N	35,884	28,269	35,884
Panel C. All musculoskeletal surgeries			
$\text{Treated}_h \times \text{Post}_t$	0.000	-56.324***	-0.596***
	(0.002)	(6.769)	(0.041)
$\text{mean}(y_{imht}\|\text{Post}_t = 0)$	0.062	150.935	2.340
N	418,090	294,198	418,090
Surgery type FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓
Control function residuals	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using equation (1). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality ($N = 326$).

[a] Emergency readmission (to any hospital) within 30 days of departing from the last hospital in the treatment spell.

[b] Waiting time in days. Some of the values are missing, which results in lower N compared to other columns (see online Appendix Section A1.6 for more details).

[c] Number of days.

Table A23: Marginal Effects of Concentration Without Controlling First-Stage Residuals

	Readmission[a] (1)	Waiting time[b] (2)	Length of stay[c] (3)
Panel A. Hip replacement surgeries			
PredictedHHI$_h$ × Post$_t$	0.111	291.854	-19.215***
	[-0.070, 0.267]	[-276.330, 783.864]	[-25.489, -12.526]
mean(y_{imht}|Post$_t = 0$)	0.071	212.851	7.756
N	6,974	6,393	6,974
Panel B. Knee replacement surgeries			
PredictedHHI$_h$ × Post$_t$	0.085	598.836	-21.464***
	[-0.227, 0.410]	[-35.440, 1208.133]	[-26.506, -15.690]
mean(y_{imht}|Post$_t = 0$)	0.103	265.332	7.485
N	8,276	7,541	8,276
Panel C. All musculoskeletal surgeries			
PredictedHHI$_h$ × Post$_t$	-0.016	-147.248*	-2.481***
	[-0.060, 0.032]	[-260.873, -28.611]	[-3.244, -1.723]
mean(y_{imht}|Post$_t = 0$)	0.056	160.512	2.763
N	72,483	56,307	72,483
Surgery type FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓
Control function residuals			

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using equation (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=74). HHI measured on a 0–1 scale, where higher value indicates more concentration.
[a] Emergency readmission (to any hospital) within 30 days of discharge.
[b] Number of days. Some of the values are missing, which results in smaller N compared to other columns (see online Appendix Section A1.6 for more details).
[c] Number of days.

Table A24: Marginal Effect of Concentration on Emergency Care Quality and Length of Stay

	Death within 30 days[a] (1)	Readmission[b] (2)	Length of stay[c] (3)
Panel A. AMI			
PredictedHHI$_h$ × Post$_t$	0.197**	0.256	7.915
	[0.031, 0.387]	[-0.069, 0.576]	[-6.689, 24.473]
mean(y_{imht}|Post$_t = 0$)	0.074	0.154	9.499
N	5,347	5,347	5,347
Panel B. Stroke			
PredictedHHI$_h$ × Post$_t$	-0.137	1.024***	-25.749
	[-0.462, 0.143]	[0.795, 1.319]	[-67.701, 5.545]
mean(y_{imht}|Post$_t = 0$)	0.070	0.201	15.625
N	9,312	9,312	9,312
Panel C. Hip fracture			
PredictedHHI$_h$ × Post$_t$	0.024	0.316	-28.012
	[-0.218, 0.213]	[-0.566, 1.026]	[-64.740, 9.183]
mean(y_{imht}|Post$_t = 0$)	0.034	0.211	21.166
N	1,833	1,833	1,833
Diagnosis code FEs	✓	✓	✓
Municipal FEs	✓	✓	✓
Age & sex	✓	✓	✓

Notes: t-test level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Estimated using equation (4). Includes 18–74-year-old patients in Q1/2004–Q4/2010. Standard errors clustered at the level of patient's home municipality (N=69–74 depending on the sample).
[a] Death (before or after discharge) within 30 days of admission.
[b] Emergency readmission (to any hospital) within 30 days of discharge.
[c] Number of days.