Discussion Paper Series

IZA DP No. 19022

October 2026

Hospital Bed Reductions, Congestion and Mortality

Vincenzo Carrieri
University of Calabria,
RWI and IZA@LISER

Antonella Rita Ferrara
University of Calabria

Raffaele Lagravinese
University of Bari Aldo Moro

The IZA Discussion Paper Series (ISSN: 2365-9793) ("Series") is the primary platform for disseminating research produced within the framework of the IZA@LISER Network, an unincorporated international network of labour economists coordinated by the Luxembourg Institute of Socio-Economic Research (LISER). The Series is operated by LISER, a Luxembourg public establishment (établissement public) registered with the Luxembourg Business Registers under number J57, with its registered office at 11, Porte des Sciences, 4366 Esch-sur-Alzette, Grand Duchy of Luxembourg.

Any opinions expressed in this Series are solely those of the author(s). LISER accepts no responsibility or liability for the content of the contributions published herein. LISER adheres to the European Code of Conduct for Research Integrity. Contributions published in this Series present preliminary work intended to foster academic debate. They may be revised, are not definitive, and should be cited accordingly. Copyright remains with the author(s) unless otherwise indicated.

IZA DP No. 19022 October 2026

Hospital Bed Reductions, Congestion and Mortality*

Abstract

Concentrating public services in fewer facilities has become an increasingly common response to cost-containment pressures. We study the consequences of such consolidation in hospital care, exploiting a national reform in Italy that imposed a binding cap on hospital beds and generated reductions in ward-level capacity. Using unique administrative data covering a large panel of Italian hospitals and a difference-in-differences design, we find that surviving same-specialty departments increase activity by 13–32 percent following downsizing elsewhere in their local networks. Total pre-operative patient days increase by 30 percent, indicating greater pressure on remaining capacity. Finally, we find an increase in mortality outcomes for time-sensitive conditions (myocardial infarction and stroke) and no persistent effect for conditions in which short delays matter less such as hip fracture and cancer. The adjustment is concentrated in public hospitals, with no statistically significant response among private providers. Our findings highlight a trade-off in public service consolidation: concentrating activity increases utilization of remaining capacity but can worsen outcomes when timely service provision is critical.

JEL classification

H51, H75, I11, I18

Keywords

hospital bed reductions, public service retrenchment, spatial access, mortality, difference-in-differences

Corresponding author

Raffaele Lagravinese
raffaele.lagravinese@uniba.it

* Earlier versions of this paper were presented at the 37th Annual Conference of the Italian Society of Public Economics (SIEP, Naples, 2025), Dondena Workshop, Bocconi University (Milan 2025), the 3rd Rome Health Economics Workshop (RomHEW, Rome, 2026), the 47th Annual Conference of the Italian Regional Science Association (AISRe, Florence, 2026), the Annual Workshop of the Italian Society of Industrial Economics and Policy (SIEPI, Bari 2026), and the Conference of the European Health Economics Association (EuHEA, Rotterdam, 2026). We thank participants for helpful comments and suggestions. All remaining errors are our own.

1 Introduction

Governments across developed economies have retrenched spatially distributed public services over the past two decades, consolidating schools, courts, local government offices, and hospitals into fewer, larger units. School consolidation may reduce per-pupil costs but increases commuting distances and can harm rural and disadvantaged students (Hoxby, 2000; Duncombe and Yinger, 2007; Humlum and Smith, 2015). Court consolidation can speed up case resolution but raises access costs and reduces the demand for justice (Canzian and Ferrara, 2025; Mocetti et al., 2026). In each case, system-level efficiency gains come with localized access costs (Tiebout, 1956; Oates, 1972; Gaigné et al., 2016), and the net welfare effect depends on which of the two dominates. Answering that question requires studying both sides together, yet most evidence on retrenchment identifies one side of the trade-off at a time.

Hospitals offer a setting where this trade-off is unusually consequential and unusually well measured. Hospital bed capacity fell by more than 10 percent across OECD countries in the last twenty years (OECD/European Commission, 2024), renewing the debate over whether the resulting efficiency gains come at the expense of quality and access (Richards and Whaley, 2026). Most of the evidence on hospital capacity reductions studies mergers (Gaynor et al., 2026), which bundle a capacity change with managerial integration, common information systems, and often a change in market power. We study a different and underexplored channel. Italy's national bed cap forced hospitals to cut ward-level capacity directly, without any change in ownership, reallocating the caseload of downsized wards toward surviving same-specialty departments elsewhere in the network. Our setting isolates the capacity channel on its own.

Italy entered this process with among the fewest hospital beds per capita in Europe. The bed cap introduced in 2015 reduced public hospital beds by roughly 20 percent, to 3.1 per 1,000 inhabitants, in under a decade. The national decree set a target without prescribing which wards would close, when, or how, and implementation unfolded ward by ward through regional and local decisions rather than through a planned sequence (Section 2.1). We exploit the staggered timing and location of these downsizing episodes using ward-level administrative data for a large panel of Italian hospitals over 2015 to 2022. Because the data record every ward of each hospital in the panel, we can identify, for each downsized ward, the same-specialty departments of hospitals operating within the same local health authority, and track how their activity, patient days, and mortality evolve after a nearby downsizing episode.

We first show that bed reductions are concentrated among peripheral, low-hierarchy facilities, consistent with the hub-and-spoke logic of the reform, and that pre-treatment utilization and network characteristics do not predict downsizing. Its incidence differs across specialties, but these differences are absorbed by specialty-by-LHA and specialty-by-year fixed effects. This pattern is consistent with a supply-side shock to local healthcare markets. We then estimate its effects on surviving departments in a difference-in-differences setting. We also implement the Extended Two-Way Fixed Effects (ETWFE) estimator of Wooldridge (2025) to examine treatment dynamics while allowing for heterogeneous effects across cohorts and over time under staggered adoption. These event studies are estimated at the LHA–specialty level, capturing the adjustment of hospital networks to capacity reductions. Randomization inference and leave-one-hospital-out estimates provide

additional robustness checks.

Our results reveal a clear but asymmetric efficiency-accessibility trade-off. Downsizing reallocates patients toward surviving same-specialty departments, raising total discharges by around 20 percent and surgical procedures by around 31 percent. Part of this increase is the physiological counterpart of the reform's intent to displace patients toward other facilities; part also appears to reflect the downsizing of specialist care that manages chronic conditions, whose patients increasingly end up in hospital. Bed reductions, however, achieve their stated efficiency objective. But these gains come at a cost. Total pre-operative days in surviving departments increase by 26 percent, driven by the larger caseload rather than by longer stays per patient, and mortality outcomes for myocardial infarction (AMI hereafter) and for stroke, time-sensitive conditions, rises after downsizing, while cancer and hip-fracture mortality show no persistent increase. This pattern is consistent with delays in access to acute care within the local network, rather than with a general decline in quality. The burden of adjustment is highly unequal. Public hospitals absorb the displaced demand and the associated congestion, while accredited private providers show no significant response on any margin.

Our paper makes two contributions. First, we contribute to the literature on the spatial organization of public services and on access to healthcare. Distance to hospital care worsens outcomes for time-sensitive conditions (Buchmueller et al., 2006; Avdic, 2016; Ghislandi et al., 2025; Gravelle et al., 2024; Durrance et al., 2024), consolidation may improve outcomes when volume-outcome relationships are strong (Avdic et al., 2019; Fischer et al., 2024), and a separate strand studies the cost consequences of bed downsizing (Gaynor and Anderson, 1995; Piacenza et al., 2010; Siverskog and Henriksson, 2022), generally without linking capacity reductions to outcomes in the departments that absorb the displaced caseload. These strands are rarely connected, and most of their evidence comes from the United States and Sweden. Within Italy, Ghislandi et al. (2025) show that emergency department closures raised in-hospital AMI mortality through longer travel times, and Balia et al. (2026) show that merging Local Health Authorities into larger units raised municipal mortality. We study a ward-level capacity shock rather than a facility closure or an administrative merger, and we study the full chain from the supply shock to the reallocation of patients, the resulting congestion, and cause-specific mortality. The mortality response is concentrated in conditions for which delay in treatment is a leading determinant of survival, as shown by evidence on ambulance response times (Lucchese, 2024), door-to-needle times for stroke (Man et al., 2020), and door-to-balloon times for myocardial infarction (McNamara et al., 2006). Our results suggest that bed reductions generate reallocation and congestion externalities that may extend to cause-specific mortality, a welfare cost that aggregate efficiency indicators are likely to miss.

Second, we contribute to the literature on the public-private mix in healthcare provision. Mixed systems are often assumed to buffer supply shocks, with private capacity absorbing demand that public providers cannot meet, an assumption that echoes the argument that private provision is efficient only insofar as noncontractible quality dimensions do not suffer (Hart et al., 1997). We find no evidence that private capacity buffers a public capacity shock under prospective payment, as the absorption of displaced demand is a dimension of provision that is difficult to contract upon. Under DRG-based reimbursement, providers have an incentive to select low-cost cases and

avoid complex ones (Ellis and McGuire, 1986, 1990; Ellis, 1998), a mechanism that Berta et al. (2026) formalize and test in the Italian setting. Duggan (2000) documents an early version of this behavior in California, and Duggan et al. (2023) show that hospital privatization reduces the intake of unprofitable Medicaid patients without this demand being picked up elsewhere in the market, with an analogous pattern in school choice (Altonji et al., 2015). We add a setting in which private providers do not take up demand displaced by a public capacity shock, leaving the public sector to absorb the caseload displaced by downsizing. This contrasts with Coube et al. (2026), who find that outsourcing the management, but not the ownership, of Brazilian public hospitals to private non-profit operators under performance-linked contracts raised both efficiency and quality. Read together, the two results suggest that whether private involvement substitutes for or merely skims public capacity depends on the contractual design governing it rather than on privateness per se, with direct implications for bed-reduction reforms in mixed public-private systems.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting, the data, and the variables. Section 3 presents the empirical strategy. Section 4 reports the results on reallocation, congestion, and mortality, together with heterogeneity by provider type and geography. Section 5 presents robustness checks, and Section 6 concludes.

2 Institutional Setting, Data, and Variables

2.1 Hospital Network Reorganisation in Italy

The Italian National Health Service (Servizio Sanitario Nazionale, SSN) is a universal, tax-financed system that guarantees uniform coverage nationwide while devolving the organization and planning of healthcare to the regional level. Care is organized and purchased by Local Health Authorities (Aziende Sanitarie Locali, LHAs), which contract with both public hospitals and accredited private hospitals (*ospedali privati convenzionati*) operating within their catchment areas. National authorities set overall hospital capacity standards, but regions retain substantial discretion over how those standards are implemented, including which wards reduce capacity, when these reductions occur, and how hospital networks are reorganized in response.

Ministerial Decree 70/2015 (DM 70/2015) introduced a binding national ceiling of 3.7 hospital beds per 1,000 inhabitants and established national standards for the reorganization of hospital networks. The reform was intended to rationalize hospital capacity and concentrate hospital care in a smaller and more specialized network. It set the overall constraint at the national level, but left regions substantial discretion over how to reorganize capacity across hospitals and specialties.

This decentralized implementation is a key feature of the reform. Regions hold formal constitutional authority over healthcare provision, and the national standards therefore had to be translated into regional hospital plans and regulations. Rather than producing a simultaneous nationwide adjustment, the reform unfolded gradually across regions, hospitals, and specialties. As Fregoni and Leonardi (2025) document, more than 600 hospital branches significantly reduced their bed capacity between 2010 and 2019. Downsizing occurred across all Italian macro-areas but was disproportionately concentrated in peripheral facilities.

The reform does not provide an administrative list of wards that were downsized, nor does it

identify the year in which individual capacity reductions occurred. We therefore reconstruct these episodes directly from administrative records on ward-level bed capacity. Because hospital capacity also exhibits small year-to-year adjustments that need not reflect meaningful retrenchment, we focus on large changes and classify a ward as downsized when its bed capacity falls by at least 50 percent. The cutoff identifies episodes in which at least half of a ward's capacity is removed. This definition differs from the one adopted in the closest Italian evidence in nature, not only in magnitude. Ghislandi et al. (2025), building on Gujral and Basu (2019), infer closures from patient flows and treat a hospital as closed for acute myocardial infarction when its emergency admissions fall by at least 90 percent and remain low thereafter, and Fregoni and Leonardi (2025) consider hospital branches whose bed supply falls to zero. Both definitions identify exit. In Ghislandi et al. (2025) the 90 percent cutoff applies to a flow and allows for residual admissions recorded after a hospital has stopped treating these patients. Our cutoff applies to a stock of installed beds, the unit in which DM 70/2015 sets its binding ceiling, and identifies a contraction in capacity that need not involve exit. Measuring treatment in beds rather than admissions also matters for identification. Admissions fall when patients bypass a ward for reasons of demand or perceived quality, so an admissions-based definition would classify as treated precisely those episodes in which patients had already moved to neighboring providers. Section 2.3 describes the construction of the measure.

Figure 1 maps the resulting downsizing episodes over the 2015–2022 period. They are widely distributed across the country and occur in all Italian macro-areas, including regions without a history of exceptional fiscal pressure. This geographic dispersion is consistent with the national scope of the capacity constraint rather than with retrenchment being confined to regions facing particular fiscal distress.

The timing and location of these reductions provide further evidence on the nature of the consolidation process. The reform did not establish a common implementation schedule beyond the requirement that regions comply with the national bed cap. Figure 2 examines whether subsequent downsizing is systematically related to observable pre-treatment characteristics. It reports coefficients from a logit model of ward downsizing on pre-period hospital characteristics, clinical activity, and geographical and network position. Pre-treatment utilization, hospital ownership, macro-area and network characteristics do not significantly predict subsequent downsizing. In particular, wards with higher pre-treatment utilization are no more likely to experience a capacity reduction. The incidence of downsizing differs a bit across specialties, a difference however absorbed by the specialty-by-LHA and specialty-by-year fixed effects in our specification. Thus, the observed timing and location of downsizing do not appear to reflect pre-existing differences in local demand or hospital activity. Combined with the institutional organization of the reform, this pattern is consistent with capacity reductions being driven primarily by supply-side decisions over how to meet the national capacity constraint.

The Italian reimbursement system provides a further institutional feature relevant to how the hospital network adjusts to capacity reductions. Both public and accredited private hospitals providing care within the SSN are reimbursed through a prospective Diagnosis-Related Group (DRG) system, and patients face no direct charges. The two types of providers, however, face different constraints and incentives. Public hospitals operate under budget constraints set by the LHA,

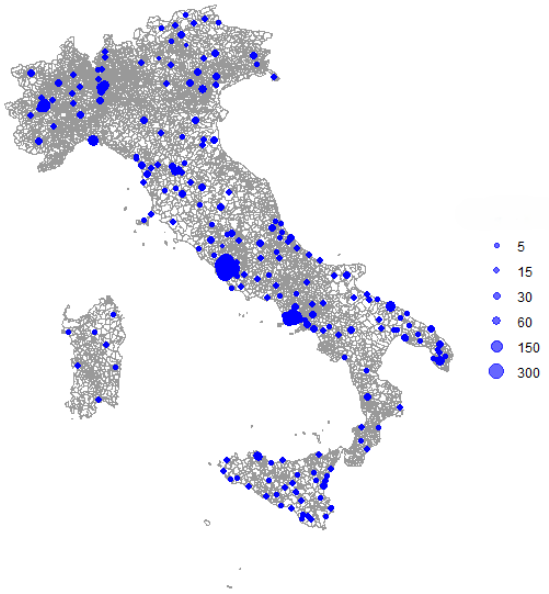


Figure 1: Downsized wards by municipality, 2015–2022.

Note: The map shows the number of downsized wards per muncipality. Larger dots signal a higher number of treated wards in the respective municipality.

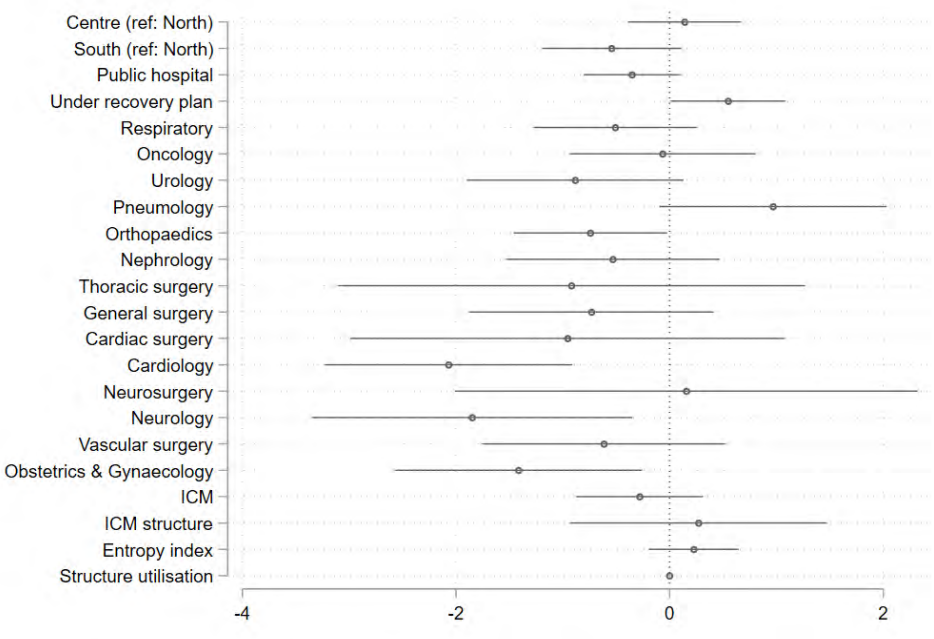


Figure 2: Determinants of ward downsizing: logit coefficients

Notes: Log-odds from a logit model of ward downsizing on pre-period hospital characteristics, clinical activity, geographical and network position. Estimation sample restricted to ward-years prior to treatment. Horizontal bars denote 95% confidence intervals. *Source:* Authors' calculations based on Ministero della Salute administrative data.

cannot select patients on the basis of expected profitability, and are required to treat all patients who present to them, including emergency admissions. Accredited private hospitals operate under volume-based contracts and retain some discretion over which cases to accept. Because DRG payments poorly distinguish short-stay, low-complexity cases from more resource-intensive ones, private providers face an incentive to concentrate activity in the former.

These institutional differences imply that public and private providers need not respond symmetrically when capacity is reduced elsewhere in the local hospital network. Public hospitals may have to accommodate displaced demand across a broad range of patients, whereas accredited private providers have greater scope to respond selectively according to case complexity. This asymmetry provides a natural basis for examining how the public–private composition of hospital provision shapes the adjustment to capacity reductions.

2.2 Data Sources

We assemble an original dataset covering 255 Italian hospitals over the period 2015–2022. The panel includes the hospitals for which ward-level activity data can be linked to the other sources described below. The estimation samples range from 17,750 to 20,394 ward-year observations, depending on the outcome. The dataset combines information gathered from multiple administrative sources: the National Health Service administrative database, available for different years on the website of the Italian Ministry of Health; the National Agency for Regional Health Services (AGENAS) of the Italian Ministry of Health. These data are also complemented by municipal-level demographic and socioeconomic data from the Atlante dei Comuni, provided by the Italian National Institute of Statistics (ISTAT).

All sources are matched using hospital identification codes and further refined through probabilistic record linkage, to address inconsistencies in naming conventions and reporting units across datasets. This procedure allows us to harmonize information coming from different administrative and statistical sources. The final dataset is organized at the ward-by-hospital-year level and is integrated with municipality-level demographic characteristics of each ward's catchment area. This structure lets us analyze hospital activity, congestion, and clinical outcomes together with the institutional determinants of downsizing.

2.3 Treatment Definition

As discussed in Section 2.1, we reconstruct downsizing episodes directly from administrative records on ward-level bed capacity. We track each ward over time and identify large reductions in the number of beds relative to the previous year.

Our objective is to capture substantial retrenchment rather than the smaller year-to-year adjustments that routinely occur across the hospital network. We define a ward as downsized when its bed capacity falls by at least 50 percent relative to the previous year. This threshold identifies episodes in which at least half of a ward's capacity is removed and is intended to separate economically meaningful contractions from marginal fluctuations in reported bed capacity. The treatment year is defined as the year preceding the observed fall in bed capacity.

For each downsized ward, we identify the other wards in the sample operating in the same clinical specialty and Local Health Authority (LHA). These surviving same-specialty wards constitute the local hospital network potentially exposed to the capacity reduction. Wards in the same specialty and LHA that are not yet exposed to a qualifying capacity reduction provide the comparison group.

2.4 Outcome Variables

Our outcomes capture three margins through which surviving departments may adjust to a reduction in nearby same-specialty capacity, namely activity and capacity use, congestion, and quality of care. Because our object of interest is the response of the hospital network that remains after downsizing, each ward's network outcome excludes the ward itself. For the downsized ward, this removes its own capacity reduction from the outcome. For the other wards of the network, the downsized ward remains in the average, since in most episodes it continues to operate with reduced capacity and remains part of the local network to which patients are directed.

We therefore construct leave-one-out measures across the other same-specialty wards of sample hospitals in the local health authority. Because the panel does not include every hospital, the local network is measured on the hospitals in the sample. For each ward i in hospital h and LHA j, the corresponding network outcome is

$$\bar{Y}_{-ihj} = \frac{1}{N_{hj} - 1} \sum_{\substack{k=1 \\ k \neq i}}^{N_{hj}} Y_{khj} = \frac{\sum_{k=1}^{N_{hj}} Y_{khj} - Y_{ihj}}{N_{hj} - 1},$$

where N_{hj} denotes the number of wards in the relevant same-specialty local network. This construction allows us to measure directly how the capacity and activity of the remaining providers evolve when one part of the network is downsized.

The estimated effects should therefore be read as changes in the activity of the same-specialty network surrounding each department, excluding the department itself. Because the downsized ward remains in the average with reduced capacity, these changes are net of its own contraction. The first group of outcomes measures activity and capacity use. These include surgical procedures, utilized beds, total discharges, one-day discharges, and surgical discharges. Together, they capture whether surviving providers absorb additional activity and how intensively their existing capacity is used.

The second group captures congestion, which we measure as the load on bed capacity. We consider two pairs of indicators. The first pair refers to the pre-operative phase: the average pre-operative stay, defined as the mean number of days surgical patients spend in hospital between admission and surgery, and total pre-operative days, which sums these days across all surgical patients. The second pair refers to the whole hospital stay: the average length of stay per ordinary admission and total inpatient days, which sums inpatient days across all ordinary admissions. We refer to these four indicators jointly as patient days. Averages measure the time each patient occupies a bed, whereas totals increase with the number of patients treated as well as with the time each of them spends in hospital. Comparing the two allows us to assess whether the reallocation of activity toward surviving providers is accompanied by greater pressure on available bed capacity, and

whether this pressure stems from a larger volume of patients or from longer stays. The third group measures quality of care and patient outcomes. We use mortality outcomes for acute conditions for which timely hospital treatment is particularly important, including myocardial infarction and stroke. We also consider cancer and hip-fracture outcomes, which provide comparisons across clinical pathways with different degrees and forms of time sensitivity. Cancer mortality is measured as the average of the AGENAS indicators of 30-day mortality after surgery for eight tumour sites (brain, colon, rectum, stomach, liver, pancreas, lung and kidney), computed over the sites available for each hospital and year. Because these are elective surgical pathways, they are exposed to the same pre-operative bottleneck documented below but are less sensitive to delays measured in minutes or hours. This broader set of outcomes allows us to assess whether any mortality effects are concentrated in specific clinical pathways rather than reflecting a general deterioration in hospital performance. Table 1 reports the complete set of outcomes and their sources.

Variable	Interpretation	Level	Source
Activity volumes and capacity use			
Surgical procedures	Inpatient and day-hospital interventions capturing changes in activity levels.	Ward-year	Ministry of Health
Utilized hospital beds	Average number of occupied beds, proxying capacity pressure.	Ward-year	Ministry of Health
Total discharges	Total number of patients discharged, capturing hospital demand absorption.	Ward-year	Ministry of Health
Discharged in 1 day	Short stays, informative about case mix and patient selection.	Ward-year	Ministry of Health
Surgical discharges	Discharges involving surgery, reflecting surgical caseload.	Ward-year	Ministry of Health
Congestion (patient days)			
Average pre-operative stay	Average number of days between admission and surgery per surgical patient, capturing the time patients wait for surgery once admitted.	Ward-year	Ministry of Health
Total pre-operative days	Sum of days between admission and surgery across surgical patients, capturing the load of patients awaiting surgery on bed capacity.	Ward-year	Ministry of Health
Average length of stay	Average number of inpatient days per ordinary admission, capturing the time each patient occupies a bed.	Ward-year	Ministry of Health
Total inpatient days	Sum of inpatient days across ordinary admissions, capturing the overall load on bed capacity.	Ward-year	Ministry of Health
Quality of care and patient outcomes			
MI	Acute myocardial infarction: raw 30-day mortality rate (deaths over admissions, %), number of 30-day deaths and number of index admissions.	Ward-LHA	PNE, AGENAS
Stroke	Stroke: raw mortality rate (deaths over admissions, %), number of deaths and number of index admissions.	Ward-LHA	PNE, AGENAS
Hip fracture	Hip fracture: raw 30-day mortality rate (deaths over admissions, %), number of 30-day deaths and number of index admissions, providing an additional acute-care comparison.	Ward-LHA	PNE, AGENAS
Cancer	Average 30-day mortality after surgery for malignant tumours of the brain, colon, rectum, stomach, liver, pancreas, lung and kidney (risk-adjusted rate), number of 30-day deaths and number of surgical admissions, capturing a surgical pathway with a different form of time sensitivity.	Ward-LHA	PNE, AGENAS

Table 1: Main outcome variables

3 Empirical Strategy

Our empirical design exploits variation in the timing and location of ward downsizing to estimate how surviving departments respond to reductions in same-specialty capacity within their local hospital network. We compare outcomes in departments exposed to a nearby same-specialty downsizing episode with those in departments that are not yet exposed. We estimate the following generalized difference-in-differences specification

$$Y_{jhit} = \alpha_0 + \alpha_1(Treat_{jhi} \times Post_{jhit}) + \delta_{s(i)j} + \eta_h + \lambda_t + \mu_{s(i)t} + \epsilon_{jhit}, \tag{1}$$

where Y_{jhit} denotes the outcome for neighbouring same-specialty department i in hospital h, belonging to local health authority (LHA) j, at time t. As described in Section 2.4, outcomes are constructed as leave-one-out measures that exclude the downsized ward itself. $Treat_{jhi} \times Post_{jhit}$ equals one once a department is exposed to the downsizing of at least one other ward of the same specialty within its local health authority, and remains equal to one thereafter. The coefficient α_1 captures the average change in outcomes among surviving departments following a reduction in same-specialty capacity elsewhere in their local hospital network.

The specification includes a rich set of fixed effects. Specialty-by-LHA fixed effects ($\delta_{s(i)j}$), where $s(i)$ denotes the specialty of department i, account for persistent differences in specialty mix and local healthcare-system structure. Hospital fixed effects (η_h) absorb time-invariant differences across hospitals. Year fixed effects (λ_t) absorb national trends common to all departments, and specialty-by-year fixed effects ($\mu_{s(i)t}$) absorb national specialty-specific shocks, such as changes in clinical guidelines, technology, or reimbursement rules. Standard errors are clustered at the local health authority level.

To examine heterogeneity, we interact the treatment indicator with mutually exclusive group indicators,

$$Y_{jhit} = \alpha_0 + \sum_g \alpha_1^g \left(Treat_{jhi} \times Post_{jhit} \times D_h^g\right) + \delta_{s(i)j} + \eta_h + \lambda_t + \mu_{s(i)t} + \epsilon_{jhit}, \tag{2}$$

where D_h^g indicates whether hospital h is public or private, or whether it is located in the North, Centre or South. The main effects of these indicators are absorbed by the hospital fixed effects. Each coefficient α_1^g measures the treatment effect for departments in group g.

The identifying assumption is that, conditional on these fixed effects, outcomes in exposed and not-yet-exposed departments would have evolved similarly in the absence of downsizing. A concern would arise if capacity reductions systematically occurred in local markets already experiencing changes in demand or hospital activity. The institutional setting and the data provide little evidence of such selection. As shown in Figure 2, subsequent downsizing is not systematically related to pre-treatment utilization or network characteristics.. Moreover, the decentralized implementation of the national capacity constraint generated substantial variation in the timing of reductions across regions, hospitals, and specialties. The fixed-effect structure further absorbs persistent differences across local networks and hospitals and contemporaneous shocks at the specialty level.

We complement the difference-in-differences estimates with an event-study analysis based on

the Extended Two-Way Fixed Effects (ETWFE) estimator of Wooldridge (2025). This approach allows us to trace treatment dynamics and assess potential anticipation effects while accommodating treatment-effect heterogeneity across cohorts and over time in a staggered adoption setting. We use never-exposed networks as the comparison group and estimate lead and lag coefficients relative to the year preceding exposure. The event-study analysis is conducted at the LHA–specialty level, reflecting both the definition of treatment and the organization of hospital care. Because wards within the same specialty and LHA share the same capacity shock, the first qualifying downsizing episode defines a common treatment date for the network. Aggregating outcomes at this level also captures the reallocation of patients across hospitals within the network, rather than attributing changes in activity to individual departments. This is particularly relevant for mortality, as emergency care in Italy is organized through integrated hospital networks that coordinate patient allocation across providers, as discussed in Section 4.2. Outcomes are aggregated within each network and expressed in levels, so the event-study coefficients measure changes in beds, discharges, patient days, and mortality indicators relative to the year preceding exposure. The event-study estimates, presented in Section 4, allow us to examine pre-treatment trends and the evolution of outcomes following downsizing. We complement this analysis with randomization inference (Section 5.1), which compares the observed estimates with those obtained under placebo assignments of treatment timing and location, and leave-one-out exercises (Section 5.2), which exclude one hospital or, for mortality, one LHA at a time.

4 Results

4.1 Activity volumes and capacity utilisation

We first examine how ward downsizing affects activity in the surviving departments of the local hospital network. Table 2 reports the difference-in-differences estimates for five outcomes: utilized beds, total discharges, one-day discharges, surgical discharges, and surgical procedures. These measures capture complementary margins of adjustment, ranging from bed utilization and overall patient throughput to the composition and intensity of hospital activity. All outcomes are expressed in logs, so the estimated coefficients can be interpreted as semi-elasticities; throughout the text, we report the corresponding percentage changes. Figure 3 then presents the event-study estimates for the same outcomes, allowing us to assess pre-treatment trends and trace the dynamics of the response following downsizing.

The estimates in Table 2 show a sizeable increase in activity among departments exposed to a same-specialty downsizing within the local health authority. Since outcomes are leave-one-out averages, these estimates capture the change in activity of the other same-specialty departments in the local network. Utilized beds increase by around 12 percent, indicating greater use of bed capacity in the surviving departments. Total discharges rise by 18.7 percent, providing direct evidence that these departments absorb a substantial additional patient caseload. The response is particularly pronounced for surgical activity, with surgical discharges increasing by 24.1 percent and surgical procedures by 27.4 percent. One-day discharges, by contrast, rise by 5 percent, and the estimate is not statistically significant. The close correspondence between surgical discharges and surgical

procedures indicates that the increase in surgical caseload reflects additional operations rather than a change in the classification of discharges. The smaller response of one-day discharges further suggests that the reallocation is not disproportionately concentrated in short-stay episodes.

	(1) Beds used	(2) Discharges	(3) Discharged in 1 day	(4) Surgical discharges	(5) Surgical procedures
Treatment effect	0.119**	0.187***	0.050	0.241***	0.274***
	(0.045)	(0.068)	(0.071)	(0.075)	(0.080)
Observations	20,394	20,217	19,963	18,371	18,371

Notes: Outcomes are computed as LOO mean (Section 2.4) and are in log. Treatment captures ward's retrenchment and it is measured as described in Section 2.3. All the models include year, hospital, specialty-by-year and specialty-by-LHA fixed effects.
Clustered robust standard errors at the local health authority level are in parentheses. Percentage effects reported in the text are computed as $100 \times (e^{\hat{\beta}} - 1)$. *** p<0.01, ** p<0.05, * p<0.1.

Table 2: Results on activity volumes and capacity utilisation

The event-study estimates in Figure 3 trace the dynamics of this response. As explained in Section 3, the event study is estimated on outcomes expressed in levels, so its coefficients measure changes in the number of beds, discharges and procedures and are not directly comparable in magnitude with the semi-elasticities in Table 2. The response is immediate for most activity measures: utilized beds, total discharges, one-day discharges and surgical discharges rise significantly in the year of exposure, while surgical procedures increase significantly from $t + 1$. Surgical discharges and surgical procedures continue to grow over the following years and peak at $t + 2$; at $t + 3$, the coefficient on surgical procedures remains significant, whereas that on surgical discharges does not. Total discharges and utilized beds display a less regular profile, with significant increases at t and $t+2$ and imprecise estimates at $t+1$ and $t+3$, while one-day discharges show no systematic pattern after the year of exposure. None of the pre-exposure coefficients is statistically different from zero, providing no evidence of anticipation effects.

Taken together, the estimates show adjustment along both the number of patients treated and the intensity with which existing hospital capacity is used. Surviving departments treat substantially more patients and perform more procedures, while utilized beds also increase. Moreover, patient activity increases proportionally more than bed utilization. Total discharges rise by 18.7 percent, surgical discharges by 24.1 percent and surgical procedures by 27.4 percent, compared with a 11.9 percent increase in utilized beds. The pattern therefore indicates a substantial intensification of activity in the departments that remain in operation.

These findings are consistent with the reallocation channel documented in the hospital-closure literature. Lindrooth et al. (2003) show that surviving hospitals absorb patients displaced by nearby closures through higher admissions and emergency-room visits, while treatment costs per admission fall by 2–4 percent overall and by 6–8 percent among patients who would otherwise have been treated at the closed hospital. Our results point to a similar reallocation mechanism, but identify it at the department-within-hospital level and within the same specialty, where substitution is more direct. They also provide a useful contrast with the hospital-merger evidence. Gaynor et al. (2026)

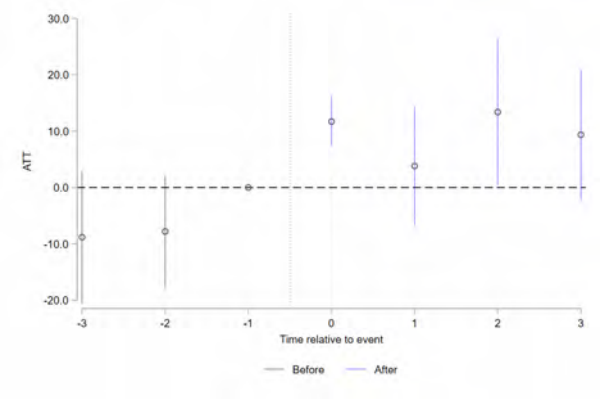
find limited effects on profitability and patient outcomes following a large merger despite substantial integration of management and information systems. A plausible reason for the stronger adjustment in our setting is that mergers need not entail the removal of treatment capacity, whereas ward downsizing constitutes a direct specialty-specific supply shock. Displaced activity is consequently more likely to be concentrated among a relatively small set of surviving same-specialty departments rather than dispersed across a broader hospital market.

Finally, it is important to note that not all of this additional activity should be read as pure reallocation. Part of it is the physiological counterpart of the reform's intent to displace patients toward surviving facilities, but part may also signal additional hospital use stemming from worse management of chronic conditions. Ward downsizing typically entails a contraction of the specialist care that hospital departments deliver for chronic conditions such as diabetes and cardiovascular disease. A growing literature shows that the quality of chronic-condition management in specialist care shapes hospitalization rates: when specialist services are withdrawn, effects spill over onto hospital use (Kozhimannil et al., 2018), while improvements in the quality of care driven by technological progress reduce avoidable admissions (Carrieri et al., 2026). Distinguishing between these two interpretations requires examining whether downsizing is also associated with worse treatment outcomes. If downsizing merely displaced patients, outcomes should be largely unaffected; if it also eroded the quality of hospital care, we should observe a deterioration. We investigate this mechanism in Section 4.2.

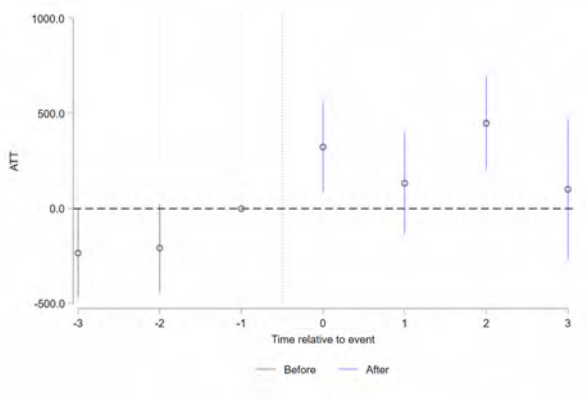
4.2 Congestion and quality of care

We next examine whether the increase in activity documented above is accompanied by greater congestion and, ultimately, worse patient outcomes. We proceed in two steps. We first examine the load on bed capacity, using total and average measures of both pre-operative days and inpatient days, and then turn to mortality outcomes, distinguishing between conditions for which timely treatment is particularly important and conditions for which short-run delays operate differently. Table 3 reports the difference-in-differences estimates for the four measures of patient days. Columns (1) and (3) refer to averages, namely the average pre-operative stay and the average length of stay, which measure the time each patient spends in hospital. Columns (2) and (4) refer to totals, namely total pre-operative days and total inpatient days, which rise with the number of patients treated as well as with the time each of them spends in hospital, and therefore measure the load that the reallocated caseload places on the bed capacity of surviving departments. The estimates on totals show substantial increases: following a same-specialty downsizing within the local health authority, total pre-operative days rise by 26 percent and total inpatient days by 23 percent. These increases are close to those in the corresponding activity measures in Table 2. Consistently, the coefficients on the average pre-operative stay and the average length of stay are not statistically different from zero. Taken together, these results indicate that the greater pressure on bed capacity is driven by the larger number of patients treated in surviving departments rather than by longer stays per patient. Figure 4 presents the associated event-study estimates.

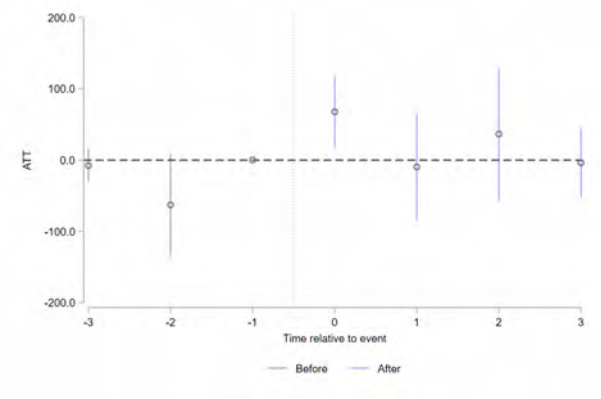
Total pre-operative days are the outcome most closely related to the surgical bottleneck. They count the bed-days occupied by patients awaiting their procedure, and their 26 percent increase



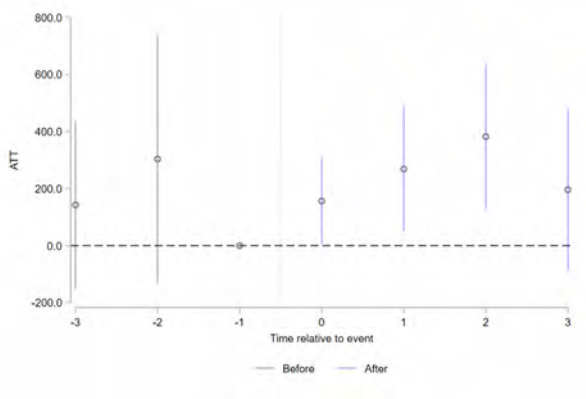
(a) Beds used



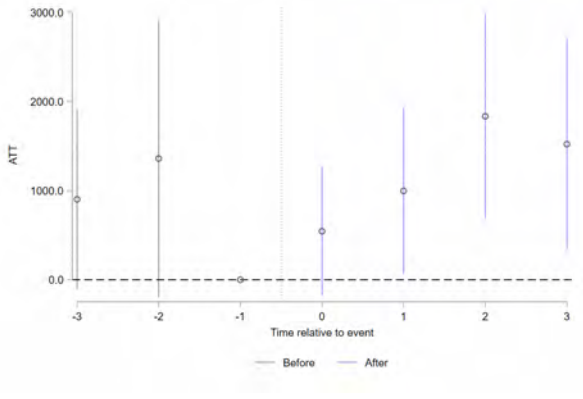
(b) Total discharges



(c) Discharged in 1 day



(d) Surgical discharges



(e) Surgical procedures

Figure 3: Event study results: activity volumes and capacity utilisation

Note: The vertical dotted line separates pre- and post-exposure periods; year $t-1$ is the reference year. Dots are the point estimates of the average treatment effects on the treated, estimated with the Extended Two-Way Fixed Effects (ETWFE) estimator of Wooldridge (2025) and aggregated by time relative to exposure; the vertical whiskers represent the 95% confidence intervals. Treatment cohorts are defined by the year of the first downsizing episode, as described in Section 2.3. Outcomes are expressed in levels. All models include year and ward fixed effects. Standard errors are clustered at the LHA level.

	(1) Average pre-operative stay	(2) Total pre-operative days	(3) Average length of stay	(4) Total inpatient days
Treatment effect	0.131	0.262***	−0.038	0.228***
	(0.152)	(0.082)	(0.122)	(0.076)
Observations	17,750	18,371	20,293	20,217

Notes: Outcomes are computed as LOO mean (Section 2.4) and are in log. Treatment captures ward's retrenchment and it is measured as described in Section 2.3. All the models include year, hospital, specialty-by-year and specialty-by-LHA fixed effects. Clustered robust standard errors at the local health authority level are in parentheses. Percentage effects reported in the text are computed as $100 \times (e^{\hat{\beta}} - 1)$. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Results on patient days

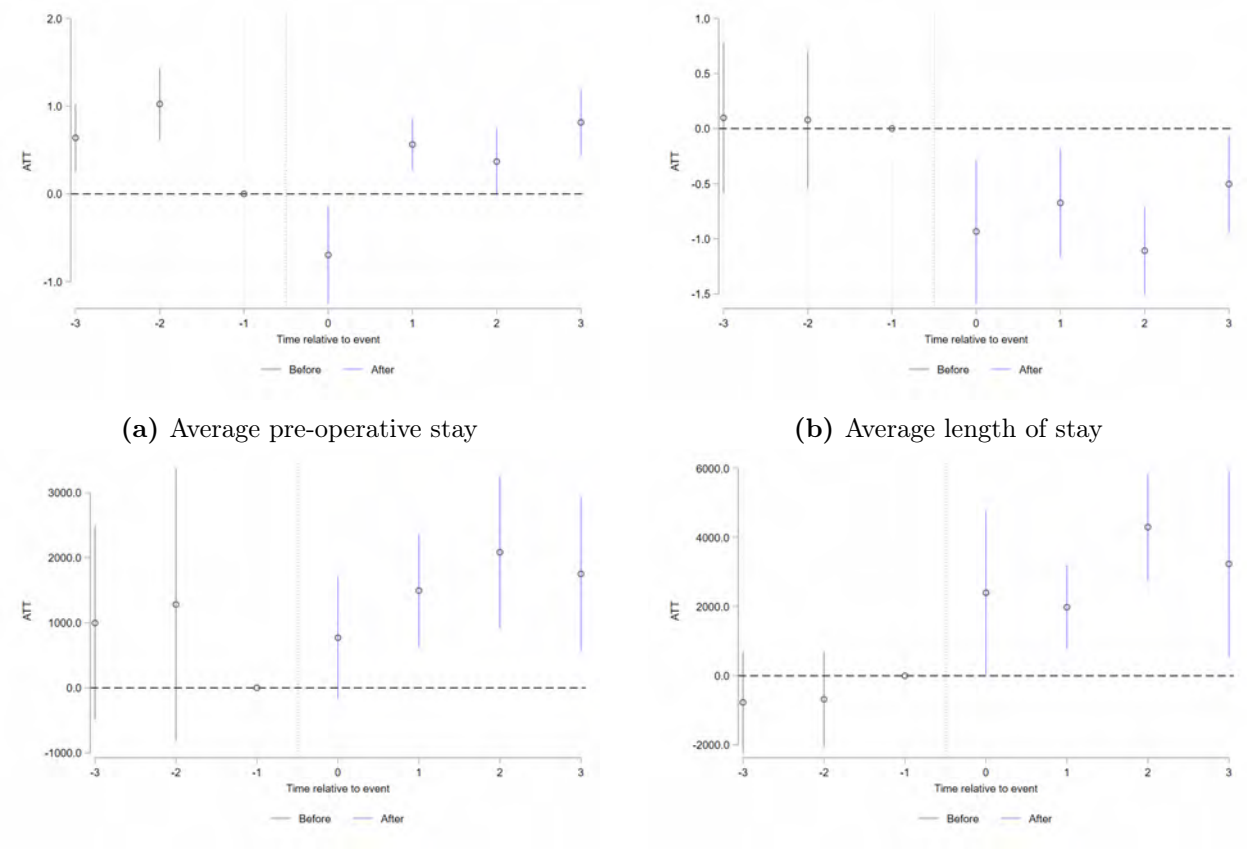
indicates that surviving departments devote substantially more bed capacity to patients before surgery. Because this increase closely tracks the rise in surgical discharges, it reflects primarily the larger surgical caseload rather than longer waits for each patient. We therefore interpret the increase in the load on the bed and surgical capacity of surviving departments as a proxy for congestion.

Figure 4 presents the event-study estimates for the four measures of patient days. Total pre-operative days rise gradually after exposure: the coefficient in the year of exposure is positive but imprecisely estimated, while from $t + 1$ onward the estimates are positive and statistically significant, peaking at $t + 2$. Total inpatient days increase significantly from the year of exposure onward, with the largest effect at $t + 2$. The evidence on averages points in the opposite direction. The average length of stay falls by between 0.5 and 1.1 days in each post-exposure year. The average pre-operative stay is the only outcome with statistically significant pre-exposure coefficients, and its post-exposure estimates change sign over time, showing no clear pattern. For the other three measures, none of the pre-exposure coefficients is statistically different from zero.

Overall, the dynamic evidence indicates that the load on the bed capacity of surviving departments rises after exposure and remains elevated, driven by a larger caseload rather than by longer stays per patient.

These findings are consistent with related evidence on hospital closures and capacity constraints. Ghislandi et al. (2025) document longer hospital stays following AMI ward closures in Italy. Our results show that the load on surviving providers rises outside emergency care as well, and that it is concentrated among surgical patients before their procedure, although it operates through a larger caseload rather than through longer stays. The congestion mechanism is also consistent with Francetic et al. (2024), who show that unexpected demand surges in English emergency departments worsen outcomes for higher-severity patients when capacity constraints bind, and with Hoe (2022), who finds that a one-standard-deviation admission shock in English trauma and orthopaedics departments raises unplanned readmissions by 4.1 percent, with larger effects in departments with fewer beds. Closest to our setting, Avdic et al. (2024) show that Swedish maternity ward closures increased maternal obstetric trauma through overcrowding in the remaining wards rather than through longer travel distances. Although the English studies exploit short-run demand shocks and ours a persistent reduction in local supply, these studies point to the same underlying adjustment. When additional demand is concentrated on providers with limited capacity to expand, the pressure

on their existing capacity rises.



(a) Average pre-operative stay

(b) Average length of stay

(c) Total pre-operative days

(d) Total inpatient days

Figure 4: Event study results on patient days

Note: The vertical dotted line separates pre- and post-exposure periods; year $t-1$ is the reference year. Dots are the point estimates of the average treatment effects on the treated, estimated with the Extended Two-Way Fixed Effects (ETWFE) estimator of Wooldridge (2025) and aggregated by time relative to exposure; the vertical whiskers represent the 95% confidence intervals. Treatment cohorts are defined by the year of the first downsizing episode, as described in Section 2.3. Outcomes are expressed in levels. All models include year and ward fixed effects. Standard errors are clustered at the LHA level.

We next examine whether the congestion documented above translates into worse health outcomes. For mortality, we shift the unit of analysis from the individual hospital to the LHA– ward network. This aggregation is particularly appropriate for acute and emergency conditions. In the Italian National Health Service, emergency care is organized as an integrated network rather than as a set of independent hospitals. Emergency calls, territorial emergency services, emergency departments, and hospital facilities are coordinated to direct patients to the provider within the network that is best equipped to deliver the required care. The relevant margin for mortality outcomes is therefore not necessarily what happens at the individual hospital in which capacity is reduced, but how the emergency-care network as a whole responds when same-specialty capacity is removed from one of its providers.

Accordingly, we aggregate the AGENAS deaths indicators of sample hospitals at the ward–LHA

level and estimate event-study specifications around the first qualifying downsizing episode within the corresponding network. This approach has two advantages. First, it matches the institutional organization of emergency care and captures potential reallocation of acute patients across hospitals following downsizing. Second, it avoids assigning changes in mortality to a particular surviving hospital when the capacity shock may alter where emergency patients are treated within the local network.Table 4 reports pooled ATT estimates while Figures 5 and 6 show the event-study dynamic. We distinguish between acute, time-sensitive conditions, including myocardial infarction and stroke, and cancer and hip fracture, which provide comparison outcomes for which short-run delays operate differently.

In Table 4 we report three outcomes: the condition-specific in-hospital mortality rate, the total number of deaths, and the number of admissions for each condition. This distinction is useful for investigating the mechanism discussed above, namely whether downsizing merely displaces patients or also worsens health outcomes. Admissions increase for all conditions, although the estimates are statistically significant only for stroke, at the 10 percent level. This pattern is consistent with the displacement of patients intended by the reform. It may also partly reflect the contraction of the specialist services provided in treated departments, which are relevant for the management of chronic conditions. Outcomes also generally worsen. Total deaths rise for all conditions albeit significantly only for stroke and cancer, pointing to a general deterioration of mortality indicators at the network level. In-hospital mortality is the outcome most informative about the quality of hospital care delivered. We find that it increases significantly for AMI and, at the 10 percent level, for stroke, two time-sensitive conditions for which timely and effective treatment is critical. In both cases the effect is about 2.6 percentage points. By contrast, we find no significant effect on in-hospital mortality for hip fracture and cancer. Ghislandi et al. (2025) find that AMI ward closures in Italy increased in-hospital mortality by 0.7 percentage points, around 10 percent of the mean mortality rate. Our estimate for AMI is larger, at about 2.6 percentage points. Part of this gap may reflect differences in the outcome, since our 30-day mortality indicator also captures deaths occurring after discharge or transfer, and in the level of analysis, since we measure outcomes for the local network as a whole.

Figures 5 and 6 report the corresponding event-study estimates. Pre-exposure coefficients are generally close to zero, with the exception of cancer deaths and admissions, for which the earliest pre-exposure coefficient is negative. For time-sensitive conditions (Figure 5), the post-exposure patterns are broadly in line with the pooled estimates. In-hospital mortality for AMI and stroke rises after downsizing, with point estimates that grow larger in later periods. This pattern is suggestive of a gradual deterioration in the quality of care as surviving departments absorb displaced patients, rather than of a one-off disruption at the time of the shock. Admissions and total deaths are also generally positive in the post-exposure period, although estimated with less precision, consistent with the displacement of patients toward surviving structures and with some worsening of mortality indicators at the network level, particularly for stroke. For the comparison conditions (Figure 6), the evidence is weaker. For hip fracture, post-exposure estimates show no clear pattern for mortality and deaths, while admissions are slightly positive but imprecisely estimated. For cancer, in-hospital mortality and admissions show no clear post-exposure response, whereas total deaths increase after

exposure, in line with the pooled estimate. Taken together, the event-study evidence is consistent with downsizing worsening hospital outcomes mainly for conditions in which timely and effective treatment is critical, while we find little evidence of effects on in-hospital mortality for less time-sensitive conditions.

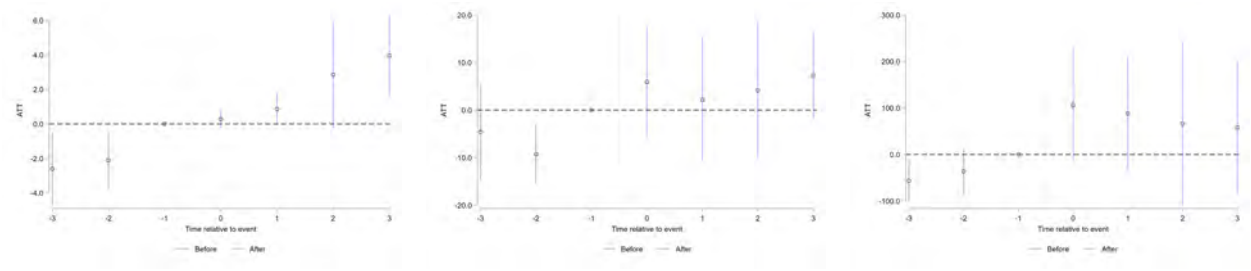
	(1)	(2)	(3)
	Mortality rate	Deaths	Admissions
Panel A: Myocardial infarction (MI)			
Pooled ATT	2.572***	5.513	73.697
	(0.735)	(5.460)	(71.476)
Observations	14855	15671	15671
Panel B: Stroke			
Pooled ATT	2.593*	14.034**	71.978*
	(1.484)	(7.013)	(42.565)
Observations	13669	15671	15671
Panel C: Hip fracture (30 days)			
Pooled ATT	0.017	0.658	1.353
	(0.144)	(0.443)	(5.987)
Observations	14911	14911	14911
Panel D: Cancer			
Pooled ATT	0.436	4.015***	10.817
	(0.534)	(1.291)	(36.058)
Observations	13451	15671	15671

Notes: The table reports the average treatment effect on the treated estimated with the Extended Two-Way Fixed Effects (ETWFE) estimator of Wooldridge (2025), aggregated across all treated cohorts and post-exposure periods. The specification is the same as in the event-study estimates in Figures 5 and 6. The mortality rate is the raw rate (deaths over admissions, %). Cancer is computed as the averge mortality rate for malignant tumours of the brain, colon, rectum, stomach, liver, pancreas, lung and kidney. All models include year and ward fixed effects.
Clustered robust standard errors at the LHA level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

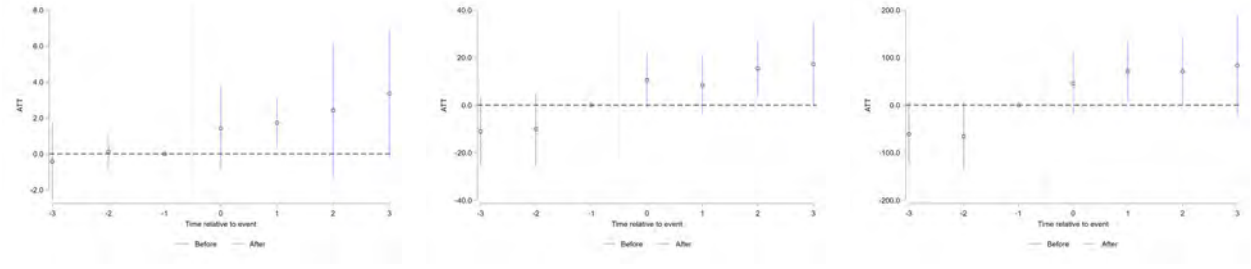
Table 4: Pooled ATT

Combined with the activity and patient-day results, these estimates suggest that downsizing reduces same-specialty capacity at one point in the local hospital network, reallocates activity to the departments that remain open, and increases the load on their bed capacity. At the network level, this adjustment is accompanied by higher mortality for AMI and stroke, conditions in which timely and effective treatment is critical, while we find little evidence of adverse effects on mortality for the comparison conditions. Unlike in the case of facility closures, where the main channel is longer travel times (Ghislandi et al., 2025), the evidence points to a congestion channel operating within the local network: displaced patients are absorbed by surviving departments whose capacity comes under greater pressure. This interpretation is consistent with Ghislandi et al. (2025), who find that among patients experiencing only moderate increases in travel time, mortality effects appear to be driven by channels other than travel time, such as congestion. The LHA–specialty analysis is important for this interpretation, because the estimated mortality effect captures the consequence

of the capacity shock for the emergency-care network as a whole, irrespective of which hospital ultimately treats the patient.



(a) MI – Mortality rate **(b)** MI – Deaths **(c)** MI – Admissions



(d) Stroke – Mortality rate **(e)** Stroke – Deaths **(f)** Stroke – Admissions

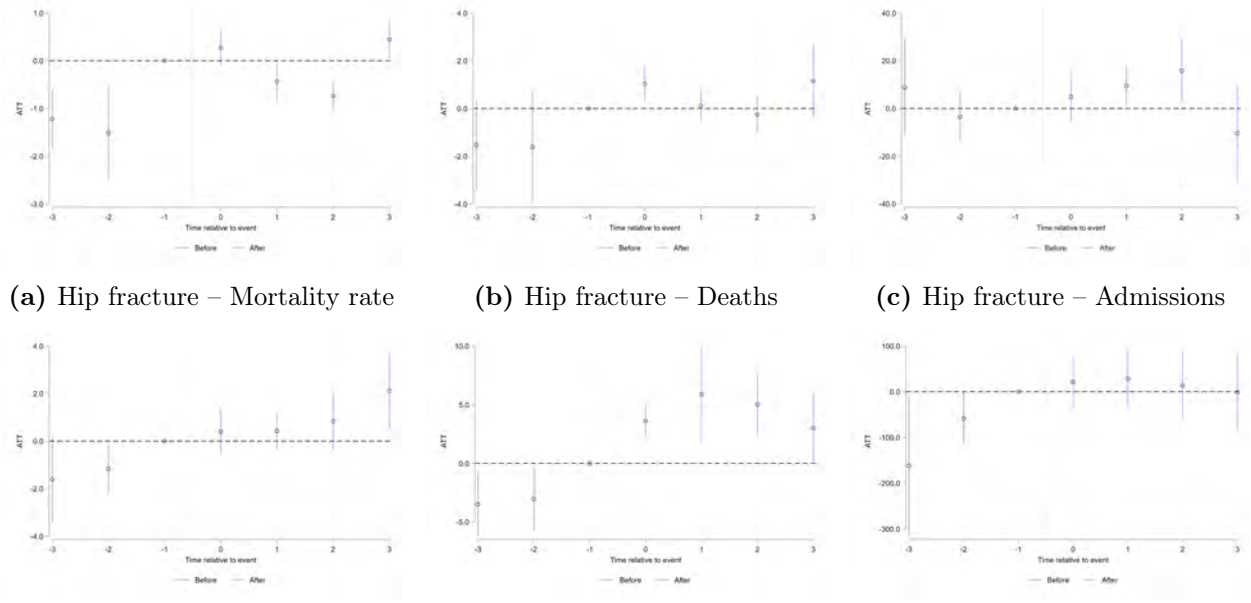
Figure 5: Event study results on mortality, deaths and admissions:time sensitive conditions

Note: The vertical dotted line separates pre- and post-exposure periods; year $t-1$ is the reference year. Dots are the point estimates of the average treatment effects on the treated, estimated with the Extended Two-Way Fixed Effects (ETWFE) estimator of Wooldridge (2025) and aggregated by time relative to exposure; the vertical whiskers represent the 95% confidence intervals. Treatment cohorts are defined by the year of the first downsizing episode, as described in Section 2.3. Outcomes are expressed in levels. All models include year and ward fixed effects. Standard errors are clustered at the LHA level.

4.3 Heterogeneity

We next examine whether the adjustment to ward downsizing differs across providers and geographical areas. These dimensions of heterogeneity are informative about the mechanisms through which local hospital networks accommodate reductions in capacity. We first distinguish between public and private providers and then examine differences across the North, Centre, and South of Italy.

Tables 5 and 6 report the estimates separately for public and private hospitals. The response is statistically significant only for public hospitals. The increase in activity following same-specialty downsizing is concentrated in public hospitals, where surgical discharges increased by 31.3 percent and surgical procedures by 33 percent and are statistically significant at 1 percent. Total discharges rise by 17.5 percent and utilized beds by 10.6 percent, both significant at the 10 percent level, consistent with public providers absorbing most of the additional caseload generated by downsizing elsewhere in the local network. Private hospitals, by contrast, show no statistically significant change in any activity measure. Their largest point estimate is for one-day discharges, at 15.3 percent, but it is imprecisely estimated, and the estimates for surgical discharges and surgical procedures are negative.



(a) Hip fracture – Mortality rate **(b)** Hip fracture – Deaths **(c)** Hip fracture – Admissions

(d) Cancer – Mortality rate **(e)** Cancer – Deaths **(f)** Cancer – Admissions

Figure 6: Event study results on mortality, deaths and admissions: comparison conditions

Note: The vertical dotted line separates pre- and post-exposure periods; year $t-1$ is the reference year. Dots are the point estimates of the average treatment effects on the treated, estimated with the Extended Two-Way Fixed Effects (ETWFE) estimator of Wooldridge (2025) and aggregated by time relative to exposure; the vertical whiskers represent the 95% confidence intervals. Treatment cohorts are defined by the year of the first downsizing episode, as described in Section 2.3. Outcomes are expressed in levels. All models include year and ward fixed effects. Standard errors are clustered at the LHA level.

The composition of this response is informative. Private hospitals do not appear to substitute for the capacity removed elsewhere in the network. Their point estimates are positive only for bed use, discharges and one-day discharges, and negative for surgical activity, while the increase in patient volumes and surgical activity is absorbed by public hospitals. Although public and accredited private hospitals operate within the same regional health systems, they respond differently to the same local supply shock.

Table 6 reinforces this interpretation. Patient days increase significantly among public providers, by 22.8 percent for total inpatient days and 35.5 percent for total pre-operative days, whereas the private estimates are small and not statistically significant. The provider absorbing the additional activity is therefore also the provider experiencing the associated increase in congestion. Taken together, the volume and patient-day estimates indicate that private capacity provides only limited substitution for reductions in public hospital capacity, and the adjustment occurs predominantly within the public sector.

This asymmetry is consistent with evidence that public and private hospitals respond differently to the profitability of different types of patients. Duggan et al. (2023) show that when public hospitals in the United States are privatized, the new operators reduce admissions of relatively unprofitable Medicaid patients, without these patients being fully absorbed by neighboring providers. Our setting differs because ownership and reimbursement rules do not change, and the shock is instead a reduction in local hospital capacity. Yet the resulting allocation of activity points in a similar direction. Private providers do not absorb the displaced caseload, while public hospitals take on the broader increase in demand. In our setting, therefore, provider heterogeneity emerges even in the absence of a change in ownership.

	(1) Beds used	(2) Discharges	(3) Discharged in 1 day	(4) Surgical discharges	(5) Surgical procedures
Private	0.029	0.051	0.142	−0.096	−0.142
	(0.117)	(0.096)	(0.146)	(0.082)	(0.106)
Public	0.101*	0.161*	0.003	0.272***	0.285***
	(0.055)	(0.081)	(0.082)	(0.096)	(0.099)
Observations	20,394	20,217	19,963	18,371	18,371

Notes: Outcomes are computed as LOO mean (Section 2.4) and are in log. Treatment captures ward's retrenchment and it is measured as described in Section 2.3.All the models include year, hospital, specialty-by-year and specialty-by-LHA fixed effects. *Public* and *Private* report the treatment effect for public and private structures, respectively.
Clustered robust standard errors at the local health authority level are in parentheses. Percentage effects reported in the text are computed as $100 \times (e^{\hat{\beta}} - 1)$. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Public vs Private: Activity volumes and capacity

We next examine geographical heterogeneity. Italy provides a particularly informative setting because hospital capacity and the organization of hospital networks differ substantially across macro-areas. Tables 7 and 8 report the effects separately for the North, Centre, and South [1].

[1]Macro-area dummies follow the ISTAT classification of Italian geographical areas and are assigned according to

	(1) Average pre-operative stay	(2) Total pre-operative days	(3) Average length of stay	(4) Total inpatient days
Private	−0.086	−0.105	−0.051	0.036
	(0.072)	(0.078)	(0.047)	(0.091)
Public	0.156	0.304***	−0.064	0.205**
	(0.206)	(0.108)	(0.165)	(0.099)
Observations	17,750	18,371	20,293	20,217

Notes: Outcomes are computed as LOO mean (Section 2.4) and are in log. Treatment captures ward's retrenchment and it is measured as described in Section 2.3. All the models include year, hospital, specialty-by-year and specialty-by-LHA fixed effects. *Public* and *Private* report the treatment effect for public and private structures, respectively.
Clustered robust standard errors at the local health authority level are in parentheses. Percentage effects reported in the text are computed as $100 \times (e^{\hat{\beta}} - 1)$. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Public vs Private: Patient days

The activity point estimates are positive in all three macro-areas, but their precision differs. Total discharges rise by 28.7 percent in the Centre, the only macro-area where the estimate is statistically significant, while the estimates for the North and South are positive but imprecise. Utilized beds increase by 18.2 percent in the Centre, with a similar but insignificant point estimate in the North. Surgical activity responds in the Centre and South, where surgical discharges increase by 43.8 and 31.5 percent and surgical procedures by 48.4 and 27.6 percent, respectively, whereas the North shows no significant change. The geographical estimates therefore confirm the reallocation result most clearly in the Centre and show that networks in the South accommodate the additional caseload mainly through surgical activity.

The differences become more pronounced when we examine patient days. Total pre-operative days increase by 52.8 percent in the Centre and 33 percent in the South, significant at the 1 and 10 percent levels respectively. In the North, by contrast, the estimate is close to zero (1.5 percent) and not statistically significant, so the increase in the pre-operative load is concentrated in the Centre and South. Total inpatient days increase by 39.5 percent in the Centre, while the point estimates for the North and South are positive but not statistically significant.

Taken together, these results indicate that the reallocation of activity following downsizing is most clearly identified in the Centre, with surgical activity also responding in the South, while the extent to which it translates into congestion differs geographically. The absence of a response in total pre-operative days in the North contrasts with the sizeable increases in the Centre and South. This pattern is consistent with differences in the ability of local hospital networks to accommodate additional activity. Where the surviving network has less scope to absorb the displaced caseload, the same supply shock is more likely to translate into a heavier load on surgical capacity.

the region in which the hospital is located. *North* includes Piedmont, Aosta Valley, Liguria, Lombardy, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia and Emilia-Romagna; *Centre* includes Tuscany, Umbria, Marche and Lazio; *South* includes Abruzzo, Molise, Campania, Apulia, Basilicata, Calabria, Sicily and Sardinia.

	(1) Beds used	(2) Discharges	(3) Discharged in 1 day	(4) Surgical discharges	(5) Surgical procedures
North	0.184	0.278	0.127	0.061	0.206
	(0.152)	(0.199)	(0.181)	(0.150)	(0.196)
Centre	0.167***	0.252**	0.074	0.363***	0.395***
	(0.062)	(0.122)	(0.115)	(0.108)	(0.118)
South	0.101	0.119	0.042	0.274*	0.244*
	(0.082)	(0.102)	(0.133)	(0.141)	(0.141)
Observations	20,394	20,217	19,963	18,371	18,371

Notes: Outcomes are computed as LOO mean (Section 2.4) and are in log. Treatment captures ward's retrenchment and it is measured as described in Section 2.3. All the models include year, hospital, specialty-by-year and specialty-by-LHA fixed effects. *North*, *Centre* and *South* report the treatment effect estimated through separate interactions of the treatment variable with macro-area dummies.
Clustered robust standard errors at the local health authority level are in parentheses. Percentage effects reported in the text are computed as $100 \times (e^{\hat{\beta}} - 1)$. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Geographical heterogeneity: activity volumes and capacity

	(1) Average pre-operative stay	(2) Total pre-operative days	(3) Average length of stay	(4) Total inpatient days
North	−0.089	0.015	−0.512	0.253
	(0.082)	(0.127)	(0.538)	(0.164)
Centre	0.033	0.424***	−0.058	0.333**
	(0.349)	(0.126)	(0.166)	(0.162)
South	0.263	0.285*	0.187	0.145
	(0.180)	(0.159)	(0.183)	(0.118)
Observations	17,750	18,371	20,293	20,217

Notes: Outcomes are computed as LOO mean (Section 2.4) and are in log. Treatment captures ward's retrenchment and it is measured as described in Section 2.3. All the models include year, hospital, specialty-by-year and specialty-by-LHA fixed effects. *North*, *Centre* and *South* report the treatment effect estimated through separate interactions of the treatment variable with macro-area dummies.
Clustered robust standard errors at the local health authority level are in parentheses. Percentage effects reported in the text are computed as $100 \times (e^{\hat{\beta}} - 1)$. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Geographical heterogeneity: patient days

5 Robustness & sensitivity

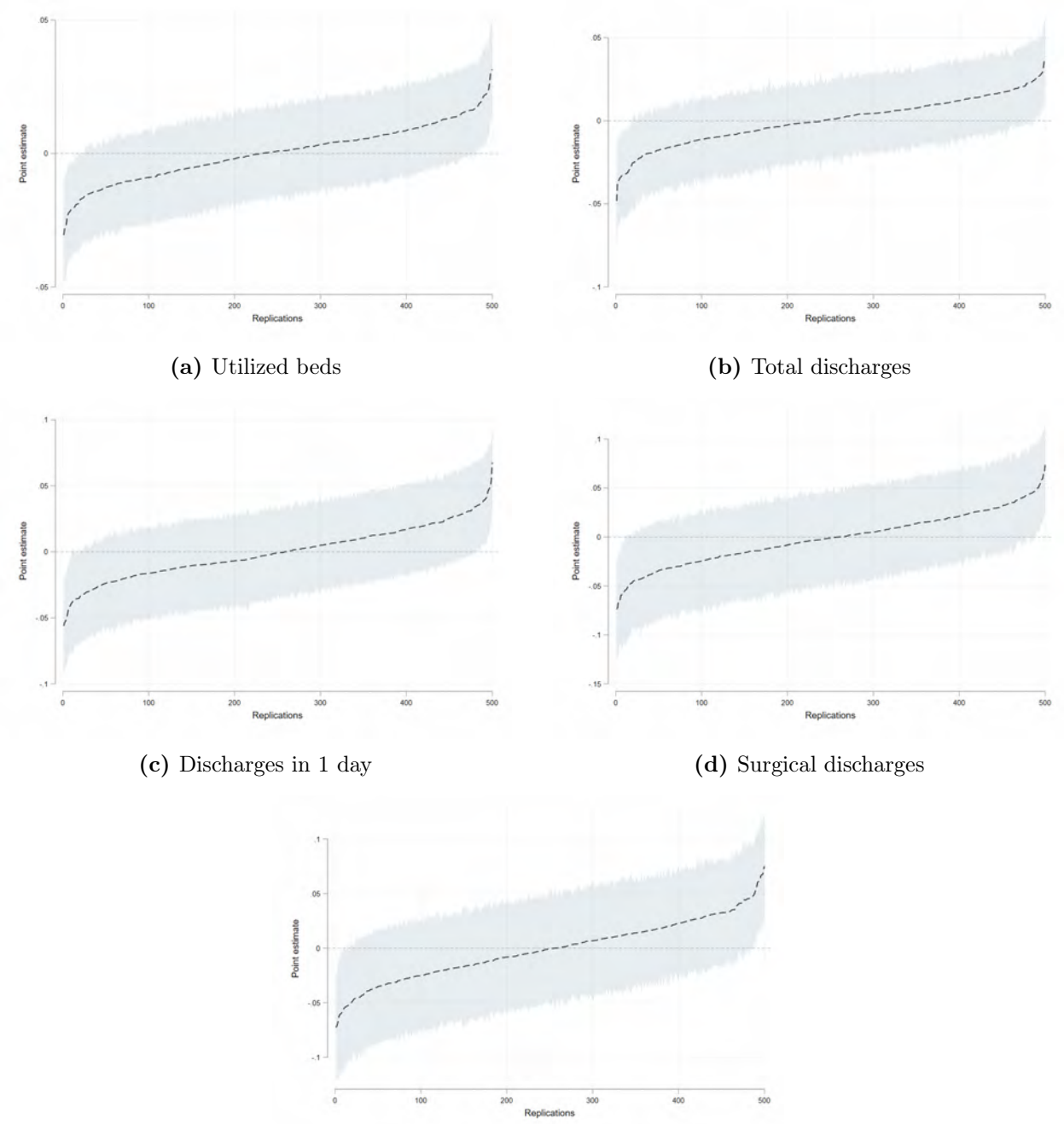
5.1 Randomisation inference

To assess whether our findings could arise from features of the panel structure rather than from the actual timing and location of downsizing, we re-estimate the baseline model under placebo assignments of treatment. In each of 500 replications, we randomly reassign treatment to ward-within-hospital-year observations, drawing in each year as many observations as are treated in the actual data among those not in the actual treatment group. The placebo post-treatment indicator switches on in the first year of placebo exposure and remains on thereafter. Each replication retains the specification, fixed effects and clustering structure of the baseline model (Section 3). Under the null of no treatment effect, placebo coefficients should be centred around zero, and estimates as large in absolute value as the baseline ones should occur infrequently.

This exercise also addresses the concern raised by Bertrand et al. (2004) that conventional standard errors may overstate precision in difference-in-differences designs with serially correlated outcomes. The empirical distribution of the placebo estimates provides a benchmark for the baseline estimates that does not rely on the asymptotic properties of clustered standard errors. Figures 7–9 plot, for each outcome, the sorted placebo estimates with their 95% confidence intervals, together with the baseline estimate.

The placebo coefficients are centred around zero for all outcomes. For the outcomes with statistically significant baseline estimates, namely utilised beds, total discharges, surgical discharges, surgical procedures, total pre-operative days and total inpatient days, the baseline estimates lie in the tails of the placebo distributions. By contrast, the baseline estimates for one-day discharges, the average pre-operative stay and the average length of stay, which are not statistically significant in Tables 2 and 3, fall within the placebo distributions, consistent with the absence of a systematic effect on these margins.

For mortality, the exercise is based on the pooled ETWFE estimate reported in Table 4, estimated on the panel at the ward–LHA level. Since treatment adoption is staggered at this level, placebo assignment is carried out on units rather than on observations: in each replication, we restrict the sample to never-treated ward–LHA units and randomly assign a placebo adoption year to a subset of them, keeping the share of treated units and the distribution of adoption cohorts as in the actual data, and re-estimate the pooled ETWFE model with the baseline specification. The baseline estimates for stroke and 30-day stroke mortality lie in the tails of the placebo distributions, whereas those for infarction, hip-fracture and cancer mortality fall within them. Taken together, these results indicate that the activity, congestion and mortality effects are unlikely to be explained by the panel structure alone, and that they are associated with the actual timing and location of ward downsizing.



(a) Utilized beds

(b) Total discharges

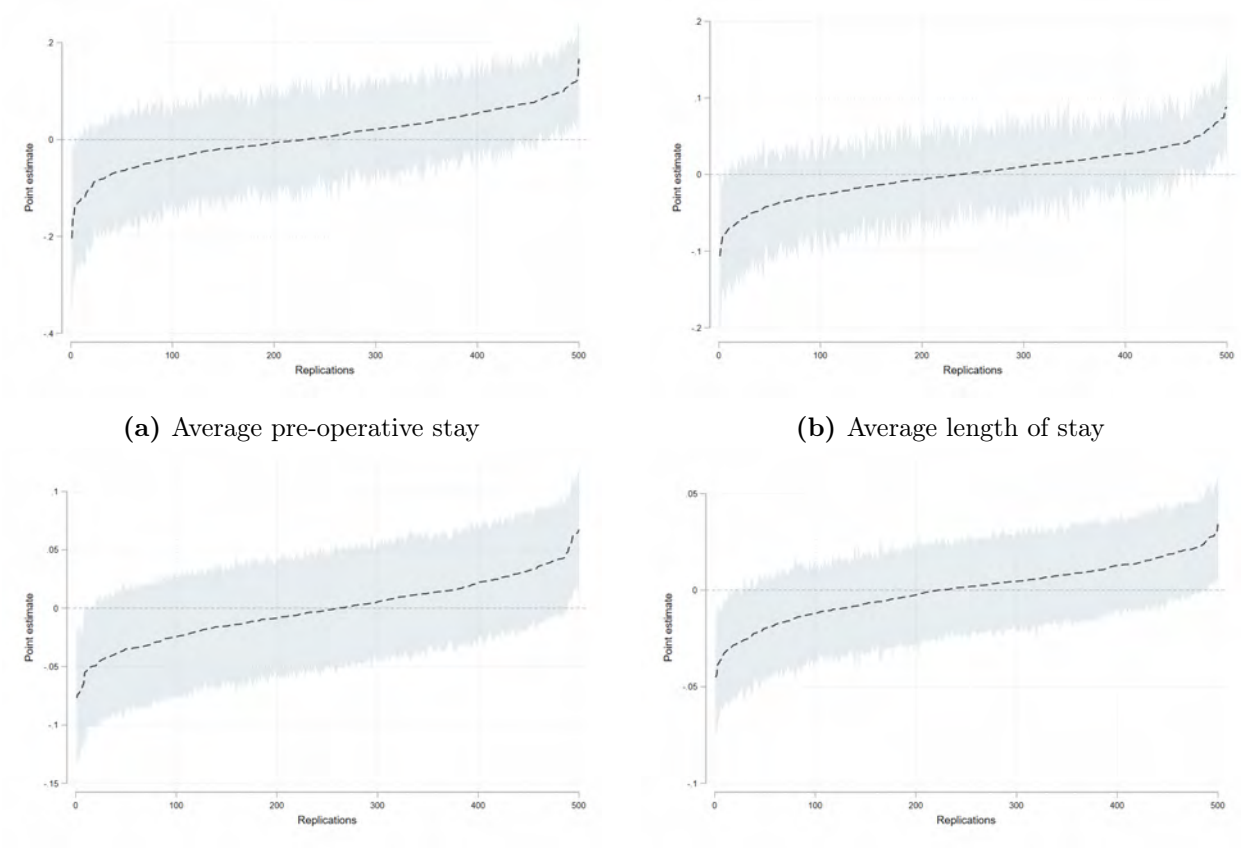
(c) Discharges in 1 day

(d) Surgical discharges

(e) Surgical procedures

Figure 7: Randomisation tests — activity volumes and capacity

Note: The figure reports the placebo coefficients (dashed line), sorted in ascending order, and their 95% confidence intervals (shaded area) from 500 replications. In each replication, a placebo treatment is randomly assigned, within each year from 2015 onwards, to as many ward-within-hospital-year observations as are treated in the actual sample, drawing only among wards that do not belong to the actual treatment group. The placebo post-treatment indicator switches on in the first year in which a ward receives the placebo treatment and remains on thereafter. In each replication, the model is re-estimated with the same specification, fixed effects and clustering of standard errors as in the baseline (Section 3), and the coefficient on the interaction between the placebo treatment and placebo post-treatment indicators is retained.



(a) Average pre-operative stay

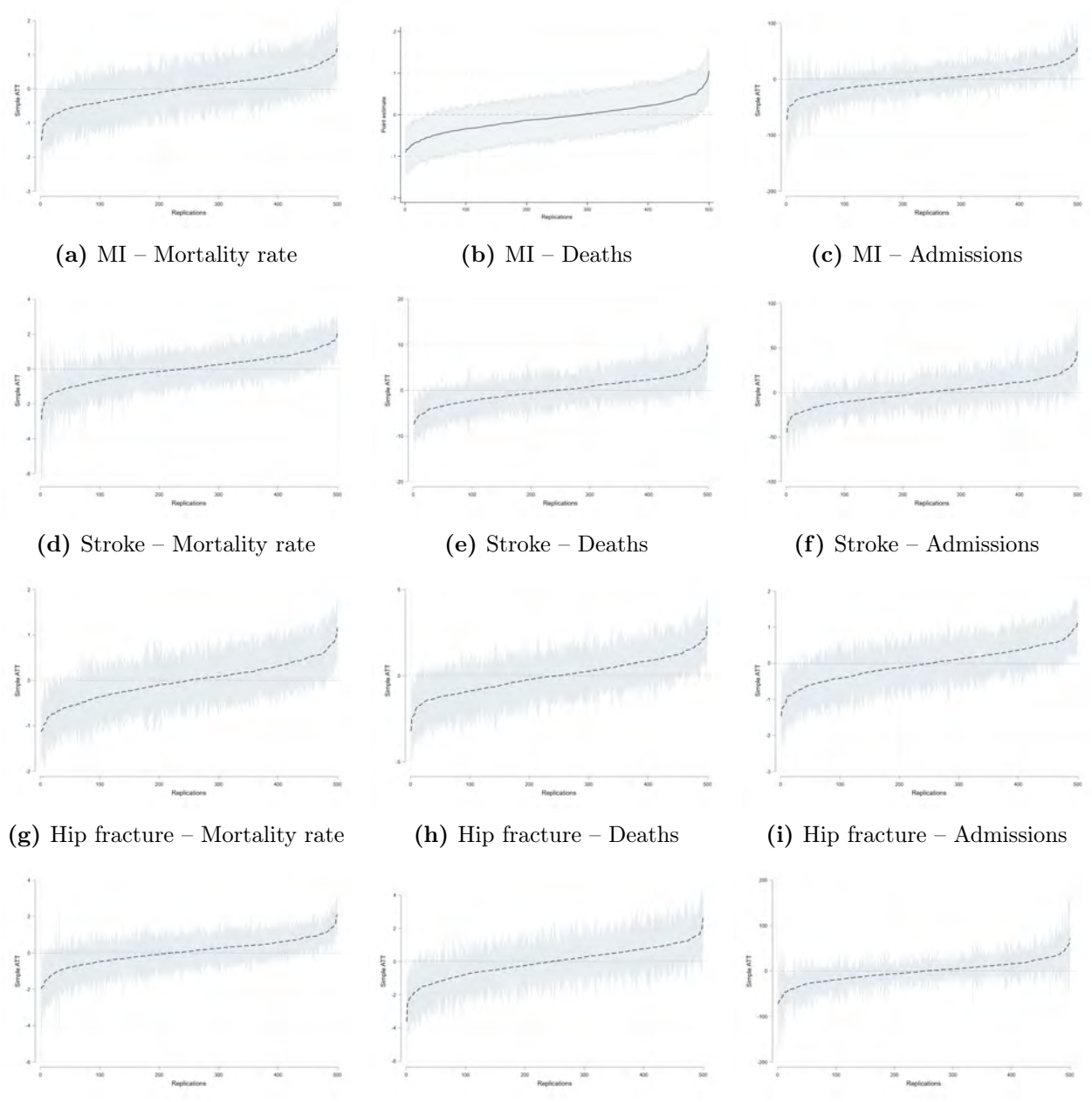
(b) Average length of stay

(c) Total pre-operative days

(d) Total inpatient days

Figure 8: Randomisation tests — patient days

Note: The figure reports the placebo coefficients (dashed line), sorted in ascending order, and their 95% confidence intervals (shaded area) from 500 replications. In each replication, a placebo treatment is randomly assigned, within each year from 2015 onwards, to as many ward-within-hospital-year observations as are treated in the actual sample, drawing only among wards that do not belong to the actual treatment group. The placebo post-treatment indicator switches on in the first year in which a ward receives the placebo treatment and remains on thereafter. In each replication, the model is re-estimated with the same specification, fixed effects and clustering of standard errors as in the baseline (Section 3), and the coefficient on the interaction between the placebo treatment and placebo post-treatment indicators is retained.



(a) MI – Mortality rate **(b)** MI – Deaths **(c)** MI – Admissions

(d) Stroke – Mortality rate **(e)** Stroke – Deaths **(f)** Stroke – Admissions

(g) Hip fracture – Mortality rate **(h)** Hip fracture – Deaths **(i)** Hip fracture – Admissions

(j) Cancer – Mortality rate **(k)** Cancer – Deaths **(l)** Cancer – Admissions

Figure 9: Randomisation tests: mortality, deaths and admissions

Note: The figure reports the placebo coefficients (dashed line), sorted in ascending order, and their 95% confidence intervals (shaded area) from 500 replications. In each replication, the sample is restricted to never-treated ward–LHA and a placebo adoption year is randomly assigned to a subset of them, keeping the share of treated units and the distribution of adoption cohorts as in the actual data; the remaining units serve as never-treated controls. The pooled ETWFE model is re-estimated with the same specification, fixed effects and clustering of standard errors as in the baseline (Section 3), and the simple ATT is retained.

5.2 Sensitivity to hospital exclusion

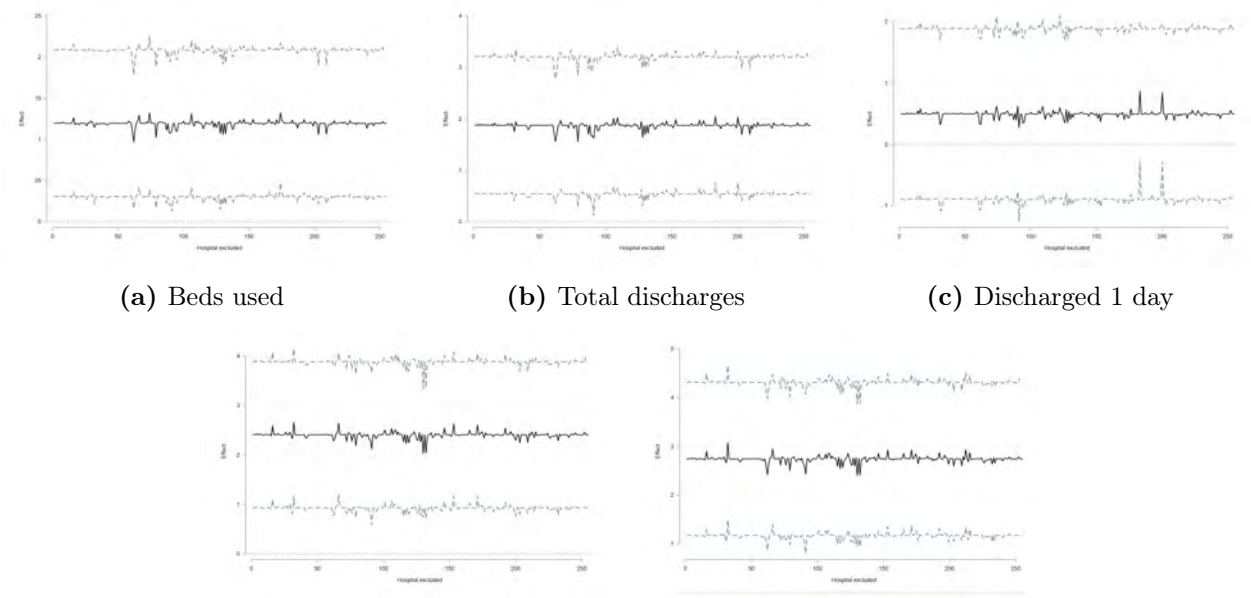
We next assess whether the baseline results are driven by individual hospitals with particularly large or influential treatment episodes. Figures 10–12 report leave-one-hospital-out sensitivity checks for the outcomes considered in the main analysis. For each outcome, we re-estimate the baseline specification repeatedly, excluding one hospital at a time. The solid line reports the resulting point estimate, while the dashed lines show the corresponding 95% confidence interval; the horizontal axis indexes the hospital excluded in each estimation. This exercise therefore allows us to assess both the stability of the magnitude of the estimated effects and their statistical precision when individual hospitals are removed from the sample.

Figure 10 shows that the activity results are highly stable across hospital exclusions. For utilized beds, total discharges, one-day discharges, surgical discharges, and surgical procedures, the point estimates display limited variation as individual hospitals are sequentially removed. The positive activity response documented in the baseline analysis is therefore not attributable to a particular hospital or to a small number of unusually influential downsizing episodes.

A similarly stable pattern emerges for the congestion outcomes in Figure 11. The estimated effects on both total inpatient and total pre-operative days remain positive across the leave-one-hospital-out specifications, with relatively limited variation in magnitude. The increase in congestion following downsizing therefore also reflects a systematic response across the hospital network rather than the experience of a single large provider.

Figure 12 extends the exercise to mortality. Because mortality is analysed at the network level, we exclude one LHA at a time rather than one hospital, and re-estimate the pooled average treatment effect reported in Table 4. Estimates for the acute, time-sensitive conditions remain on the positive side across hospital exclusions, whereas the comparison outcomes do not display a corresponding systematic adverse pattern. Statistical precision varies across the mortality outcomes, as in the baseline specification, but the direction and magnitude of the estimates are not driven by the inclusion of any single LHA. This is particularly relevant for the acute-condition results, where the sensitivity exercise indicates that the positive mortality pattern does not originate from an isolated LHA-level episode.

Overall, the leave-one-hospital-out exercise shows that the main findings are robust to the composition of the hospital sample. The increases in activity and congestion, as well as the differential mortality pattern across clinical pathways, remain qualitatively stable when individual hospitals are excluded in turn. This reduces the concern that the baseline estimates are driven by a small number of influential hospitals and supports the interpretation of the results as reflecting a broader response of local hospital networks to ward downsizing.

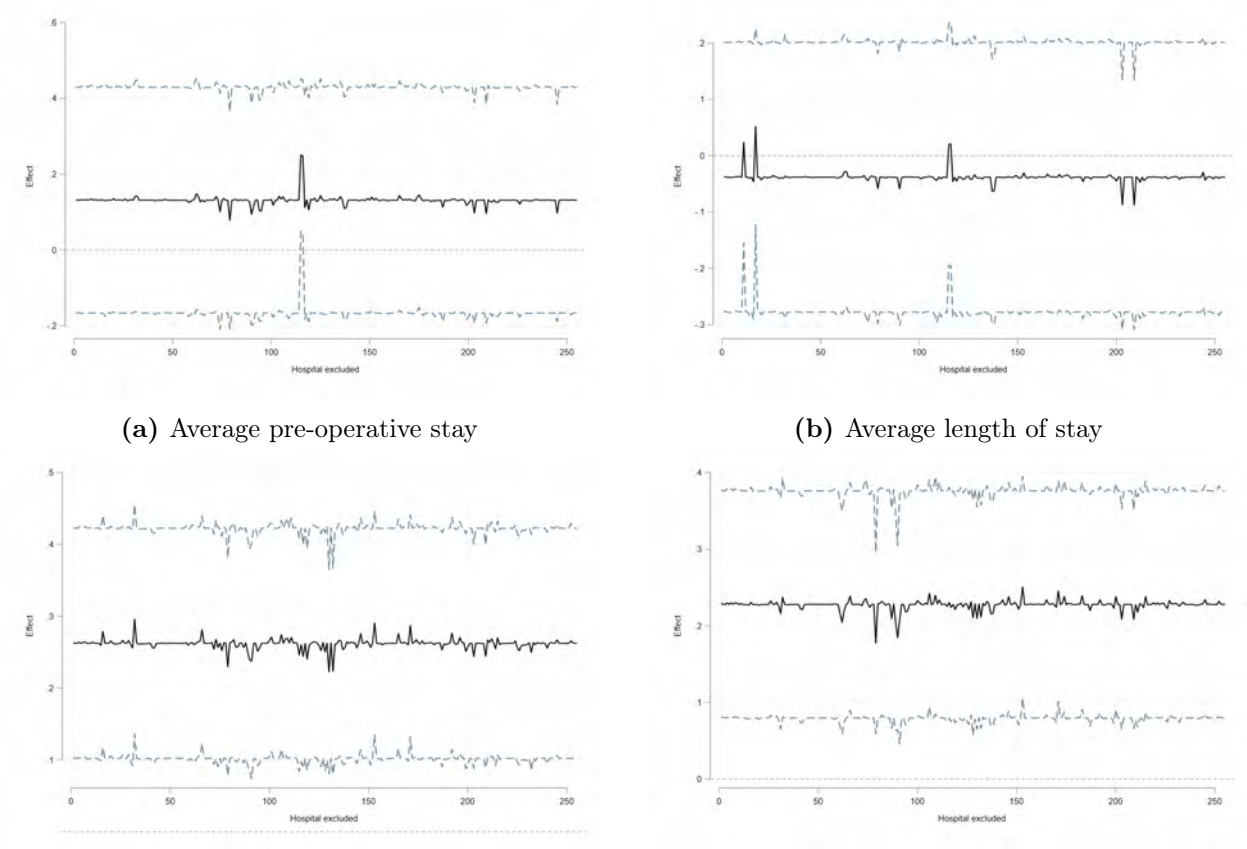


(a) Beds used **(b)** Total discharges **(c)** Discharged 1 day

(d) Surgical discharges **(e)** Surgical procedures

Figure 10: Leave-one-hospital-out estimates: volumes

Note: The figure reports the results of a leave-one-hospital-out exercise. Each point on the horizontal axis corresponds to a re-estimation of the baseline model in which all observations of one hospital are excluded from the sample, for a total of 255 replications, one for each hospital. The solid line reports the estimated coefficient on $Treat_{jhi} \times Post_{jhit}$ and the dashed lines the corresponding 95% confidence intervals. In each replication, the model is estimated with the same specification as in the main results (Table 2), including year, hospital, specialty-by-year and specialty-by-LHA fixed effects. Standard errors are clustered at the LHA level.



(a) Average pre-operative stay

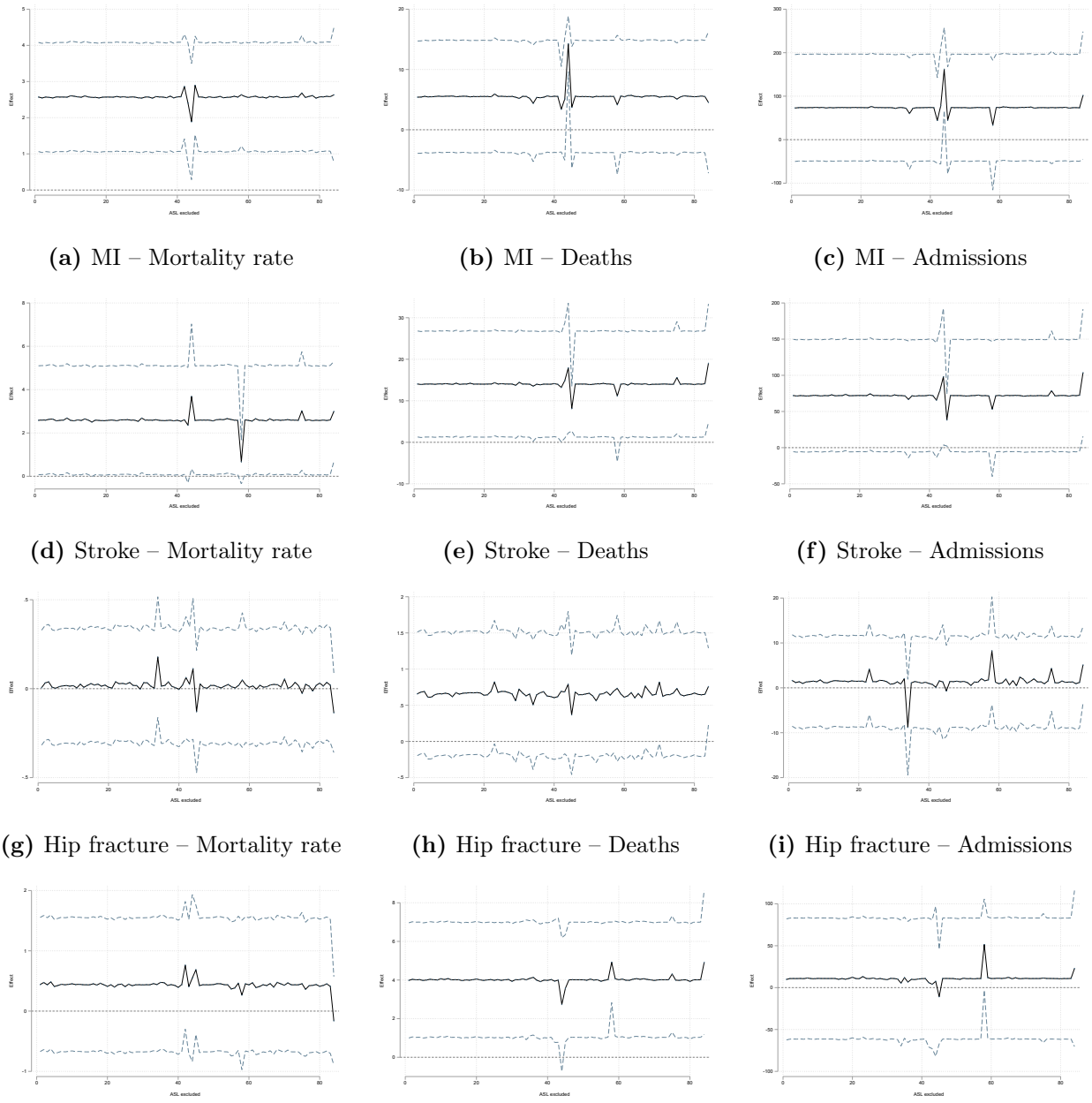
(b) Average length of stay

(c) Total pre-operative days

(d) Total inpatient days

Figure 11: Leave-one-hospital-out estimates: patient days

Note: The figure reports the results of a leave-one-hospital-out exercise. Each point on the horizontal axis corresponds to a re-estimation of the baseline model in which all observations of one hospital are excluded from the sample, for a total of 255 replications, one for each hospital. The solid line reports the estimated coefficient on $Treat_{jhi} \times Post_{jhit}$ and the dashed lines the corresponding 95% confidence intervals. In each replication, the model is estimated with the same specification as in the main results (Table 3), including year, hospital, specialty-by-year and specialty-by-LHA fixed effects. Standard errors are clustered at the LHA level.



(a) MI – Mortality rate **(b)** MI – Deaths **(c)** MI – Admissions

(d) Stroke – Mortality rate **(e)** Stroke – Deaths **(f)** Stroke – Admissions

(g) Hip fracture – Mortality rate **(h)** Hip fracture – Deaths **(i)** Hip fracture – Admissions

(j) Cancer – Mortality rate **(k)** Cancer – Deaths **(l)** Cancer – Admissions

Figure 12: Leave-one-LHA-out estimates: mortality, deaths and admissions

Note: The figure reports the results of a leave-one-out exercise applied to LHA. Each point on the horizontal axis corresponds to a re-estimation of the baseline model in which all observations of one LHA are excluded from the sample, for a total of 95 replications, one for each LHA. The solid line reports the estimated pooled ATT coefficient and the dashed lines the corresponding 95% confidence intervals. In each replication, the model is estimated with the same specification as in the event study results (Figures 5 and 6), including year and ward fixed effects. Standard errors are clustered at the LHA level.

6 Conclusions

This paper studies how hospital networks adjust to reductions in local capacity. We exploit the staggered timing and location of ward downsizing in Italy between 2015 and 2022 to examine how surviving same-specialty departments respond along three margins: activity, congestion, and patient outcomes. The setting allows us to study a central trade-off in the consolidation of spatially distributed public services. Reducing capacity may increase utilization of the remaining infrastructure, but concentrating demand on fewer providers may also generate congestion and impair access. Similar trade-offs arise in other public services, including schools and courts (Mocetti et al., 2026), but they are particularly consequential in healthcare, where delays in access can affect patient outcomes.

Our results document a clear sequence of adjustment. First, ward downsizing generates substantial reallocation of activity toward surviving same-specialty departments. Total discharges and surgical activity increase markedly, and utilized beds rise as well. The magnitude of these responses is larger than the modest or null changes often documented following hospital mergers (Gaynor et al., 2026). This difference is consistent with the nature of the shock we study. Ward downsizing directly removes specialty-specific capacity from a local hospital network, whereas mergers need not eliminate treatment capacity and may generate much more diffuse changes in patient allocation.

Second, higher activity is accompanied by greater congestion, in the sense of a heavier load on bed capacity. Total inpatient and pre-operative days increase by 25 and 30 percent, respectively, in line with the increase in discharges and surgical discharges, while average stays do not rise. Surviving departments therefore accommodate the supply shock by treating more patients with existing resources, and the additional caseload occupies substantially more bed-days, including before surgery. Part of this additional activity reflects the displacement of patients intended by the reform; part may also stem from the contraction of the specialist care that treated departments provide for chronic conditions, whose weaker management can translate into additional hospital use (Kozhimannil et al., 2018; Carrieri et al., 2026). Third, the mortality outcomes suggest that the consequences of this adjustment differ across clinical pathways. Admissions increase for all conditions, although significantly only for stroke, consistent with the displacement of patients toward surviving structures. In-hospital mortality, the outcome most informative about the quality of hospital care, rises by about 2.6 percentage points for both AMI and stroke, significantly at the 1 and 10 percent levels, respectively, and the event-study estimates suggest that these effects build up over time. Total stroke deaths also increase, pointing to a worsening of mortality indicators at the network level. By contrast, we find no effect on in-hospital mortality for hip fracture and cancer, although total cancer deaths increase. This pattern indicates that downsizing does not merely displace patients but also worsens outcomes for conditions in which timely and effective treatment is critical. It is consistent with previous evidence that reductions in access to hospital care can worsen outcomes for time-sensitive conditions. Ghislandi et al. (2025) document higher in-hospital AMI mortality following AMI ward closures in Italy, mainly through longer travel times, while Buchmueller et al. (2006) find higher mortality following increases in distance to hospital care in the United States. Our results point to a related but distinct mechanism, whereby adverse outcomes may arise even when patients remain within the local hospital network, because displaced demand increases pressure on the providers that remain in operation. Our larger estimates relative

to Ghislandi et al. (2025) are consistent with this additional congestion channel. The absence of a corresponding adverse pattern in mortality for the comparison conditions is consistent with evidence of limited health effects following closures in settings where short-run delays operate differently (Fischer et al., 2024).

The adjustment is also uneven across providers and regions. Public hospitals absorb the broad increase in activity and experience the associated rise in congestion, whereas private providers show no statistically significant response on any margin. This asymmetry is consistent with a broader literature showing that provider responses depend on the profitability and complexity of patients under prospective payment (Ellis and McGuire, 1986, 1990; Ellis, 1998) and that changes in incentives or ownership can alter the composition of patients treated across public and private providers (Duggan, 2000). Our setting shows that differential responses across provider types can also emerge following a capacity shock, even without a simultaneous change in ownership or reimbursement rules. The adjustment also differs across macro-areas. The activity response is most precisely estimated in the Centre, and the increase in total pre-operative days is statistically significant in the Centre and South but close to zero in the North.

Several robustness exercises support these findings. Randomisation inference shows that the observed activity and congestion effects are unusual relative to placebo treatment assignments. Leave-one-out estimates indicate that the main results are not driven by individual influential hospitals.

These results have implications for how hospital consolidation is evaluated. Higher utilization of surviving capacity is not sufficient to characterize the consequences of downsizing. The same reallocation that increases activity also raises the load on the departments that receive displaced patients. Evaluations based primarily on aggregate utilization or cost indicators may therefore miss an important dimension of adjustment, particularly when quality is difficult to contract upon or observe (Hart et al., 1997). Our results also suggest that the relevant capacity constraint is not simply the aggregate number of hospital beds. What matters is whether the departments receiving displaced patients have sufficient capacity to accommodate additional demand without creating bottlenecks along time-sensitive clinical pathways. Finally, the absence of a significant response among private providers indicates that private capacity need not substitute for public hospital capacity when the latter is reduced.

More generally, the evidence highlights a trade-off inherent in the consolidation of geographically distributed public services. Concentrating activity can increase the utilization of existing infrastructure, but the benefits of consolidation depend on the ability of the remaining network to absorb displaced demand. In hospital care, where the cost of delay differs sharply across patients and clinical pathways, preserving sufficient capacity in the parts of the network that receive this demand is therefore central to the design of consolidation policies.

Several avenues for future research emerge from our findings. First, patient-level data linking admissions, transfers, and outpatient contacts would allow a more direct examination of the pathways connecting local capacity reductions to treatment delays and mortality, for instance by tracing travel distances, time to treatment, and the hospitals to which displaced patients are ultimately admitted. Second, our results suggest that part of the increase in hospital activity may stem from

the contraction of specialist care for chronic conditions. Data on outpatient specialist visits and on hospitalizations for ambulatory care-sensitive conditions would help separate this channel from the pure displacement of patients intended by the reform. Third, in a regionally decentralized system such as the Italian one, capacity reductions may also induce patients to seek care outside their region of residence; studying inter-regional patient mobility would shed light on whether downsizing shifts demand, and its costs, across regional health systems. Fourth, little is known about how staff and other resources are reallocated following downsizing, and whether the reallocation of personnel to surviving departments mitigates the congestion we document. Finally, assessing how the public–private asymmetry documented here varies across reimbursement systems and institutional settings, and combining the activity, congestion, and outcome margins into a broader welfare evaluation, would help inform the design of consolidation policies.

References

Altonji, J. G., Huang, C.-I., and Taber, C. R. (2015). Estimating the cream skimming effect of school choice. *Journal of Political Economy*, 123(2):266–324.

Avdic, D. (2016). Improving efficiency or impairing access? Health care consolidation and quality of care: Evidence from emergency hospital closures in Sweden. *Journal of Health Economics*, 48:44–60.

Avdic, D., Lundborg, P., and Vikström, J. (2019). Estimating returns to hospital volume: Evidence from advanced cancer surgery. *Journal of Health Economics*, 63:81–99.

Avdic, D., Lundborg, P., and Vikström, J. (2024). Does health care consolidation harm patients? Evidence from maternity ward closures. *American Economic Journal: Economic Policy*, 16(1):160–189.

Balia, S., Brau, R., and Pau, S. (2026). Consolidation of local health authorities and mortality in the Italian NHS. *Journal of Health Economics*, page 103174.

Berta, P., De Fraja, G., and Verzillo, S. (2026). Optimal healthcare contracts: Theory and empirical evidence. *Journal of Economics & Management Strategy*. Forthcoming, doi:10.1111/jems.70035.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Buchmueller, T. C., Jacobson, M., and Wold, C. (2006). How far to the hospital? The effect of hospital closures on access to care. *Journal of Health Economics*, 25(4):740–761.

Canzian, G. and Ferrara, A. R. (2025). Judicial reform and banks credit risk exposure. *Oxford Bulletin of Economics and Statistics*, 87(2):414–447.

Carrieri, V., Gyrd-Hansen, D., and Li Donni, P. (2026). Dynamic healthcare demand with informative dropout: evidence from register data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, page qnag108.

Coube, M., Fontes, L. F., and Rocha, R. (2026). Public services under private management. *The Quarterly Journal of Economics*, 141(3):2597–2673.

Duggan, M., Gupta, A., Jackson, E., and Templeton, Z. S. (2023). The impact of privatization: Evidence from the hospital sector. NBER Working Paper 30824, National Bureau of Economic Research.

Duggan, M. G. (2000). Hospital ownership and public medical spending. *Quarterly Journal of Economics*, 115(4):1343–1373.

Duncombe, W. and Yinger, J. (2007). Does school district consolidation cut costs? *Education Finance and Policy*, 2(4):341–375.

Durrance, C. P., Guldi, M., and Schulkind, L. (2024). The effect of rural hospital closures on maternal and infant health. *Health Services Research*, 59(2):e14248.

Ellis, R. P. (1998). Creaming, skimping and dumping, provider competition on the intensive and extensive margins. *Journal of Health Economics*, 17(5):537–555.

Ellis, R. P. and McGuire, T. G. (1986). Provider behavior under prospective reimbursement. *Journal of Health Economics*, 5(2):129–151.

Ellis, R. P. and McGuire, T. G. (1990). Optimal payment systems for health services. *Journal of Health Economics*, 9(4):375–396.

Fischer, S., Royer, H., and White, C. (2024). Health care centralization: The health impacts of obstetric unit closures in the United States. *American Economic Journal: Applied Economics*, 16(3):113–141.

Francetic, I., Meacock, R., and Sutton, M. (2024). Free-for-all: Does crowding impact outcomes because hospital emergency departments do not prioritise effectively? *Journal of Health Economics*, 95:102881.

Fregoni, M. and Leonardi, M. (2025). A map of discontent: economic distress or public service deprivation? *Regional Studies*, 59(1):2577920.

Gaigné, C., Riou, S., and Thisse, J.-F. (2016). How to make the metropolitan area work? Neither big government, nor laissez-faire. *Journal of Public Economics*, 134:100–113.

Gaynor, M. and Anderson, G. (1995). Uncertain demand, the structure of hospital costs, and the cost of empty hospital beds. *Journal of Health Economics*, 14(3):291–317.

Gaynor, M., Sacarny, A., Sadun, R., Syverson, C., and Venkatesh, S. (2026). The anatomy of a hospital system merger: The patient did not respond well to treatment. *The Review of Economics and Statistics*, 108(1):272–281.

Ghislandi, S., Renner, A.-T., and Varghese, N. E. (2025). The impact of budget cuts on individual patient health: Causal evidence from hospital closures. *Journal of Health Economics*, 101:102975.

Gravelle, H., Moscelli, G., Santos, R., and Siciliani, L. (2024). Hospital closure in urban and rural areas and patients' welfare. *Regional Science and Urban Economics*, 109:104064.

Gujral, K. and Basu, A. (2019). Impact of rural and urban hospital closures on inpatient mortality. NBER Working Paper 26182, National Bureau of Economic Research.

Hart, O., Shleifer, A., and Vishny, R. W. (1997). The proper scope of government: Theory and an application to prisons. *Quarterly Journal of Economics*, 112(4):1127–1161.

Hoe, T. P. (2022). Does hospital crowding matter? Evidence from trauma and orthopedics in England. *American Economic Journal: Economic Policy*, 14(2):231–262.

Hoxby, C. M. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review*, 90(5):1209–1238.

Humlum, M. K. and Smith, N. (2015). Long-term effects of school size on students' outcomes. *Economics of Education Review*, 45:28–43.

Kozhimannil, K. B., Hung, P., Henning-Smith, C., Casey, M. M., and Prasad, S. (2018). Association between loss of hospital-based obstetric services and birth outcomes in rural counties in the united states. *Jama*, 319(12):1239–1247.

Lindrooth, R. C., Lo Sasso, A. T., and Bazzoli, G. J. (2003). The effect of urban hospital closure on markets. *Journal of Health Economics*, 22(5):691–712.

Lucchese, E. (2024). How important are delays in treatment for health outcomes? The case of ambulance response time and cardiovascular events. *Health Economics*, 33(4):652–673.

Man, S., Xian, Y., Holmes, D. N., Matsouaka, R. A., Saver, J. L., Smith, E. E., Bhatt, D. L., Schwamm, L. H., and Fonarow, G. C. (2020). Association between thrombolytic door-to-needle time and 1-year mortality and readmission in patients with acute ischemic stroke. *JAMA*, 323(21):2170–2184.

McNamara, R. L., Wang, Y., Herrin, J., Curtis, J. P., Bradley, E. H., Magid, D. J., Peterson, E. D., Blaney, M., Frederick, P. D., and Krumholz, H. M. (2006). Effect of door-to-balloon time on mortality in patients with ST-segment elevation myocardial infarction. *Journal of the American College of Cardiology*, 47(11):2180–2186.

Mocetti, S., Pesenti, O., and Roma, G. (2026). The effects of the reform of the judicial map on the functioning of civil justice. *Journal of Economic Behavior & Organization*, 248.

Oates, W. E. (1972). *Fiscal Federalism*. Harcourt Brace Jovanovich, New York.

OECD/European Commission (2024). *Health at a Glance: Europe 2024*. OECD Publishing, Paris.

Piacenza, M., Turati, G., and Vannoni, D. (2010). Restructuring hospital industry to control public health care expenditure: The role of input substitutability. *Economic Modelling*, 27(4):881–890.

Richards, M. R. and Whaley, C. M. (2026). The current era of health care consolidation. *Journal of Economic Perspectives*, 40(2):93–116.

Siverskog, J. and Henriksson, M. (2022). The health cost of reducing hospital bed capacity. *Social Science & Medicine*, 313:115399.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy*, 64(5):416–424.

Wooldridge, J. M. (2025). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Empirical Economics*, 69(5):2545–2587.