Funded by the European Union
Grant Agreement No 101095424

D6.10 Report with results of empirical analysis of impact of incentives on quality in Italy

(Version 1.0, 30-06-2025)

DELIVERABLE: D6.10 Report with results of empirical analysis of impact of incentives on quality in Italy
WORK PACKAGE: WP6. Incentives for efficiency and quality of health care
AUTHOR(S): Paola Bertoli, University of Verona (UNIVR), Emanuele Bracco (UNIVR), Clara Fertonani (UNIVR), Paolo Pertile (UNIVR)
DUE DATE: 30/06/2025
ACTUAL SUBMISSION DATE: 30/06/2025
DISSEMINATION LEVEL [x] PU: Public (must be available on the website)
GRANT AGREEMENT No: 101095424
PROJECT STARTING DATE: 01/01/2023
PROJECT DURATION: 48 months
COORDINATOR UNIVR - University of Verona

REVISION HISTORY			
Version	Date	Modified by	Comments

Contents

Glossary of terms

- DRG: diagnosis-related-groups
- GP: General Practitioner
- HHI: Herfindahl-Hirschman index
- LEA: Essential Levels of Care
- NHS: National Health Service
- PNE: Programma Nazionale Esiti
- PRS: Prospective Reimbursement System
- wNH: weighted number of hospitals

Executive Summary

This report, developed as part of the FLASH project under Work Package 6, delivers an empirical analysis of the relationship between the quality of hospital care and Diagnosis-Related Group (DRG) reimbursement tariffs in Italy. The research contributes to the broader discussion on how financial incentives in healthcare influence provider behavior. Italy's National Health Service (NHS) provides a particularly rich environment for such analysis, as its regional governments enjoy large organizational autonomy, including in tariff setting.

The main objective of the study is to assess whether and how variations in DRG tariffs affect the quality of care, with a particular focus on selected process and outcome indicators. This includes a detailed examination of the interplay between financial incentives and provider behavior, while also accounting for the role of market competition.

A key achievement of this research lies in the creation of two unique datasets. The first includes DRG tariffs from all Italian regions across time and clinical areas, covering the period from the introduction of DRGs in 1995 to the present. This dataset stands out for its comprehensiveness, overcoming challenges related to non-digitized and fragmented information. The second dataset contains quality indicators from the National Clinical Outcome Programme (PNE), administered by AGENAS, covering all accredited public and private hospitals since 2012. Together, these datasets enable novel, robust analyses linking reimbursement policies to clinical quality measures.

The report first analyzes the effects of public reporting and competition on healthcare quality, using the process indicator "percentage of femur fracture surgeries performed within two days" as a case study. Results show that publishing hospital performance data improved care quality. Importantly, the quality improvement associated with the release of information to the general public was significantly greater in hospitals operating in more competitive environments, highlighting the role of market dynamics in amplifying the effects of transparency.

To explore the relationship between DRG tariffs and quality, the study then shifts to analyzing indicators related to both femur fractures and deliveries. Two complementary empirical approaches are used: a fixed-effects panel model and a difference-in-differences design leveraging a national policy change that imposed a cap on DRG tariffs. Overall, our results suggest that higher DRG tariffs are associated with improvements in process indicators—such as prompt femur fracture surgeries and increased rates of vaginal deliveries after previous C-sections—but, not with improvements in outcome indicators like readmission rates.

These findings suggest that financial incentives do influence provider behavior, primarily in ways that are measurable and administratively straightforward. However, such changes do not necessarily lead to better health outcomes. Furthermore, the positive effects of higher tariffs on quality indicators are stronger in more competitive settings, reaffirming the importance of market context.

Some relevant policy implications can be drawn from our analysis. First, the release of information regarding the comparative performance of hospitals can lead to significant behavioral responses by providers as measured by process indicators. Moreover, the competitive setting can affect those behavioral responses, meaning that it is among the factors that should be considered in predicting the impact of quality related policies. Second, higher tariffs can lead to better performance as measured by process indicators, whereas we find generally no impact on outcome indicators. This suggests that it is crucial for policy-makers to select quality indicators that are closely associated with better outcomes, in order to avoid providing incentives that encourage providers to focus on activities with easily measurable consequences that have a weak or no association with patient outcomes.

1 Introduction

This report is part of the research activity of the FLASH project under Work Package 6, which is devoted to the analysis of the impact of several types of incentives both in primary care (Tasks 6.1-6.3) and hospital care (Tasks 6.4 and 6.5). As part of the research activity of Task 6.5, the main objective of this report is an empirical analysis of the relationship between quality of care and reimbursement (DRG) tariffs. In particular, the report aims to contribute to the existing literature and to the policy debate, by exploiting some unique features of the Italian NHS, namely the fact that Italian Regions, which are largely responsible for the organization of the public provision of health care within their territories, also enjoy significant autonomy in tariff setting. For the sake of our empirical investigation, this allows us to exploit heterogeneity in the level of tariffs set in different Italian regions to study the relationship between tariffs and quality, which is a long-lasting debate in the literature, with crucial policy implications. To answer our research questions, we combine two original dataset. The first dataset was created by collecting information on DRG tariffs set in all Italian regions since the DRG system was introduced in 1995. Although all DRG tariffs are public information, we are not aware of any other dataset sharing the same characteristics as our dataset in terms of coverage over space (regions), time and clinical areas. The main challenge in creating this dataset is that, although the relevant information is public, it is often not digitised, particularly for the early periods. Furthermore, the information required to compile such a dataset is highly fragmented.

The second dataset contains information on quality indicators as collected and published since 2012 by the Italian Healthcare Ministry through its agency AGENAS as part of the "National [Clinical] Outcome Programme" (Programma Nazionale Esiti, PNE). This program provides a number of clinical performance measures at the hospital level for benchmarking purposes. The first edition included hospital-level data from 2007-2012 with around 100 indicators on volumes, outcomes and care process. The program currently publishes almost 200 indicators annually and covers the universe of accredited public and private hospitals. Whenever it is appropriate, quality indicators are risk-adjusted. To the best of our knowledge, research conducted so far used single releases of the PNE data, each of which includes indicators for a limited number of years (see, e.g. Cavalieri et al., 2013).

By combining these two datasets, we can observe for each hospital in a given year all the quality indicators reported by the PNE for that hospital and the relevant tariff. While in some cases a DRG tariff can be easily associated with a quality indicator, in other cases the matching is more challenging, because several tariffs are potentially relevant for the same indicator. It is worth mentioning that variability in DRG tariffs exists not only between regions, but also within regions. This is because different tariffs may apply to different providers depending on characteristics that change from one region to another. This additional layer of heterogeneity further contributes to the potential of our empirical analysis.

The main focus of the present report is on the relationship between quality and DRG tariffs. Attention will also be paid to understanding the role of competition among providers as a factor potentially affecting quality and the response to changes in tariffs.

The remaining part of the deliverable is structured as follows. In Sections 1.1 and 1.2 we provide an overview of the related literature and of the relevant institutional setting, respectively. In Section 2, we consider quality and tariffs separately. In particular, in Section 2.1 we investigate the impact that the publication of the PNE indicators had on quality of healthcare provided. With the aim of identifying the behavioral response of providers to the release of quality related information, we focus on a process indicator: the fraction of femur fractions receiving surgery within two days. This indicator is among those receiving the greatest attention in the media at the time when quality indicators are published each year. Moreover, it is available for the large majority of hospitals, meaning that the analysis can be conducted on a large sample. An interesting characteristic of the dataset is that it allows us to investigate the potentially heterogeneous impact of the release of information to two different target populations: to health professionals in 2012 and to the public in 2016. Our analysis covers both the empirical analysis of the impact of the publication of the quality indicators at the two different levels and an analysis of how the behavioral response of hospitals to the publication of the indicators changes with the degree of competition faced by the hospital.

Section 2.3 describes the DRG tariff dataset and provides some descriptive statistics, with a focus on the analysis of the heterogeneity in tariff setting across regions. Section 3.1 exploits the combination of the two dataset mentioned above to investigate the relationship between quality and level of DRG tariffs. This is done by relying on panel data regressions for a group of indicators in Section 3.1, whereas an attempt to identify the causal impact of tariffs on quality is made in Section 3.2, by exploiting a plausibly exogenous change in the maximum tariff defined at the national level, which was implemented in 2013. In Section 3.3 the role played by competition in shaping the relationship between tariffs and quality is explored. Finally, in Section 3.4 we briefly discuss the relationship between tariffs, quality and waiting times.

1.1 Background literature

There is an extensive research on changes in provider reimbursement schemes focusing specifically on prospective reimbursement systems (PRS), which approaches the problem from a theoretical perspective (see for instance, Ellis, 1998a; Ellis and Mcguire, 1993; Hafsteinsdottir and Siciliani, 2010). Nevertheless, empirical evidence on this topic has not yet found equally fertile ground: findings are sparse, and many studies fail to adequately address endogeneity problems (Moreno-Serra and Wagstaff (2010); Messerle and Schreyögg (2023)).

According to Dafny (2005), studies on this topic could focus on two types of hospital behavior: real and nominal responses. The former refer to the substantial changes in healthcare outcomes that the payment system typically aims to affect. These include changes in the actual volume of admissions and the intensity of care provided by hospitals (e.g. adjustments in total costs, length of stay, number of surgical procedures, intensive-care-unit days, and in-hospital death rates). On the other hand, nominal responses identify instead hospital practices with the objective of inflating hospital reimbursement without affecting the actual quality or intensity of care provided. Examples in this category include upcoding or adjustments in the combination of different DRG caseloads (Liang, 2015). Upcoding refers to the practice of assigning patients to more remunerative DRGs. Jürges Köberlein (2015) classify upcoding into three different categories: legal, semi-legal and illegal. Legal practices involve more accurate coding, reducing instances of downcoding (e.g., including existing co-morbidities that increase treatment costs) to ensure appropriate hospital reimbursement. Semi-legal practices entail, for instance, swapping the primary and secondary diagnoses when co-morbidities are present. Illegal practices encompass instead manipulating medical documentation to make patients' condition appear worse. Differently, altering the combination of different DRG caseloads (as estimated by hospitals' case-mix index) consists in adjusting the hospital's output mix in order to maximize profits (Liang, 2015).

In order to review the literature on hospital DRG-based PRS, we build on the recent contribution by (Messerle and Schreyögg, 2023), and expand the classification they propose. According to their analysis, three distinct streams can be identified within this body of literature.

The first stream investigates hospitals' nominal and real responses to price-changes and reforms within established DRG-based PRS. These studies move from the consideration that, once the reimbursement system is fully implemented and demand for healthcare is considered to be exogenous, changing DRG-prices and weights is the only mechanism to affect hospital behavior. Moreover, reforms in the post DRG-based PRS introduction era allow to bypass the issue typical of data covering the period of transition from one reimbursement system to another, namely the difficulty of disentangling the effects of simultaneous changes in both average and marginal payment incentives (Verzulli et al., 2017). This literature was formally initiated by Dafny (2005), who exploited an exogenous reform involving a change in 43 per cent of Medicare DRG prices in 1988. More specifically, the policy altered the description of pairs of DRG codes that previously shared the same diagnosis but were distinguished by age and complications restrictions (i.e., “with complications or age over 69 years” and “without complications and age under 70 years”). With the reform, age qualifiers were removed, and complication criteria kept (i.e., “with complications” or “without complications”), along with an appropriate recalibration of DRG weights. Dafny (2005) exploits this refinement to identify price effect in relation to changes in caseloads and to upcoding. The latter was detected by creating a variable that represents the share of admissions to a specific DRG in a certain year assigned to the higher-weighted DRG code in that pair and regressing it on the difference between the DRG weights of the higher and lower codes within each pair (i.e., the “DRG spread”). Results show that hospitals upcoded more when the price increase was greater and that this behavior was stronger in for-profit hospitals rather than nonprofit or government-owned hospitals. Dafny’s empirical approach was applied by researchers to detect upcoding behavior in various countries, including Italy (Di Giacomo et al., 2017; Verzulli et al., 2017), Portugal (Barros and Braun, 2017) and Norway (Januleviciute et al., 2016; Anthun et al., 2017). Most of these studies have concluded that hospitals primarily respond to price changing by upcoding patients into diagnoses that are more lucrative due to higher reimbursement rates (Barros and Braun, 2017; Verzulli et al., 2017). However, whether hospitals also adjust their case mix to maximize profits remains debated (Bäuml, 2021). Verzulli et al. (2017) observed an increase in discharges for surgical DRGs, but not for medical ones, in line with the theoretical results by Hafsteinsdottir and Siciliani (2010). Conversely, Januleviciute et al. (2016) found an increase in the number of patients treated for medical DRGs, but no effect for surgical ones. Finally, Liang (2015), who employed a fixed-effect three-stage least squares on panel data from 268 hospitals in Taiwan, found an increase in orthopaedic department cases in response to DRG profitability.

The second stream revolves around the effects of DRG-based PRS introduction, using as level of observation individual hospitals, DRGs or population subgroups. This literature mainly focuses on real responses, though some studies report results on nominal responses as well (e.g., Shigeoka and Fushimi (2014); Jürges and Köberlein (2015)). Real responses are often investigated on a descriptive basis, and these studies often suffer from selection bias or endogeneity problems (Meng et al., 2020; Valentelyte et al., 2021). However, more rigorous analyses also exist. Some works have focused on vulnerable newborn (i.e. infants who have a low birth weight or who may need intensive care), since the unpredictability of newborns’ health status before delivery rules out endogeneity descending from selection of patients. Building on this, Shigeoka and Fushimi (2014) used insurance claim data for in-hospital births across 188 Japanese hospitals between April 2004 and December 2008 to test whether the introduction of a partial DRG-based PRS led to an increase in the length of stay in neonatal intensive care units (real response) and to hospital gaming in reporting birth weights (nominal response). Their difference-in-difference strategy relied on the fact that reimbursements sharply reared up at certain birthweights cutoffs and with longer hospitalizations. Similarly, Jürges and Köberlein (2015) found extensive upcoding in German neonatology, where birth weights were frequently shifted from slightly above DRG-relevant thresholds to just below them to inflate hospital reimbursement. They contribute to this literature by providing evidence not only of the influence of financial incentives on upcoding, but also of the link between patient health (assessed by indicators not involved in DRG coding) and upcoding. Indeed, hospital doctors and midwives are inclined to avoid upcoding newborns with low survival chances, and they tend to upcode more those who are expected to require higher medical expenses.

A key limitation of studies in this field is that they are unable to track admission-specific reimbursements across years because they only observe actual reimbursements for a single year (repeated cross-sections). This prevents comparisons from being made across different time periods (Bäuml (2021)).

The third stream of literature relates to the introduction of DRG-based PRS on a country-wide scale. The ultimate goal of this research is to find evidence of a causal link between the implementation of the reimbursement system and consequent trends in relevant outcomes using country-level panel data. The health outcomes considered in this type of works are real responses, as the empirical investigation generally concentrate on variables proxying either healthcare cost, the quality of hospital treatment, or both (Aragón et al. (2022)). Moreno-Serra and Wagstaff (2010) were pioneers in this field, identifying a key limitation of earlier research: the failure to account for broader healthcare system impacts beyond just the hospitals directly affected by payment reforms. To address this gap, they analyse the effects of reforms in hospital funding mechanism on healthcare output and mortality using country-level panel data of 28 Central and Eastern European and Central Asian countries (1990 - 2004). Employing a staggered difference-in-differences approach, they found that the introduction of DRG-based PRS led to increased healthcare expenditure and a decrease in length of stay. A similar study was conducted by Wubulihasimu et al. (2016), who considered the same outcomes, but with different geographical scopes and time spans. Indeed, focusing on OECD countries in the period 1980-2009, they implement a staggered difference-in-differences model allowing for "structural changes" in the level of healthcare output and mortality to account for the long-lasting effects of such reforms, showing that DRG-based PRS positively affect life expectancy. Nevertheless, their findings are dependent on model specifications.

Since recent advances in difference-in-differences methods reveal that this methodology is inappropriate in presence of heterogenous effects (e.g.,de Chaisemartin and D'Haultfœuille, 2020), more rigorous results in this literature are those presented by Aragón et al. (2022) and Messerle and Schreyögg (2023).Aragón et al. (2022) expanded the research by Farrar et al. (2009) regarding the impact of DRG-based PRS system introduction on length of stay in the British NHS. More specifically, they compare three quasi-experimental estimations: difference-in-difference, synthetic difference-in-differences (building a synthetic England using data from Scottish Health Boards) and interrupted time series. Similarly, Messerle and Schreyögg (2023) applied three quasi-experimental approaches (difference-in-difference, synthetic control and synthetic Diff-in-Diff) to analyse a comprehensive DRG-based payment scheme introduced in Germany in 2004. Overall, this stream of literature seems to suggest that the introduction of DRG-based payments led to a reduction in the average length of stay (Farrar et al., 2009; Aragón et al., 2022) and an increase in hospital activity (Messerle and Schreyögg, 2023).

1.2 Institutional setting

The Italian Healthcare System is a socialized system, free at the point of consumption for inpatient procedures and primary care services; small co-payments are charged for outpatient visits and diagnostic tests.

Italy's nineteen regions and two Autonomous Provinces are responsible for organizing the healthcare system in their territories, within a national legal framework.[1] The Central Govern-

[1] Italy is subdivided in 20 regions. One of these, Trentino-Alto Adige enjoys special autonomy, is subdivided into two separate Autonomous Provinces (Trento and Bolzano) each of which has the same powers of a region.

ment sets the "Essential Levels of Care" (LEA) to which each regional healthcare system has to comply, monitors compliance, defines the size of the "National Healthcare Fund" allocates it to the regions. Regional autonomy has brought some regions to run large healthcare spending deficits, which lead them into a "debt-recovery plan" procedure. When this procedure is triggered, spending and hiring are strongly limited and the regional healthcare system is run by a special administrator nominated by the central government.

Each region is usually divided into several local healthcare authorities (usually called Authorità Sanitaria Locale, ASL, but with different regions choosing different name variations), which are in charge of organizing primary and hospital care within their territories. Patients are typically referred by GPs for outpatient visits and clinical exams and then booked in by the hospital were the outpatient visit occurred in case they need inpatient procedures. Patients are free to choose in which hospital to seek specialist care or undergo inpatient procedures. In case of emergency, an ambulance would bring the patient to the nearest available hospital (given the situation at hand) according to the procedures of the local Regional Emergency Agency (AREU).

Most hospital are public, but there is also an important presence of private (both profit and non-profit) accredited hospitals, which are free at the point of consumption for patients. Public hospitals are often directly run by the relevant local healthcare authority, in which case purchase and provision are integrated. In other cases, public hospitals are autonomous entities (Aziende Ospedaliere – Hospital Trusts) providing services that are paid for by the relevant local healthcare authority. In this case providers and purchasers are separated. Public-hospital management is always directly or indirectly appointed by the regional government.

Private hospitals also play an important role in the system. Some of them are "accreditated", meaning that healthcare services are privately provided, but publicly financed. When patients are treated by these hospitals under the public regime, co-payments, if any, are the same as in public hospitals and those treatments are paid mainly by the relevant local health authorities. Finally, there are also fully-private hospitals, where treatments are fully paid by patients. Private out-of-pocket expenditure is mostly non-mediated by insurance, but directly paid by patients, even though private health insurance is currently becoming more common than it was in the past. Private insurance cannot replace but only complement public insurance.

The Italian NHS was significantly reformed in teh 1. In that period, the growing pressure to improve the efficiency of the hospital sector and the strong urge to contain public health expenditure, which had reached critical levels, called for a major reform of the Italian National Health System (*Sistema Sanitario Nazionale*, or *SSN*). Those reforms introduced three key elements in hospital settings: a higher degree of managerialism, quasi-market arrangements and a new reimbursement system for hospitals. (Taroni, 1997; Louis et al., 1999; Fattore and Torbica, 2006). Regions enjoyed significant autonomy in organizing their regional systems, including the definition of the relationship between public and private hospitals and the degree of separation between purchasers and providers.

The reimbursement system shifted from retrospective to mainly prospective, based on the DRG system. This new reimbursement system, intended to be adopted by each region between 1995 and 1997, was formally based on DRG with rates set at the national level. However, since hospital financing remained under the responsibility of regional authorities, regions were granted the freedom to set their own set of tariffs, as long as they were lower than those established nationally (Verzulli et al., 2017).

2 Quality and tariffs

2.1 Quality

The quality of health care is a central concept in the analysis of the effectiveness and efficiency of health care systems. With most healthcare systems undertaking significant reforms to enhance efficiency, quality has received growing attention, due to the possibility that a trade-off between costs and quality exists (Ellis and McGuire, 1986; Hodgkin and McGuire, 1994).

Starting approximately from the 1980s, the role of competition as a determinant of quality has been widely explored in the literature both theoretically (Ma, 1994a; Chalkley and Malcomson, 1998a) and empirically (Propper et al., 2009). More recently, the increasing availability of widely accessible information on many quality indicators of health services led to several contributions investigating the impact of their availability on quality. Although the observability and verifiability of quality can crucially affect the operation of competitive mechanisms (Dranove and Satterthwaite, 1992; Ellis, 1998b), with few exceptions, the roles of information availability and competition have been studied separately.

This section focuses on two main mechanisms through which quality-enhancement can be pursued, and the interaction between them: (1) the measurement and release of indicators of clinical quality (2) competition among providers. In particular, we investigate how different levels of diffusion of information about quality performance affect hospitals' quality-related behavior. The timing of information release for the PNE, which occurred first for health professionals in 2012 and then for the general public in 2016, allows us to explore this specific aspect. To the best of our knowledge, this is the first study to investigate how the level of diffusion of information interacts with competition in determining the impact of the availability of quality indicators on hospitals' performance.

2.2 The impact of reporting on quality

The implementation of PNE and the subsequent release of information on hospital quality may have affected hospital performance through a reputational mechanism (Hibbard et al., 2005) first channeled by healthcare professionals and policy makers, and later by patients. Hospital managers and doctors are generally aware of their own institution's performance irrespective of centralized release of quality indicators; they may also have imperfect information on other hospitals' performance through professional contacts. The implementation of the PNE allowed any physician to access high-quality information on every hospital and to compare and rank institutions, with potential repercussions on hospital reputation among professionals. This reputational effect may also have reverberated on patient choice through the advice of GPs, who have access to these data. The release of clinical quality indicators may have affected hospital reputation also among policy makers, as well as the reputation of policy makers themselves, by making quality information more readily available to politicians and top civil-servants. Such reputational effects are expected to be stronger when comparisons are more salient, e.g. in case of geographical proximity and lessened where hospitals have no obvious close-by comparable institution. Reputational concerns become even more severe once the totality of the information was released to the general public, allowing both potential patients and the media to make direct comparisons.[2]

Given these considerations, we expect the release of clinical quality indicators through the PNE to positively influence hospital performance. The impact of the information release may

[2]AGENAS publishes a yearly report based on the PNE data which receives significant press attention in both local and national outlets and is used by the press to draft "hospital rankings" for various common procedures.

by driven by several mechanisms, which can be different for different levels of dissemination of the information. In what follows, we pay particular attention to the role of the competitive mechanism.

The exact importance of each treatment effect and their interaction with competition in ultimately an empirical question, as it depends on the relative strength of different reputational and information channels and is the very object of our empirical analysis.

Finally, it is not ex-ante clear what would be the effect of information and competition on inequality of care and heterogeneity of provider's quality, as this depends on the interplay of our treatment with the ex-ante distribution of quality of care. Equity implications are also potentially relevant. For instance, it is interesting to investigate whether releasing quality-related information has led to wider or narrower gaps between the hospitals that were originally performing best and worst.

2.2.1 Data

To select our hospital quality indicator, we leverage data from the PNE itself for the period 2008-2020 for all public and private accredited hospitals focusing on surgery for hip fracture. Hip fracture surgery is an ideal case-study, because differences among hospitals in the implicit clinical threshold for hospital admission, and consequently systematic differences in unobserved patient severity, are virtually non-existing. In fact, hip fractures are a common clinical problem among older people, who are more prone than younger people to fall and are more likely to suffer from an important risk factor such as osteoporosis. They are usually life-altering events, being a leading cause of morbidity and reduced quality of life (Ballane et al., 2014; Johnell and Kanis, 2004). Hip fracture patients are almost always hospitalized and definitive treatment requires surgery. According to international guidelines, the operation ought to be performed within 48 hours after admissions (Klestil et al., 2018; Roberts et al., 2015).[3] Early surgery is associated with lower rates of perioperative and postoperative complications (eg., pneumonia, urinary tract infection, thrombosis, embolism, stroke) as well as with better postoperative functioning and a higher likelihood of returning to an independent lifestyle (Roberts et al., 2015; Seong et al., 2020).[4]

Given the crucial role of the timing of surgical treatment, we capture hospital quality in treating hip fractures through the risk-adjusted share of patients receiving surgery within two days of hospitalization, which is our primary outcome of interest.[5]

Patients with femur fractures are likely to reach hospital by ambulance, which will take them to the closest and most appropriate hospital (OrthoInfo, 2024). Still, even in this scenario, mobility is possible. More importantly, this very simple performance indicator has a potentially large reputational effect as it is the most important quality indicator of hospital performance in Orthopedics and Traumatology. AGENAS (the central healthcare agency in charge of PNE) and the Ministry of Health use it as benchmark to capture hospital quality in Orthopedics. The index is also used in news outlets to rank hospital Orthopedics and Traumatology departments.

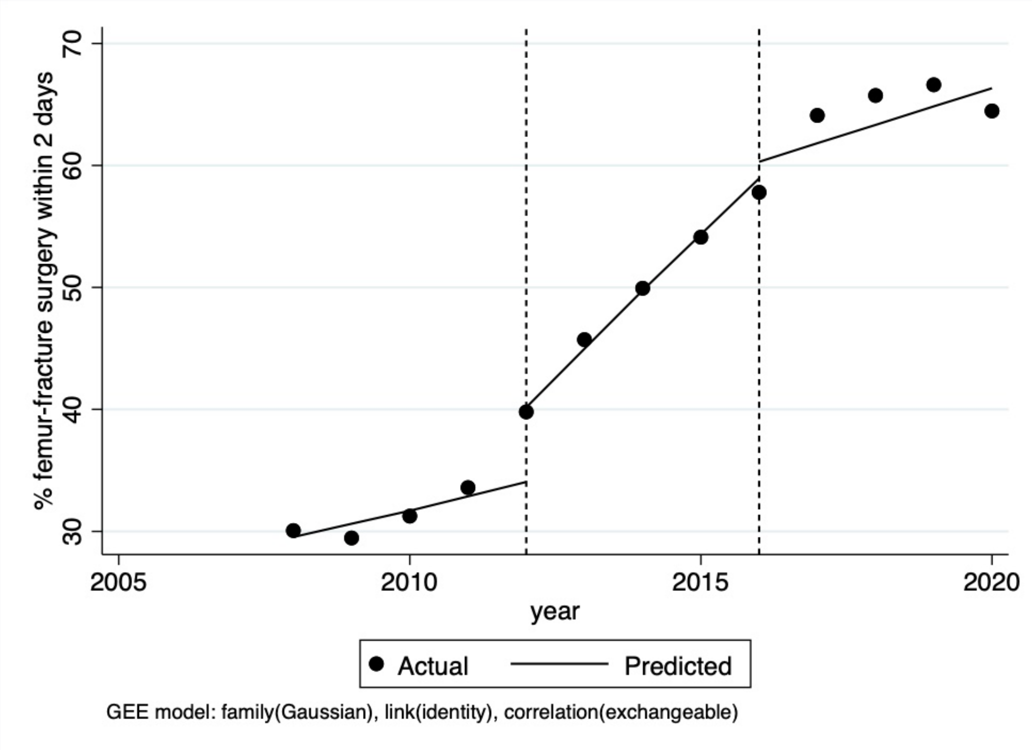
As additional control variables we include the share of elderly people within the relevant local health authority and a dummy variable indicating whether or not the region where the hospital

[3]Non-surgical treatment is reserved for debilitated patients in poor general health and is accompanied by higher in-hospital, 30-day, and 1-year mortality compared surgically treated patients (Kim et al., 2020).

[4]Systematic reviews have found that surgery within 48 hours is associated with reductions in one-month and one-year mortality (Roberts et al., 2015; Shiga et al., 2008; Seong et al., 2020), although not all studies support this view (Vidán et al., 2011; Khan et al., 2009).

[5]The protocol for calculating the quality index used in this paper excludes patients with important comorbidities (e.g. trauma patients, patients with malignant tumor in the previous two years) and risk-adjusts for other conditions (e.g. diabetes, Parkinson, dementia). Details are available here (in Italian).

Figure 1: Interrupted time series: Indicator and Information Treatments.



is located is under a "debt-recovery plan" in a given year.

We define two treatments by taking advantage of the different timing of access to PNE indicators for health professionals and the public.:

- *T1*: Indicator **treatment** (2012-2020). This treatment refers to the collection of information on specific clinical outcomes and the dissemination among health professionals.
- *T2*: Information **treatment** (2016-2020). This treatment refers to the availability of information to the public.

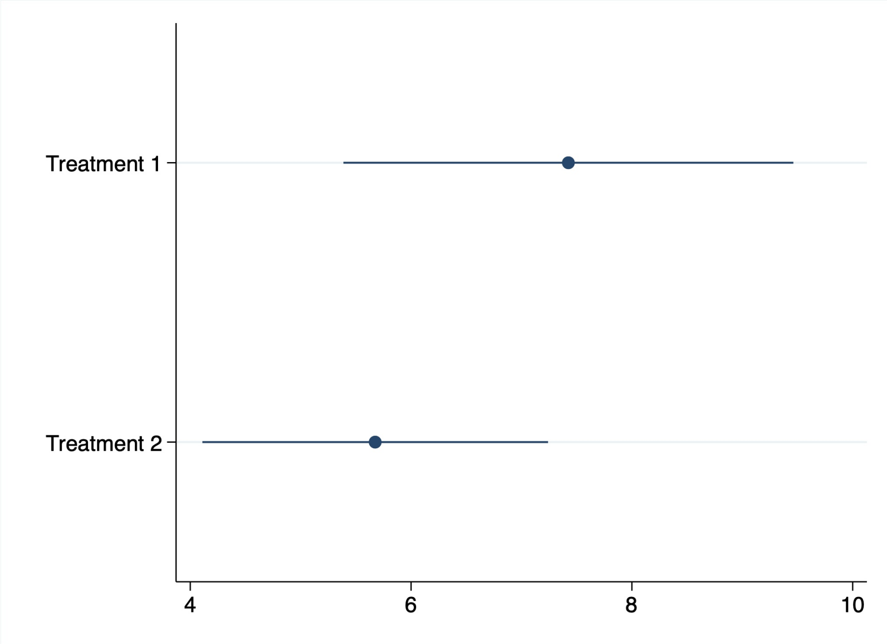
As a first step, we analyze the impact of the two treatments on our quality index through an interrrupted time series methodology. Figure 1 clearly shows that both stages of release of the information are associated with a jump in the performance level.

The result is similar if we regress our index of interest on the two treatment dummies, our control variables, macroarea time trends and hospital fixed effects.

As we can see from Figure 2, both treatments have a positive and significant effect. Treatment 1 is associated with an increase in the quality indicator by 7.4 percentage points, while Treatment 2 is associated with a further increase[6] of 5.7 percentage points.

[6]Note that the dummy variable associated with Treatment 1 is equal to one for all years starting from 2012, meaning that the coefficient of the dummy variable associated with Treatment 2 indicates an impact that adds to that of the first treatment.

Figure 2: Effect of Treatments on Quality



2.2.2 Quality reporting and competition

The second focus of this work is on the interaction between the availability of quality related information and competition, which has received limited attention in the literature so far. The two contributions closest to ours in terms of objectives are Chou et al. (2014) and Strumann et al. (2022). Chou et al. (2014) use data related to the publication of Coronary Artery Bypass Graft (CABG) report cards in Pennsylvania to test the hypothesis that the provision of quality-related information has a greater impact on quality in areas with higher levels of competition. The authors report an estimated reduction in mortality of 5% to 10% in more competitive areas, which is also associated with increased expenditure. Strumann et al. (2022) study the heterogeneous impact of the public reporting of indicators of quality of acute care implemented in Germany in 2008 on 30-day mortality due to stroke. They find a grater performance improvement in areas with higher competition, as measured by the Herfindahl-Hirschman index. An important difference between our contribution and these analyses is that our data allow us to estimate separately the impact of two different levels of diffusion of quality information, i.e. to health professionals and to the public.

Measuring the effect of competition on clinical outcomes is particularly challenging because of endogeneity and reverse causality: on the one hand, patients in busier hospitals may have worse clinical outcomes because of congestion; on the other hand a higher number of procedures per hospital (or per physician) may lead to better results. Also the case mix may be very different between (say) a rural hospital serving the whole population in the surrounding area, and city hospitals, which may specialize in a particular type of case mix.

To overcome this problems, first, we calculate our competition measure on data from the period before the release of information on hospital performance. Second, our preferred measure of competition is not affected by the distribution of procedures across hospitals, but relies only on the geographical distribution of the population.

Assuming that the population does not move within their geographical area systematically to be closer to specific hospitals, this should protect us from endogeneity. Third, our specification uses hospital fixed effects, so that our findings rely on within-hospital variation.

From the National Institute of Statistics, we collect population census data at the municipality level for all municipalities below 150,000 inhabitants and at the census-tract level for larger municipalities, together with the geolocation of each hospital. Using this information, we obtain the crow's flight distance from each location (municipality or census tract) to each hospital and between hospital pairs.

Our competition measure is calculated censoring the data to observations (hospitals × year) with at least 5 procedures, and our results are unchanged if other censoring thresholds are chosen.

For our baseline analysis, we calculate a competition index defined as the *weighted average number of accessible hospitals within a catchment area, wNH.* To this end, we define first the catchment area of a hospital as the area within a certain radius, corresponding to 20km for the baseline. Then, we associate each locality with the number of hospitals within reach (i.e. to how many different catchment areas the locality belongs to). Finally, the competition index of a given hospital is obtained as the population-weighted average of the number of hospitals within reach for all the localities within its catchment area. wNH can also be thought of as the number of hospitals accessible to the average person within a given hospital catchment area. This index is calculated annually, considering only active hospitals.

The design of our competition index is reminiscent of the approach many other authors: we define catchment areas around hospital (as in Propper et al., 2004; Guida et al., 2019; Strumann et al., 2022), we develop a count measure (as in Propper et al., 2004; Guida et al., 2019). As part of our robustness checks we use the Herfindahl-Hirschman index as an alternative measure of the intensity of competition (Gaynor et al., 2013; Guida et al., 2019; Strumann et al., 2022) and investigate the sensitivity of our results to changes in the radius relevant to the definition of the catchment area (see Section 2.2.5).

Table 1 shows summary statistics for the variables included in our main regressions, while Table 2 reports summary statistics competition indices used as robustness checks.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	N
Femur fracture: % surgery within 2 days	48.919	25.821	5830
Femur fracture: number of procedures	113.334	127.245	11974
Femur fracture: 30-day mortality	6.174	4.263	5809
Competition index: wNH 20-km straight	9.794	11.962	8258
Competition (median)	0.463	0.499	8258
CQ	2.527	1.101	10320
% elderly	21.745	2.848	16248
debt-recovery plan	0.364	0.481	16255

Our main empirical specification is the following:

$$\begin{aligned} y_{it} = \alpha + \delta C_i \quad &+ \quad \beta_1 T1_{it} + \gamma_1 T1_{it} * C_i \\ &+ \quad \beta_2 T2_{it} + \gamma_2 T2_{it} * C_i + \delta X_{it} + a_a * t + h_i + \epsilon_{it} \end{aligned} \tag{1}$$

where subscripts i and t indicate hospital and year, respectively, C is a measure of competition and X is a vector of controls. The specification also involves hospital fixed effects, h_i, to control for time-invariant unobservables at the hospital level. We also include area-specific linear time

Table 2: Summary statistics: Competition indices

Variable	Mean	Std. Dev.	N
wNH 20km straight [baseline]	9.794	11.962	8258
wNH 20km travel dist	6.576	9.210	8258
wNH 30 min travel time	7.41	8.881	8258
HHI 20km straight	0.44	0.358	7738
HHI2 20km straight	0.479	0.363	7683

trends.[7] Standard errors are clustered at the hospital level.

For the reasons already outlined we construct our competition measures using pre-treatment data (years 2008-2011). The values of these indexes are used to compute the value of C in equation 1. In particular, we consider two alternative definitions of C: (1) a dummy $COMP$ that takes value 1 in localities with a pre-treatment competition index above the median; (2) a vector of three dummies CQ_2, CQ_3, CQ_4, which take value 1 if the hospital belongs to the pre-treatments second, third or fourth competition quartile.

2.2.3 Descriptive evidence

From the early 2010s, hospitals underwent a dramatic improvement in their performance of our outcome of interest, with a substantial contraction of variability in the quality of care across hospitals. Figure 3 shows the time series of the performance index split by pre-treatment performance (panel [a]) and competition (panel [b]) quartiles: the fraction of femur fractures receiving surgery within two days increased from 30% in 2008 to 64% in 2020. Figure 3a shows that performance improved across the performance distribution, with the most significant gains observed in the previously lowest-performing ones. This panel highlights an interesting trend: pre-treatment, worst-performing hospitals were predominantly in the most competitive areas (yellow line), while those in the least competitive areas performed relatively well. By the end of our observation period, the fourth quartile is among those with the best performance. The same story can be gathered from panels (c) and (d), which show the performance kernel density plot by competition quartile pre- and post-treatments. Also in this graphs we can see how hospitals in the most competitive areas (yellow lines) had the worst pre-treatments and the best post-treatments performance outcomes. All panels of Figure 3 tell the same story: the most competitive areas saw the largest improvement in performance and the across-hospital heterogeneity in performance levels decreased over time.

Finally, Figure 4 shows the regional level geographic dispersion of performance in two chosen years, pre-treatments (2008) and post-treatments, just before the Covid pandemics struck (2019). Both in 2008 and 2019 one can observe a geographic gradient, with best performing regions in the Center-North and worse performing regions in the South. One can also notice that between 2008 and 2019 there has been an important improvement in performance across the whole country, which also shrunk the performance differential across regions.

2.2.4 Empirical results

Our main interest is on how the interaction between the two treatments and competition affects the share of femur-fracture patients receiving surgery within two days. Table 3 includes our main

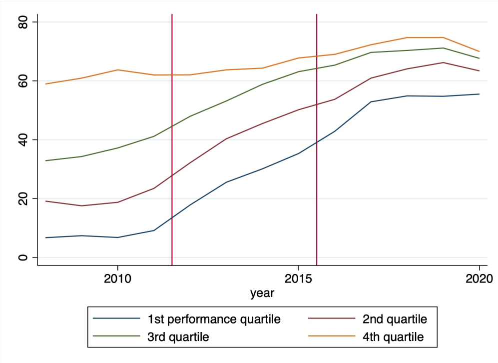
[7]The five statistical macro-areas are North-West, North-East, Center, South and Islands (i.e., Sicily and Sardinia). Results adding quadratic or logarithmic trends or using regional-level trands are qualitatively unchanged and available upon request.

Table 3: Main results: Effect of Treatment and Competition on hospital performance

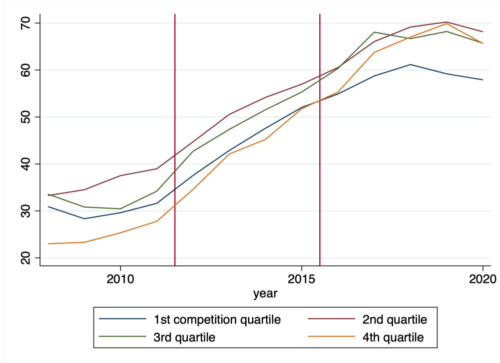
	(1) b/se	(2) b/se	(3) b/se
Treatment 1	7.425***	6.623***	5.021***
	(1.04)	(1.23)	(1.67)
T1 x COMP		2.741*	
		(1.64)	
T1 x CQ_2			1.860
			(2.25)
T1 x CQ_3			2.335
			(2.26)
T1 x CQ_4			5.552**
			(2.30)
Treatment 2	5.675***	3.601***	1.842
	(0.80)	(1.01)	(1.27)
T2 x COMP		5.288***	
		(1.45)	
T2 x CQ_2			3.524*
			(2.00)
T2 x CQ_3			3.472*
			(1.92)
T2 x CQ_4			9.724***
			(1.89)
Observations	5830	5781	5830
R^2	0.473	0.490	0.484

Note: panel data regression with hospital fixed effect and robust standard errors. Standard errors in parenthesis. Linear macro-area time trends and controls included in all regressions. Dependent variable is: share of femur fracture procedures withing two days of hospitalization.

Figure 3: Descriptive evidence: share of femur-fracture operated within 2 days.

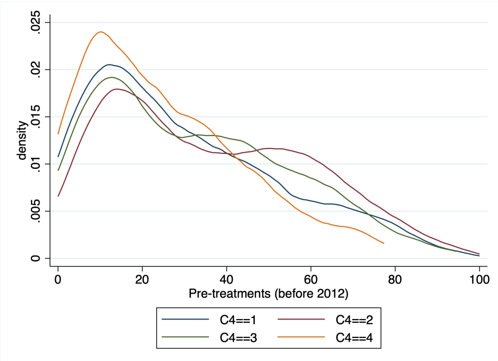


(a) Performance Quartiles.

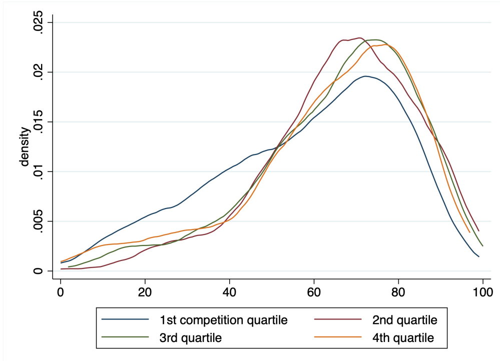


(b) Competition Quartiles

.



(c) Pre-treatments kernel performance distribution (until 2011).



(d) Post -treatments kernel performance distribution (from 2016).

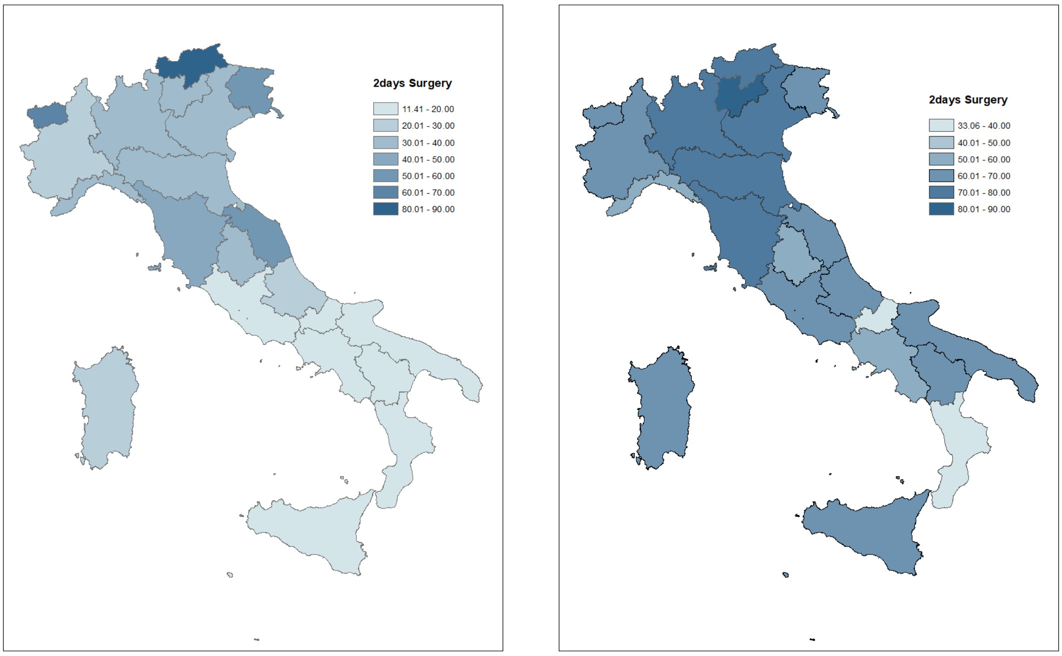
results.

Column (1) reports the same information that was discussed in commenting Figure 2: Treatment 1 is associated with a 7.4 percentage point increase in the share of patients receiving surgery within two days, while with Treatment 2 this effect increases by an extra 5.7 percentage points. In Columns 2 and 3 we introduce the competition variable(s), and interact them with treatment dummies.[8] Introducing competition allows us to tell a richer story. Even with these additional controls, the impact of $T1$ is still statistically significant, with performance improving by 5-6.7 percentage points depending on how we define competition. The impact of competition during this first phase, in which the indices are not available to the general public yet, is limited: when we define as competitive the hospital above the index median, we see a weak effect on competitive hospitals of 2.7 percentage points, only significant at the 10% level (Column 2); when we define competition in quartiles (Column 3) we can see that only hospitals in the most competitive quartile further improve their performance in the presence of competition.

Turning our attention to the effect of making information accessible to the public ($T2$) and its interplay with competition, the effect of the sole revelation of the information remains positive and significant in Column 2, but not in Column 3. The combined effect of competition and

[8] As our competition indices are time-invariant, we cannot separately include the competition variables as sole regressors.

Figure 4: Geographic distribution of hospital performance, 2008 and 2019.



(a) (b)

information is instead strongly significant: from Column 2 we can see that hospitals whose competition index is above the median gain an extra 5.3 percentage points in performance. When we split competition into quartiles, we can observe a gradient, with hospitals in the middle of the competition distribution showing a weakly significant improvement in their performance, while hospitals at the top of the competition distribution show a 9.7 percentage point increase in performance, with respect to those at the bottom of the competition distribution.

2.2.5 Heterogeneity Analysis

The results from the previous section highlight a significant impact of both treatments on performance and a significant role of competition in determining the size of the impact of the *information treatment*. A further question is whether these impacts widened or narrowed the gap between better and worse performing hospitals.

In Table 4 we perform a heterogeneity analysis, by splitting the sample according to the median value of the pre-treatments performance. We show the effect on low-performing hospitals in Columns 1 and 3, and on high-performing hospitals in Columns 2 and 4. All regressions include controls, time trends and hospital fixed effects. Columns 1 and 2 only include the two treatments as regressors of interest, whereas Columns 3 and 4 also include the competition (median) dummy and its interaction with the treatments.

The impact of T1 appears to be largely independent of the pre-treatments performance in all specifications of Table 4: the coefficient of Treatment 1 in Column 1 (Column 3) is not statistically different from that in Column 2 (Column 4). The impact of Treatment 2 is instead much weaker for hospitals with a better pre-treatments performance. The indicator increases by 9.7 percentage points for low-performing hospitals (Column 1) and by only 2.4 percentage

Table 4: Heterogeneity analysis by pre-treatment performance.

	(1) Low b/se	(2) High b/se	(3) Low b/se	(4) High b/se
Treatment 1	7.521***	6.934***	6.746***	6.630***
	(1.59)	(1.29)	(1.82)	(1.50)
T1 x COMP			1.411	1.925
			(2.27)	(2.20)
Treatment 2	9.676***	2.464**	9.168***	-0.530
	(1.22)	(0.98)	(1.60)	(1.15)
T2 x COMP			0.957	8.538***
			(2.30)	(1.63)
Observations	2779	2815	2768	2803
R^2	0.609	0.370	0.616	0.396

Note: panel data analysis with hospital fixed effect and robust standard errors. Dependent variable is: share of femur fracture procedures within two days of hospitalization. Sample split between hospitals with "low" and "high" pre-treatments performance (year 2011). Standard errors in parenthesis. Linear macro-area time trends and controls included in all regressions.

points for high-performing hospitals when treatments are not interacted with the competition variable; this difference is significant at the 1% level. This result also holds when interacting the treatment and competition dummies (Columns 3 and 4). Columns 3 and 4 also highlight that among hospitals with a better initial performance, the *information treatment* produces an impact only for those exposed to more competition. The estimated increase for this group is of 8.5 percentage points, whereas it is non-significant for hospitals in less competitive areas.

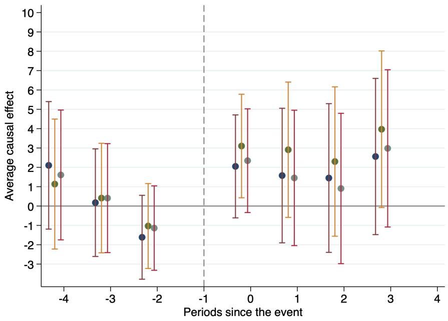
This table highlights a number of important points. First of all, collecting performance indicators (T1) affect all hospitals equally independently on performance or competitive environment. Secondly, better-performing hospitals are more sensitive to competition when information is shared with the general public (T2). Finally, sharing information to the general public reduces the variability in the quality of care, with worse-performing hospitals increasing their performance more than high-performance ones.

We also conduct two separate event-studies on Treatments 1 and 2. In the first event study, we drop from the dataset the years from 2016 onwards, so to include only years with no treatments or only Treatment 1 (event study for T1). In the second event study we perform the same analysis dropping years 2008-2011, comparing the situation with only Treatment 1 (years 2012-2015) with the situation when also Treatment 2 kicks in (years 2016-2020).

In Figures 5 we plot the yearly effect of the *indicator treatment* (T1) and competition on our clinical outcome of interest, while in Figure 6 we plot the yearly effect of the *information treatment* (T2) and competition on our clinical outcome of interest. These figures report results obtained using competition dummies computed using alternative distance measures (left-hand side panels) and competition indices (right-hand side panels). These results are obtained using our preferred specification, the competition (median) dummy, the full set of controls and hospital fixed effects.

From Figure 5 we can see that the combination of the indicator treatment (T1) with the competition dummy is typically positive, but not consistently significant. Conversely, in Figure 6 the interaction with the information treatment (T2) is positive and statistically significant

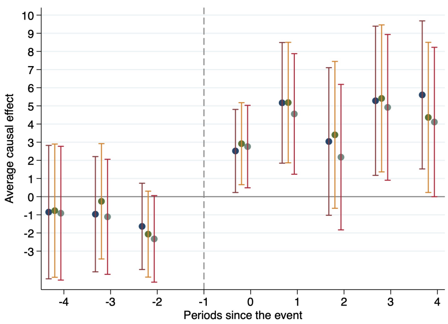
Figure 5: Event study. Combined effect of T1 and competition (median dummy) across different competition indices and distance measures.



(a) Different distances: 20km crow's flight (baseline), 30min drive, 20km drive

(b) Different competition indices: wNH (baseline), HHI, HHI2.

Figure 6: Event study. Combined effect of T2 and competition (median dummy) across different competition indices and distance measures.



(a) Different distances: 20km crow's flight (baseline), 30min drive, 20km drive

(b) Different competition indices: wNH (baseline), HHI, HHI2.

across all specifications. Overall, these results are fully consistent with those of our main analysis (Table 3). Point estimates are also comparable, with the combined effect of T2 and competition ranging between 3 and 6 percentage points with respect to hospitals in low-competition areas after information is released to the general public. These figures further confirm the robustness of our results across a number of measures of competitiveness.

2.3 Tariffs

Since the early 1980s, several countries have adopted prospective reimbursement systems (PRS) based on diagnosis related groups (DRGs) as primary source of hospital financing. The pioneering experience of the USA, which was the first to implement this funding mechanism in 1983, was gradually followed by many other European and non-European countries, albeit with adjustments tailored to the distinct aspects of each country's healthcare system (Messerle and Schreyögg, 2023).

The key features of DRG-based PRS, which have generally replaced either total cost reimbursement or fee-for-service hospital payment systems, are two:

- the classification of patients into economically homogeneous groups on the basis of information regarding patient diagnostic characteristics, such as primary diagnosis, secondary diagnosis, procedures, age or discharge status (Fattore and Torbica, 2006; Messerle and Schreyögg, 2023);
- the setting of a fixed tariff that hospitals will receive for treating patients within a specific DRG category (Aragón et al., 2022).

Hence, under a DRG-based PRS, hospital reimbursement is linked to the number of patients treated and discharged within each specific DRG, regardless of the actual costs incurred by the provider. Each DRG is assigned either a specific weight or a specific tariff (i.e., "DRG price"), reflecting the average cost of treating patients within that group (Dafny, 2005). This formula may be adjusted considering a hospital-specific rate, which takes into account aspects such as additional costs related to teaching, the proportion of indigent patients treated, the presence of an emergency department, or simply hospital ownership (Barbetta et al., 2007; Di Giacomo et al., 2017). In addition to that, the reimbursement amount can be adjusted to account for the complexity of the treated case and particularly long lengths of stay (Taroni, 1997).

The decentralisation of the Italian healthcare system, as discussed above, also applies to decisions relating to hospital reimbursement. This results in significant variation in DRG tariff levels across Italian regions. This heterogeneity makes Italy a particularly interesting case-study to explore the implications of changes in the level of DRG tariffs for potentially many outcomes. In this report the dimension of interest is quality of healthcare. To the best of our knowledge, up to now there has been no systematic collection of data on the regional tariffs implemented since the introduction of DRG-based PRS. Our work addresses this gap by compiling a comprehensive panel dataset of regional DRG tariffs from 1995 onward.

2.3.1 DRG tariffs in Italy

The Italian government released the first DRG tariff list with Ministerial Decree 169/1994 (*D.M. 15/04/1994*) and the reimbursement system was formally in place from January 1st, 1995. Since then, national interventions on the matter have been limited to the revision of the national list of tariffs (or of the national DRG-weights)[9] and the updating of the various versions of the DRG-Grouper [10]. Concerning the latter, Italy has adopted three of them so far: version 10 (1994 - 2005), version 19 (2006-2008).[11] and version 24 (2009 on)[12]

National tariffs serve as a benchmark, while regional governments are responsible for defining their own DRG rates. There are three main aspects in which the setting of DRG tariffs varies

[9] The national tariffs for acute hospital admissions have been defined and updated four times: (i) in 1994, with the D.M. 14/12/1994, with reference to the DRG 10.0 classification system; (ii) in 1997, with the D.M. 30/06/1997, still with reference to the DRG 10.0 classification system; (iii) in 2006, with the D.M. 12/09/2006 with reference to the DRG 10.0 classification system and (iv) in 2012 with the D.M. with reference to the 24.0 Grouper version.

[10] DRG Grouper versions are updates to the software systems used to classify hospital cases into DRGs for billing and reimbursement purposes. Each version incorporates changes in medical coding, treatment practices, and healthcare regulations, ensuring that classifications follow medical advancements and policy updates.

[11] 29 new DRGs were introduced; 15 DRGs were abolished; 2 DRGs changed their MDC or description with tariff adjustment and 19 DRGs undergone a change in their MDC, without tariff adjustment.

[12] 54 DRGs were added (those numbered from 524 to 579, except for 526 and 527, which were not included); 22 DRGs were abolished and 12 DRGs were modified in the description or MDC, but with no tariff adjustment.

significantly between regions: the criteria used for tariff determination, the adjustment of tariffs according to the different types of providers and the role that tariffs play according to the type of hospital.

Tariff determination: as previously mentioned, the Italian central government periodically releases DRG tariff lists that serve as a benchmark, which regions may choose to adopt or not. Alternatively, they can set their own rates. In this case, they can use two methods:

- the *"weight-method"*, which involves maintaining the national DRG weights and modifying only the DRG point value, or manually constructing the standard production costs for selected DRGs and then deriving the tariff for the remaining DRGs applying the relative weight system indicated by the national government.
- the *"analytical method"*, which is based on cost assessments from representative samples of hospitals within the reference territory. These costs are adjusted to exclude outpatient-related services, and each DRG is assigned a tariff through specific algorithms that consider the observed case mix.

Tariff differentiation: each region is free to differentiate tariffs across hospitals, regardless of whether national or regional tariffs are adopted. The criteria for differentiation typically include factors such as hospital ownership, hospital type, or the presence of an emergency department. Regions also enjoy broad autonomy in setting these criteria.

Role of tariffs: tariffs serve different purposes depending on the hospital market structure of each region. For accredited private hospitals and AOs, DRG-tariffs represent regional reimbursement for their inpatient care activities. In contrast, POs, which are directly controlled by ASLs, are reimbursed through DRG tariffs only for hospitalizations involving patients who are not residents of the ASL's catchment area (Table 5).

Table 5: Hospital type and funding rule in the Italian NHS

Ownership	**Hospital type**	**Funding**
Public	*Presidi Ospedalieri* (PO)	(1) Expenditure covered by the *Azienda Sanitaria Locale*'s (i.e., the local health autority) budget; (2) DRG-based PRS only for inpatient care towards non residents in the *ASL*
	Aziende ospedaliere (AO) (public trusts)	(1) DRG-based PRS ; (2) block assignments for specific health services
Private	Not-for-profit	Almost fully dependent on DRG-based PRS
	For-profit	

2.3.2 Description of the dataset

The dataset of the Italian DRG tariffs contains all the tariffs of the standard hospitalizations from 1995 to 2021 of all Italian regions. Data collection began at the end of January 2023. We started our research by reviewing all officially available documentation on the Italian DRG-based hospital reimbursement system. This documentation primarily consisted of reports written by working groups of doctors and economists, such as those from Assobiomedica (e.g., Mazzei, 2005) and the Observatory on Healthcare Organizations and Policies in Italy (OASI). Our first step was to review the national legislation on the subject and systematically collect the tariffs reported in the Ministerial Decrees. After completing the review of national legislation, we proceeded to analyze the regional regulations. When specific provisions were not available online, we contacted the relevant offices. We completed the reconstruction of the legislative process for each region and the collection of all tariffs by the end of June 2023.

Our dataset resembles the national DRG-tariff fee schedule. The latter, which was first defined by Article 2 of D.M. 15/04/1994, defines three tariff systems related to day hospital care, ordinary hospital admissions, admissions to rehabilitation and long-term post-acute care departments or hospitals. In our dataset, we focus exclusively on ordinary admissions in acute care wards (i.e., inpatient care).

Each row of the tariff schedule reports the DRG identification code and its description, the Major Diagnostic Category (MDC) it belongs to, and whether it is a medical (M) or surgical (C) DRG or neither. The difference between the latter lies in the presence of a surgical procedure or a less invasive medical treatment.

For each DRG, we collect information about:

- the tariff for "standard" admissions (*ricovero ordinario*) with a length of stay greater than one day, but less than the length-of-stay threshold for the relevant DRG (*valore soglia*).
- the tariff for one-day admissions of patients who neither died nor were transferred from other hospitals (*ricoveri di un giorno*), i.e., the tariffs associated with one-day admissions (which include both those of patients admitted and discharged on the same day and those of patients discharged the day after admission).
- the tariff for hospitalizations longer than a threshold length (*valore soglia*). In this case, the hospital reimbursement is computed by adding to the tariff for the "standard" admission this extra-payment for each day in excess of the threshold length of stay.
- the threshold value (*valore soglia*) for an hospitalization to be considered "with abnormal length of stay".

In the analyses conducted in the remaining part of this report, attention will be restricted to ordinary admission. Table 6 summarizes the main variables contained in the DRG dataset.

Table 6: Information available in the DRG dataset

Variable	**Description**
Id Region	Identification code of the region as assigned by ISTAT.
Region	Name of the region.
Year	Year the tariff refers to.
ID DRG	Identification code of the DRG according to an international numbering system.
MDC	Major Diagnostic Category the DRG refers to.
DRG Type	Type of the DRG (Surgical, medical or neither).
DRG Description	Description of the DRG.
Tariff	Tariff of standard hospitalization.

Since our dataset contains about 500 DRGs per region per year (and even more when regions differentiate between providers) for the 1995-2021 period, it is quite challenging to describe the dataset exhaustively. As a consequence, we present a series of descriptive statistics for a restricted sample of DRGs. Specifically, we selected one DRG for each of the main Major Diagnostic Categories (MDCs), choosing those with the highest admission volumes within each MDC and that were not significantly affected by changes in the DRG Grouper version. Descriptive statistics for the selected DRGs are reported in Table 7.

A comparison between the minimum and the maximum values of a given DRG makes apparent the great variation in the level of reimbursement and this is further confirmed by observing the standard deviations. To better dig into this heterogeneity and have a better understanding of the evolution of DRG tariffs over time, we produce as an example the graphical analysis of DRG 395 which concerns red blood cell disorders for patients age 17 or older.

Table 7: Selective sample of DRGs: Descriptive statistics

	Obs.	*Mean*	*Std. Dev.*	*Min*	*Max*
Degenerative Nervous System Diseases	1,319	3,397.073	785.755	1,938.01	6,117.61
Retina Procedures	1,319	2,345.346	677.860	1,053.33	4,489.17
Miscellaneous ENT Procedures	1,319	1,667.446	360.144	1,009.8	3,065
Pulmonary Edema and Respiratory Failure	1,319	3,161.887	606.274	1,670.61	5,086
Heart Failure and Shock	1,319	2,860.488	474.309	1,569	7,522.71
Esophagitis, gastroenteritis, and miscellaneous digestive disorders, age >17 without CC	1,319	1,340.087	362.822	652.12	3,864
Malignant Neoplasms of Hepatobiliary System or Pancreas	1,319	3,489.832	611.954	1,995.59	5,337.07
Femur neck fracture: surgery w/o CC	1,319	4,603.157	1,123.012	2,431	9,559
Subtotal Mastectomy for Malignant Neoplasms, w/o CC	1,319	2,297.722	495.197	1,285.96	4,425
Thyroid Procedures	1,319	2,582.462	564.308	1,292	4,249.8
Renal Failure	1,319	3,711.208	684.638	1,242.84	7,408
Transurethral Prostatectomy w/o CC	1,319	2,497.879	438.974	1,347.12	4,053
Uterus and Adnexa Procedures for Non-Malignant Neoplasms, w/o CC	1,319	2,578.338	524.062	1,410.05	5,516.28
Vaginal Delivery w/o CC	1,319	1,365.057	331.674	373	2,717
Newborns with Other Significant Conditions	1,319	1,380.485	751.671	490.66	10,419
Red Blood Cell Disorders, Age >17	1,319	2,472.685	767.353	1,139.68	4,721
Chemotherapy w/o Secondary Diagnosis of Acute Leukemia	1,319	1,784.506	397.364	1,002.96	4,307.25
Viral Illnesses and Fever of Unknown Origin, Age <18	1,319	1,645.735	459.711	901.69	6,810
Psychoses	1,319	2,690.951	741.232	1,320.57	5,659

Figure 7 presents the evolution of such DRG over time and by region and two relevant dynamics emerge. First, regions have diverged more and more in tariff levels over time. At the time of DRGs introduction there were already marked differences on the set levels with tariffs ranging from approximately 1,800 euros to slightly more than 4,000 euros. By the end of our observation period, these differences are even more pronounced with tariffs being within approximately 1,600 euros and 4,600 euros. Second, most regions substantially decreased their tariffs in 2013 after the issuing of the 2012 Ministerial Decree that updated the maximum national tariffs, pushing regions to revise their own DRG schedule (for more details on the Ministerial

Decree, see Section 3.2).

Figure 7: DRG 395 - Tariffs' Trend

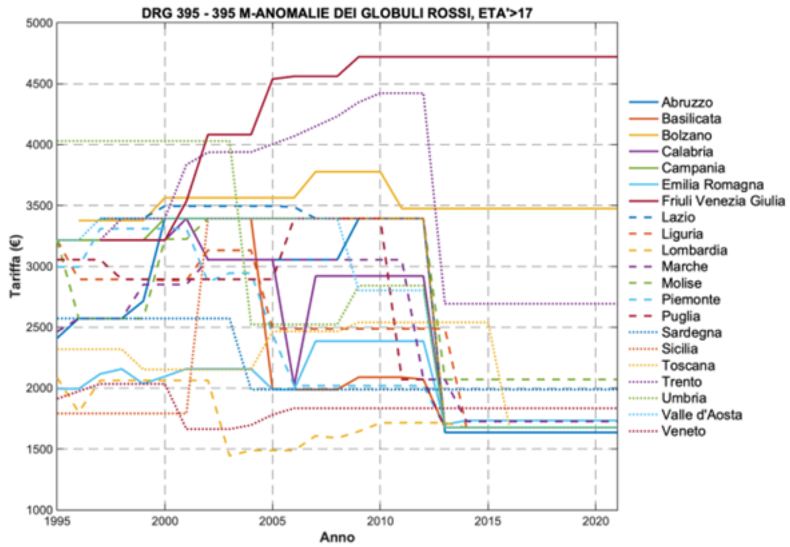
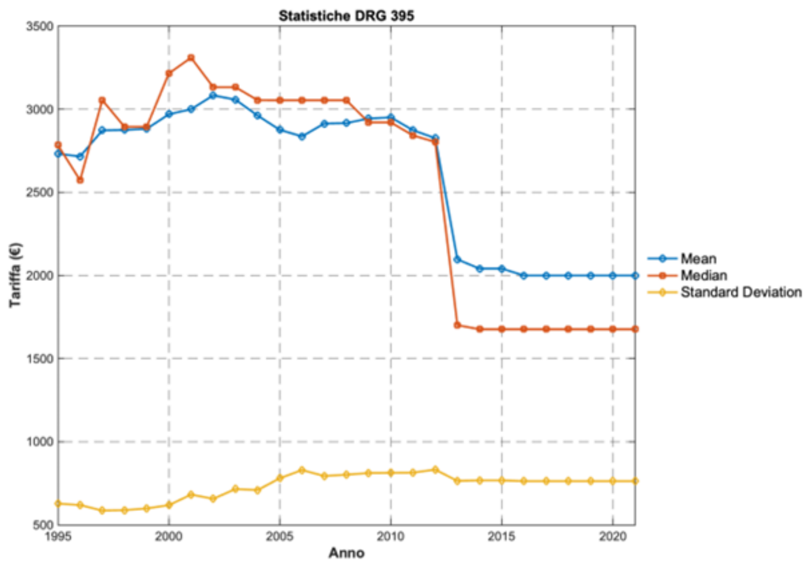


Figure 8: DRG 395 - Statistics' Trend



These dynamics are confirmed also when looking at the evolution of the mean, median and standard deviation over time. As depicted in Figure 8, the mean and the median exhibit substantially similar behaviors, except from 2013 onward, when after a considerable common drop the mean remains consistently higher. This is consistent with the fact that regardless of the reduction in tariffs the most regions implemented following the Ministerial Decree, some regions decided to keep much higher levels (i.e., Friuli Venezia Giulia and Bolzen). Finally, the trend of the standard deviation shows that the dispersion has tended to increase over time stabilizing in the last decade.

These dynamics are common to most DRG tariffs, regardless of the medical areas and of the fact of being surgical or medical DRGs, as shown by the same graphical analyses performed for additional DRGs in Appendix A.

3 The interaction between tariffs and quality

As an increasing number of countries moved away from retrospective payment to adopt prospective payment, the possibility that a trade-off between quality and (cost-)efficiency exists has been widely debated. A central issue in this debate is that healthcare quality is typically non-verifiable, according to the definition adopted in the contract theory. Ellis (1998b) identifies a number of departures for a hypothetical first-best potentially implied by the adoption of a system based on a fixed reimbursement per case (as opposed to cost-reimbursement): incentives not to treat high cost patients and / or over-treat low cost ones; upcoding; the provision of a sub-optimal level of quality of healthcare. This last effect is clearly the one that is most narrowly related to the objectives of the present report.

Of course, it is not only the choice between retrospective and prospective payment that matters, but also the level of payment. Conditional on using a DRG based payment system, an increase in the DRG tariff is expected to imply higher quality in equilibrium, although at the cost of increased expenditure. This is an intuitive result and it is also in line with the findings of Deliverable 6.6 (see, e.g., Eqs. 11 and 13). In this framework, competition can play a role, because an increase in the level of quality of care provided allows providers to attract more patients. Ma (1994b) identifies conditions under which a first-best outcome can be achieved in a competitive setting with unverifiable quality. However, competition alone is not sufficient to ensure that an efficient level of quality is achieved when quality of care is multi-dimensional (Chalkley and Malcomson, 1998b).

In Sections 3.1 and 3.2, we use the data described above to empirically investigate the role of DRG tariffs in affecting quality and appropriateness of the provision of healthcare.

3.1 Empirical analysis

To investigate the relationship between tariffs and quality, we focus on two clinical conditions: femur fractures and deliveries. This choice is justified by the fact that these clinical conditions offer several benefits for our empirical analyses.

First, both are characterized by low patient mobility. For instance, between 2001 and 2013, on average, only 3% of Italian mothers gave birth in a region other than their region of residence (Ministero della Salute (2006, 2012)). Similarly, as seen in Section 2.2.1, femur-fracture patients usually require ambulance transportation which means being directed to the closest and available hospital.

Second, inpatient treatment is the rule. This is evident in the case of femur fracture since its definitive treatment requires surgery. Still, it holds also in the case of deliveries which basically refer only to inpatient practices as the share of mothers giving birth outside the Italian healthcare system is negligible.[13] Hence, for both conditions under study, there are no concerns regarding patient selection bias among patients treated inside and outside hospitals.

Third, deliveries and femur fractures require specific medical interventions which are associated with unique DRG tariffs. Specifically, we work with three DRGs: DRG211 for femur neck facture surgeries, DRG371 for c-sections and DRG373 for vaginal deliveries.[14] These tariffs vary widely across regions and types of hospitals.[15] For instance, in 2000, a vaginal delivery in Tuscany was paid on average almost 30% more than in the nearby Emilia Romagna, while the amount paid in Tuscany was almost 40% less than that in nearby Umbria. However, Umbria, Tuscany, and Emilia Romagna have similar socio-economic characteristics and population health. Similarly, in 2015, in Tuscany a cesarean delivery was paid on average 36% more than in Emilia Romagna and 21% more than in Umbria. If we consider the surgery for femur fracture, in 2015 the same procedure of surgery for neck fracture in Emilia is paid about 10% more than in Tuscany, whereas in Emilia the tariff is paid 13% less than in Umbria. Moreover, economic incentives have proven to affect doctors and hospitals behavior in both obstetrics and orthopedics (eg., Barili et al. (2021a); Papanicolas and McGuire (2015); Gruber and Owings (1996)).

Finally, when quality is concerned, the PNE provides clear straightforward quality measures specific for deliveries and femur fractures. The successful treatment of femur fractures requires a prompt surgical intervention since this decreases the rate of both perioperative and postoperative complications (eg., pneumonia, embolism, stroke) and increases the likelihood of better postoperative functioning and of returning to an independent lifestyle (Roberts et al., 2015; Seong et al., 2020). In particular, the 48 hours following admission are internationally recognized as the crucial time-window to perform a timely operation (Klestil et al., 2018; Roberts et al., 2015). As a result, the timeliness of the intervention becomes the best proxy of healthcare quality in this field. Consistently, our first outcome of interest is the risk-adjusted share of patients receiving surgery within two days since hospitalization.

Differently, obstetric quality is commonly linked to two different, but related, aspects. On the one hand, c-sections are considered one of the most over, thus inappropriately, used procedures in developed countries and Italy is no exception.[16] Then, treatment selection (i.e., vaginal delivery vs. cesarean delivery) is seen as an important determinant of obstetric quality being a signal of appropriate clinical practices. On the other hand, quality is also associated with patients' health outcomes following childbirths[17]. Consistently, the PNE provides the following three indicators that complement our set of outcomes of interest:

- the risk-adjusted incidence of vaginal deliveries on women who previously experienced a c-section;

[13]In 2023, for instance, this share amounted to 0.13% (Ministero della Salute, 2023).

[14]For each procedure, there is also a specific DRG in case of complications. Specifically, DRG210 for femur fracture, DRG370 for c-sections and DRG372 for vaginal deliveries. Still, we do not consider them as their incidence is negligible. For instance, during 2001 and 2016, less than 3% of both cesarean and vaginal deliveries were reimbursed according to the DRG370 and DRG372 respectively.

[15]Such differences cannot be fully explained by differences in the costs of providing a procedure. For instance, the costs of certain inputs, such as personnel, do not vary across regions to an extent that would justify these differences. In fact, both Italian physicians and nurses are civil servants and are paid according to a so-called collective labor agreement (*Contratto collettivo nazionale - CCNL*) such as the CCNL 2002-2005 and the CCNL 2006-2009.

[16]Since 1985, the World Health Organization (WHO) has established a range of 10% -15% as an acceptable incidence of c-sections, and, later, *Healthy People 2010* confirmed this view by establishing a new target of 15% for the performance of c-sections in the US. Regardless of the decreasing trend, c-section incidence in Italy is well known for being above these figures and what is generally suggested by obstetric indications (Barili et al., 2021b; Bertoli and Grembi, 2019). In 2023, the average c-section incidence was still 30.3% (Ministero della Salute, 2023). This figure indicates a very high use of c-section that cannot be explained by the risk profile of mothers.

[17]The PNE does not provide the readmission rate in case of femur fracture.

- readmission rate after a c-section computed as the risk-adjusted share of c-sections followed within 42 days from the date of childbirth by at least one hospitalization with a stay of minimum 2 days;
- readmission rate after a vaginal delivery computed as the risk-adjusted share of vaginal deliveries followed within 42 days from the date of childbirth by at least one hospitalization with a stay of minimum 2 days.

These quality-related information at the hospital level from PNE is matched with information on relevant DRGs and with additional information on hospital characteristics, as explained above. Table 8 provides the main summary statistics of the resulting dataset which include 1273 hospitals.

Table 8: Quality indicators and DRG tariffs: Descriptive statistics

	Obs.	*Mean*	*Std. Dev.*	*Min*	*Max*
DRG Tariffs					
DRG211 - Femur neck fracture surgery	12,836	5,068.054	853.036	2,430.99	8,562.00
DRG371 - Cesarean	6,640	2,240.452	336.463	1,345.02	3,940.89
DRG373 - Vaginal	6,640	1,521.142	383.417	848.99	2,600.40
Outcomes: Quality indicators					
% of femur fracture surgeries within 2 days	6,285	49.923	25.829	0	100.00
Readmission rate after a c-section	5,653	0.855	0.794	0	15.4
Readmission rate after a vaginal delivery	5,599	0.532	0.475	0	9.37
% of vaginal deliveries on women with previous c-section	4,558	8.645	10.872	0	98.55

The first step of our empirical approach leverages the heterogeneous institutional setting provided by Italy and estimates the panel fixed effects model described by Equation (2)

$$Q_{it} = \alpha + \beta DRG_{it} + \gamma CASE_{it} + \pi_i + \delta_r + \eta_t + \epsilon_{it} \tag{2}$$

where Q_{it} refers to the quality indicator for hospital i at time t, $CASE_{it}$ is the number of procedures performed by hospital i in year t, π_i are hospital fixed effects, δ_r are regional fixed effects, η_t are year fixed effects and ϵ_{it} is the error term clustered at the hospital level. In this first model, we focus on the role of DRG tariffs by including them individually through DRG_{it} which coincides with the DRG tariff for a given procedure (i.e., c-section, vaginal delivery or femur fracture surgery) for hospital i at time t .

To ensure comparability between variables expressed on different scales, both the dependent variable, i.e. the quality indicator (expressed as a percentage), and the main independent variable, i.e. the DRG tariff specific to the clinical condition at stake, were standardized (z-score transformation). This standardization is relevant since DRG tariffs, which are expressed in euros, vary on a much larger scale than quality indicators. Hence, coefficients in Tables 9 and 10 should be interpreted as the effect of a one standard deviation increase in the tariff on the quality indicator, measured in standard deviations.

As shown in Table 9, a higher tariff for femur fracture surgery seems to lead to higher quality, since it is associated with an increased fraction of surgeries performed within 2 days. A higher reimbursement may allow hospitals to invest more in their orthopedic wards which in turn improve the timeliness of their interventions.

Table 9: Femur fracture

	% femur fracture surgeries within 2 days
	(1)
DRG211	0.160***
	(0.027)
Observations	6,282
R^2	0.6927
R^2 adjusted	0.6642

Notes: The model includes hospital FE, year FE and regional FE while also controlling for the yearly number of femur surgeries at the hospital level. SE clustered at the hospital level in parentheses.* ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$).

The analysis of obstetrics quality, however, provides less straightforward results. Table 10 shows the results for one process indicator -the incidence of vaginal deliveries on women with a previous c-section (Column 1) – and two outcome indicators – readmission rates after a vaginal delivery (Column 2) and c-section (Column 3). We find a positive and statistically significant relationship for the process indicator, no effect for readmission rates after a vaginal delivery, and a positive relationship for readmission rates after a c-section.

Overall, these results still weigh in favor of a link between quality and tariffs but they stress how a given reimbursement may be affecting quality differently depending on the quality dimension considered. The higher DRG tariff for vaginal deliveries, the more hospitals seem willing to opt for such a procedure even in a pool of risky patients (i.e., women with a previous c-section). Higher tariffs end up bringing an improvement in clinical practice. This finding is consistent with what we found for the process indicator considered for femur fractures.

The picture is less clear for outcome indicators. We find no impact on readmission rates following a vaginal delivery. This is not necessarily disappointing. Combining these two results together, one appreciates the fact that even if there are more potentially risky women giving birth naturally, there is no worsening of health outcomes, as measured by this indicator.

On the side of c-sections, a higher DRG tariff seems to lead to lower health outcomes as the related admission rate increases. However, we cannot identify the mechanism behind this increase. One plausible explanation may be that a higher tariff changes the incidence of c-sections, thus the risk profile of those who receive a c-section, in the first place. Hence, when different dimensions of quality are interrelated, the relationship with DRG tariffs may be more complex. In the case of obstetric practices, it would be important to examine these relationships by leveraging individual patient data, including information on newborns.

Table 10: Deliveries

	% of vaginal deliveries on women with previous c-section	Readmission rate after a vaginal delivery	Readmission rate after a c-section c-section
	(1)	(2)	(3)
DRG371 (C-section)			0.048*
			(0.028)
DRG373 (Vaginal)	0.091**	0.022	
	(0.040)	(0.026)	
Observations	4,557	5,895	5,652
R^2	0.8327	0.2139	0.2276
R^2 adjusted	0.8157	0.1323	0.1483

Notes: The model includes hospital FE, year FE and regional FE while also controlling for the yearly number of deliveries at the hospital level. SE clustered at the hospital level in parentheses. * ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$).

3.2 An analysis of the causal impact of tariffs on quality

The results presented in the previous section suggest that DRG tariffs are associated with the quality of healthcare provided at the hospital level. In this section, we further explore this relationship and attempt to identify the causal impact of tariffs on healthcare quality.

In order to do so, we exploit an exogenous change in hospital tariffs in a Difference-in-Differences framework. At the end of 2012, a new national set of DRG tariffs was introduced by Ministerial Decree 18/10/2012. The new tariffs implied a significant downward adjustment compared to the previous national schedule, resulting in estimated savings of around € 200 million (Ministero della Salute, 2013; Quotidiano Sanità, 2013).

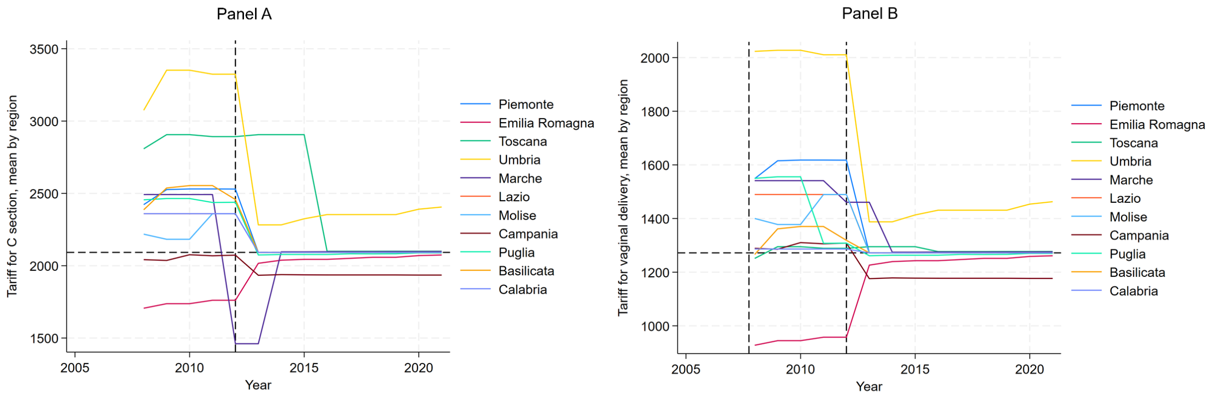
Our identification strategy builds on the fact that national tariffs define the maximum reference DRG rates for the reimbursement of inpatient care. Hence, regions whose tariffs were above the new national benchmarks were forced to revise their tariff rates downward, whereas regions where DRG tariffs were already below the national benchmark were not affected by the policy. However, exceptions were possible, because regions were allowed not to reduce their DRG rates below the new thresholds provided that the additional expenditure implied by this decision was funded with regional financial resources. To ensure a clear definition of the treatment, for this analysis we drop from the sample hospitals located in regions that had a pre-treatment tariff above the new thresholds and did not reduce it after the implementation of the policy (Abruzzo,[18] Liguria, Lombardy,[19] and Veneto). We also exclude from the sample regions with the special autonomous status, which face fewer constraints and are less tightly subjected to national policy [20]

[18]During those years, Abruzzo issued a series of measures modifying regional tariffs, but these were subsequently challenged by private hospital care providers and the Italian association of private hospitals, leading to their annulment by the Administrative Court (TAR) of L'Aquila. As such, the region represents a special case.

[19]Since 2005, Lombardy has implemented a regional policy that equalized the tariffs for cesarean and vaginal delivery. The aim was to promote clinical appropriateness and curb opportunistic behavior. However, the existence of this prior policy may confound our results.

[20]In Italy, regions with special autonomous status are Friuli Venezia Giulia, Sardegna, Sicily, Valle d'Aosta and Trentino-Alto Adige.

Figure 9: Average regional DRG Tariffs for deliveries



Compared to what we did in Section 3.1, we cannot adopt this approach for femur fractures, because DRG tariffs for femur surgery in all regions were already below the new threshold before the policy implementation. Hence, we will focus only on deliveries.

Figure 9 shows, in Panel A, the regional-level average tariff for c-sections in the sample of interest, whereas Panel B reports the regional-level average tariff for vaginal deliveries. After the policy implementation in 2012, we observe a sharp drop in the two tariffs for regions above the national tariff, whereas those regions below this level remain basically unaffected.

We estimate the model separately for cesarean and vaginal deliveries; since the DRG tariffs are different for the two procedures, this affects the assignment to treatment status based on pre-reform tariff levels. In both cases, we define as treated those hospitals whose reimbursement level for the respective procedure (i.e., the cesarean section tariff in the case of cesarean sections, and the vaginal delivery tariff in the case of vaginal deliveries) was above the corresponding national tariff at the time of policy implementation. Differently, control hospitals are those whose tariffs were already below the national threshold prior to the reform.

Our empirical strategy relies on a basic difference-in-differences approach as shown in Equation 3:

$$Q_{it} = \beta_0 + \beta_1 Post_t \times Treatment_i + \beta_2 Post_t + \gamma CASE_{it} + \pi_i + \delta_r + \eta_t + \epsilon_{it} \quad (3)$$

where Q_{it} denotes the quality indicator for hospital i at time t; $Post_t$ is a dummy equal to 1 for years after 2012; $CASE_{it}$ represents the number of deliveries at hospital i in year t. Hospital fixed effects (π_i), regional fixed effects (δ_r), and year fixed effects (η_t) are included. The error term ϵ_{it} is clustered at the hospital level. Our coefficient of interest is β_1, which captures the differential change in quality for treated hospitals with respect to control ones, after the national policy implementation.

Consistently with the previous empirical analysis, our outcomes of interest are the PNE indicators of obstetric quality considered in Section 3.1.

Results for the estimate of the model in Equation 3 using readmission rates after a c-section as dependent variable are reported in Table 11. Our results seem to suggest that the policy implementation had no effect on readmission rates after a c-section for treated hospitals with respect to controls. This result seems to diverge from the one found in Section 3.1 where the associative evidence suggested a positive relationship between the DRG tariff for c-sections and the quality indicator. This discrepancy between the two results may be due to what we previously hypothesized: the associative analysis may be partially driven by selection effects in the risk profile of those who receive a c-section. Once we focus on the exogenous change in tariffs, the relationship between the financial incentive and readmissions may be weaker than expected or mediated through channels not captured by this outcome alone.

Table 11: Difference-in-Differences estimate of the effect of the drop in tariff for cesarean delivery

	Readmission rate after a c-section
	(1)
Diff-in-Diff	-0.040
	(0.054)
Observations	3111
R^2	0.2453
R^2 Adjusted	0.1668

Notes: SE clustered at the hospital level in parentheses.
* ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$).

Table 12 reports the results for the indicators for which the treatment group is defined based on the reduction in the DRG tariff for vaginal deliveries. We find a statistically significant decrease in the proportion of vaginal deliveries on women with a previous c-section (-2.4 percentage points) and no effect on readmission rates after a vaginal delivery. These results are both consistent with the associative evidence presented in Section 3.1.

Concerning the process indicator, we previously registered a positive relationship with the level of the DRG tariff for vaginal delivery. Once the tariff for vaginal delivery is exogenously reduced, we observe a significant drop in the indicator, which should be interpreted as a reduction in quality.

Also consistent with the results of Table 10 is the lack of a statistically significant impact on readmissions following a vaginal delivery.

Table 12: Difference-in-Differences estimates of the effect of the drop of the tariff for vaginal delivery

	% of vaginal deliveries on women with previous c-section	**Readmission rate after a vaginal delivery**
	(1)	(2)
Diff-in-Diff	-2.413**	-0.024
	(0.990)	(0.044)
Observations	2588	2987
R^2	0.8185	0.1718
R^2 Adjusted	0.7998	0.0844

Notes: SE clustered at the hospital level in parentheses.
* ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$).

3.3 The role of competition

In Section 3.1, we studied the relationship between pricing and quality. In this setting, competition can play an important role, as better-performing hospitals in terms of quality may attract more patients. Therefore, in this part of the report, we aim to refine our results presented by analyzing how the interaction between DRG tariffs and quality is influenced by the competitive structure of the market.

While there is an extensive literature on the impact of competition on quality of healthcare provision, the evidence on the relationship between quality and DRG tariffs across different competitive settings is quite scarce. The motivation for hospitals to compete may depend on several factors, including their objectives (e.g, maximize the patients' benefit, altruism etc.) and the financial incentive to attract patients (see for instance Brekke and Straume (2017)). For this reason, the economic reasoning behind the relationship between pricing and quality in different competitive settings may not be straightforward. On the one hand, higher DRG tariffs may provide stronger incentives for hospitals to improve quality in order to attract patients. This incentive should be even stronger for hospitals operating in highly competitive areas. This would be consistent also with the predictions of some theoretical models, suggesting that competition and positive financial incentives reinforce both the altruistic and financial motivations of the hospital to improve quality (see, e.g., Brekke et al., 2021). On the other hand, this mechanism may not apply to all quality measures.

For our empirical analysis, we rely on the competition measures that were introduced in Section 2.1. In particular, we measure competition considering the *weighted average of accessible hospitals within a catchment area* (*wNH*) and the Herfindahl-Hirschman index (*HHI*). For both competitive indexes, we define a hospital's catchment area as the geographical span within a 20 km radius. This index is computed annually, considering only active hospitals. For each measure, we compute the yearly median and define a dummy variable that equals one if the given competition index is above the median level. Our empirical specification is the same as in Equation 2. In this case, we explore the heterogeneity of the impacts by running the regression separately for hospitals facing competition above and below the yearly median level.

Table 13 reports the results for the fraction of femur fractures performed within two days since hospitalization. In line with the results presented in Section 3.1, the coefficient on the tariff for femur fracture surgery is positive and statistically significant across all specifications. In addition to that, the coefficient is substantially larger for hospitals facing comparatively high competition; this result is consistent across both measures of competition. This is in line with the expectation that for hospitals facing more competition, the positive impact of higher tariffs on quality is stronger.

Table 13: % femur fracture surgeries within 2 days by level of competition

% femur fracture surgeries within 2 days				
	Panel A: Above-median competition		**Panel B:** Below-median competition	
	wNH	HHI	wNH	HHI
	(1)	(2)	(3)	(4)
DRG211	0.197***	0.177***	0.091**	0.124***
	(0.040)	(0.043)	(0.038)	(0.036)
Observations	2828	2500	3189	3506
R^2	0.7325	0.7483	0.6865	0.6881
R^2 Adjusted	0.7018	0.7181	0.6516	0.6538

Notes: SE clustered at the hospital level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14 shows the results for the proportion of vaginal deliveries on women with a previous c-section. Coefficients are all positive and statistically significant across all four specifications. As before, the impact tends to be greater for hospitals located in more competitive areas, although the difference is smaller in this case.

Table 14: % vaginal deliveries on women with previous c-section by level of competition

% vaginal deliveries on women with previous c section				
	Panel A: Above-median competition		**Panel B:** Below-median competition	
	wNH	HHI	wNH	HHI
	(1)	(2)	(3)	(4)
DRG373 (Vaginal)	0.129**	0.118***	0.100**	0.092*
	(0.057)	(0.045)	(0.050)	(0.054)
Observations	2409	2160	2038	2278
R^2	0.8367	0.8361	0.8461	0.8506
R^2 Adjusted	0.8181	0.8169	0.8267	0.8320

Notes: SE clustered at the hospital level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Concerning readmissions after childbirth, results are shown in Table 15 and in Table 16. For these indicators, the combination of the results from Tables 10, 11 and 12 suggested essentially no association with the levels of the relevant tariffs. This results still hold when we split the sample between more and less competitive areas. The only exception is the unexpected positive impact on readmissions after a vaginal deliveries of an increase in the tariff for hospitals exposed to more competition when competition is measured using the weighted average of accessible hospitals within a catchment area (*wNH*). Selection mechanisms not captured by our analysis could potentially explain this result.

Table 15: Readmissions after vaginal delivery by level of competition

	% of readmissions after a vaginal delivery			
	Panel A		**Panel B**	
	Above-median competition		Below-median competition	
	wNH	HHI	wNH	HHI
	(1)	(2)	(3)	(4)
DRG373 (Vaginal)	0.095**	0.051	-0.023	0.007
	(0.045)	(0.041)	(0.033)	(0.034)
Observations	2729	2335	2701	3082
R^2	0.2234	0.3060	0.2135	0.2062
R^2 Adjusted	0.1281	0.2213	0.1141	0.1097

Notes: SE clustered at the hospital level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: Readmissions after c section by Level of Competition

	% of readmissions after a c-section			
	Panel A		**Panel B**	
	Above-median competition		Below-median competition	
	wNH	HHI	wNH	HHI
	(1)	(2)	(3)	(4)
DRG371(C-section)	0.059	0.064	0.054	0.049
	(0.042)	(0.044)	(0.034)	(0.034)
Observations	2825	2426	2657	3037
R^2	0.2616	0.2698	0.2225	0.2173
R^2 Adjusted	0.1752	0.1836	0.1246	0.1215

Notes: SE clustered at the hospital level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.4 The role of waiting times

A theoretical analysis of the relationship between quality and waiting times under different hospital reimbursement rules was provided in Deliverable 6.6 of the FLASH project. That report presents a competitive model where hospitals set the quality level of their supply of services, knowing that quality enhancement allows them to attract more patients. Patients are heterogeneous in terms of disease severity, and more sever patients enjoy higher marginal benefits of quality. Patients also suffer from a utility loss associated with increased waiting times. The existence of waiting times is a results of the fact that hospital capacity is limited.

An increase in the level of quality provided by one hospital has two opposing effects. On the one hand, the direct implication of the quality improvement is an increase in the number of patients attracted. However, this increase leads to longer waiting times, which offsets at least part of the utility increase directly implied by the quality increase. Nash equilibria are characterised both for the case where providers are symmetric and asymmetric in terms of relative weight of patients' benefits and hospital net revenues in their objective function.

In line with the predictions from the existing theoretical literature, higher reimbursements are associated with higher quality levels in equilibrium. Moreover, comparatively severe patients are more likely to be treated in high quality hospitals, which also have longer waiting lists in equilibrium. Testing empirically the predictions for the aforementioned model would be highly valuable. However, this turns out not to be possible given the current data availability for waiting times in Italy. Due to the strong decentralisation of the Italian health system, there is no single, consistent data source for waiting times. Moreover, data published at the regional level are not hospital specific, which would prevent us from merging them with the data used for the empirical analysis presented in this report.

The problem of long waiting lists became apparent in Italy during the Covid-19 pandemic and immediately after it. Additional resources were made available from the central government to Regional Health Systems with the aim of making up for the delays accumulated during the pandemic. During this phase, the lack of a harmonized data base of waiting times, which is essential for the monitoring of the impact of the initiatives undertaken, became apparent. With the aim of filling this gap, a national platform (”Piattaforma nazionale delle liste di attesa (PNL)”) has been launched at the beginning of 2025, but has not yet been implemented. If the platform is successfully implemented, it should finally allow for comprehensive empirical analysis of the impact of policies aimed at reducing waiting times.

With reliable data at the hospital level, the combination of theoretical and empirical results could have important policy implications. For example, implications could be derived on criteria for the revision of the distribution of capacities across hospitals, with the aim of enhancing the overall efficiency of the health system. Hopefully, the collection of sufficiently granular, harmonized data on waiting times will make these analyses possible in the future.

In this framework, a natural question to ask is whether financial incentives related to DRG tariffs could be used to reduce waiting times. In this respect, a policy recently implemented by region Lombardy provides a particularly interesting case study. Lombardy has been very proactive in implementing measures to recover from the delays that accumulated during the pandemic. Since 2022, the region has expanded the financial resources available to address waiting lists (*Delibera della Giunta Regionale 2022/6002*), e.g. by expanding the number of days and hours of healthcare service provision (*Delibera della Giunta Regionale 2022/6279*).

Moreover, a new performance-based remuneration model linking DRG tariffs to waiting times was introduced for oncological surgical procedures.[21] Starting in April 2022, Lombardy introduced tariff cuts for hospitals that fail to meet waiting time targets (and rewards for those that do), based on the priority classes defined at the national level:

- Class A – hospitalization within 30 days, for clinical cases that may rapidly worsen or become emergencies.
- Class B – hospitalization within 60 days, for clinical cases involving intense pain or severe dysfunctions, including serious disabilities that are not likely to worsen quickly.
- Class C – hospitalization within 180 days, for cases with minimal pain, dysfunction, or disability, with no tendency to deteriorate rapidly.
- Class D – hospitalization within 365 days, covering clinical cases with no pain or dysfunction.

The penalties are clearly outlined in the resolution and involve progressively reduced reimbursement rates for services, based on the discrepancy between the expected and actual delivery times, as detailed in Table 17.

[21] The clinical areas affected by the policy are: colorectal, diagnostic and endoscopic surgery, cranio-maxillofacial surgery, otorhinolaryngology, melanoma-sarcoma surgery, plastic-reconstructive surgery, thoracic surgery, urology, gynecology, breast care, head and neck tumors, clinical immunotherapy for cancer and innovative therapies, inpatient radiotherapy

Table 17: Tariff cuts based on waiting times as implemented in Lombardy

	within time	exceeding by 1–9 days	exceeding by 10–19 days	exceeding by 20–29 days	exceeding by 30–39 days	exceeding by 40–49 days	exceeding by 50–59 days	exceeding by 60–89 days	exceeding by 90–119 days	exceeding by 120–179 days	exceeding by 180 days or more
Class A hospitalization within 30 days	100% DRG	100% DRG	95% DRG	90% DRG	85% DRG	80% DRG	75% DRG	50% DRG	50% DRG	50% DRG	50% DRG
Class B hospitalization within 60 days	100% DRG	100% DRG	95% DRG	90% DRG	85% DRG	80% DRG	75% DRG	50% DRG	50% DRG	50% DRG	50% DRG
Class C hospitalization within 180 days	100% DRG	100% DRG	95% DRG	95% DRG	90% DRG	90% DRG	90% DRG	85% DRG	80% DRG	75% DRG	50% DRG
Class D hospitalization within 365 days	100% DRG	100% DRG	95% DRG	95% DRG	90% DRG	90% DRG	90% DRG	85% DRG	80% DRG	75% DRG	50% DRG

Approximately four months after the introduction of this policy for oncological surgeries, in July 2022, the same performance mechanisms were extended to non-oncological surgical procedures and to outpatient specialist services.

The goal of these resolutions is to use monetary incentives to encourage hospitals to reduce waiting times, without increasing expenditure. The underlying idea seems to be that there are sources of technical inefficiency or suboptimal prioritisation that could be overcome to reduce waiting lists. Alternative approaches could consider making more resources available to reduce waiting lists. The data currently available do not allow us to estimate the impact of this policy. Additional insights on the impact of this policy could be gained when this is possible.

4 Conclusion

The present report investigates the relationship between quality and hospital care reimbursement. Understanding how providers respond to changes in the incentives related to the payment system is essential, for example, to understand to what extent reducing DRG tariffs with the aim of improving efficiency is a threat in terms of quality of care provision.

Our empirical analysis uses data from the Italian NHS. Italy is a particularly interesting case due to the fact that regions, which are largely responsible for the organization of the public provision of health care within their territories, also enjoy significant autonomy in setting DRG tariffs.

A major achievement is the creation of two large datasets, one including all DRG tariffs for all Italian regions and all providers since the introduction of DRG payment in Italy. The other dataset collects all the quality indicators published by Agenas under the PNE programme for all years since its start. For this report, we focused on a limited number of quality indicators relating to femur fractures and deliveries. However, merging the two datasets has huge potential for future research, enabling additional research questions to be explored in many more clinical areas.

Our analysis starts by focusing on quality per se, irrespective of its interaction with tariffs. We study the impact of two mechanisms which, according to the literature, could enhance quality of care – reporting of quality performance and competition – and the interaction among them. We use data on the fraction of femur fractures treated with surgery within two days to explore the behavioral response of providers to the publication of the performance of all Italian hospitals at two different levels: among health professionals in 2012 and for the general public in 2016. Both changes in the level of dissemination of the information had a positive impact on quality. However, the impact of making quality information available to the general public shows a clear gradient in relation to competition: only hospitals facing more competition significantly improve their performance in this case.

To study the interaction between tariffs and quality we consider quality indicators related to deliveries, in addition to the indicator related to femur fractures referred to above. Our empirical analysis employs two alternative approaches. First, we use a panel fixed effects model to investigate the relationship between quality of care and DRG tariffs. Then, we exploit a plausibly exogenous reduction in DRG levels due to the reduction of the reference tariff set at the national level which should provide a maximum threshold for regional tariffs. As this policy change only affected those regions with pre-policy tariffs above the new threshold, we define a treatment and a control group, which allows us to study the problem in a difference-in-differences framework. The results from these two alternative analyses are largely in line with each other. We find that higher tariffs tend to lead to better performance in terms of process indicators (femur fracture surgeries within 2 days, and % of vaginal deliveries on women with a previous c-section), whereas we tend to find no impact on outcome indicators (% of readmissions after a vaginal delivery, % of readmissions after a c-section). This suggests that providers respond to financial incentives related to changes in tariffs, e.g. by reallocating resources, but, at least for the set of indicators that we consider, this seems not to be necessarily translated into better patient outcomes.

Process indicators are also those for which competition seems to matter more. In particular, we find that the impact of an increase in the relevant tariff on the fraction of femur fractures receiving surgery within two days is significantly higher in hospitals exposed to more competition.

Some relevant policy implications can be drawn from our analyses. First, the release of information regarding the comparative performance of hospitals can lead to significant behavioral responses by providers as measured by process indicators. Moreover, the competitive setting can affect those behavioral responses, meaning that it is among the factors that should be considered in predicting the impact of quality related policies. Second, higher tariffs can lead to better performance as measured by process indicators, whereas we find generally no impact on outcome indicators. This suggests that the selection of quality indicators to consider from the perspective of policy-makers is crucial, as they should be narrowly associated with better outcomes so to avoid providing incentives that lead providers to exert great effort in activities with easily measurable consequences, but weak or no association with patient outcomes. These results should, of course, be interpreted with caution because they are based on a limited set of indicators. However, collecting the dataset described above will enable us to expand the scope of the present analyses in future research.

References

Anthun, K. S., Bjørngaard, J. H., and Magnussen, J. (2017). Economic incentives and diagnostic coding in a public health care system. *International Journal of Health Economics and Management*, 17(1):83–101.

Aragón, M. J., Chalkley, M., and Kreif, N. (2022). The long-run effects of diagnosis related group payment on hospital lengths of stay in a publicly funded health care system: Evidence from 15 years of micro data. *Health Economics*, 31(6):956–972.

Aragón, M. J., Chalkley, M., and Kreif, N. (2022). The long-run effects of diagnosis related group payment on hospital lengths of stay in a publicly funded health care system: Evidence from 15 years of micro data. *Health Economics (United Kingdom)*, 31(6):956–972.

Ballane, G., Cauley, J. A., Luckey, M. M., and Fuleihan, G. E.-H. (2014). Secular trends in hip fractures worldwide: opposing trends east versus west. *Journal of bone and mineral research*, 29(8):1745–1755.

Barbetta, G. P., Turati, G., and Zago, A. M. (2007). Behavioral differences between public and private not-for-profit hospitals in the italian national health service. *Health Economics*, 16(1):75–96.

Barili, E., Bertoli, P., and Grembi, V. (2021a). Fee equalization and appropriate health care. *Economics Human Biology*, 41:100981.

Barili, E., Bertoli, P., and Grembi, V. (2021b). Neighborhoods, networks, and delivery methods. *Journal of Health Economics*, 80:102513.

Barros, P. and Braun, G. (2017). Upcoding in a national health service: the evidence from portugal. *Health Economics (United Kingdom)*, 26(5):600–618.

Bertoli, P. and Grembi, V. (2019). Malpractice risk and medical treatment selection. *Journal of Public Economics*, 174:22–35.

Brekke, K. R., Canta, C., Siciliani, L., and Straume, O. R. (2021). Hospital competition in a national health service: Evidence from a patient choice reform. *Journal of Health Economics*, 79:102509.

Brekke, K. R. and Straume, O. R. (2017). Competition policy for health care provision in norway. *Health Policy*, 121(2):134–140.

Bäuml, M. (2021). How do hospitals respond to cross price incentives inherent in diagnosis-related groups systems? the importance of substitution in the market for sepsis conditions. *Health Economics (United Kingdom)*, 30(4):711–728.

Cavalieri, M., Gitto, L., and Guccio, C. (2013). Reimbursement systems and quality of hospital care: An empirical analysis for italy. *Health Policy*, 111:273–289.

Chalkley, M. and Malcomson, J. M. (1998a). Contracting for health services with unmonitored quality. *The Economic Journal*, 108(449):1093–1110.

Chalkley, M. and Malcomson, J. M. (1998b). Contracting for health services with unmonitored quality. *The Economic Journal*, 108(449):1093–1110.

Chou, S.-Y., Deily, M. E., Li, S., and Lu, Y. (2014). Competition and the impact of online hospital report cards. *Journal of Health Economics*, 34:42–58.

Dafny, L. S. (2005). How do hospitals respond to price changes? *American Economic Review*, 95(5):1527–1547.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Di Giacomo, M., Piacenza, M., Siciliani, L., and Turati, G. (2017). Do public hospitals respond to changes in drg price regulation? the case of birth deliveries in the italian nhs. *Health Economics (United Kingdom)*, 26:23–37.

Dranove, D. and Satterthwaite, M. (1992). Monopolistic competition when price and quality are imperfectly observable. *The RAND Journal of Economics*, 23(4):518–534.

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

Ellis, R. P. (1998b). 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. cost sharing and supply. *Journal of Health Economics*, 5(2):129–151.

Ellis, R. P. and Mcguire, T. G. (1993). Supply-side and demand-side cost sharing in health care. *Journal of Economic Perspectives*, 7.

Farrar, S., Yi, D., Sutton, M., Chalkley, M., Sussex, J., and Scott, A. (2009). Has payment by results affected the way that english hospitals provide care? difference-in-differences analysis. *BMJ (Online)*, 339:554–556.

Fattore, G. and Torbica, A. (2006). Inpatient reimbursement system in italy: How do tariffs relate to costs? *Health Care Management Science*, 9(3):251–258.

Gaynor, M., Moreno-Serra, R., and Propper, C. (2013). Death by market power: Reform, competition, and patient outcomes in the national health service. *American Economic Journal: Economic Policy*, 5(4):134–66.

Gruber, J. and Owings, M. (1996). Physician financial incentives and cesarean section delivery. *The RAND Journal of Economics*, 27(1):99–123.

Guida, P., Iacoviello, M., Passantino, A., and Scrutinio, D. (2019). Measures of hospital competition and their impact on early mortality for congestive heart failure, acute myocardial infarction and cardiac surgery. *International Journal for Quality in Health Care*, 31(8):598–605.

Hafsteinsdottir, E. J. G. and Siciliani, L. (2010). Drg prospective payment systems: Refine or not refine? *Health Economics*, 19(10):1226–1239.

Hibbard, J. H., Stockard, J., and Tusler, M. (2005). Hospital performance reports: impact on quality, market share, and reputation. *Health Affairs*, 24(4):1150–1160.

Hodgkin, D. and McGuire, T. (1994). Payment levels and hospital response to prospective payment. *Journal of Health Economics*, 13(1):1–29.

Januleviciute, J., Askildsen, J. E., Kaarboe, O., Siciliani, L., and Sutton, M. (2016). How do hospitals respond to price changes? evidence from norway. *Health Economics (United Kingdom)*, 25(5):620–636.

Johnell, O. and Kanis, J. (2004). An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. *Osteoporosis international*, 15:897–902.

Jürges, H. and Köberlein, J. (2015). What explains drg upcoding in neonatology? the roles of financial incentives and infant health. *Journal of Health Economics*, 43:13–26.

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

Kim, S.-J., Park, H.-S., and Lee, D.-W. (2020). Outcome of nonoperative treatment for hip fractures in elderly patients: a systematic review of recent literature. *Journal of Orthopaedic Surgery*, 28(2):2309499020936848.

Klestil, T., Röder, C., Stotter, C., Winkler, B., Nehrer, S., Lutz, M., Klerings, I., Wagner, G., Gartlehner, G., and Nussbaumer-Streit, B. (2018). Impact of timing of surgery in elderly hip fracture patients: a systematic review and meta-analysis. *Sci. Rep.*, 8(1):13933.

Liang, L. L. (2015). Do diagnosis-related group-based payments incentivise hospitals to adjust output mix? *Health Economics (United Kingdom)*, 24(4):454–469.

Louis, D. Z., Yuen, E., Braga, M., Cicchetti, A., Rabinowitz, C., Laine, C., and Gonnella, a. S. (1999). Impact of a drg-based hospital financing system on quality and outcomes of care in italy. *Health Services Research*, 34(1 Pt 2):405–415.

Ma, C.-t. A. (1994a). Health care payment systems: Cost and quality incentives. *Journal of Economics & Management Strategy*, 3(1):93–112.

Ma, C.-t. A. (1994b). Health care payment systems: cost and quality incentives. *Journal of Economics & Management Strategy*, 3(1):93–112.

Mazzei, L. (2005). I sistemi tariffari per le prestazioni di assistenza ospedaliera: Un esame della normativa nazionale e regionale. primo aggiornamento.

Meng, Z., Hui, W., Cai, Y., Liu, J., and Wu, H. (2020). The effects of drgs-based payment compared with cost-based payment on inpatient healthcare utilization: A systematic review and meta-analysis. *Health Policy*, 124(4):359–367.

Messerle, R. and Schreyögg, J. (2023). Country-level effects of diagnosis-related groups: evidence from germany's comprehensive reform of hospital payments. *European Journal of Health Economics*.

Ministero della Salute (2006). Certificato di assistenza al parto. analisi dell'evento nascita - anno 2006. `https://www.salute.gov.it/new/it/pubblicazione/certificato-di-assistenza-al-parto-cedap-analisi-dellevento-nascita-anno-2006/`. Accessed: 2025-06-20.

Ministero della Salute (2012). Certificato di assistenza al parto. analisi dell'evento nascita - anno 2012. `https://www.salute.gov.it/new/it/pubblicazione/certificato-di-assistenza-al-parto-cedap-analisi-dellevento-nascita-anno-2012/`. Accessed: 2025-06-20.

Ministero della Salute (2013). Relazione tecnica. PDF document.

Ministero della Salute (2023). Certificato di assistenza al parto. analisi dell'evento nascita - anno 2023. `https://www.salute.gov.it/new/it/scheda-statistica/certificato-di-assistenza-al-parto-cedap-analisi-dellevento-nascita/`. Accessed: 2025-06-20.

Moreno-Serra, R. and Wagstaff, A. (2010). System-wide impacts of hospital payment reforms: Evidence from central and eastern europe and central asia. *Journal of Health Economics*, 29(4):585–602.

OrthoInfo (2024). Hip fractures. `https://orthoinfo.aaos.org/en/diseases--conditions/hip-fractures/`. Accessed: 2024-12-14.

Papanicolas, I. and McGuire, A. (2015). Do financial incentives trump clinical guidance? hip replacement in england and scotland. *Journal of Health Economics*, 44:25–36.

Propper, C., Burgess, S., and Gossage, D. (2009). Competition and quality: Evidence from the nhs internal market 1991–9. *The Economic Journal*, 118(525):138–170.

Propper, C., Burgess, S., and Green, K. (2004). Does competition between hospitals improve the quality of care?: Hospital death rates and the NHS internal market. *Journal of Public Economics*, 88(7-8):1247–1272.

Quotidiano Sanità (2013). Tariffe massime del ssn. in gazzetta il decreto con i nuovi drg. Accessed: 2025-06-03.

Roberts, K. C., Brox, W. T., Jevsevar, D. S., and Sevarino, K. (2015). Management of hip fractures in the elderly. *J. Am. Acad. Orthop. Surg.*, 23(2):131–137.

Seong, Y. J., Shin, W. C., Moon, N. H., and Suh, K. T. (2020). Timing of hip-fracture surgery in elderly patients: Literature review and recommendations. *Hip Pelvis*, 32(1):11–16.

Shiga, T., Wajima, Z., and Ohe, Y. (2008). Is operative delay associated with increased mortality of hip fracture patients? systematic review, meta-analysis, and meta-regression. *Can. J. Anaesth.*, 55(3):146–154.

Shigeoka, H. and Fushimi, K. (2014). Supplier-induced demand for newborn treatment: Evidence from japan. *Journal of Health Economics*, 35(1):162–178.

Strumann, C., Geissler, A., Busse, R., and Pross, C. (2022). Can competition improve hospital quality of care? a difference-in-differences approach to evaluate the effect of increasing quality transparency on hospital quality. *Eur. J. Health Econ.*, 23(7):1229–1242.

Taroni, F. (1997). *DRG/ROD e nuovo sistema di finanziamento degli ospedali.* Il Pensiero Scientifico Editore.

Valentelyte, G., Keegan, C., and Sorensen, J. (2021). Analytical methods to assess the impacts of activity-based funding (abf): a scoping review. *Health Economics Review*, 11(1).

Verzulli, R., Fiorentini, G., Lippi Bruni, M., and Ugolini, C. (2017). Price changes in regulated healthcare markets: Do public hospitals respond and how? *Health Economics (United Kingdom)*, 26(11):1429–1446.

Vidán, M. T., Sánchez, E., Gracia, Y., Maranón, E., Vaquero, J., and Serra, J. A. (2011). Causes and effects of surgical delay in patients with hip fracture: a cohort study. *Annals of internal medicine*, 155(4):226–233.

Appendix A

Figure A1: DRG 183 (Esophagitis, gastroenteritis, and miscellaneous digestive system disorders, for patients over 17 years old without complications) - Tariffs' Trend

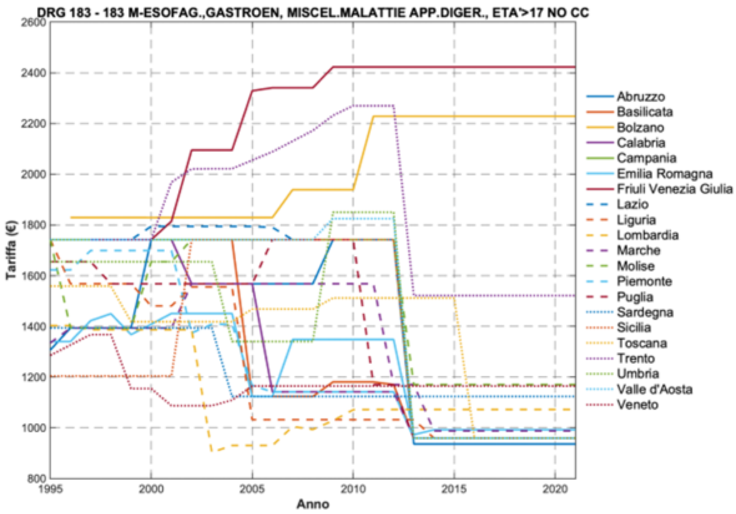


Figure A2: DRG 183 (Esophagitis, gastroenteritis, and miscellaneous digestive system disorders, for patients over 17 years old without complications) - Statistics' Trend

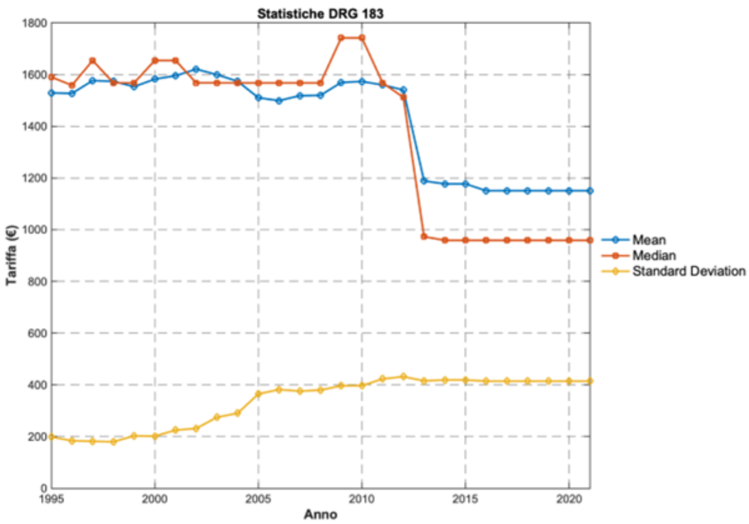


Figure A3: DRG 162 (Procedures for inguinal and femoral hernia, for patients over 17 years old without complications) - Tariffs' Trend

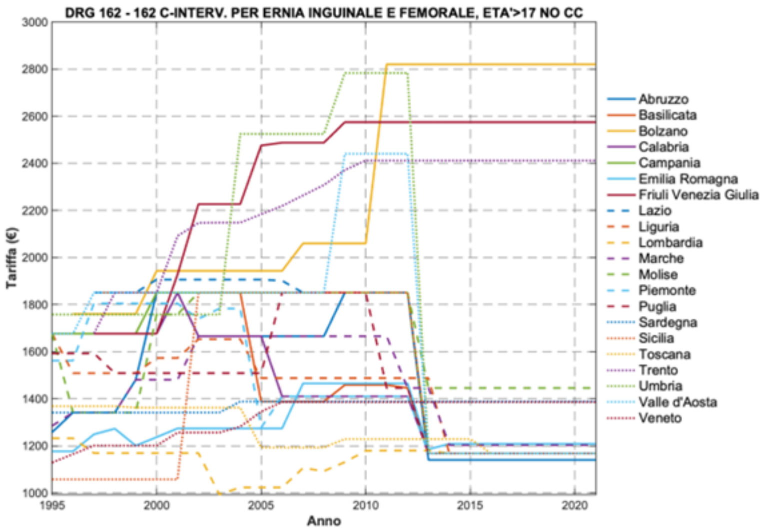


Figure A4: DRG 162 (Procedures for inguinal and femoral hernia, for patients over 17 years old without complications) - Statistics' Trend

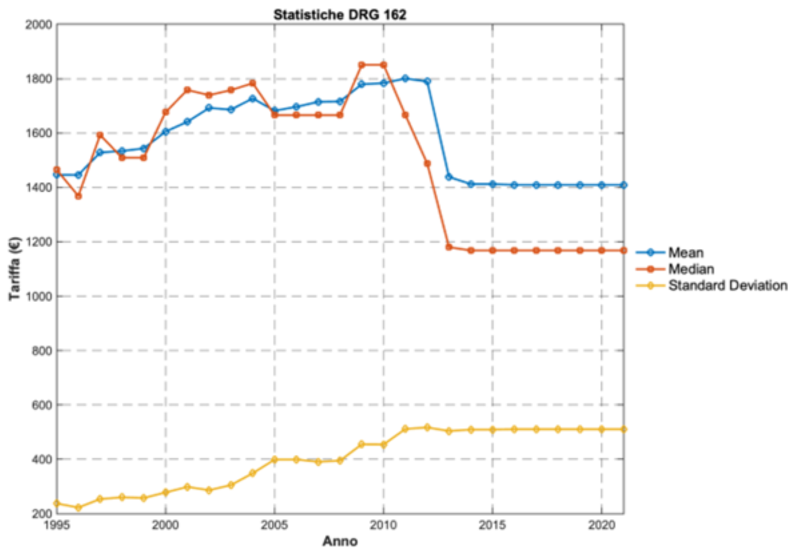


Figure A5: DRG 88 (Chronic obstructive pulmonary disease) - Tariffs' Trend

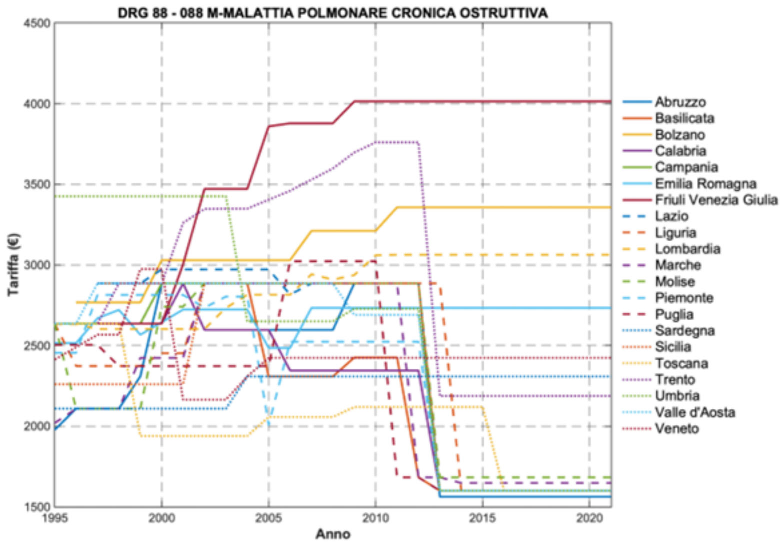


Figure A6: DRG 88 (Chronic obstructive pulmonary disease) - Statistics' Trend

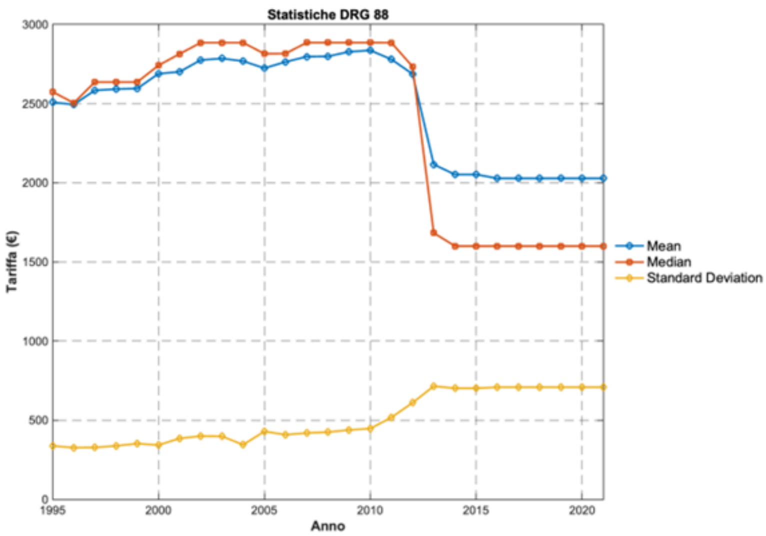


Figure A7: DRG 229 (Procedures on the hand or wrist, except major joint procedures, without complications) - Tariffs' Trend

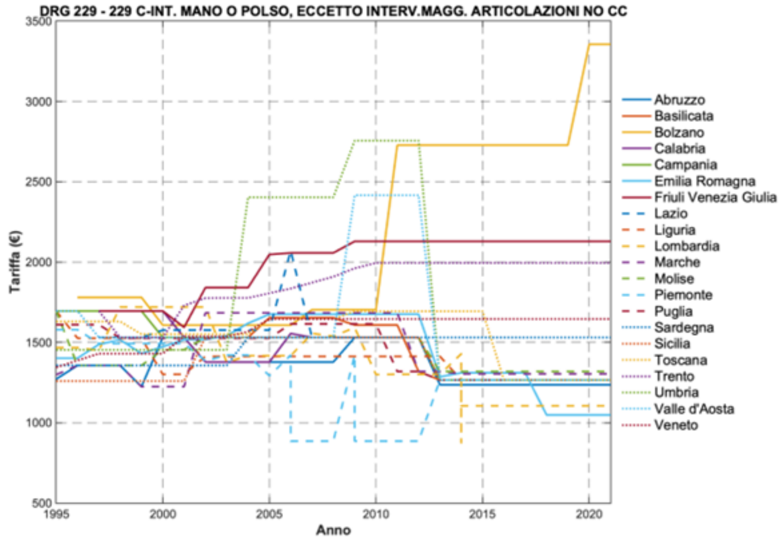


Figure A8: DRG 229 (Procedures on the hand or wrist, except major joint procedures, without complications) - Statistics' Trend