Article:

De Luca, Giacomo Davide orcid.org/0000-0001-5376-9663, Lisi, Domenico, Martorana, Marco et al. (1 more author) (2021) Does Higher Institutional Quality Improve the Appropriateness of Healthcare Provision? JOURNAL OF PUBLIC ECONOMICS. 104356. ISSN 0047-2727

https://doi.org/10.1016/j.jpubeco.2020.104356

Does higher Institutional Quality improve the Appropriateness of Healthcare Provision?

Giacomo De Luca[a c] Domenico Lisi[b] Marco Martorana[b] Luigi Siciliani[a]

[a] *Department of Economics and Related Studies, University of York, United Kingdom*

[b] *Department of Economics and Business, University of Catania, Italy*

[c] *LICOS, KU Leuven, Belgium*

June 2020

Abstract

We study the effect of institutional quality on the appropriateness of healthcare provision in Italian hospitals. We focus on cesarean section rates for first-time mothers, which is a common indicator of appropriateness in healthcare and is vulnerable to providers' opportunistic behaviors. To identify the causal effect of institutional quality we rely on an IV strategy based on historical instruments, exploiting the idea that current differences in institutional quality across regions have been shaped by different cultural and political histories. We find that a standard deviation increase in our indicator of institutional quality leads to a decrease of about 10 percentage points in cesarean section rates. Our results are robust to different measures of institutional quality and samples.

JEL Classification: I11, D73, C26.

Keywords: institutional quality, appropriateness, healthcare provision, IV, cesarean sections.

1. Introduction

Institutions matter. From the seminal work of North (1981, 1990, 1991), the role of institutions in affecting behaviors in society has been widely recognized. Institutions influence economic performance through various mechanisms.[1] The institutional framework affects investment and growth (Mauro, 1995), the effectiveness of foreign aid and the expected benefits of natural resource endowments (Burnside and Dollar, 2000; Mehlum et al., 2006), firms' efficiency (Dal Bò and Rossi, 2007), and even the emergence of organized crime (Acemoglu et al., 2020).

In the public sector, where public officials need to be kept accountable with implicit incentives (e.g., Alesina and Tabellini, 2007, 2008), the role of the institutional framework is of primary importance. Mauro (1998) studies the composition of public expenditure across over 100 countries, and finds that countries where the perception of corruption is higher spend less on sectors which provide less lucrative opportunities for public officials. Similarly, Gupta et al. (2001) report that corruption is associated with higher military spending, suggesting the presence of rent-seeking behavior. In the political sector, Nannicini et al. (2013) show that the electoral punishment of politicians' misbehaviors (such as absenteeism in Italian Parliament) is considerably larger in districts with higher social capital. Wong et al. (2017) examine the effects of an institutional reform implemented in China, and find that a higher quality of governance in rural villages increases the quality of public infrastructure provision.

In this study we investigate the effect of institutional quality on the appropriateness of healthcare provision in Italian hospitals. Healthcare spending in OECD countries represents about 9% of GDP and the majority of health expenditure is publicly financed (OECD, 2017). Healthcare accounts for about 15% of total public spending, though in some countries such as the United States and Germany more than 20% of public spending is for healthcare (OECD, 2017). Health spending is expected to grow further driven by an ageing population and technological innovation, raising challenges to the sustainability of public finances. Increasing the appropriateness in healthcare provision is thus critical in making health systems sustainable.

Due to the asymmetry of information which characterizes the doctor-patient relationship (Arrow, 1963), healthcare is vulnerable to providers' opportunistic behaviors, which may lead to under- or over-treatment (Ellis and McGuire, 1986). This is also the case within the hospital sector, where

[1] See Barro (1991), Acemoglu et al. (2001), Rodrik et al. (2004), Dell (2010), Engerman and Sokoloff (2012), Borcan et al. (2018), Acemoglu et al. (2019). For a recent comprehensive review of the empirical evidence on the impact of institutions on economic development and growth, see Durlauf (2018).

providers respond to financial incentives across a range of health systems, including those based on a National Health Service where hospital physicians are typically salaried (see Section 3 for more detailed discussion). The asymmetry of information is further exacerbated by the difficulty for public insurers to measure providers' performance, which further reduces their accountability (European Commission, 2017).

We focus on cesarean section rates during childbirth in Italy as a measure of appropriateness of healthcare provision. Cesarean section rates for first-time mothers are recognized as a valid indicator of appropriateness in the provision of healthcare services both in the literature and by policymakers (Baicker et al., 2006; OECD, 2009). In absence of clinical reasons or complications (e.g., uterine rupture), vaginal delivery is the recommended mode of delivery. Cesarean section is more invasive, it involves risks during surgery, has longer recovery times following birth and is more costly. For example, cesarean delivery has been found to increase the risk of maternal mortality and re-hospitalization, infant morbidity, and is negatively associated with child cognitive development (Lydon-Rochelle et al., 2000; Villar et al., 2006; Li et al., 2014; Polidano et al., 2017).

Crucially, the decision making in childbirth falls in the gray area of medicine, implying that it is possible for physicians to argue that a cesarean section is appropriate when it could be avoided (Chandra et al., 2011; Johnson and Rehavi, 2016). This makes the choice of childbirth delivery particularly exposed to the discretion of physicians (Gruber and Owings, 1996; Hopkins, 2000; Di Giacomo et al., 2017) and to what in the health economics literature is known more broadly as "supplier-induced demand" (Rice, 1983).[2]

There are several mechanisms through which local institutional quality can affect the appropriateness of hospital services. Regional authorities can measure and monitor caesarean sections for each hospital, publish them in the public domain to "shame" providers with excessive rates or to trigger audits and ask for detailed accounts of their treatment choices. They can also reduce the price differential between a cesarean section and a natural delivery, to dampen or eliminate the financial incentives. The extent to which these measures are introduced will depend on the quality of local administration, their accounting systems, and their determination and culture to reduce waste of public resources. Aside from formal measures, a local institutional environment

[2] Supplier-induced demand refers to the extent to which a doctor provides or recommends the provision of medical services that differ from what the patient would choose if he or she had available the same information and knowledge as the physician. This asymmetry of information can lead to over-treatment, under-treatment and over-charging (Rice, 1983).

characterized by widespread corruption, poor governance, and weak rule of law can reinforce hospital incentives to over-treat.

Italy exhibits large variations in cesarean section rates across regions that cannot be explained by sociodemographic and clinical factors (e.g., Francese et al., 2014; Guccio and Lisi, 2016). In turn, this has led to concerns by the Italian Ministry of Health that regularly monitors cesarean section rates across regions and providers (Ministry of Health, 2017). The cesarean section rate in Italy was 35% in 2016 and varied between 20% in Trentino Alto Adige and 59% in Campania, well beyond the level deemed appropriate by the World Health Organization (1985) of around 20%, as for instance observed in Nordic countries (OECD, 2017).

We measure institutional quality in Italian regions with the Institutional Quality Index, a multidimensional quantitative indicator (Nifo and Vecchione, 2014) based on the hierarchy framework employed by the World Bank Worldwide Governance Indicators (Kaufmann et al., 2010). As an alternative measure, we also employ a perception-based multidimensional indicator, the European Quality of Government Index (EQI) (Charron et al., 2014).[3] Indicators of the quality of institutional environment (e.g., corruption, quality of government) display large differences across Italian regions (Golden and Picci, 2005; Charron et al., 2014). For example, Charron et al. (2014, p. 74) notice that "the gap between Bolzano and Campania in the data is much larger than the gap in the national averages between Denmark and Portugal". Such regional variation makes Italy a particularly suitable case study to investigate the effect of institutional quality (e.g., Nannicini et al., 2013; Castro et al., 2014; Lasagni et al., 2015).

To identify the causal effect of institutional quality on the appropriateness of healthcare services, we employ an instrumental variable approach based on historical data. We follow the influential literature which argues that current institutional backwardness is also a byproduct of history (e.g., North, 1981; Acemoglu et al., 2001; Tabellini, 2010; Guiso et al., 2016): current differences in institutional quality across Italian regions have been shaped by different cultural and political histories. Specifically, our instruments are an indicator of quality of political institutions in the period from 1600 to 1850 and the percentage of population over age 6 able to read in 1881, provided in Tabellini (2010).

[3] Multidimensional indicators of institutional quality (based on similar underlying dimensions) are also employed in the cross-country literature on the impact of institutional quality (e.g., Knack and Keefer, 1995; Bosworth and Collins, 2003; Easterly and Levine, 2003; Tan, 2010).

Our results show that institutional quality improves the appropriateness in the provision of childbirth services. A standard deviation increase in our indicator of institutional quality leads to a reduction of about 10 percentage points in cesarean section rates. The results are robust to alternative indicators of institutional quality and samples. We also find that, amongst the different dimensions of institutional quality, corruption and government effectiveness in the local area appear to be the most relevant in affecting the provision of childbirth services.

We provide evidence on possible mechanisms through which institutional quality affects cesarean section rates. Among them, we find that higher institutional quality maps into pro-active policies (i.e. hospitals' payment policies) to reduce inappropriate provision, which in turn lead to lower cesarean section rates. We also show that institutional quality does not lead to worse mothers' health outcomes as measured by emergency readmissions. Finally, we investigate if institutional quality affects not only appropriateness of care, as measured by cesarean section rates, but also clinical quality. We show that higher institutional quality reduces heart attack mortality rates, but does not affect stroke mortality or mortality for chronic obstructive pulmonary disease.[4] Overall, our results suggest that a worse institutional environment weakens hospitals' incentives to select treatments that are cost-effective.

Our study brings together two strands of the literature. First, we contribute to the literature on the effect of the quality of institutional environment in the public sector (e.g., Mauro, 1998; Nannicini et al., 2013; Castro et al., 2014; Wong et al., 2017) by highlighting the importance of the health sector. Second, we contribute to the literature on the causes of inappropriateness in healthcare provision (e.g., Brown, 1996; Dubay et al., 1999; Gruber et al., 1999; Currie and MacLeod, 2008; Epstein and Nicholson, 2009; Francese et al., 2014; Johnson and Rehavi, 2016; Foo et al., 2017) by highlighting the role of institutional quality in affecting physician behavior.

Only very few studies have investigated the role of institutional quality in the healthcare sector. Di Tella and Schargrodsky (2003) analyze the prices paid by public hospitals for basic inputs during a crackdown on corruption in Buenos Aires, finding that prices decreased by about 15%

[4] The results are consistent with the extensive hospital competition literature, discussed in more detail in Section 4.1, which finds that hospital competition reduces heart attack mortality in the US for Medicare patients and in England for patients covered by the National Health Service (Kessler and McClellan, 2000; Kessler and Geppert, 2005; Cooper et al., 2011; Gaynor et al., 2013; Bloom et al., 2015), though Gowrisankaran and Town (2003) find that competition increases heart attack mortality for Medicare patients in California. Similarly, Moscelli et al. (2018) finds that competition has no effect on stroke mortality, and also shows that quality measures have generally low correlation across hospitals, therefore suggesting that findings for one condition does not necessarily translate to other health conditions.

relative to the pre-crackdown period. Azfar and Gurgur (2008) study healthcare provision in the Philippines, and report that a higher corruption in municipalities decreases immunization rates, delays vaccination of newborns and increases waiting time. Cavalieri et al. (2017) study the execution of public contracts for infrastructure in the hospital sector, and show that the performance in the provision of infrastructure is negatively affected by environmental corruption. With respect to this literature, we focus on inappropriate behavior in healthcare provision rather than prices or the execution of infrastructures in the hospital sector. Moreover, we also show that, while corruption is among the most relevant environmental factors, it is not the only source of waste in healthcare.

The study, which is closest to ours, is Francese et al. (2014). It investigates the impact of institutional features, such as supply and pricing policies, and some political economy indicators (namely, occupation and years of experience of the regional president), on regional cesarean section rates in Italy over the period 1998–2005. In their regional level analysis, they include regional fixed effects and thus exploit variation over time to identify the determinants of cesarean sections. In our study, we investigate a range of dimensions of local institutional *quality* in which hospitals operate. As shown below, such institutional quality varies little over time, while it varies significantly across regions.[5] We therefore exploit variation across regions and employ an instrumental variable strategy to address the potential omitted-variable bias. Our unit of analysis is at the more disaggregated hospital level rather than regional level (though institutional quality varies across regions), and we use the more recent data for the period 2007–2012.

The rest of the study is organized as follows. Section 2 describes the Italian healthcare system. Section 3 provides a conceptual framework which discusses the mechanisms through which local institutional quality may affect the appropriateness of healthcare provision. Section 4 describes the data. Section 5 provides the empirical strategy. Section 6 gives the results. Section 7 concludes and draws implications for healthcare policy.

2. The Italian healthcare system

The Italian National Health Service (NHS) was established in 1978 as a public health system providing universal access for a large basket of healthcare services (known as *Livelli Essenziali di Assistenza*). During the 1990s, a major reform was undertaken which separated the purchasers

[5] This is consistent with the fact documented in the literature that institutional factors change slowly over time (e.g., Acemoglu et al., 2005; Acemoglu and Robinson, 2008; Kaufmann et al., 2010).

(local health authorities) and the providers of health services, a form of quasi-market model, and gave providers more autonomy in managing their costs (France et al., 2005). Following the reform, patients have free choice of the hospital. Most hospitals are paid through an activity-based payment system set on fixed tariffs related to the diagnosis-related group (henceforth DRG) classification of discharges. Specialists working within the NHS are mostly salaried. Lump-sum transfers are also used to finance specific services, such as emergency services, prevention activities, and integrated care programs.

A key feature of this reform was the devolution of responsibilities to regional governments for the implementation of the quasi-market reform and, more generally, for the delivery of healthcare services. Only one out of the twenty regions, Lombardy, has implemented a full separation between purchasers and providers with public hospitals having managerial autonomy and independence. Many regions have opted for a partial separation with many public hospitals being managed and directly controlled by the local health authorities. Regions also differ in the degree of involvement of private providers to treat publicly-funded patients. Public and private hospitals are paid through the tariff system, which pays a fixed tariff for each additional patient treated. Instead, (local) public hospitals are funded by the global budget of the local health authority in which they fall. The budget of the health authority is based on a capitation rule, which implies that local public hospitals have less incentives to increase activity.

Regional governments are also free to choose between applying the national DRG tariffs provided by the Italian Ministry of Health and establishing their own DRG tariffs. While many regions still use the national tariffs, some regions have set their own fees, such as Lombardy, Emilia-Romagna, Sicily, and the Autonomous Province of Bozen. Furthermore, they are responsible for the appropriateness and the quality in the provision of hospital services.

Overall, while the funding of the Italian NHS is decided by the central government, decisions over its provision are decentralized to the regions (France et al., 2005). Therefore, the quality of the institutional environment in each region plays an important role in affecting providers' behavior. Regions characterized by inefficient or corrupt public administrations could lead to sub-optimal provision of health care especially for medical decisions where providers' discretion is large, as in the context of childbirth.

3. Conceptual framework

An extensive theoretical literature has modeled provider behavior in the healthcare sector. The seminal paper by Ellis and McGuire (1986) shows how healthcare providers, such as hospitals, choose the intensity of care to trade-off altruistic concerns against profit. Hospitals are complex organizations. The payoff function is generally modeled with a reduced-form approach that encompasses the utility functions of key workers, such as doctors, managers (Chalkley and Malcomson, 1998a, 1998b). Such payoff function can be thought as the result of a bargaining process between managers and doctors within the hospital (Chone and Ma, 2011; Galizzi and Miraldo, 2011). The literature shows that activity-based payments of the DRG type, with a fixed price for each patient treated, induce hospitals to contain costs, but can also incentivize quality provision if demand responds to quality and there is excess capacity so that hospitals compete. However, whether quality is under-provided or over-provided depends on the demand elasticity and price levels (Ma, 1994).

Hospital payment systems have been constantly revised over time, and DRG tariffs have been split for a given diagnosis to reflect the cost of different treatments. However, splitting DRGs for different treatments raises incentive issues on its own. The key theoretical insight from the literature is that in the presence of multiple treatments, the provider may have an incentive to over-provide the more intensive treatment to patients with intermediate severity of illness (Malcomson, 2005; Siciliani, 2006; Hafsteinsdottir and Siciliani, 2010). One example within neonatal departments is giving birth with hospitals being paid a higher fixed tariff for a caesarean section and a lower tariff for vaginal delivery. Another example is within cardiovascular conditions with a higher tariff being paid for a coronary bypass (a surgical procedure) relative to less invasive procedures (e.g. angioplasty) or alternative medical treatments. The larger the difference in the price mark up between the more and the less intensive treatment, the stronger the incentive to provide the more intensive treatment.

These predictions apply across a range of health systems, including those based on a National Health Service. Although these are publicly funded (except for small patient co-payments), hospitals compete on quality for patients under an activity-based payment of the DRG type. Examples include England and Italy (in particular Lombardy) where “internal markets” have been created, and hospitals have been split from purchasers (health authorities, commissioners, etc.) for the last thirty years. Although doctors are generally salaried within NHSs, hospital managers are still responsible for the financial viability of the hospital. As mentioned above, the treatment provided by doctors within hospitals can be thought as the outcome of a bargaining process where managers have to ensure that doctors take treatment costs into account. There is an extensive empirical literature which shows that NHS hospitals respond to provider incentives. For example, the introduction of the DRG system in England, which replaced hospital fixed budgets, increased volumes (Farrar et al., 2009). Similarly, the higher DRG tariff for cemented hip replacement, relative to uncemented one, led English hospitals to increase the proportion of the more remunerative cemented procedure (Papanicolas and McGuire, 2015).

In summary, the extent to which a hospital has an incentive to provide a more intensive treatment, such a caesarean section, is the result of a trade-off between profit considerations against non-profit ones. The latter include several factors that may limit the number of inappropriate treatments, such as ethical and altruistic concerns towards patients, and their reputation with peers, patients and the community. The weight given to profit versus other considerations will vary depending on circumstances and the institutional context. In poorer regions where funders have fewer resources to distribute to hospitals, managers may exercise more pressure to doctors to increase revenues, though public hospitals may be less sensitive to profit considerations relative to for-profit or non-profit hospitals (Dafny, 2005; Brekke et al., 2012).

Local authorities and purchasers may also differ in the instruments used to address possible over-treatment. As already mentioned, one instrument is the price difference between treatments, which can be dampened or eliminated (as, for instance, in Lombardy for childbirth delivery). But there are others. Regional authorities could measure and monitor caesarean sections for each hospital, and either make these available in the public domain (to "shame" the providers with excessive rates) or to trigger audits where providers are required to give a detailed account of the treatment choices. A higher probability of an audit makes it more costly to the provider to over-treat, and therefore reduces the probability of providing the more intensive treatment, such a caesarean section (Kuhn and Siciliani, 2013; Guccio and Lisi, 2016). In turn, local authorities may be more likely to introduce such measures when they perceive higher scope of over-treatment. Whether such formal measures are introduced also depends on the quality of local administration, accounting systems they have in place, widespread culture of corruption and misallocation of public resources.

Aside from such formal measures, an environment characterized by low institutional quality further reinforces hospital incentives to over-treat. Weak rule of law and a lack of good governance foster opportunistic behavior, as the risks of being caught in wrongdoing is perceived as negligible (Coviello et al., 2018). Local governments may also exert low effort in monitoring local hospitals or health authorities, if monitoring implies a transfer of resources from the regional to the national budget.[6] Similarly, low participation of citizens in the public and cultural aspects of life induces individuals to strengthen their ties within their family, kin, or friends in restricted circles, even at the expenses of the rest of the society. In contrast, an active citizens' attitude in the public domain is deemed crucial to make public officials accountable, thereby discouraging their misbehaviors (Nannicini et al., 2013).

Finally, where corruption is widespread or does not carry a stigma among the local population, physicians may find it less morally taxing to implement an unnecessary procedure, if this increases their personal benefit (or an expected future reward from their employer), and only marginally affects patients health. This holds in particular for childbirth services where treatment choice falls in the gray area of medicine , which makes it difficult for both public insurers and citizens to assess, even *ex-post*, the appropriateness of childbirth (Chandra et al., 2011; European Commission, 2017).

4. Data

4.1 Hospital level data

We merge data from different sources. The primary source is the National Program for Outcome Assessment (*Programma Nazionale Esiti*, PNE), which is funded by the Italian Ministry of Health and the National Agency for Regional Health Services (AGENAS). The PNE was developed to provide reliable comparative assessment of healthcare providers with respect to the effectiveness, equity and appropriateness of treatments.[7] We draw from this source the risk-adjusted cesarean section rate, which is measured annually at the hospital level for the period 2007–2012. For each hospital with at least 10 childbirths in the selected year, the PNE provides data on total number of births, vaginal births, cesarean sections for first-time mothers (resident in Italy) aged 10-55 years, as well as the risk-adjusted rates.

The method for risk adjustment is the direct standardization. The predictive model is estimated through a logistic regression (with dependent variable equal to 1 if delivery involves a cesarean

---

[6] For a theory model on public officials in the field of public works "looking the other way" in environments with low institutional quality, see e.g. Guccio et al. (2019).

[7] The identification of the risk factors affecting health outcomes is an explicit objective of the program to ensure reliable comparisons of providers. Detailed information on the PNE can be found on the program's website at the url: http://82.112.223.85/PNEed14_EN/index.php

section and 0 for a natural birth), which controls for maternal age, 39 comorbidities (main and secondary diagnoses over the last two years) and other risk factors selected through bootstrap stepwise procedures, and includes hospital fixed effects. Then, for each hospital the predicted probability of a cesarean section is evaluated at the sample mean of covariates. Risk factors include malignant cancers, clotting disorder, heart disease, HIV, diabetes, high blood pressure, tuberculosis, smoking, overweight, ante-partum haemorrhage, foetal anomalies, cord prolapse, and malpresentation. The complete list of factors used in the risk adjustment is available in Appendix A.1.

As a secondary outcome for maternal care, we also include mother readmissions rates, i.e. the proportion of hospitalizations for cesarean delivery followed within 42 days by at least one ordinary admission with length of stay ≥2 days (*Mother hospital readmission*) but only for 2012, the last year of our sample, as data on readmission rates were not collected in previous years. These are also risk-adjusted following the same methodology used for cesarean sections.

In a robustness check, we use additional quality indicators also measured at the hospital level: risk-adjusted 30-days mortality rates for Acute Myocardial Infarction (*AMI-MR*) – more commonly known as a "heart attack" –, Chronic Obstructive Pulmonary Disease (*COPD-MR*), and Stroke (*STK-MR*). These data are also provided by the PNE, but are available only for 2012, the last year of our sample period. As for cesarean section, the precise risk-adjustment procedure is reported in Appendix A.1. AMI mortality is a validated clinical quality indicator, and it is the indicator that is most commonly used in the literature investigating the effect of hospital competition on quality. This includes the seminal studies by Kessler and McClellan (2000) and Kessler and Geppert (2005) in the US, but also more recently by Cooper et al. (2011), Gaynor et al. (2013) and Bloom et al. (2015) in the UK. These five studies find that competition reduces AMI mortality, and therefore supports that hospital competition increases quality. The other indicators have also been used. Gowrisankaran and Town (2003) studies the effect of competition on pneumonia mortality (in addition to AMI mortality) in California, and finds that competition increases mortality for both pneumonia and AMI. Moscelli et al. (2018) finds that competition reduces mortality for hip fracture and AMI, but does not affect stroke mortality.

We control for a range of hospital types following a classification provided by the Ministry of Health. Autonomous public hospitals (*Public*) are large hospitals that have managerial independence from local health authorities. Local public hospitals (*Local*) are instead directly managed by the local health authorities. Private hospitals can be for-profit (*Private For-Profit)* and not-for-profit (*Private Nor-For-Profit*). We further distinguish between teaching hospitals (*Teaching*), and research hospitals (*Research*). The latter, known as IRCCS (Istituto di Ricovero e Cura a Carattere Scientifico), are hospitals that strongly engage in clinical research.

We measure hospital characteristics, such as the total number of hospital beds (*Beds*), and the number of births (*Births*). Hospital beds are a proxy for hospital size and control for possible scale economies and hospital ability to make large capital investments. The volume of births captures possible *learning-by-doing* effects, in line with previous studies (Birkmeyer et al., 2002; Gaynor et al., 2005; Chandra et al., 2011). We also control if the hospital is located in a province that is the regional capital (*Regional Capital*) and, within a province, if the hospital is located in a municipality that is the provincial capital (*Province Capital*), identified by the information on hospital location. This may act as a proxy of the cost of capital.

All hospital data have been cleaned for outliers and missing values.[8] The final sample consists of 492 hospitals over about 600 Italian hospitals providing childbirth services for the period 2007-2012, yielding 2952 observations.

4.2 Institutional quality

We measure institutional quality at the regional level with the Institutional Quality Index (IQI) proposed by Nifo and Vecchione (2014).[9] IQI is a composite indicator, ranging from 0 to 1 (with 1 representing the best institutional quality) that summarizes different dimensions of the quality of institutional environment. It is based on the same hierarchy framework used by the World Bank Worldwide Governance Indicators (WGI) proposed by Kaufmann et al. (2010). Similar to the WGI, it combines several relevant aspects of local governance, each measured aggregating the regional performance along different dimensions.

---

[8] Specifically, outliers refer to those hospitals for which the PNE has laid down an inspection on the quality of data regarding childbirth services, because of belonging to the extreme percentiles of the frequency distribution; at the most, about 30 hospitals have been subject to scrutiny in a given year over the time-period. Missing values instead refer to about 90 Italian hospitals (with less than 10 childbirths in the selected year) for which the PNE did not provide risk-adjusted cesarean rate. A full list of Italian hospitals excluded from our sample is available on request.

[9] Recent examples of studies using IQI are Boschma et al. (2015) and Lasagni et al. (2015) which look at firm performance in the private sector, and Baldi et al. (2016) and Guccio et al. (2019) which study performance in public procurement. Similar indices or subcomponents of IQI feature regularly in the recent political economy literature on Italy (e.g., Guiso et al., 2004, 2016; Del Monte and Papagni, 2007; Nannicini et al., 2013; Castro et al., 2014; Coviello et al., 2018).

More specifically, the IQI is based on five dimensions: 1) *voice and accountability*, which combines scores in citizens' participation in public elections, the number of associations, the number of social cooperatives and cultural liveliness measured in terms of books published and purchased in bookshops; 2) *government effectiveness*, which measures the endowment of social (education, healthcare, and leisure) and economic infrastructure (roads, railroads, ports, airports, energy, ICT, and banking), health deficit per-capita, the proportion of separate waste collection on total waste collection, and environmental protection; 3) *regulatory quality*, which combines information on economy openness, local government employees, business density, business start-ups/mortality, and business environment; 4) *rule of law*, which aggregates crime levels, tax evasion, shadow economy, magistrate productivity, and trial times; and 5) *corruption*, which is based on the crime rates against the public administration, the number of local administrations overruled by central government, the difference between a measure of quantities of public infrastructure and the cumulative price government paid for public capital stocks (Golden and Picci, 2005). The data used to build these indicators are collected from Italian institutional sources and research institutes.

The Analytic Hierarchy Process (AHP) introduced by Saaty (1980, 1992) is then applied to combine the five components of the IQI. The AHP uses a predetermined multi-layer framework with a hierarchy among elementary (e.g., corruption) and aggregate indexes (e.g., IQI), and it allows to derive a weight for each index of a given layer (see Nifo and Vecchione, 2014, for more details). Specifically, the weights for the five components in the IQI resulting from the AHP are (Nifo and Vecchione, 2014, p. 1635): 1) 0.168 for *voice and accountability*, 2) 0.312 for *government effectiveness*, 3) 0.107 for *regulatory quality*, 4) 0.345 for *rule of law*, 5) 0.067 for *corruption*. In addition to the IQI, we employ in our analysis each of the five components. We also construct two alternative indices. The first assigns equal weights to the five components. The second excludes regional health deficits in the government effectiveness component.[10]

Finally, we use an alternative measure of institutional quality known as the European Quality of Government Index (EQI). EQI is a regional perception-based indicator of government quality developed by the Quality of Government Institute at the University of Gothenburg (Charron et al., 2014, 2015), ranging from -2.88 to 1.75.[11] Higher values mean better perceived quality of governance. It is based on a questionnaire asking respondents to rate three public services (i.e.

[10] The expenditure for childbirth services for first-time mother is less than 1% of total health expenditure in Italy, thus the inclusion of regional health deficit in IQI per se does not involve a simultaneity problem in our estimates.
[11] EQI is built so that the average value of the indicator across European regions is equal to 0 (Charron et al., 2014).

education, healthcare, law enforcement) with respect to three related concepts developed along the WGI categories: quality, impartiality, and corruption. EQI is available only for two years, 2010 and 2013.

4.3 Other regional controls

On the demand side, we measure the proportion of women in the region with only primary school (*Low Education*), which is available from the Italian National Institute of Statistics (ISTAT).[12] This accounts for some demand factors (not controlled in the risk-adjustment model) because education may be negatively associated with a cesarean section (e.g., due to preferences, more detailed information of cesarean section consequences, and less educated women being more exposed to supplier discretion).

We have three variables related to regional payment policies for childbirth. The first variable is an indicator equal to 1 if the region has set its own DRG tariffs in that year and 0 if instead the national DRG tariffs apply (*Regional_DRG*). The information on regional DRG tariffs are provided by *AGENAS* and the Regional Health Authorities. The second variable (*DRG_tariff_VD*) measures the tariff paid to hospitals for a vaginal delivery (DRG 373) in each region. It represents the baseline revenue for a childbirth; thus, when it is low, it could provide an incentive for hospitals to perform a cesarean section. Then, the third variable (*DRG_tariff_CS*) measures the tariff paid for a cesarean section (DRG 371) in each region. Tariffs for cesarean deliveries are higher than those for vaginal ones, since the former involves a surgical intervention performed in an operating room.[13]

In some specifications, we also consider whether regions were subject to what are known as "recovery plans" (*Piani di Rientro*) under which they were obliged to reduce their health deficit. The variable (*Recovery_Plan*) is equal to 1 if the region was subject to recovery plan in that year and 0 otherwise.[14]

---

[12] Available at http://dati.istat.it/?lang=en.

[13] The costs of childbirth delivery are relatively homogeneous at the national level, at least among the same type of providers (e.g., Francese et al., 2014; Cavalieri et al., 2014).

[14] In the sample period 2007-2012, the following Italian regions were subject to recovery plans: Abruzzo, Campania, Lazio, Molise and Sicily for the whole period; Liguria and Sardinia for the period 2007-2009; Calabria, Piedmont and Puglia for the period 2010-2012. Full details on recovery plans in Italy, can be found in the dedicated website: http://www.salute.gov.it/portale/pianiRientro/dettaglioContenutiPianiRientro.jsp?lingua=italiano&id=5022&area=pianiRientro&menu=vuoto.

Finally, we measure the regional urbanization rate in 1860 (*Urbanization₁₈₆₀*), available from Tabellini (2010),[15] which is a proxy of local economic development at the time of Italy unification. It is employed as an additional control to back up our exclusion restriction, as described in more detail below.

4.4 Instruments

We employ two instruments based on historical variables. Both are computed at the regional level and provided in Tabellini (2010). The first is an indicator of quality of political institutions in the period from 1600 to 1850. The quality of political institutions is measured by the "Constraints on the executive" as defined by the dataset Polity IV, namely "institutionalized constraints on the decision making powers of chief executives". Higher values (more constrained executives) correspond to better institutions. This is measured at the regional level in five points in time: 1600, 1700, 1750, 1800, 1850. The five measures are then aggregated into a single index of political institutions for each region. We follow Tabellini (2010) and define $Institutions_{1600\text{-}1850}$ as the first principal component of the variable "Constraints on the executive" at the five points in time.

The second is the percentage of population over six years old who are able to read in 1881 ($Literacy_{1881}$), twenty years after the Italian unification. It gives a good proxy of the different cultural traits in Italian regions pre-dating the Kingdom of Italy. We discuss the validity of these instruments in Section 5.

4.5 Descriptive statistics

Table 1 reports the summary statistics. The average cesarean section rate for first-time mothers is 31%. To investigate how much cesarean section rates vary between regions as opposed to within regions across hospitals, Table 2 shows the results of one-way random-effects ANOVA model for risk-adjusted cesarean section rates.[16] Unlike the simple ANOVA, this allows us not only to decompose the variance between and within regions, but also to estimate the intraclass correlation coefficient (ICC), which can be interpreted as the proportion of variance that is attributable to variation between regions (Donner, 1986). The F-ratio ($\text{MeanSquare}_{\text{between}}/\text{MeanSquare}_{\text{within}}$) suggests that the variance between regions is remarkably larger than the variance within regions.

---

[15] See Acemoglu et al. (2005) for a discussion of using urbanisation rate as a proxy for economic development in historical studies.

[16] Specifically, our one-way random-effects ANOVA model assumes that we observe C-section rate y_{ij}, in n_i hospitals within k regions (i.e. our group variable) such that: $y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1,2, \dots, k, \quad j = 1,2, \dots, n_i$, where μ is the overall mean, α_i and ϵ_{ij} are independent random variables with variance σ_α^2 and σ_ϵ^2, respectively.

Furthermore, the ICC of 0.49 indicates that about half of total variation in cesarean section rates is attributable to variation between regions, with the rest attributable to the (unobservable) hospital characteristics.

– Tables 1 and 2 about here –

We also estimated a random intercept model controlling for hospitals' observable characteristics (see Section 4.1),[17] which gives an ICC of about 0.65. This suggests that, once we control for hospitals' observable characteristics, 65% of residual variance in hospitals' cesarean section rates is due to variation between regions. Finally, if we further include the IQI as covariate at the regional level in the random intercept model, this inclusion decreases the covariate-adjusted variance between regions, which suggests that IQI is a significant factor of between regions variation in cesarean section rates (see Appendix, Table A.1). Overall, these tests confirm a large variation in cesarean section rates across regions, which is the variation we exploit in our empirical strategy.

From Table 1, hospitals have on average 398 beds and assist 831 birth deliveries, respectively. 57% are local public hospitals and 18% are autonomous public hospitals. Only 5% are teaching hospitals, 14% are private for-profit hospitals and 4% are private not-for-profit hospitals. About 2% are research hospitals. 36% are located in the regional capital, and 47% in the provincial capital. The average mother readmission rate after a cesarean section in the Italian hospitals is 0.78%. The risk-adjusted mortality rate for AMI, COPD and stroke was respectively, 12%, 10% and 13%.

IQI summary statistics reveal remarkable differences in institutional quality across the 20 Italian regions. The average IQI is 0.59. Figure 1 shows the geographical distribution of our two indicators of institutional quality and raw and risk-adjusted cesarean rates.[18] The region with the highest institutional quality is Tuscany while the region with the lowest one is Calabria. There appears to be a negative association between them, as regions with the highest institutional quality exhibit low cesarean section rates.

- Figure 1 about here -

[17] Specifically, our random intercept model is: $y_{ij} = \mu + \beta x_{ij} + \alpha_i + \epsilon_{ij}, \quad i = 1,2,\ldots,k, \quad j = 1,2,\ldots,n_i$, where x_{ij} is a vector of hospitals' observable characteristics and β is a vector of slopes fixed across regions.

[18] The geographical distributions of IQI sub-components are reported in the Appendix.

Similar differences across regions are suggested by EQI. The average EQI in Italy is -0.69. This indicates that, on average, Italian regions have a perceived institutional quality lower than the average in Europe. The region with the highest institutional quality is Trentino-Alto Adige and the region with the lowest quality is again Calabria.

28% of women have only primary school education, ranging from 20% to 35% across regions. Concerning payment policies for childbirth, over the time period considered about 60% of Italian regions have developed their own DRG tariffs instead of applying the national one. The DRG tariff paid to hospitals for a vaginal delivery is on average € 1544, though with substantive variations across Italian regions. Veneto pays the lowest price (€ 923), instead the Autonomous Province of Bozen pays the highest one (€ 2226). On the other hand, the DRG tariff for cesarean section is on average € 2514, with Emilia Romagna paying the lowest tariff (€ 1806) and again the Autonomous Province of Bozen paying the highest one (€ 3941). One region, Lombardy, pays the same tariff for both modes of delivery (€ 2097).

The average urbanization rate in 1860 was 11%, ranging from 2% to 24% across regions. As for our two instruments, the average literacy rate at the end of the 19th century was 37%, with substantial regional variation ranging from 68% in Piedmont to 15% in Calabria. As for the quality of past political institutions, the region with the highest quality is Liguria. Those with the lowest level are the regions belonging to the Kingdom of the Two Sicilies (i.e. Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, and Sicily)[19].

5. Empirical specification

Our empirical strategy aims at estimating the causal effect of institutional quality on cesarean section rates. Our regression model is

$$Cesarean\ section_{irt} = \alpha + \beta\ IQI_{rt} + \boldsymbol{x}_{irt}^{1\prime}\boldsymbol{\gamma}_{1} + \boldsymbol{x}_{rt}^{2\prime}\boldsymbol{\gamma}_{2} + \gamma_{3}\, x_{r}^{3} + \boldsymbol{d}_{t} + \varepsilon_{irt} \qquad (1)$$

where $Cesarean\ section_{irt}$ is the risk-adjusted cesarean rate in hospital *i* in region *r* at time *t*, IQI_{rt} is the institutional quality in region *r* at time *t*, $\boldsymbol{x}_{irt}^{1}$ is a vector of control variables at the hospital level (number of beds, birth deliveries, hospital type, whether the hospital is located in the

[19] The Kingdom of the Two Sicilies was the largest of the states in Italy before the unification of 1861 (e.g., Zamagni, 1993).

provincial and regional capital), $\boldsymbol{x}_{rt}^{2}$ are variables that vary only at the regional level and over time (proportion of women with only primary school, hospital DRG pricing, recovery plan) and x_r^3 is the regional level degree of urbanization. $\boldsymbol{d}_t$ are year dummies.

We employ a pooled regression approach and exploit variation in institutional quality across regions rather than over time. This is because institutional quality varies little over time, while it varies significantly across regions. ANOVA shows that variation over time explains about 1% of total variance of institutional quality. This is not surprising because institutions change slowly over time (e.g., Acemoglu et al., 2005; Acemoglu and Robinson, 2008; Kaufmann et al., 2010).

We first estimate model (1) by OLS. However, despite the extensive list of controls, we cannot exclude the presence of omitted factors affecting both caesarean rates and the quality of institutions. Moreover, there may remain measurement error in the institutional quality variable. To overcome these problems, we use an instrumental variable (IV) strategy similar to that proposed by Tabellini (2010).

Current institutions, and their quality, are largely the product of historical dynamics (e.g., North, 1981; La Porta et al., 1997; Acemoglu et al., 2001; Persson and Tabellini, 2009; Guiso et al., 2016). We therefore exploit the effect of past cultural traits and institutional quality as an exogenous source of variation in current institutional quality (e.g., Tabellini, 2010; Guiso et al., 2016). Formally, our first-stage regression is:

$$IQI_{rt} = a + \boldsymbol{z}_r' \, \boldsymbol{b} + \boldsymbol{x}_{irt}^{1\prime} \, \boldsymbol{g_1} + \boldsymbol{x}_{rt}^{2\prime} \, \boldsymbol{g_2} + g_3 \, x_r^3 + \boldsymbol{\delta}_t + e_{irt} \qquad (2)$$

where $\boldsymbol{z_r}$ is our vector of instruments including a) the constraints on the executive of local political institutions in the period from 1600 to 1850 and b) the regional literacy rates at the end of the 19th century. e_{irt} is an error term potentially correlated with ε_{irt} in (1).

The intuition behind our first IV rests on two main mechanisms through which past political institutions influenced the functioning of current institutions: culture and governance. First, institutions shape *culture* (Banfield, 1958; Putnam et al., 1993; Boix and Posner, 1998). According to Aristotle's Nicomachean Ethics "Lawgivers make the citizen good by inculcating habits in them, and this is the aim of every lawgiver; if he does not succeed in doing that, his legislation is a failure. It is in this that a good constitution differs from a bad one" (as quoted by Bowles, 2001).

Therefore, Aristotle argued that good political institutions induce citizens to substitute opportunism for trustworthiness, transmitting this change in individual moral values to their descendants. The study by Tabellini (2010) supports this argument by showing that more autocratic governments left a legacy of low generalized trust across Europe (and Italy). The changes in local culture are persistent and ultimately influence both the choice and the functioning of political and legal institutions (Roland, 2004). Hence, more autocratic political regimes, exercising their power without constraints, might have developed a set of values for which corruption, inefficient uses of public resources, weak enforcement of law are accepted as part of the institutional system.

Second, institutions affect *governance*. Poor quality institutions in the past generated extensive clientelistic networks (Shefter, 1977; Charron and Lapuente, 2013). Weakly constrained executives generally resort to patronage to survive in power. Relying on clientelistic networks both reinforces the support and offers the opportunity to coopt challengers. Ruling elites distributed jobs in the public sector and public procurement contracts, and privileged the distribution of private over public goods. Even though the local power structure may have changed over time, notably after Italy unification, what persisted across time is the informal rule of engaging in patronage networks (Piattoni, 2001). Whenever politicians retain some discretionary power in law implementation, even good-minded centrally-designed policies can be used locally for patronage purposes. Importantly, the existence of patronage influences incentives of future generations, whose investments in their academic and professional growth, for instance, may be reduced based on the perception that what matters to get a good job is "knowing somebody" (Chubb, 1981). Acemoglu (1995) models a similar mechanism resulting in a sort of poverty trap, in which talents are misallocated in a society in which rent-seeking activities are more rewarding than productive activities. A vicious cycle would therefore result in which relatively unconstrained past executive increase local clientelism and ultimately perturbs the functioning of current democratic institutions.

The idea behind our second IV is that low literacy reinforces both mechanisms described above. It enables manipulation from the elites because it constrains the ability of individuals to understand and control their social environment, it reduces culture dissemination and it isolates the local population from better functioning external environment. Conversely, widespread literacy increases the level of socialization and the participation of individuals to the public life, strengthening democratic institutions (Lipset, 1959; Almond and Verba, 1963).

This IV strategy is valid if the error term in the second-stage equation is orthogonal to the excluded instruments, *Institutions*$_{1600\text{-}1850}$ and *Literacy*$_{1881}$. Two main channels may threaten our exclusion restriction. Current cesarean section rates may depend on the current level of regional economic development, which is directly influenced by past economic development. The latter itself has been shaped by past political institutions and past education. Similarly, cesarean section rates are likely to depend on the average regional education, notably among women. For instance, more educated women may have greater control on their birth delivery decisions. Again, current education is likely to be shaped by both past political institutions and past education. We therefore include in our model a measure of past regional economic development (as proxied by *Urbanisation*$_{1860}$) and a measure of contemporaneous female education (the proportion of women with only primary school, *Low Education*) to control for these additional channels. We argue that controlling for contemporaneous female education and past regional development, along with a full set of hospital control variables, our exclusion restriction is plausible. Additionally, given that we have two instruments for one endogenous variable, we can test our overidentifying restrictions.

As a preliminary evidence, Table 3 reports pairwise correlations between institutional quality indicators and historical instruments in our dataset. Our instruments are significantly correlated with fairly all institutional quality indicators (except for *IQI_rule*). As expected, correlation coefficients are always positive and not negligible. This evidence supports, in the first place, our empirical strategy of using historical values as instruments for current institutional quality.

- Table 3 about here -

We estimate (1) and (2) by 2SLS approach. However, because our dependent variable is a fraction, we also estimate model (1) with the generalized linear model (GLM) with logit link function and binomial distribution for the error term, as suggested by Papke and Wooldridge (1996). Differently from linear regression models, fractional response models account for the bounded nature of the dependent variable and ensure that fitted values stay within the unit interval; furthermore, the marginal effect of any explanatory variable is not constant throughout the range, as it is (and it cannot be given the bounded nature of the dependent variable) in linear models. To address the endogeneity in this non-linear model, we employ a control function (CF) approach given that the standard 2SLS is no longer valid. We therefore include the fitted residual from the first-stage as a regressor in the non-linear second-stage regression (Wooldridge, 2015).

6. Results

6.1 Baseline estimates

Table 4 provides our baseline results. For each estimate, we report bootstrapped standard errors clustered at regional level. Column (1) in Table 4 provides the OLS model. It suggests that higher institutional quality, as measured by IQI, reduces cesarean section rates. The effect is statistically significant at 1% level. The coefficient of -0.35 implies that a standard deviation increase in our indicator of institutional quality (equal to 0.21) reduces cesarean section rates by about 7 percentage points (the average cesarean section rate is 31%).

The coefficients on the control variables are in line with our expectations and previous literature. On the supply side, a larger volume of *Births* is negatively associated with cesarean section rates, which is consistent with *learning-by-doing* effects in the provision of childbirth services. Larger hospitals with more *Beds* are not associated with cesarean section rates as found in previous studies (Gruber et al., 1999; Guccio and Lisi, 2016).

Relative to autonomous public hospitals (the base category), cesarean rates are higher for private for-profit hospitals and teaching hospitals and are similar to non-profit and research hospitals. Cesarean rates are also higher in larger cities, as proxied by the dummy on whether the hospital is located in a regional capital. Finally, the coefficient on the proportion of women with *Low education* is not statistically significant, and thus it does not support the idea that less educated women are more exposed to supplier induced demand.

Estimates in column (1) in Table 4 do not take into account the presence of possible omitted factors affecting both cesarean rates and institutional quality, as well as measurement errors in IQI, implying that OLS estimates cannot be interpreted as causal. To address these, we employ the IV strategy described in Section 5, exploiting the idea that past institutional and cultural traits are correlated with current institutional quality. In columns (2) to (4) in Table 4 we report the reduced-form models in which IQI is replaced by the constraints faced by political institutions in 1600-1850 (column 2), literacy rates at the end of the 19th century (column 3), and both instruments (column 4). In line with the logic of our IV strategy: more constrained past executives and higher historical literacy rates map into lower cesarean section rates today. According to the estimates reported in column (2), a standard deviation increase in the constraints faced by the executive in 1600-1850 (equal to 1.26) is predicted to reduce cesarean section rates by 6 percentage points.

Likewise, the estimates in column (3) imply that a standard deviation increase in the historical literacy rates (equal to 0.17) is predicted to reduce cesarean section rates by 6 percentage points.

- Table 4 about here -

Table 5 reports our 2SLS estimates. The estimated effects of IQI from the second-stage regressions are larger than the OLS estimates, ranging from -0.51 to -0.61. Therefore, one standard deviation increase in institutional quality now reduces cesarean section rates by about 10-12 percentage points. The magnitude and statistical significance of other covariates are in line with OLS estimates.

Standard diagnostic tests support our IV identification strategy. Specifically, endogeneity tests reject the null hypothesis that IQI can be treated as exogenous at conventional significance levels (p-value = 0.009). Then, F-statistics of the Kleibergen-Paap test for weak identification range from 14.80 to 18.66, indicating that our instruments are not weak. The first-stage coefficients 0.079 and 0.757 in column (1) and (2), respectively, are positive and statistically significant, consistently with the idea that past cultural and political traits shape current institutional quality. When we include both instruments in column (3), first-stage coefficients are still positive, though only $Literacy_{1881}$ results statistically significant. Finally, the Hansen's J statistic does not reject our overidentifying restrictions (p-value = 0.163), thus supporting the orthogonality condition (i.e. the error term in the second-stage equation is orthogonal to our historical instruments).[20]

- Table 5 about here -

In Table 6 we report the results when we use the GLM model which specifically takes into account the fractional nature of the dependent variable. Column (1) provides the GLM model without addressing possible endogeneity of IQI, and is thus analogous to the OLS model in Table 4, column (1). Columns (2) to (4) display the GLM estimates where we follow the CF approach to address possible endogeneity of IQI.[21] The marginal effects are reported in the square brackets and are very similar to those reported in Table 5. Therefore, the CF approach corroborates the 2SLS

[20] Following Cameron et al. (2008), in Table A.2 we also produce our standard errors clustered at the regional level using a wild bootstrap-t procedure, performing better in presence of small number of clusters. The IQI coefficients (both OLS and 2SLS) remain statistically significant.
[21] As suggested by Wooldridge (2015), bootstrapped standard errors are employed in both first- and second-stage regressions of the CF estimates.

results. The coefficient of the fitted residuals from the first-stage (i.e. $\hat{e}_{First\ Stage}$) is statistically significant, again confirming the presence of endogeneity.[22]

- Table 6 about here –

6.2 Robustness to sample definition

A potential concern in our estimates is that, while cesarean section rates vary also over time (though little), the identification of the causal effect of institutional quality through historical instruments relies on variation across regions. As discussed in Section 4, this is because variation over time in institutional quality is negligible. Still, one may worry that we are inflating our sample size by employing a pooled regression. Table 7 reports the results from a robustness check in which the IV model (employing both instruments) is estimated separately for each year. Results in Table 7, including significance and magnitude of the IQI coefficients as well as diagnostic tests, are fully in line with previous estimates.

- Table 7 about here -

A similar concern arises from the fact that institutional quality varies only at the regional level, while our dependent variable is at the hospital level. In Table 8 we provide IV estimates in which cesarean section rates are aggregated at the regional level. The magnitude of the IQI coefficient is still comparable to previous estimates.

- Table 8 about here -

6.3 Alternative indicators of institutional quality

The IQI gives different weights to the five domains described in the data section. We conduct two robustness checks. First, we construct a version of the IQI index which gives equal weight to each of the five domains of institutional quality (*IQI_equal_weight*). Second, we construct a version of the IQI index which excludes regional health deficit in the government effectiveness component (*IQI_without_deficit*) to address potential simultaneity concern.

[22] In the Appendix, we report the same 2SLS and CF estimates for unadjusted C-section rates. The IQI coefficients in Table A.3 and Table A.4 are in line respectively with those in Table 5 and Table 6, though risk-adjustment appears to be relevant.

The first two columns in Table 9 show the results obtained using these two alternative indices. The coefficient of *IQI_equal_weight* is negative and statistically significant. It is somewhat higher (in absolute terms) than in Table 5. Since the original IQI assigns a weight higher (lower) than 0.2 to the rule of law (corruption) component, the higher effect of *IQI_equal_weight* is consistent with the evidence in the next section on specific IQI sub-indicators, which shows that corruption (rule of law) has a stronger (negligible) effect on cesarean section rates. The coefficient of *IQI_without_deficit* is in line with the baseline IV estimate.

- Table 9 about here -

The last three columns in Table 9 report IV estimates when we replace the IQI with the European Quality of Government Index, which is a perception-based indicator of government quality. Since EQI is available only for 2010 and 2013, in column (3) we restrict the sample to 2010 and 2012. Instead, in column (4) we estimate the model on the full sample, assigning EQI 2013 to 2011 and 2012, and assigning EQI 2010 to the other years. Finally, in column (5) we estimate the model on the two years sample 2011-2012 in which we assign EQI 2010 to later years, so that in this specification EQI is predetermined with respect to cesarean section rates. All three columns display negative and significant coefficients. The results in column (3) suggest that a standard deviation increase of EQI (equal to 0.88) reduces cesarean section rates by about 9 percentage points, in line with the effect of IQI. Unlike in the context of IQI analysis, the endogeneity tests do not reject the null hypothesis of exogeneity of EQI. This difference might be due to EQI being a perception indicator based on survey data, potentially less exposed to the presence of omitted factors affecting both cesarean sections and institutional quality.

6.4 Specific dimensions of institutional quality

We investigate whether specific dimensions of institutional quality are more relevant than others. Table 10 reports IV estimates when we include each of the five IQI sub-indicators individually (columns 1-5) and simultaneously (column 6). When considered individually, all IQI sub-indicators (corruption, government effectiveness, regulatory quality, rule of law, and voice and accountability) have a negative effect on cesarean section rates with similar order of magnitude. The effect is also statistically significant at 1% level except for rule of law (*IQI_rule*). The coefficients are higher relative to the regression with the composite IQI, probably due to the inclusion of the non-significant indicator on rule of law. The diagnostic tests are in line with the results for the composite IQI. Once we include all sub-indicators in the same specification in column (6)[23], government effectiveness (*IQI_govern*) and corruption (*IQI_corrupt*) are the only significant dimensions. Therefore, the level of corruption and government effectiveness in the region appear to be the most important dimensions of institutional quality in affecting the appropriateness of healthcare provision.

- Table 10 about here -

6.5 Possible mechanisms

In this section we explore some possible mechanisms through which institutional quality affects the provision of healthcare services. The first mechanism relates to regional payment policies. Because in the Italian NHS regional governments are free to set their own DRG tariffs (see Section 2), differences in the quality of institutional environment may translate into different payment policies which, in turn, may affect childbirth delivery (see Section 3). Column (1) in Table 11 reports the IV estimates (based on 2SLS) when our three indicators of regional payment policy are included, i.e. whether the region sets its own tariff as opposed to adopting the national tariff (*Regional_DRG*), and the two tariffs for vaginal delivery and for cesarean section (*DRG_tariff_VD* and *DRG_tariff_CS* in thousands of Euro).

The coefficient of *Regional_DRG* is negative and statistically significant. It suggests that regions who set their own tariffs have lower cesarean section rates. This is consistent with the view that regions that exercise more discretion in setting the tariffs are also more committed to improving the appropriateness of healthcare provision (Francese et al., 2014). The coefficient of *DRG_tariff_VD* is negative and statistically significant, indicating that regions that set a higher tariff for vaginal delivery, and therefore have a lower difference in tariff between vaginal delivery and cesarean section, have fewer cesarean sections.[24] Instead, for a given price for vaginal delivery, the DRG tariff for cesarean section (i.e. *DRG_tariff_CS*) does not affect cesarean sections.[25]

- Table 11 about here -

The coefficient of IQI is still negative and strongly significant, but the coefficient is about half the coefficient estimated in Table 5. This suggests that regional payment policies, and the financial

[23] This specification is estimated through OLS given that we have only two instruments with five potentially endogenous variables.
[24] Lombardy, for instance, sets the same DRG tariff for C-sections and vaginal deliveries.
[25] Estimating an IV model with the tariff difference between cesarean and vaginal delivery produces similar results.

incentives which they provide, are one important mechanism through which institutional quality affects the appropriateness of healthcare provision. To study the extent to which DRG policies and institutional quality are related, we compare in Table 12 the distributions of IQI and IQI sub-indicators in the two groups of regions defined by *Regional_DRG* (i.e. regions applying national tariffs and regions setting their own tariffs). Regions who adopt their own regional tariffs display higher institutional quality, on average, than regions using the national ones. This result holds for all sub-indicators. To formally test for statistically significant differences in the distributions, we run the Mann-Whitney and the Kolmogorov-Smirnov two tails tests. As shown in Table 12, the null hypothesis of equality of distributions can be rejected at conventional levels of significance for IQI as well as for all the sub-indicators, suggesting that higher institutional quality maps into pro-active DRG policies to reduce inappropriate provision.

- Table 12 about here -

A second possible mechanism relates to whether lower institutional quality generates financial constraints for hospitals that influence healthcare provision. To test this we exploit the fact that some regions in the sample period were subject to recovery plans (see Section 4.3) under which they were obliged to reduce year after year their health deficit under strict monitoring by the Ministry of Health. These recovery plans imposed stringent financial restrictions to regional health spending. Hospitals in a tighter financial climate may be under higher pressure to increase revenues through more cesarean sections, and this effect may be exacerbated in regions with lower institutional quality.

To account for regional health deficits, we therefore include the variable *Recovery_Plan* (equal to 1 if region r was subject to recovery plan in year t and 0 otherwise), and its interaction with institutional quality (*IQI*Recovery_Plan*). Column (2) of Table 11 reports the IV estimate once these two variables are included, and in which we also instrument the interaction term exploiting the fact that we have two historical instruments. Both variables on recovery plans are not statistically significant. A similar result is obtained in column (3) where we just control for *Recovery_Plan*, which again is not statistically significant. This indicates that the effect of institutional quality on appropriateness in healthcare provision is not due to greater difficulties for hospitals to obtain resources in regions with lower institutional quality.

6.6 Mother health outcomes

We may wonder if higher inappropriateness of health services through higher cesarean sections affects mothers' health outcomes, e.g. through more post-birth complications, which we measure by the risk-adjusted mother readmissions within 42 days from cesarean delivery (*Mother hospital readmission*) in 2012 (the only available year). Column (1) of Table 13 reports the IV estimate. The coefficient of IQI shows that institutional quality does not affect mothers' hospital readmission rates. This is also supported by the IV estimate in column (2) of Table 13 which shows that the effect of institutional quality on cesarean section rate is robust to controlling for mother readmission rates.

- Table 13 about here –

6.7 Hospital quality indicators

In this sub-section, we test whether higher institutional quality reduces not only appropriateness of care, as measured by cesarean sections above, but also clinical quality in the provision of healthcare services, as measured by hospital risk-adjusted mortality rates for conditions with high mortality risk, such as a heart attack or a stroke. In Table 14 we provide OLS and 2SLS estimates using as dependent variables risk-adjusted 30-days mortality rates for AMI (heart attack), COPD (chronic obstructive pulmonary disease), and Stroke at the hospital level. OLS estimates show that a higher institutional quality is negatively associated with each of the three mortality rates, but it is statistically significant only for AMI and Stroke mortality rates.

The 2SLS estimates instead suggest that institutional quality reduces mortality only for AMI (heart attack). The coefficient of -0.188 implies that one standard deviation increase in IQI reduces AMI mortality by about 4 percentage points. Standard diagnostic tests support the IV strategy. Although statistically not significant, the effect on stroke mortality is comparable in magnitude.

- Table 14 about here –

7. Conclusions

We have investigated the effect of institutional quality on the appropriateness of healthcare provision, as measured by the excess use of cesarean sections in childbirth delivery. We found that higher institutional quality does improve the appropriateness in the provision of childbirth services in Italy, and the effect is both statistically and economically significant.[26] The estimates suggest that a standard deviation increase in institutional quality reduces cesarean section rates by about 10 percentage points, and this result is robust to a range of alternative specifications.

Healthcare services are characterized by asymmetry of information between patients and physicians about the appropriate treatment or course of action. Our findings suggest that institutional quality can play a significant role in shaping provider incentives and in improving the efficiency of health systems. As for the specific dimensions of institutional quality, we find that the level of corruption and government effectiveness appear to be the most relevant in affecting the appropriateness of healthcare provision.

Our findings highlight the importance of paying attention to the quality of the institutional framework in which healthcare providers operate, and call for long term policies enhancing their accountability. This involves expanding the scope of benchmarking across hospitals and local administrations through performance measures, and of public reporting of cesarean section rates in an accessible and user-friendly format. Previous evidence has shown that improving transparency can boost accountability of public officials and, in turn, the efficiency in the use of public resources (Vadlamannati and Cooray, 2016). This is in line with the second generation theory of fiscal federalism in which the effect of devolution rests on the accountability of public officials in the local areas (Bardhan, 2002). Our analysis also suggests that in decentralized systems, the central government could play a coordinating role or introduce minimum standards to contain inappropriateness in healthcare provision.

While our empirical analysis focuses on the healthcare sector, similar mechanisms may apply for other publicly-funded services characterized by high discretion of providers and difficulties in the measurement of performance which blur accountability (e.g., education, procurement). Indeed, whenever formal rules are missing or generic, the institutional framework is important in shaping the rules of the game and, thus, in affecting the behavior of public officials.

[26] Back-of-the-envelope calculations suggest that the costs of the excessive caesarean sections amount to about € 50 million in health expenditure (see Appendix A.2, Table A.9).

References

Acemoglu, D. (1995). Reward structures and the allocation of talent. *European Economic Review*, 39(1), 17-33.

Acemoglu, D., De Feo, G., & De Luca, G. D. (2020). Weak states: Causes and consequences of the Sicilian Mafia. *The Review of Economic Studies*, 87(2), 537-581.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369-1401.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). The rise of Europe: Atlantic trade, institutional change, and economic growth. *American Economic Review*, 95(3), 546-579.

Acemoglu, D., Naidu, S., Restrepo, P., & Robinson, J. A. (2019). Democracy Does Cause Growth. *Journal of Political Economy*, 127(1), 47-100.

Acemoglu, D., & Robinson, J. A. (2008). Persistence of power, elites, and institutions. *American Economic Review*, 98(1), 267-93.

Alesina, A., & Tabellini, G. (2007). Bureaucrats or politicians? Part I: a single policy task. *American Economic Review*, 97(1), 169-179.

Alesina, A., & Tabellini, G. (2008). Bureaucrats or politicians? Part II: Multiple policy tasks. *Journal of Public Economics*, 92(3-4), 426-447.

Almond, G., & Verba, V. (1963). *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton University Press.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53(5), 941-973.

Azfar, O., & Gurgur, T. (2008). Does corruption affect health outcomes in the Philippines?. *Economics of Governance*, 9(3), 197-244.

Baicker, K., Buckles, K. S., & Chandra, A. (2006). Geographic variation in the appropriate use of cesarean delivery. *Health Affairs*, 25(5), w355-w367.

Baldi, S., Bottasso, A., Conti, M., & Piccardo, C. (2016). To bid or not to bid: That is the question: Public procurement, project complexity and corruption. *European Journal of Political Economy*, 43, 89-106.

Banfield, E. C. (1958). *The Moral Basis of a Backward Society*. New York: The Free Press.

Bardhan, P. (2002). Decentralization of governance and development. *Journal of Economic Perspectives*, 16, 185-205.

Barro, R. J. (1991). Economic growth in a cross section of countries. *The Quarterly Journal of Economics*, 106(2), 407-443.

Birkmeyer, J. D., Siewers, A. E., Finlayson, E. V., Stukel, T. A. et al. (2002). Hospital volume and surgical mortality in the United States. *New England Journal of Medicine*, 346(15), 1128-1137.

Bloom, N., Propper, C., Seiler, S., Van Reenen, J. (2015). The impact of competition on management quality: evidence from public hospitals. *The Review of Economic Studies*, 82, 457-489.

Boix, C., & Posner. D. (1998). The Origins and Political Consequences of Social Capital. *British Journal of Political Science*, 28(4), 686-93.

Borcan, O., Olsson, O., & Putterman, L. (2018). State history and economic development: evidence from six millennia. *Journal of Economic Growth*, 23(1), 1-40.

Bosworth, B., & Collins, S. M. (2003). The empirics of growth: An update. *Brookings Papers on Economic Activity*, 2003(2), 113-206.

Boschma, R., Marrocu, E., & Paci, R. (2015). Symmetric and asymmetric effects of proximities. The case of M&A deals in Italy. *Journal of Economic Geography*, 16(2), 505-535.

Bowles, S. (2001). Individual Interactions, Group Conflicts, and the Evolution of Preferences. In: Durlauf, S. N., Young, H. P. (Eds.), *Social Dynamics*, Brookings Institution.

Brekke, K. R., Siciliani, L., & Straume, O. R. (2012). Quality competition with profit constraints. *Journal of Economic Behavior & Organization*, 84(2), 642-659.

Burnside, C., & Dollar, D. (2000). Aid, policies, and growth. *American Economic Review*, 90(4), 847-868.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3), 414-427.

Castro, M. F., Guccio, C., & Rizzo, I. (2014). An assessment of the waste effects of corruption on infrastructure provision. *International Tax and Public Finance*, 21(4), 813-843.

Cavalieri, M., Guccio, C., Lisi, D., & Pignataro, G. (2014). Financial Incentives and Inappropriateness in Health Care: Evidence from Italian Cesarean Sections. *FinanzArchiv: Public Finance Analysis*, 70(3), 430-457.

Cavalieri, M., Guccio, C., & Rizzo, I. (2017). On the role of environmental corruption in healthcare infrastructures: An empirical assessment for Italy using DEA with truncated regression approach. *Health Policy*, 121(5), 515-524.

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

Chalkley, M., & Malcomson, J. M. (1998b). Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics*, 17(1), 1-19.

Chandra, A., Cutler, D., & Song, Z. (2011). Who Ordered That? The Economics of Treatment Choices in Medical Care. In: Mark, V, Pauly, TGM, Barros, PP (Eds.), *Handbook of Health Economics*, Vol. 2, 397-432, Elsevier.

Charron, N., Dijkstra, L., & Lapuente, V. (2014). Regional governance matters: Quality of government within European Union member states. *Regional Studies*, 48(1), 68-90.

Charron, N., Dijkstra, L., & Lapuente, V. (2015). Mapping the regional divide in Europe: A measure for assessing quality of government in 206 European regions. *Social Indicators Research*, 122(2), 315-346.

Charron, N., & Lapuente, V. (2013). Why Do Some Regions in Europe Have a Higher Quality of Government? *The Journal of Politics*, 75(3), 567-582.

Choné, P., & Ma, C. T. A. (2011). Optimal health care contract under physician agency. *Annales d'Économie et de Statistique*, 229-256.

Chubb, J. (1981). The Social Bases of an Urban Political Machine: The Case of Palermo. *Political Science Quarterly*, 96(1), 107-25.

Cooper, Z., Gibbons, S., Jones, S., McGuire, A., (2011). Does hospital competition save lives? Evidence from the NHS patient choice reforms. *The Economic Journal*, 121, 228-260.

Coviello, D., Moretti, L., Spagnolo, G., & Valbonesi, P. (2018). Court efficiency and procurement performance. *The Scandinavian Journal of Economics*, 120(3), 826-858.

Currie, J., & MacLeod, W. B. (2008). First do no harm? Tort reform and birth outcomes. *The Quarterly Journal of Economics*, 123(2), 795-830.

Dafny, L. S. (2005). How do hospitals respond to price changes? *American Economic Review*, 95(5), 1525-1547.

Dal Bó, E., & Rossi, M. A. (2007). Corruption and inefficiency: Theory and evidence from electric utilities. *Journal of Public Economics*, 91(5-6), 939-962.

Del Monte, A., & Papagni, E. (2007). The determinants of corruption in Italy: Regional panel data analysis. *European Journal of Political Economy*, 23(2), 379-396.

Dell, M. (2010). The persistent effects of Peru's mining mita. *Econometrica*, 78(6), 1863-1903.

Di Giacomo, M., Piacenza, M., Siciliani, L., & Turati, G. (2017). Do public hospitals respond to changes in DRG price regulation? The case of birth deliveries in the Italian NHS. *Health Economics*, 26, 23-37.

Di Tella, R., & Schargrodsky, E. (2003). The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires. *The Journal of Law and Economics*, 46(1), 269-292.

Donner, A. (1986). A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *International Statistical Review*, 54(1), 67-82.

Dubay, L., Kaestner, R., & Waidmann, T. (1999). The impact of malpractice fears on cesarean section rates. *Journal of Health Economics*, 18(4), 491-522.

Durlauf, S. N. (2018). Institutions, Development and Growth: Where Does Evidence Stand? *Economic Development and Institutions Working Paper 18/04.1*.

Engerman, S. L., & Sokoloff, K. L. (2012). *Economic Development in the Americas Since 1500: Endowments and Institutions*. New York: Cambridge University Press.

Epstein, A. J., & Nicholson, S. (2009). The formation and evolution of physician treatment styles: an application to cesarean sections. *Journal of Health Economics*, 28(6), 1126-1140.

Easterly, W., & Levine, R. (2003). Tropics, germs, and crops: how endowments influence economic development. *Journal of Monetary Economics*, 50(1), 3-39.

European Commission (2017). *Updated Study on Corruption in the Healthcare Sector*. European Commission, Bruxelles.

Farrar, S., Yi, D., Sutton, M., Chalkley, M., Sussex, J., & Scott, A. (2009). Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis. *British Medical Journal*, 339, b3047.

Foo, P. K., Lee, R. S., & Fong, K. (2017). Physician Prices, Hospital Prices, and Treatment Choice in Labor and Delivery. *American Journal of Health Economics*, 3(3), 422-453.

France, G., Taroni, F., & Donatini, A. (2005). The Italian health-care system. *Health Economics*, 14(S1), S187-S202.

Francese, M., Piacenza, M., Romanelli, M., & Turati, G. (2014). Understanding inappropriateness in health spending: The role of regional policies and institutions in caesarean deliveries. *Regional Science and Urban Economics*, 49, 262-277.

Galizzi, M. M., & Miraldo, M. (2011). The effects of hospitals' governance on optimal contracts: Bargaining vs. contracting. *Journal of Health Economics*, 30(2), 408-424.

Gaynor, M., Seider, H., & Vogt, W. B. (2005). The volume-outcome effect, scale economies, and learning-by-doing. *American Economic Review*, 95(2), 243-247.

Gaynor, M., Moreno-Serra, R., Propper, C. (2013). Death by market power: Reform, competition and patient outcomes in the British National Health Service. *American Economic Journal: Economic Policy*, 5, 134-166.

Goel, R. K., & Nelson, M. A. (1998). Corruption and government size: A disaggregated analysis. *Public Choice*, 97(1-2), 107-120.

Golden, M. A., & Picci, L. (2005). Proposal for a new measure of corruption, illustrated with Italian data. *Economics & Politics*, 17(1), 37-75.

Gowrisankaran, G., & Town, R. (2003). Competition, payers, and hospital quality. *Health Services Research*, 38, 1403-1422.

Gruber, J., Kim, J., & Mayzlin, D. (1999). Physician fees and procedure intensity: the case of cesarean delivery. *Journal of Health Economics*, 18(4), 473-490.

Gruber, J., & Owings, M. (1996). Physician Financial Incentives and Cesarean Section Delivery. *The Rand Journal of Economics*, 27(1), 99-123.

Guccio, C., & Lisi, D. (2016). Thus do all. Social interactions in inappropriate behavior for childbirth services in a highly decentralized healthcare system. *Regional Science and Urban Economics*, 61, 1-17.

Guccio, C., Lisi, D., & Rizzo, I. (2019). When the purchasing officer looks the other way: on the waste effects of debauched local environment in public works execution. *Economics of Governance*, 20(3), 205-236.

Guiso, L., Sapienza, P., & Zingales, L. (2004). The role of social capital in financial development. *American Economic Review*, 94(3), 526-556.

Guiso, L., Sapienza, P., & Zingales, L. (2016). Long-term persistence. *Journal of the European Economic Association*, 14(6), 1401-1436.

Gupta, S., De Mello, L., & Sharan, R. (2001). Corruption and military spending. *European Journal of Political Economy*, 17(4), 749-777.

Hafsteinsdottir, E. J. G., & Siciliani, L. (2010). DRG prospective payment systems: refine or not refine?. *Health Economics*, 19(10), 1226-1239.

Hopkins, K. (2000). Are Brazilian women really choosing to deliver by cesarean?. *Social Science & Medicine*, 51(5), 725-740.

Johnson, E. M., & Rehavi, M. M. (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, 8(1), 115-41.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010), The worldwide governance indicators: a summary of methodology, data and analytical issues. *World Bank Policy Research Working Paper No. 5430*. Washington, DC: World Bank.

Kessler, D., & McClellan, M. (2000). Is hospital competition socially wasteful?. *The Quarterly Journal of Economics*, 115, 577-615.

Kessler, D. P., & Geppert, J. J. (2005). The effects of competition on variation in the quality and cost of medical care. *Journal of Economics & Management Strategy*, 14, 575-589.

Knack, S., & Keefer, P. (1995). Institutions and economic performance: cross-country tests using alternative institutional measures. *Economics & Politics*, 7(3), 207-227.

Kuhn, M., & Siciliani, L. (2013). Manipulation and auditing of public sector contracts. *European Journal of Political Economy*, 32, 251-267.

Lasagni, A., Nifo, A., & Vecchione, G. (2015). Firm productivity and institutional quality: Evidence from Italian industry. *Journal of Regional Science*, 55(5), 774-800.

Li, W. Y., Liabsuetrakul, T., Stray-Pedersen, B., Li, Y. J., Guo, L. J., & Qin, W. Z. (2014). The effects of mode of delivery and time since birth on chronic pelvic pain and health-related quality of life. *International Journal of Gynecology & Obstetrics*, 124(2), 139-142.

Lipset, S. M. (1959). Some Social Requisites of Democracy: Economic Development and Political Legitimacy. *American Political Science Review*, 53(1), 69-105.

Lydon-Rochelle, M., Holt, V. L., Martin, D. P., & Easterling, T. R. (2000). Association between method of delivery and maternal rehospitalization. *Journal of the American Medical Association*, 283(18), 2411-2416.

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

Malcomson, J. M. (2005). Supplier discretion over provision: theory and an application to medical care. *The RAND Journal of Economics*, 412-432.

Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 110(3), 681-712.

Mauro, P. (1998). Corruption and the composition of government expenditure. *Journal of Public Economics*, 69(2), 263-279.

Mehlum, H., Moene, K., & Torvik, R. (2006). Institutions and the resource curse. *The Economic Journal*, 116(508), 1-20.

Ministero della Salute (2017). *Rapporto annuale sull'attivita' di ricovero ospedaliero*. Roma.

Moscelli, G., Gravelle, H., Siciliani, L., & Santos, R. (2018). Heterogeneous effects of patient choice and hospital competition on mortality. *Social Science and Medicine*, 216, 50-58.

Nannicini, T., Stella, A., Tabellini, G., & Troiano, U. (2013). Social capital and political accountability. *American Economic Journal: Economic Policy*, 5(2), 222-50.

Nifo, A., & Vecchione, G. (2014). Do institutions play a role in skilled migration? The case of Italy. *Regional Studies*, 48(10), 1628-1649.

North, D. C. (1981). *Structure and change in economic history*. Norton.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press, Cambridge.

North, D. C. (1991). Institutions. *Journal of Economic Perspectives*, 5(1), 97-112.

OECD (2009). *Health at a Glance 2009: OECD Indicators*. OECD Publishing, Paris

OECD (2017). *Health at a Glance 2017: OECD Indicators*. OECD Publishing, Paris.

Papanicolas, I., & McGuire, A. (2015). Do financial incentives trump clinical guidance? Hip Replacement in England and Scotland. *Journal of Health Economics*, 44, 25-36.

Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, 11(6), 619-632.

Persson, T., & Tabellini, G. (2009). Democratic Capital: The Nexus of Political and Economic Change. *American Economic Journal: Macroeconomics*, 1(2), 88-126

Piattoni, S. (2001). *Clientelism, Interests and Democratic Representation*. Cambridge: Cambridge University Press.

Polidano, C., Zhu, A., & Bornstein, J. C. (2017). The relation between cesarean birth and child cognitive development. *Scientific Reports*, 7(1), 11483.

Putnam, R., Leonardi, R., & Nanetti, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.

Rice, T. (1983). The impact of changing Medicare reimbursement rates on physician-induced demand. *Medical Care*, 21, 803-815.

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of Economic Growth*, 9(2), 131-165.

Roland, G. (2004). Understanding Institutional Change: Fast-Moving and Slow-Moving Institutions. *Studies in Comparative International Development*, 38, 109-131.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. McGraw Hill, New York.

Saaty, T. L. (1992). *The Decision Maker for Leaders*. RWS, Pittsburgh, PA.

Shefter, M. (1977). Party and Patronage: Germany, England, and Italy. *Politics & Society*, 7, 403-51.

Siciliani, L. (2006). Selection of treatment under prospective payment systems in the hospital sector. *Journal of Health Economics*, 25(3), 479-499.

Tabellini, G. (2010). Culture and institutions: economic development in the regions of Europe. *Journal of the European Economic Association*, 8(4), 677-716.

Tan, C. M. (2010). No one true path: uncovering the interplay between geography, institutions, and fractionalization in economic development. *Journal of Applied Econometrics*, 25(7), 1100-1127.

Vadlamannati, K. C., & Cooray, A. (2016). Do freedom of information laws improve bureaucratic efficiency? An empirical investigation. *Oxford Economic Papers*, 68, 968-993.

Villar, J., Valladares, E., Wojdyla, D., Zavaleta, N., Carroli, G., Velazco, A., & Langer, A. (2006). Caesarean delivery rates and pregnancy outcomes: the 2005 WHO global survey on maternal and perinatal health in Latin America. *The Lancet*, 367(9525), 1819-1829.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2), 420-445.

World Health Organization (1985). Appropriate technology for birth. *Lancet*, 2, 436-437.

Wong, H. L., Wang, Y., Luo, R., Zhang, L., & Rozelle, S. (2017). Local governance and the quality of local infrastructure: Evidence from village road projects in rural China. *Journal of Public Economics*, 152, 119-132.

Zamagni, V. (1993). *The Economic History of Italy 1860-1990*. OUP Catalogue, Oxford University Press.

TABLES AND FIGURES

Table 1: Descriptive statistics

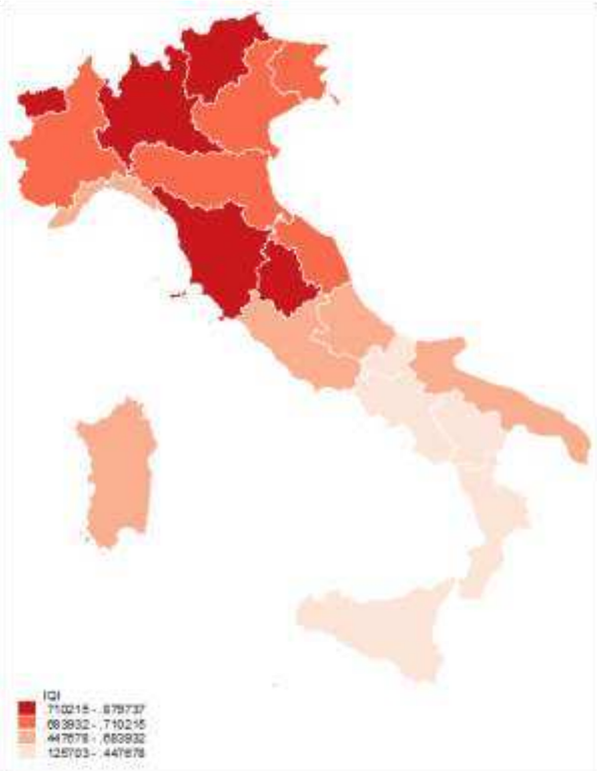
Variables	Obs.	Mean	Std. Dev.	Min	Max
Risk-adjusted cesarean rate (*Cesarean section*)	2952	0.31	0.16	0.03	0.94
Number of beds (*Beds*)	2952	397.97	328.54	25	1719
Number of deliveries (*Births*)	2952	830.97	662.85	90	7313
Autonomous public hospitals (*Public*)	2952	0.18	0.38	0.00	1.00
Local public hospitals (*Local*)	2952	0.57	0.49	0.00	1.00
Private for-profit hospitals (*Private For-Profit*)	2952	0.14	0.35	0.00	1.00
Private not-for-profit hospitals (*Private Not-For-Profit*)	2952	0.04	0.19	0.00	1.00
Teaching hospitals (*Teaching*)	2952	0.05	0.22	0.00	1.00
Research hospitals (*Research*)	2952	0.02	0.12	0.00	1.00
Hospital in regional capital (*Regional Capital*)	2952	0.36	0.48	0.00	1.00
Hospital in provincial capital (*Province Capital*)	2952	0.47	0.50	0.00	1.00
Risk-adjusted mother readmissions (*Mother hospital readmission*)[a]	492	0.78	0.54	0.00	4.43
AMI risk-adjusted mortality rate (*AMI MR*)	333	0.12	0.05	0.02	0.32
COPD risk-adjusted mortality rate (*COPD MR*)	365	0.10	0.05	0.01	0.34
Stroke risk-adjusted mortality rate (*STK MR*)	274	0.13	0.06	0.01	0.44
Institutional Quality Index (*IQI*)	120	0.59	0.21	0.09	0.90
IQI sub-indicator control of corruption (*IQI_corrupt*)	120	0.83	0.17	0.20	0.98
IQI sub-indicator government effectiveness (*IQI_govern*)	120	0.40	0.14	0.07	0.72
IQI sub-indicator regulatory framework (*IQI_regulat*)	120	0.47	0.17	0.10	0.75
IQI sub-indicator rule of law (*IQI_rule*)	120	0.55	0.18	0.13	0.83
IQI sub-indicator voice and accountability (*IQI_voice*)	120	0.44	0.14	0.15	0.72
European quality of Government Index (*EQI*)	120	-0.69	0.88	-2.28	1.04
Proportion of females with only primary school (*Low Education*)	120	0.28	0.04	0.20	0.35
Regional setting DRG tariffs (*Regional_DRG*)	120	0.59	0.49	0.00	1.00
DRG tariff for vaginal delivery (*DRG_tariff_VD*)	120	1543.58	345.54	923.00	2226.00
DRG tariff for cesarean section (*DRG_tariff_CS*)	120	2513.99	465.56	1806.00	3941.00
Regions officially under a recovery plan (*Recovery_Plan*)	120	0.37	0.48	0.00	1.00
Urbanization rate in 1860 ($Urbanization_{1860}$)	20	0.11	0.06	0.02	0.24
Political institutions constraints in 1600-1850 ($Institutions_{1600\text{-}1850}$)	20	-1.19	1.26	-2.09	2.05
Literacy rate in 1881 ($Literacy_{1881}$)	20	0.37	0.17	0.15	0.68

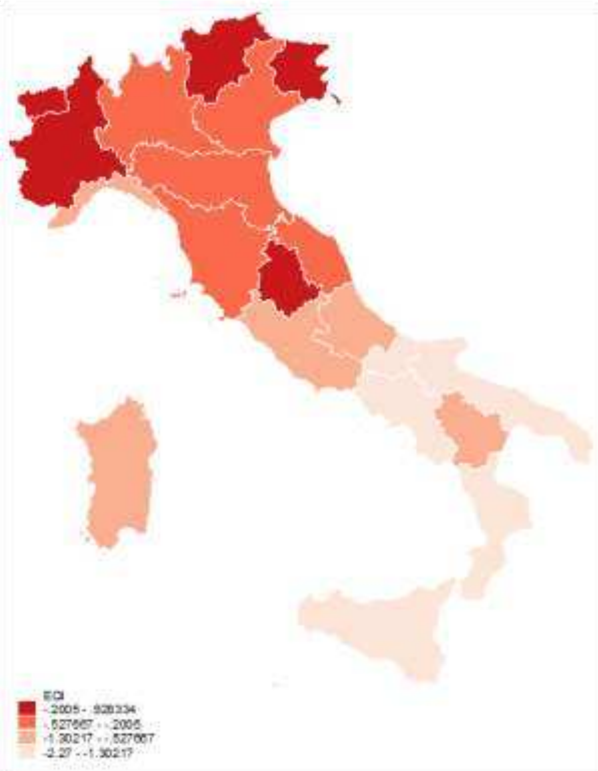
[a] The risk-adjusted mother readmission rates are multiplied by 100, and are available only from 2012.

Figure 1: Geographical distribution of cesarean rates and institutional quality

Average cesarean rates in the years 2007-2012

Average institutional quality in the years 2007-2012





Source: our elaboration on data provided by PNE, Nifo and Vecchione (2014), and Charron et al. (2014, 2015).
Note: The figure shows the geographical distribution of raw and risk-adjusted cesarean rates (above) and the two indicators of institutional quality (below), in the years 2007-2012.

Table 2: Random-effects ANOVA model of Risk-Adjusted Cesarean Section Rates

Source of variation	*SS* [a]	*df* [b]	*MS* [c]	*F-ratio* [d]
Between regions	38.72	19	2.04	136.13*
Within regions	43.89	2932	0.01	
Total	82.61	2951	0.03	
Intra-class correlation (ICC)	0.49			
Standard deviation between regions	0.12			
Standard deviation within regions	0.12			

Note: The table reports the estimates of random-effects ANOVA model with intraclass correlation (ICC) for risk-adjusted cesarean rates computed in Italian hospitals. [a] Sum of squares. [b] Degrees of freedom. [c] Mean squares (*SS*/*df*, the sample variance). [d] F-ratio is the ratio between $MS_{between}$/ MS_{within}. * significant at the 5% level.

Table 3: Pairwise correlation matrix of institutional quality and historical values

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
IQI	1.000								
IQI_corrupt	0.676*	1.000							
IQI_govern	0.800*	0.426*	1.000						
IQI_regulat	0.873*	0.728*	0.651*	1.000					
IQI_rule	0.765*	0.361*	0.371*	0.566*	1.000				
IQI_voice	0.748*	0.641*	0.634*	0.830*	0.330*	1.000			
EQI	0.815*	0.737*	0.737*	0.726*	0.494*	0.575*	1.000		
$Institutions_{1600-1850}$	0.483*	0.266*	0.668*	0.385*	0.109	0.415*	0.704*	1.000	
$Literacy_{1881}$	0.634*	0.348*	0.706*	0.568*	0.243*	0.659*	0.612*	0.683*	1.000

Source: our elaboration on data provided by Nifo and Vecchione (2014), Charron et al. (2014, 2015) and Tabellini (2010).
Note: * coefficients are significant at the 5% level.

Table 4: Risk-Adjusted Cesarean Rates (OLS) – Baseline and reduced-form estimates

	(1)	(2)	(3)	(4)
	Cesarean section	*Cesarean section*	*Cesarean section*	*Cesarean section*
IQI	-0.350			
	(0.087)***			
Institution$_{1600\text{-}1850}$		-0.049		-0.033
		(0.014)***		(0.020)*
Literacy$_{1881}$			-0.384	-0.207
			(0.164)**	(0.224)
Beds	-0.003	0.001	-0.002	0.001
	(0.003)	(0.003)	(0.005)	(0.003)
Births	-0.003	-0.005	-0.004	-0.004
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
Local	-0.018	-0.017	0.031	0.007
	(0.019)	(0.023)	(0.033)	(0.024)
Teaching	0.072	0.042	0.069	0.042
	(0.021)***	(0.018)**	(0.029)**	(0.023)*
Private For-Profit	0.151	0.162	0.178	0.169
	(0.038)***	(0.042)***	(0.038)***	(0.036)***
Private Not-For-Profit	-0.012	0.011	0.014	0.018
	(0.023)	(0.023)	(0.024)	(0.023)
Research	0.053	0.118	0.111	0.124
	(0.061)	(0.081)	(0.077)	(0.084)
Low education	0.468	-1.895	-0.637	-1.229
	(1.014)	(1.123)*	(1.166)	(1.239)
Urbanization$_{1860}$	0.002	0.004	0.004	0.004
	(0.003)	(0.004)	(0.003)	(0.004)
Regional Capital	0.039	0.029	0.045	0.031
	(0.014)***	(0.014)**	(0.014)***	(0.013)**
Province Capital	0.011	0.013	0.008	0.006
	(0.011)	(0.014)	(0.012)	(0.012)
Constant	0.459	0.420	0.490	0.438
	(0.097)***	(0.108)***	(0.114)***	(0.095)***
Year dummies	YES	YES	YES	YES
Observations	2952	2952	2952	2952

Bootstrapped standard errors clustered at the regional level in round brackets.
* significant at 10%, ** significant at 5%, *** significant at 1%.

Table 5: Risk-Adjusted Cesarean Rates (2SLS)

	(1)	(2)	(3)
	Cesarean section	*Cesarean section*	*Cesarean section*
IQI	-0.613	-0.508	-0.544
	(0.245)**	(0.122)***	(0.145)***
Beds	0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.002)
Births	-0.002	-0.003	-0.003
	(0.001)*	(0.001)***	(0.001)***
Local	-0.024	-0.022	-0.023
	(0.027)	(0.023)	(0.021)
Teaching	0.045	0.056	0.052
	(0.022)**	(0.022)**	(0.021)**
Private For-Profit	0.138	0.143	0.141
	(0.046)***	(0.041)***	(0.039)***
Private Not-For-Profit	-0.009	-0.010	-0.010
	(0.037)	(0.020)	(0.019)
Research	0.038	0.044	0.042
	(0.068)	(0.062)	(0.049)
Low education	2.196	1.502	1.742
	(1.892)	(1.293)	(1.306)
Urbanization$_{1860}$	0.001	0.001	0.001
	(0.006)	(0.003)	(0.004)
Regional Capital	0.034	0.039	0.035
	(0.017)**	(0.015)**	(0.013)***
Province Capital	0.003	0.003	0.001
	(0.021)	(0.011)	(0.011)
Constant	0.424	0.438	0.433
	(0.150)***	(0.102)***	(0.111)***
Institution$_{1600-1850}$ [d]	0.079		0.039
	(0.029)***		(0.023)
Literacy$_{1881}$ [d]		0.757	0.552
		(0.337)**	(0.211)**
Year dummies	YES	YES	YES
Endogeneity test[a]	p-value = 0.004	p-value = 0.029	p-value = 0.009
F-statistics[b]	18.660	14.842	14.798
Overidentification test[c]			p-value = 0.163
Observations	2952	2952	2952

[a] The endogeneity test is the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments and one for the equation with the larger set of instruments. Unlike the Hausman tests, this statistic is robust to heteroskedasticity and serial correlation. [b] F-statistic of the Kleibergen-Paap rk Wald test for weak identification. [c] The overidentification test reports the p-value of the Hansen's J statistic of overidentifying restrictions. [d] First-stage coefficients of historical instruments.
Bootstrapped standard errors clustered at the regional level in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 6: Risk-Adjusted Cesarean Rates (GLM and CF)

	(1)	(2)	(3)	(4)
	Cesarean section	*Cesarean section*	*Cesarean section*	*Cesarean section*
IQI	-1.658 [-0.334]	-3.109 [-0.624]	-2.518 [-0.507]	-2.698 [-0.542]
	(0.417)***	(0.810)***	(0.870)***	(0.575)***
Beds	-0.016 [-0.003]	-0.004 [-0.001]	-0.003 [-0.001]	-0.001 [-0.000]
	(0.014)	(0.015)	(0.018)	(0.017)
Births	-0.020 [-0.004]	-0.013 [-0.003]	-0.016 [-0.003]	-0.016 [-0.003]
	(0.005)***	(0.006)**	(0.005)***	(0.005)***
Local	-0.110 [-0.022]	-0.147 [-0.029]	-0.147 [-0.030]	-0.151 [-0.030]
	(0.083)	(0.103)	(0.171)	(0.013)
Teaching	0.365 [0.074]	0.229 [0.046]	0.277 [0.056]	0.263 [0.053]
	(0.104)***	(0.109)**	(0.123)**	(0.118)**
Private For-Profit	0.624 [0.126]	0.560 [0.112]	0.581 [0.117]	0.574 [0.115]
	(0.127)***	(0.147)***	(0.141)***	(0.138)***
Private Not-For-Profit	-0.036 [-0.007]	-0.027 [-0.005]	-0.032 [-0.006]	-0.032 [-0.006]
	(0.131)	(0.114)	(0.097)	(0.112)
Research	0.291 [0.059]	0.222 [0.045]	0.237 [0.048]	0.234 [0.047]
	(0.326)	(0.446)	(0.445)	(0.418)
Low education	2.636 [0.531]	12.664 [2.543]	8.464 [1.703]	9.813 [1.971]
	(6.429)	(7.476)*	(6.782)	(5.092)*
$Urbanization_{1860}$	0.007 [0.001]	-0.001 [-0.000]	0.004 [0.001]	0.002 [0.001]
	(0.013)	(0.015)	(0.015)	(0.013)
Regional Capital	0.194 [0.039]	0.168 [0.034]	0.174 [0.035]	0.172 [0.034]
	(0.061)***	(0.049)***	(0.059)***	(0.061)***
Province Capital	0.066 [0.013]	0.061 [0.012]	0.022 [0.005]	0.015 [0.003]
	(0.062)	(0.063)	(0.068)	(0.056)
$\hat{e}_{First\ Stage}$		2.069 [0.415]	1.425 [0.287]	1.848 [0.371]
		(0.990)**	(1.142)	(0.831)**
Constant	-0.136	-0.366	-0.269	-0.305
	(0.529)	(0.424)	(0.473)	(0.392)
Year dummies	YES	YES	YES	YES
Observations	2952	2952	2952	2952

Note: Bootstrapped standard errors clustered at the regional level in round brackets. Marginal effects (at means) in square brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 7: Risk-Adjusted Cesarean Rates (Cross-section)

	(2007)	(2008)	(2009)	(2010)	(2011)	(2012)
	Cesarean section	*Cesarean section*	*Cesarean section*	*Cesarean section*	*Cesarean section*	*Cesarean section*
IQI	-0.648	-0.516	-0.490	-0.536	-0.558	-0.584
	(0.235)***	(0.113)***	(0.112)***	(0.120)***	(0.156)***	(0.222)***
Other controls	YES	YES	YES	YES	YES	YES
Endogeneity test[a]	p-value = 0.010	p-value = 0.037	p-value = 0.051	p-value = 0.008	p-value = 0.008	p-value = 0.005
F-statistics[b]	11.661	16.325	14.659	14.942	16.817	10.381
Overidentification test[c]	p-value = 0.260	p-value = 0.315	p-value = 0.078	p-value = 0.522	p-value = 0.127	p-value = 0.205
Observations	492	492	492	492	492	492

[a] The endogeneity test is the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments and one for the equation with the larger set of instruments. Unlike the Hausman tests, this statistic is robust to heteroskedasticity and serial correlation. [b] F-statistic of the Kleibergen-Paap rk Wald test for weak identification. [c] The overidentification test reports the p-value of the Hansen's J statistic of overidentifying restrictions.
Bootstrapped standard errors clustered at the regional level in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 8: Risk-Adjusted Cesarean Rates (Regional level)

	(1)	(2)	(3)
	Cesarean section	*Cesarean section*	*Cesarean section*
IQI	-0.644	-0.595	-0.633
	(0.102)***	(0.131)***	(0.091)***
Other controls	YES	YES	YES
Excluded instrument:			
Institution$_{1600\text{-}1850}$	YES		YES
Literacy$_{1881}$		YES	YES
Endogeneity test[a]	p-value = 0.020	p-value = 0.020	p-value = 0.009
F-statistics[b]	107.729	23.534	63.577
Overidentification test[c]			p-value = 0.478
Observations	120	120	120

[a] The endogeneity test is the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments and one for the equation with the larger set of instruments. Unlike the Hausman tests, this statistic is robust to heteroskedasticity and serial correlation. [b] F-statistic of the Kleibergen-Paap rk Wald test for weak identification. [c] The overidentification test reports the p-value of the Hansen's J statistic of overidentifying restrictions.
Bootstrapped standard errors clustered at the regional level in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 9: Risk-Adjusted Cesarean Rates (alternative indicators)

	(1)	(2)	(3)	(4)	(5)
	Cesarean section	*Cesarean section*	*Cesarean section*	*Cesarean section*	*Cesarean section*
IQI_equal_weight [d]	-0.883				
	(0.401)**				
IQI_without_deficit [e]		-0.634			
		(0.272)***			
EQI			-0.101		
			(0.014)***		
EQI [f]				-0.107	
				(0.013)***	
EQI [g]					-0.107
					(0.016)***
Other controls	YES	YES	YES	YES	YES
Endogeneity test[a]	p-value = 0.012	p-value = 0.007	p-value = 0.441	p-value = 0.348	p-value = 0.135
F-statistics[b]	15.901	12.102	22.940	25.870	20.194
Overidentification test[c]	p-value = 0.318	p-value = 0.214	p-value = 0.145	p-value = 0.197	p-value = 0.846
Observations	2952	2952	984	2952	984

[a] The endogeneity test is the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments and one for the equation with the larger set of instruments. Unlike the Hausman tests, this statistic is robust to heteroskedasticity and serial correlation. [b] F-statistic of the Kleibergen-Paap rk Wald test for weak identification. [c] The overidentification test reports the p-value of the Hansen's J statistic of overidentifying restrictions. [d] IQI_equal_weight is the Istitutional Quality Index computed assigning equal weight to the five components of institutional quality. [e] IQI_without_deficit is the Institutional Quality Index computed without including the regional health deficit in the Government Effectiveness. [f] In this specification we assign EQI 2013 to 2011 and 2012 and EQI 2010 to the other years in the sample period. [g] In this specification we assign EQI 2010 to 2011 and 2012 and estimate the model in the two years (2011-2012) sample.
Bootstrapped standard errors clustered at the regional level in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 10: Risk-Adjusted Cesarean Rates (sub-indicators)

	(1)	(2)	(3)	(4)	(5)	(6)[d]
	Cesarean section	*Cesarean section*	*Cesarean section*	*Cesarean section*	*Cesarean section*	*Cesarean section*
IQI_corrupt	-0.707					-0.135
	(0.230)***					(0.052)**
IQI_govern		-0.680				-0.379
		(0.135)***				(0.081)***
IQI_regulat			-0.740			-0.159
			(0.275)***			(0.122)
IQI_rule				-0.792		0.119
				(2.493)		(0.076)
IQI_voice					-0.842	-0.004
					(0.302)***	(0.116)
Other controls	YES	YES	YES	YES	YES	YES
Endogeneity test[a]	p-value = 0.010	p-value = 0.032	p-value = 0.005	p-value = 0.172	p-value = 0.019	
F-statistics[b]	4.963	26.571	15.967	1.233	18.263	
Overidentification test[c]	p-value = 0.377	p-value = 0.879	p-value = 0.363	p-value = 0.015	p-value = 0.087	
Observations	2952	2952	2952	2952	2952	2952

[a] The endogeneity test is the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments and one for the equation with the larger set of instruments. Unlike the Hausman tests, this statistic is robust to heteroskedasticity and serial correlation. [b] F-statistic of the Kleibergen-Paap rk Wald test for weak identification. [c] The overidentification test reports the p-value of the Hansen's J statistic of overidentifying restrictions. [d] This specification is estimated by OLS. Bootstrapped standard errors clustered at the regional level in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 11: Risk-Adjusted Cesarean Rates (possible mechanisms)

	(1)	(2)	(3)
	Cesarean section	*Cesarean section*	*Cesarean section*
IQI	-0.299	-0.553	-0.508
	(0.099)***	(0.191)***	(0.155)***
Regional_DRG	-0.099		
	(0.021)***		
DRG_tariff_VD	-0.041		
	(0.020)**		
DRG_tariff_CS	0.012		
	(0.016)		
*IQI*Recovery_Plan* [d]		0.116	
		(0.307)	
Recovery_Plan		-0.057	0.007
		(0.148)	(0.050)
Other controls	YES	YES	YES
Endogeneity test[a]	p-value = 0.009	p-value = 0.036	p-value = 0.013
F-statistics[b]	10.925	5.259	10.481
Overidentification test[c]	p-value = 0.248		p-value = 0.249
Observations	2952	2952	2952

[a] The endogeneity test is the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments and one for the equation with the larger set of instruments. Unlike the Hausman tests, this statistic is robust to heteroskedasticity and serial correlation. [b] F-statistic of the Kleibergen-Paap rk Wald test for weak identification. [c] The overidentification test reports the p-value of the Hansen's J statistic of overidentifying restrictions. [d] IQI*Recovery_Plan is the interaction term between IQI and Recovery_Plan (equal to 1 if the region is officially under a recovery plan and 0 otherwise).
Bootstrapped standard errors clustered at the regional level in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 12: DRG and IQI – Mann-Whitney and Kolmogorov-Smirnov tests

	(1)	(2)	(3)	(4)	(5)	(6)
	IQI	*IQI_corrupt*	*IQI_govern*	*IQI_regulat*	*IQI_rule*	*IQI_voice*
National tariff						
Mean	0.421	0.673	0.282	0.332	0.506	0.357
Sd	0.151	0.247	0.098	0.166	0.118	0.170
Regional tariff						
Mean	0.642	0.886	0.440	0.515	0.566	0.463
Sd	0.200	0.086	0.135	0.141	0.194	0.119
MV	-5.442	-5.467	-4.915	-4.558	-1.891	-3.097
	(0.000)	(0.000)	(0.000)	(0.000)	(0.059)	(0.002)
KS	0.622	0.544	0.611	0.556	0.300	0.578
	(0.000)	(0.000)	(0.000)	(0.000)	(0.035)	(0.000)

Note: Mann–Whitney (MW) test; Kolmogorov–Smirnov (KS) two-tails test. Both tests formally assess the statistical difference in the distribution of IQI indicators between regions with national tariffs and regions with regional tariffs. *p-values* in round brackets.

Table 13: Risk-Adjusted Cesarean Rates (mother health outcomes)

	(1)	(2)
	Mother hospital readmission [d]	*Cesarean section*
IQI	0.843	-0.567
	(1.252)	(0.270)**
Mother hospital readmission [d]		-0.017
		(0.023)
Other controls	YES	YES
Endogeneity test[a]	p-value = 0.171	p-value = 0.006
F-statistics[b]	7.818	10.481
Overidentification test[c]	p-value = 0.098	p-value = 0.227
Observations	492	492

[a] The endogeneity test is the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments and one for the equation with the larger set of instruments. Unlike the Hausman tests, this statistic is robust to heteroskedasticity and serial correlation. [b] F-statistic of the Kleibergen-Paap rk Wald test for weak identification. [c] The overidentification test reports the p-value of the Hansen's J statistic of overidentifying restrictions. [d] Mother hospital readmission is the risk-adjusted mother readmission rate within 42 days from cesarean delivery.Bootstrapped standard errors clustered at the regional level in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

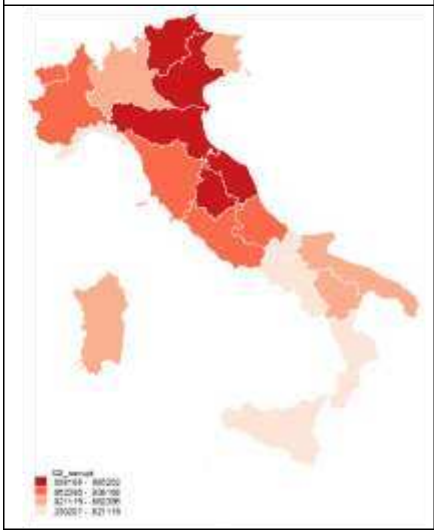
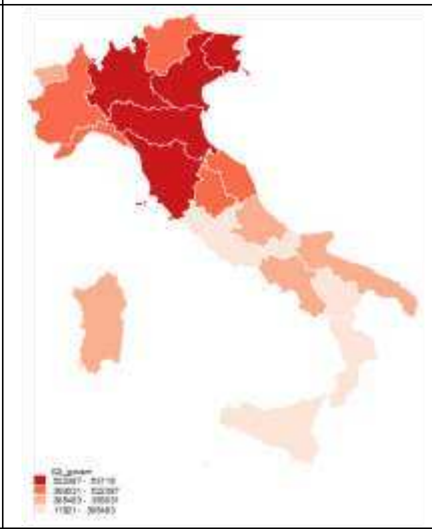
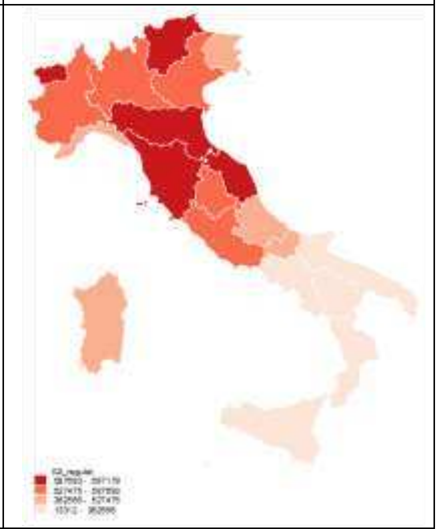
Table 14: Risk-Adjusted Mortality Rates (OLS and 2SLS) – 2012

	AMI MR *Acute Myocardial Infarction*		*COPD MR* *Chronic Obstructive Pulmonary Disease*		*STK MR* *Stroke*	
	OLS	*2SLS*	*OLS*	*2SLS*	*OLS*	*2SLS*
IQI	-0.091	-0.188	-0.028	-0.100	-0.088	-0.102
	(0.026)***	(0.061)***	(0.046)	(0.110)	(0.042)**	(0.099)
Other controls	YES	YES	YES	YES	YES	YES
Endogeneity test[a]		p-value = 0.003		p-value = 0.123		p-value = 0.417
F-statistics[b]		9.404		9.912		8.637
Overidentification test[c]		p-value = 0.538		p-value = 0.034		p-value = 0.055
Observations	385	385	365	365	294	294

[a] The endogeneity test is the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments and one for the equation with the larger set of instruments. Unlike the Hausman tests, this statistic is robust to heteroskedasticity and serial correlation. [b] F-statistic of the Kleibergen-Paap rk Wald test for weak identification. [c] The overidentification test reports the p-value of the Hansen's J statistic of overidentifying restrictions.
Bootstrapped standard errors clustered at the regional level in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

APPENDIX

Figure A.1 – Geographical distribution of institutional quality sub-indicators

IQI_corrupt	*IQI_govern*	*IQI_regulat*
		

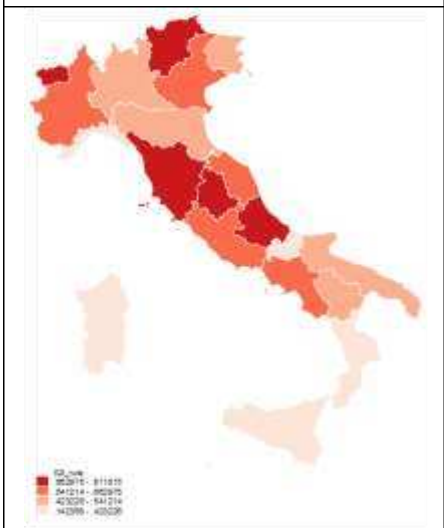
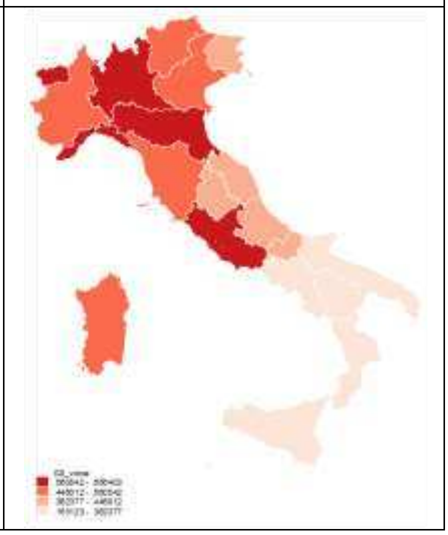
IQI_rule	*IQI_voice*
	

Table A.1: Random intercept model of Risk-Adjusted Cesarean Rates [a]

Controlling for hospital characteristics	
Source of variation	*Covariate-adjusted variance* [b]
Between regions	0.008
Within regions	0.004
Total	0.012
Covariate-adjusted intra-class correlation (ICC)	0.67
Standard deviation between regions	0.09
Standard deviation within regions	0.06
Controlling for hospital characteristics and IQI	
Source of variation	*Covariate-adjusted variance* [b]
Between regions	0.006
Within regions	0.006
Total	0.012
Covariate-adjusted intra-class correlation (ICC)	0.67
Standard deviation between regions	0.08
Standard deviation within regions	0.08

[a] The estimated coefficients of hospitals' observable characteristics are not reported for brevity, but are available upon request. [b] Residual variance after controlling for hospitals' observed characteristics.

Table A.2: Risk-Adjusted Cesarean Rates (wild bootstrap)

	(1)	(2)	(3)	(4)
	Cesarean section	*Cesarean section*	*Cesarean section*	*Cesarean section*
IQI	-0.350	-0.613	-0.508	-0.544
	(0.061)***	(0.098)***	(0.103)***	(0.076)***
Institution$_{1600\text{-}1850}$		0.079		0.039
		(0.016)***		(0.022)
Literacy$_{1881}$			0.757	0.552
			(0.184)***	(0.188)**
Other controls	YES	YES	YES	YES
Observations	2952	2952	2952	2952

Wild bootstrapped standard errors clustered at the regional level (command *cgmwildboot*) in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.3: Unadjusted Cesarean Rates (OLS and 2SLS)

	(1)	(2)	(3)	(4)
	Cesarean section	*Cesarean section*	*Cesarean section*	*Cesarean section*
IQI	-0.278	-0.505	-0.420	-0.449
	(0.079)***	(0.153)***	(0.117)***	(0.111)***
$Institution_{1600\text{-}1850}$		0.079		0.039
		(0.018)***		(0.023)
$Literacy_{1881}$			0.757	0.552
			(0.196)***	(0.211)**
Year dummies	YES	YES	YES	YES
Endogeneity test[a]		p-value = 0.003	p-value = 0.016	p-value = 0.006
F-statistics[b]		18.660	14.842	14.798
Overidentification test[c]				p-value = 0.206
Observations	2952	2952	2952	2952

[a] The endogeneity test is the difference of two Sargan-Hansen statistics: one for the equation with the smaller set of instruments and one for the equation with the larger set of instruments. Unlike the Hausman tests, this statistic is robust to heteroskedasticity and serial correlation. [b] F-statistic of the Kleibergen-Paap rk Wald test for weak identification. [c] The overidentification test reports the p-value of the Hansen's J statistic of overidentifying restrictions. Bootstrapped standard errors clustered at the regional level in round brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.4: Unadjusted Cesarean Rates (GLM and CF)

	(1)	(2)	(3)	(4)
	Cesarean section	*Cesarean section*	*Cesarean section*	*Cesarean section*
IQI	-1.346 [-0.269]	-2.561 [-0.511]	-2.097 [-0.419]	-2.241 [-0.447]
	(0.412)***	(0.595)***	(0.765)***	(0.485)***
$\hat{e}_{First\ Stage}$		1.748 [0.349]	1.250 [0.249]	1.602 [0.320]
		(0.775)**	(1.037)	(0.566)***
Year dummies	YES	YES	YES	YES
Observations	2952	2952	2952	2952

Bootstrapped standard errors clustered at the regional level in round brackets. Marginal effects (at means) in square brackets. * significant at 10%, ** significant at 5%, *** significant at 1%.

Appendix A.1 Risk adjustment

In this Appendix we list the variables used in the risk adjustment of cesarean section rates, maternal hospital readmission rates, and 30-day mortality rates for AMI, COPD, and Stroke. A detailed description of the methods can be found (in Italian) on the PNE website, at the following urls:

https://pne.agenas.it/main/doc/metodi_statistici.pdf

http://82.112.223.85/PNEed14_EN/main/doc/introduzione.pdf

Cesarean section

The following variables that relate to the mother or the fetus available from mothers' discharge records are included in the risk adjustment:

- Maternal age (≤17 years old, 18-24, 25-28, 29-33, 34-38, ≥39 years)
- Maternal and neonatal clinical factors and comorbidities, measured at the delivery admission or during hospitalizations in the previous two years. These factors, listed in Table A.5, are identified using primary and secondary diagnoses.

The full operative protocol is available at:
http://82.112.223.85/PNEed14_EN/risultati/protocolli/pro_37.pdf

Table A.5: Full list of maternal and neonatal clinical factors and comorbidities

Risk factor	**ICD-9-CM code**	
	During hospitalization for childbirth	**During previous hospitalizations**
Cancer	140.0–208.9	140.0–208.9
Anemias	280-284, 285 (excl.285.1), 648.2	280-284, 285 (excl.285.1)
Coagulation defects	286	286
Heart diseases	390-398, 410-429	390-398, 410-429
Cardiovascular diseases in pregnancy	648.5, 648.6	
Congenital anomalies of heart and circulatory system	745-747	745-747
Cerebrovascular disease	433, 437, 438	430-432, 433, 434, 436, 437, 438
Nephritis, nephritic syndrome and nephrosis	580-589	580-589
Unspecified renal disease in pregnancy, without mention of hypertension	646.2	
Diffuse diseases of connective tissue	710	710
HIV	042, 079.53, V08	042, 079.53, V08
Disorders of thyroid gland	240-246, 648.1	240-246
Diabetes	250.0-250.9, 648.0	250.0-250.9
Hypertension	401-405, 642.0-642.3, 642.9	401-405
Pre-eclampsia / eclampsia	642.4-642.7	
COPD	491-492, 494, 496	491-492, 494, 496
Asthma	493	493
Cystic fibrosis	277.0	277.0
Acute pulmonary diseases	480-487, 510-514	

Chronic pulmonary diseases	500-508, 515-517	500- 508, 515-517
Tuberculosis	010-018, 647.3	010-018
Genital herpes	054.1	
Other sexually transmitted diseases	077.98, 078.88, 079.88, 079.98, 090- 099, 647.0- 647.2	
Antepartum hemorrhage, abruptio placentae, and placenta previa	641	
Liver disorders in pregnancy	646.7	
Polyhydramnios, oligohydramnios / infection of the amniotic cavity	657, 658.0, 658.4	
Premature rupture of membranes	658.1	
Cord prolapse	663.0	
Malposition and malpresentation of fetus	652	
Fetopelvic disproportion/excessive development of the infant	653, 656.60, 656.61, 656.63	
Fetal abnormality	655	
Intrauterine growth retardation	656.5, 764	
Fetal distress	656.3, 768	
Multiple pregnancy	651, V27.2 –V27.9, V31-V37, 761.5	
Rh isoimmunization	656.1	
Maternal conditions affecting fetus or newborn	760.0, 760.1, 760.3	
Alcohol or drug dependence/abuse	303-305; 648.3 (excl.648.32/ 648.34)	
High risk pregnancy	640, 644.0, V23.0, V23.2, V23.4, V23.5, V23.7, V23.8	
Assisted fertilization	V26	

30-day mortality following acute myocardial infarction (AMI)

The variables included in the risk adjustment are available from patients' discharge records and include gender, age, and a list of comorbidities assessed during the AMI episode (within 28 days from the date of index hospitalization) and during hospitalizations occurred in the previous two years listed in Table A.6. The full operative protocol is available at: http://82.112.223.85/PNEed14_EN/risultati/protocolli/pro_1.pdf

Table A.6: Full list of patient's clinical factors and comorbidities - AMI

Risk factor	**ICD-9-CM Code**	
	Episode of AMI	**Previous hospitalizations**
Cancer	140.0–208.9, V10	140.0–208.9, V10
Diabetes	250.0-250.9	
Lipid metabolism disturbances	272	
Obesity	278.0	278.0
Blood disorders	280-285, 288, 289	280-285, 288, 289
Hypertension	401-405	
Previous myocardial infarction	412	410, 412
Other forms of ischemic heart disease	411, 413, 414	
Heart failure	428	
Ill-defined descriptions and complications of heart disease	429	
Rheumatic heart disease	393-398	391, 393-398
Cardiomyopathy	425	425
Acute endocarditis and myocarditis	421, 422	
Other heart conditions	745, V15.1, V42.2, V43.2, V43.3, V45.0	745, V15.1, V42.2, V43.2, V43.3, V45.0
Conduction disturbances and arrhythmias	426, 427	
Cerebrovascular disease	433, 437, 438	430-432, 433, 434, 436, 437, 438
Vascular disease	440-448 (escluso 441.1, 441.3, 441.5, 441.6, 444)	440-448, 557
Chronic obstructive pulmonary disease (COPD)	491-492, 494, 496	491-492, 494, 496
Chronic renal disease	582-583, 585-588	582-583, 585-588
Chronic diseases (liver, pancreas, intestine)	571-572, 577.1-577.9, 555, 556	571-572, 577.1-577.9, 555, 556
Previous coronary artery bypass graft	V45.81	36.1, V45.81
Previous coronary angioplasty	V45.82	00.66, 36.0, V45.82
Cerebral revascularization procedures	00.61, 00.62, 38.01, 38.02, 38.11, 38.12, 38.31, 38.32	
Other cardiac operations	35, 37.0, 37.1, 37.3, 37.4, 37.5, 37.6, 37.9	
Other vascular operations	38-39.5, excluding: 38.01, 38.02, 38.5, 38.11, 38.12, 38.31, 38.32, 38.93	

30-day mortality following chronic obstructive pulmonary disease (COPD)

The variables included in the risk adjustment are available from patients' discharge records and include gender, age, and a list of comorbidities assessed during the COPD episode (within 28 days from the date of index hospitalization) and during hospitalizations occurred in the previous two years listed in Table A.7. The full operative protocol is available at:

http://82.112.223.85/PNEed14_EN/risultati/protocolli/pro_21.pdf

Table A.7: Full list of patient's clinical factors and comorbidities - COPD

Risk factor	**ICD-9-CM Code**	
	Index hospitalization	**Previous hospitalizations**
Cancer	140.0–208.9, V10	140.0–208.9, V10
Diabetes	250.0-250.9	
Lipid metabolism disturbances	272	
Obesity	278.0	278.0
Blood disorders	280, 281, 285.9, 286, 287.1, 287.3-287.5	280, 281, 285.9, 286, 287.1, 287.3-287.5
Hypertension	401-405	
Disorders of thyroid gland	240-245 (excl. 245.0 245.1), 246	240-245 (excl. 245.0 245.1) 246
Ischemic diseases	412, 414	410-414 414, 429.7
Heart failure and Chronic pulmonary heart disease	428, 416.9	428, 416.9
Neurological and muscular diseases	331, 332, 333.4, 333.5, 334-335, 336.2, 340, 341, 342, 343, 344, 345, 348.1, 348.3, 356, 358, 359, 784.3	331, 332, 333.4, 333.5, 334-335, 336.2, 340, 341, 342, 343, 344, 345, 348.1, 348.3, 356, 358, 359, 784.3
Systemic diseases	446, 701.0, 710, 711.2, 714, 719.3, 720, 725	446, 701.0, 710, 711.2, 714, 719.3, 720, 725
Diseases of the digestive system	456.0- 456.2, 571-572 (excl. 571.1, 572.0- 572.2), 573.0, V42.7, 577.1-577.9, 555, 556	456.0- 456.2, 571-572 (excl. 571.1, 572.0- 572.2), 573.0, V42.7, 577.1-577.9, 555, 556
Other heart conditions	393-398, 423 (excl.423.0), 424, 425, 745, 746.3-746.6, V15.1, V42.2, V43.2, V43.3, V45.0, V45.81, V45.82	093.2, 391, 393-398, 420-425, 429 (excl.429.7), 745, 746.3-746.6, V15.1, V42.2, V43.2, V43.3, V45.0, V45.81, V45.82 procedures: 00.66, 35, 36.1, 36.0,37.0, 37.1, 37.3, 37.4,37.5, 37.6, 37.9
Conduction disturbances and arrhythmias	426.0, 426.10, 426.12, 426.13, 426.7, 426.9, 427 (excl. 427.1, 427.2, 427.5), 785.0, V45.0, V53.3	426.0, 426.10, 426.12, 426.13, 426.7, 426.9, 427, 785.0, 996.01, 996.04, V45.0, V53.3
Cerebrovascular disease	433, 437, 438, 440-448 (excl. 441.1, 441.3, 441.5, 441.6, 444), 557.1, 093.0	430-438, 440-448, 557, 093.0 procedures: 38.01, 38.02, 38.11, 38.12, 38.31, 38.32, 38.41, 38.42, 38.61, 38.62, 38.81, 38.82, 38.08, 38.18, 38.38, 38.48, 38.68, 38.88, 39.29, 38.04-38.07, 38.14-38.16, 38.34-38.37, 38.44-38.47, 38.55, 38.57, 38.64-38.67, 38.7, 38.84-38.87, 30.0, 39.1, 39.21-39.26, 39.52, 39.54

Mental and psychiatric disorders	293.8, 295-298, 299.1, 300.4, 301.12, 309.0, 309.1, 311, 290.0-290.4, 294.1, 331.0	293.8, 295-298, 299.1, 300.4, 301.12, 309.0, 309.1, 311, 290.0-290.4, 294.1, 331.0
Acute chronic respiratory conditions	518.81, 518.82, 786.0, 512, 518.0, 415, 466.0, 480-486, 487.0, 510, 511, 513, 011, 012.0, 012.1, 012.2, 012.8	
Chronic respiratory diseases	493, 495, 135, 500-505, 506-508, 515, 516, 517, 518.1-518.3, 518.89, 519	493, 495, 135, 500-505, 506-508, 515, 516, 517, 518.1-518.3, 518.89, 519
Chronic renal disease	582-583, 585-588, V42.0, V45.1, V56 procedures 38.95, 39.95, 54.98	582-583, 585-588, V42.0, V45.1, V56 procedures 38.95, 39.95, 54.98

30-day mortality following stroke (STK)

The variables included in the risk adjustment are available from patients' discharge records and include gender, age, and a list of comorbidities assessed during the stroke episode (within 28 days from the date of index hospitalization) and during hospitalizations occurred in the previous two years listed in Table A.8. The full operative protocol is available at: http://82.112.223.85/PNEed14_EN/risultati/protocolli/pro_18.pdf

Table A.8: Full list of patient's clinical factors and comorbidities - Stroke

Risk factor	**ICD-9-CM Code**	
	Index hospitalization	**Previous hospitalizations**
Cancer	V10	140.0–208.9, V10
Diabetes	250.0-250.9	
Lipid metabolism disturbances	272	
Obesity	278.0	278.0
Blood disorders	280-285, 288, 289	280-285, 288, 289
Hypertension	401-405	
Previous myocardial infarction	412	410, 412
Other forms of ischemic heart disease	411, 413, 414	
Heart failure	428	
Ill-defined descriptions and complications of heart disease	429	
Rheumatic heart disease	393-398	391, 393-398
Cardiomyopathy	425	425
Acute endocarditis and myocarditis	421, 422	
Other heart conditions	745, V15.1, V42.2, V43.2, V43.3, V45.0	745, V15.1, V42.2, V43.2, V43.3, V45.0
Conduction disturbances and arrhythmias	426, 427	
Cerebrovascular disease	430-432, 433, 434, 436, 437, 438	
Vascular disease	440-448 (escluso 441.1, 441.3, 441.5, 441.6, 444)	440-448, 557
Chronic obstructive pulmonary disease (COPD)	491-492, 494, 496	491-492, 494, 496
Chronic renal disease	582-583, 585-588	582-583, 585-588
Chronic diseases (liver, pancreas, intestine)	571-572, 577.1-577.9, 555, 556	571-572, 577.1-577.9, 555, 556
Previous coronary artery bypass graft	V45.81	36.1, V45.81
Previous coronary angioplasty	V45.82	00.66, 36.0, V45.82
Cerebral revascularization procedures	00.61, 00.62, 38.01, 38.02, 38.11, 38.12, 38.31, 38.32	
Other cardiac operations	35, 37.0, 37.1, 37.3, 37.4, 37.5, 37.6, 37.9	
Other vascular operations	38-39.5, excluding: 38.01, 38.02,	

Appendix A.2 Institutional quality and potential savings

We present some back-of-the-envelope calculations to provide a rough estimate of the potential annual saving in health expenditure resulting from an improvement in institutional quality. Our computations are illustrated in Table A.9. The first columns display for each region the institutional quality (as measured by IQI), the number of births and the number of cesarean sections in 2012. Using the national DRG tariffs, we compute the regional expenditures for childbirth provision, which lead to a national annual expenditure of about € 817 million. We then compute, based on our estimates in Table 5, the number of cesarean sections and associated health expenditure which would arise as a result of an increase of one standard deviation in institutional quality. Overall, the total potential saving in health expenditure for childbirth provision would amount to about € 50 million.

As an alternative scenario, we compute the potential savings in health expenditure that would arise if regions with a relatively low institutional quality improve quality towards the average value. To this end, we assign the average IQI (0.586) to regions displaying an IQI score below average. In this case, the total annual saving in health expenditure would be approximately equal to € 20 million. Overall, the back-of-the-envelope calculations suggest that an improved institutional quality would entail considerable savings in health expenditure.

Table A.9: IQI and potential savings

			True IQI		St. Dev. Increase IQI[a]			Improved IQI[b]		
Regions	**IQI**	**Births**	**C-sections**	**Expenditure**	**C-sections**	**Expenditure**	**Saving**	**C-sections**	**Expenditure**	**Saving**
Abruzzo	0.725	7938	2304	17382671	1510	16332124	1050547	2304	17382671	0
Basilicata	0.417	3097	700	6518677	390	6108807	409869	426	6155533	363144
Calabria	0.092	11312	2634	23911846	1503	22414771	1497075	206	20699014	3212832
Campania	0.361	33790	15932	82098894	12553	77626990	4471904	11992	76884654	5214240
Emilia-Romagna	0.727	29648	5893	61333856	2928	57410121	3923735	5893	61333856	0
Friuli-Venezia Giulia	0.727	7670	1291	15557951	524	14542872	1015078	1291	15557951	0
Lazio	0.679	36104	11350	80213196	7740	75435048	4778148	11350	80213196	0
Liguria	0.547	8365	2384	18259668	1548	17152611	1107058	2187	17998403	261266
Lombardy	0.712	70275	14357	145895002	7330	136594527	9300475	14357	145895002	0
Marche	0.733	9772	2373	20785644	1396	19492378	1293266	2373	20785644	0
Molise	0.256	1427	464	3190803	321	3001948	188855	223	2871449	319354
Piedmont	0.709	26361	5751	55210332	3115	51721612	3488720	5751	55210332	0
Prov. Auton. Bozen	0.864	4334	816	8905761	383	8332183	573579	816	8905761	0
Prov. Auton. Trento	0.864	3772	652	7673809	275	7174608	499202	652	7673809	0
Puglia	0.419	24346	7926	54450744	5491	51228696	3222047	5793	51628230	2822513
Sardinia	0.453	9381	2879	20749294	1941	19507775	1241519	2217	19872782	876512
Sicily	0.229	30128	9577	67075904	6564	63088644	3987260	4075	59795167	7280737
Tuscany	0.885	24157	4760	49919095	2344	46722061	3197034	4760	49919095	0
Umbria	0.750	6021	1392	12714360	790	11917517	796843	1392	12714360	0
Veneto	0.735	31775	6127	65484162	2949	61278931	4205231	6127	65484162	0
				817331670		76708226	50247444		796981072	20350597

[a] This scenario refers to a standard deviation increase in IQI for each region. [b] This scenario refers to an improvement in institutional quality up to the average IQI (i.e. 0.586) for those regions displaying a IQI score lower than the average.