DOCTORAL THESIS

Statistical evaluation of quality in healthcare

Author:
Paolo BERTA

Supervisor:
Veronica VINCIOTTI

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Department of Mathematics

November 29, 2018

Declaration of Authorship

I, Paolo BERTA, declare that this thesis titled, 'Statistical evaluation of quality in healthcare', and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Abstract

Governance of the healthcare systems is one of the most important challenges for Western countries. Within this, an accurate assessment of the quality is key to policy makers and public managers, in order to guarantee equity, effectiveness and efficiency. In this thesis, we investigate aspects and methods related to healthcare evaluation by focussing on the healthcare system in Lombardy (Italy), where public and private providers compete with each other, patients are free to choose where to be hospitalized, and a pay-for-performance program was recently implemented. The general aim of this thesis is to highlight the role of statistics within a quality evaluation framework, in the form of advancing the statistical methods used to measure quality, of evaluating the effectiveness of implemented policies, and of testing the effect that mechanisms of competition and cooperation can have on the quality of a healthcare system.

We firstly advance a new methodological approach for measuring hospital quality, providing a new tool for managers involved in performance evaluations. Multilevel models are typically used in healthcare, in order to account for the hierarchical structure of the data. These models however do not account for unobserved heterogeneity. We therefore propose an extension of the cluster-weighted models to the multilevel framework and focus in particular on the case of a binary dependent variable, which is common in healthcare. The resulting multilevel logistic cluster-weighted model is shown to perform well in a healthcare evaluation context.

Secondly, we evaluate the effectiveness of a pay-for-performance program. Differently from the existent literature, in this thesis we evaluate this program on the basis of five health outcomes and across a wide range of medical conditions. Availability of data pre and post-policy in Lombardy allows us to use a difference-in-differences approach. The statistical model includes multiple dependent outcomes, that allow quantifying the joint effect of the program, and random effects, that account for the heterogeneity of the data at the ward and hospital level. The results show that the policy has overall a positive effect on the hospitals' performance.

Thirdly, we study the effect of pro-competition reforms on the hospital quality. In Lombardy, competition between hospitals has been mostly driven by the adoption of a quasi-market system. Our results show that no association exists between hospital quality and competition. We speculate that this may be the result of asymmetric information, i.e. the lack of transparent information provided to citizens about the quality of hospitals. This is bound to reduce the impact of pro-competition reforms on quality and can in part explain the conflicting results found in the literature on this subject. Our results should motivate a public disclosure of quality evaluations.

Regardless of the specifics of a system, hospitals are altruistic economic agents and they cooperate in order to improve their quality. In this work, we analyse the effect of cooperation on quality, taking the network of patients' transfers between hospitals as a proxy of their level of cooperation. Using the latest network models, we find that cooperation does lead to an increase in quality and should therefore be encouraged by policy makers.

Acknowledgements

Undertaking this PhD has been a truly life experience for me and it would not have been possible without the support and guidance that I received from many people.
I would like to first say a very big thank you to my supervisor Dr Veronica Vinciotti for all the support, suggestions, help and encouragement she gave me, during all the development of this work. Without her guidance and constant feedback this PhD would not have been achievable. My wish is that this represents just a starting point for further collaborations, that would be an invaluable way to improve my knowledge and expertise in my research field.
I would also to thank my second supervisor Francesco Moscone. Francesco is first of all a friend of mine and during these years he was always supportive and stimulating for the success of this work.
My thanks also go out to my colleague Alina Peluso for the support and the friendship she gave me in this experience.
I would like also to thank Farukh Mukhamedov: I would never have expected in my life to have a Siberian friend, but this PhD makes this possible.
Thanks to all of you and all of the colleagues and friends that I omit to mention in these acknowledgements: you know that without you I never could have reached this achievement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and literature review

The most common definition of National Healthcare Service (NHS) comes from the World Health Organization (WHO) and it includes *"all the activities whose primary purpose is to promote, restore or maintain health"* (WHO 2000). The NHS governance is one of the most important challenges for the Western countries. Both ethical and spending implications, related to the health of the population, require an improvement in the quality of services provided to the citizens associated to a cost containment that guarantees the sustainability of the NHS.
The increasing in the expenditures which involved all the Western countries in the last decades, induced several Governments to introduce reforms in NHS transforming the healthcare services in a regulated market. The standard economic theory of how markets work is the model of supply and demand, where the main involved parties are the buyers and sellers. The buyers pay sellers directly for the goods and services exchanged and the prices are the main mechanism which coordinate the market. This classical theory works for many goods and services in the economy, but it is reasonable to consider the healthcare market something different. First of all third parties are involved in the process of goods and services exchange (governments, insurance, etc.), Second, patients do not really know what they need, the cost of the services and the quality of the services. Third, in several countries the prices are fixed and not involved in bargaining processes. Finally, the provided services relate to the citizens and not on stuffs. Although these characteristics distinct the healthcare market from the classical economic theory, several dynamics in the NHSs follow economic rules. The choice of a health care treatment seems purely medical, but providers increasingly evaluate and compare alternative treatments on economic aspects (Folland et al. 2007). Given that, it is easy to understand why quality assessment in healthcare becomes a powerful tool for the governance of the dynamics between

the public functions in order to protect the health of the population and the econonomic implications of the production of healthcare services. Indeed, good practices or critical points are the most important feedback for policy makers in order to promote the governance of clinical needs, use of services, and healthcare expenditures. The knowledge about critical points in the NHS is the boost for policy interventions, which are the tools adopted to manage the development of the NHS. Indeed, policy makers can modify the NHS radically by reforming the system or introducing rules in order to get small but meaningful goals. Policy interventions can affect all the organisations in the NHS such as hospitals or can interest specific individuals or units which are the targets of the policy intervention. In this direction, the growing availability and quality of administrative data has increased the opportunity for quantitative analysis and thus the need for statistical tools aimed at evaluating the quality of care.
This work deals with the role of statistics in increasing the robustness of the quality evaluation in health, and in evaluating the impact of reforms in healthcare on the quality provided to the citizens in particular with the role of multilevel models to provide the more suitable statistical tool for hospital quality evaluation. Furthermore, this work intends to study how the results of the measures of evaluation can support policy makers in their decision process.

1.1.1 Quality in healthcare

Quality in healthcare is a broad concept covering several areas, characterized by two main dimensions: *ex ante* and *ex post* evaluation. *Ex ante* evaluation relates to the processes of care and is typically defined by standards, evaluated in terms of achievement of these measures. The most important example of this approach to the evaluation is the system promoted by the Joint Commission, an independent, non-profit organization established in the USA in 1951 with the aim of improving the quality of healthcare, by identifying and verifying a set of reference quality standards. All these standards relate to the processes that characterize the hospital activity and they are specifically formulated to be used in the different cultural and institutional contexts in different countries: the standards centered on the patient monitor all the activities that directly or indirectly concern the patients, while those centered on the organization monitor the processes and actions that the providers put in place to maintain and improve the quality of care. Each standard is composed by several elements that represent the concrete measures through which it is possible to define the level of achievement of the standard. The standard proposed by the Joint Commission can be applied to evaluate the activity of both hospitals and local health authorities.
Differently from *ex ante* evaluation, the *ex post* evaluation concerns the evaluation of what happens from the moment the patient is discharged from the hospital. This approach pertains to several aspects of the hospital evaluation, such as the efficiency, the appropriateness, the customer (patient) satisfaction and, in particular, the effectiveness.

Hospital efficiency can be defined as the capacity to maximize the output (hospitalizations), given the technology and the productive factors available (input). When there is a mismatch between the output production and its optimal level, we are in the presence of inefficiency. In this case, it is important to understand and explain the causes of this deviation, and in particular if it is due to technical or allocative inefficiency. In the first case hospitals should increase their level of production, in the second case the healthcare system should re-allocate the input. Hospital efficiency may be estimated using non-parametric methods (e.g. Data Envelopment Analysis-DEA) or parametric methods (e.g. Stochastic Frontiers Analysis - SFA). Since the original contribution of Aigner et al. (1977), SFA has been widely applied to measure hospitals' efficiency because it allows to distinguish between inefficiency and random disturbances (differently from DEA). For a review of studies using stochastic frontier analysis in the health care sector see Rosko and Mutter (2008).

Appropriateness is another basic dimension for the *ex post* healthcare evaluation. It can be defined as the ability of a hospital to provide services tailored to the needs of the patient, in the right way and at the right time. As regards to hospital appropriateness, this can be assessed with respect to both the adoption of opportunistic behaviors regarding to the payment systems (Berta et al. 2010) and with respect to waiting times (Brekke et al. 2008, Siciliani and Hurst 2005). The opportunistic behaviors lead to three distortions of the DRG payment system: upcoding, cream skimming and readmissions. The upcoding practice consists in classifying a patient in a DRG that produces a higher reimbursement (Dafny 2005). The cream skimming can be defined in two ways: the selection of the more lucrative treatment or the selection of less complicated patients (Levaggi and Montefiori 2003). Last, the readmission practice implies that the same patient is discharged and admitted again after a short period, so that the hospital receives for the same treatment more than one reimbursement.

Finally, customer satisfaction measures the experience of the hospitalization from the patient point of view. Usually collected in the form of surveys, customer (patient) satisfaction can be useful in order to improve healthcare services. Moreover, the availability and use of patient satisfaction data have stimulated research investigating its relationship with other performance measures, such as the effectiveness (Grillo Ruggieri et al. 2018).

In the healthcare literature, the most analzyed dimension of the *ex post* evaluation is the effectiveness, and this is also what we consider in this thesis. Effectiveness is the expected level of outcomes achievable as a result of the application of best practices in hospital activities. In other words, effectiveness can be defined as the measure of the quality obtained from an appropriate provision of healthcare services. Effectiveness evaluation is based on the measure of the outcomes, which represent factors and conditions that approximate the clinical quality. These indicators do not measure the "true" effectiveness but feedback indications widely acceptable to identify best practices and critical points in the services provided by the healthcare system. The outcomes can be measured by exploiting the administrative data collected. This allows us to use information already available without implementing new data sources, thus avoiding an increase of costs and administrative burdens associated with this eventuality. Furthermore, there is the indisputable advantage of using information characterized by a high degree of homogeneity of contents, thanks to the collaboration between healthcare professionals and providers in the definition of the guidelines for filling the data forms (Iezzoni 1997). The scientific literature emphasizes that health outcomes can be measured in different ways. At the hospital level, examples include in-hospital mortality, post-discharge mortality or readmission within a specific time measured in days (Krumholz et al. 2013, Laudicella et al. 2013, Normand and Shahian 2007). However, differently from the most used mortality or readmission, quality evaluation considers several other outcomes such as unplanned return to the surgery room (Ansari and Collopy 1996, Leape et al. 1991) or the patients leaving hospital against the medical advice (Hwang et al. 2003).

The main issue about effectiveness evaluation concerns the statistical methods more suitable to guarantee an appropriate quality assessment. Despite this topic being studied in the literature (statistical, medical, economic, etc.) starting from the end of the eighties (Dubois et al. 1987), the debate on the statistical tools suitable for the effectiveness evaluation is still evolving. Recently in the US, the Centers for Medicare and Medicaid Services promoted a research involving Yale New Haven Health Services Corporation, Center for Outcomes Research and Evaluation and a committee appointed by the Committee of Presidents of Statistical Societies, in order to address statistical issues related to the hospital quality evaluation based on outcomes (Ash et al. 2012). Among several recommendation, this work concluded that multilevel generalized linear models are the most effective approach in order to study data characterized by a hierarchical structure. The milestone in the effectiveness evaluation based on multilevel models is the paper by Goldstein and Spiegelhalter (1996) where statistical issues in the quality evaluation are discussed. The paper presents the use of multilevel models in order to estimate league tables of the hospital performance as well as the limitations that can be faced in the risk-adjustment process and in the model assessment. Risk-adjustment is a statistical approach that allows to consider the patients' characteristics (age, sex, socio-economic status, comorbidities, etc) to measure risk, in order to provide a fair comparison of their adverse health outcomes. The most suitable statistical approach does not produce an effective evaluation if it does not consider a proper risk-adjustment. According to Iezzoni (1997), risk-adjustment is the only way to ensure an effective *ceteris paribus* evaluation among the providers, avoiding to penalize hospitals accepting more complicated patients.

All these considerations do not mean that, given approrpiate statistical tools and risk-adjustment procedures, the findings coming from an effectiveness evaluation must be treated as the "true" quality provided by each of the evaluated providers. In this sense, an illuminating discussion about this topic can be found in the paper by Lilford et al. (2004) where the authors demonstrate the caution that should be considered, when managing the results of an effectiveness evaluation. The authors state that a bad performance can depend on several factors: first, the quality of the collected data that can be heterogeneous among the providers; second, case-mix characteristics (ignored or unobservable) that are not included in the risk-adjustment models; finally, structural or institutional issues, that affect the hospital performance, but are not attributable to the hospitals (Lilford et al. 2004).

These reasons motivate the need for more advanced statistical models, such as those dealing with multiple outcomes and longitudinal data, and partly set a word of caution in the interpretation of the output of these models and their use for hospital evaluations.

Considering the growing availability of tools for quality evaluation, several Countries introduces new reforms aimed at boosting quality in healthcare: the so-called pay-for-performance (P4P) programs. The adoption of a P4P approach is intended to improve the quality of healthcare systems by supplying financial incentives to the healthcare providers that achieve specified quality benchmarks. Every P4P is based on the evaluation of the quality delivered. Although in healthcare quality is a broad concept, the most common meaning of P4P refers to the impact on effectiveness (Alshamsan et al. 2010). In the last few years there has been a growing interest in applying P4P programs to the healthcare system in many countries (see e.g. Eijkenaar (2012), Fillmore et al. (2014), Glickman et al. (2007), Pink et al. (2005)), but mixed results have been reported about the impact of these programs on the quality of care. The aim of these studies is to show that P4P schemes increase hospital quality, but often it is difficult to identify a causal effect between the performance improvement and the policy adoption. Furthermore, in several cases, due to the restricted accessibility to data, it is not possible to evidence the effect of P4P schemes on multiple health outcomes. Hence, the analysis is usually performed on one health outcome, and over a certain number of diseases. The opportunity to demonstrate the effect of a P4P program on multiple outcomes and over a wide range of hospital activities is one of the aims of this thesis.

1.1.2 Competition and cooperation in healthcare

A great debate exists, both at the national and international level, on the role of competition in different sectors of the economy, including the health care sector. In recent years, many governments have introduced competition among health care providers to meet the growing demand for health care in a climate of fiscal austerity. These interventions originate from a well-known theoretical result in economics: when prices are fixed and firms compete, a higher degree of competition is likely to produce better quality. Competition in healthcare is implemented allowing patients' choice for public services. The belief is that by stimulating patients' choice, hospitals become more responsive to patient demand driving in this way to an increase in the provided quality. Literature in health economics have gathered empirical evidence on the effects of competition in the health care sector, finding mixed results on the size and direction of these effects (Gaynor 2006). Some empirical studies demonstrated that more competition among hospitals leads to better health outcomes (Gaynor et al. 2012) whereas other studies reject this hypothesis, arguing that more competition may harm patients' health (Propper et al. 2004).

Several factors exist that may shape how competition between hospitals impacts quality. A first factor is related to how competition is implemented in healthcare systems in which prices are fixed. In some countries, such as the UK, the criteria are dependent on hospital market performances whereas other countries boost hospital competition by providing patient information on where to obtain the best treatment. Italy encourages competition by expanding patient choice sets and offering private hospitals per-treatment public reimbursement funding (Gaynor et al. 2012, Kessler and McClellan 2000, Moscone et al. 2012, Propper et al. 2004, Tay 2003). The degree of hospital competition also depends on the degree of patients' freedom of choice of where to be treated. Some markets have complete freedom (e.g. the USA) (i.e. patients can choose any hospital in the relevant market) whereas others have limited freedom (e.g., Italy) either because patients are free to choose but do not know the hospital quality or because they must select between a limited number of hospitals (Beckert et al. 2012, Cooper et al. 2011, Moscone et al. 2012, Tay 2003, Varkevisser et al. 2012). A third factor (hospital competitive strategy) describes how hospitals compete with other hospitals. In some markets, hospitals can choose both prices and quality (e.g. the USA), whereas, in others (e.g. the UK and Italy), prices are regulated by a central or local government and the providers can compete only through quality that is usually measured in terms of a set of health outcomes (Cooper et al. 2011, Kessler and McClellan 2000, Tay 2003). Finally, the degree of competition depends on the level of hospital information available to patients when they choose where to be admitted. In some markets, like the USA and the UK, patients are fully informed since hospital rankings are publicly advertised. In other countries (e.g. Italy), patients are free to choose but, as they are not privy to hospital rankings, this choice is mainly based on informal information such as word of mouth, reputation and the media (Dranove et al. 2003, Dranove and Sfekas 2008).

Under these circumstances, although top quality hospitals will attract more patients, the intensity of such an effect will depend on the hospital market structure. For example, if the hospital is a local monopolist, the effect is negligible since only those who are willing to travel long distances provide the incremental number of attracted patients. If, on the contrary, the hospital is operating in a market structure with other hospitals acting as nearby competitors, we can imagine two effects. The first is a short-run effect, whereby the top quality hospital attracts more patients (as limited by bed capacity), gains market share and is subject to less competition because it enjoys a quality difference compared with its competitors. The second is a long-term effect whereby competitors will react to the quality gap and (at least those remaining in the market) will also raise their levels of quality. This implies that, in the long run, market shares may even be unchanged compared with those existing before the quality gap.

The situation changes completely in the case of asymmetric information. Under this scenario, patients tend to choose the nearest hospital or base their decision on informal information. The latter may be based on GP referrals and neighbour assessments. For example, patients may use information about the decisions of people living in the same area and have or had the same pathology as those who must make comparable decisions. Friends, relatives or trusted people who have experienced similar health problems may also act as filters for the quality of hospitals, thus shaping individual preferences. However, as also emphasized by Moscone et al. (2012), interacting and sharing information with neighbours does not necessarily help in selecting a high quality hospital. Under this scenario, high quality hospitals may fail to attract more patients. Even if institutions have implemented measures to increase competition between hospitals, such measures may not obtain the returns from investing in quality (e.g. hiring the best physicians, buying the most expensive equipment and adopting costly control procedures in internal operations). Patients have a difficult time recognizing better quality hospitals and, hence, hospitals may not have an incentive to increase quality. To sum up, in a situation in which asymmetric information exists and prices are fixed, increasing competition may not produce an effect on health outcomes.

Despite competition is an important determinant of hospital management, hospitals, motivated by reasons such as convenience or altruism, may decide to engage in mutually beneficial cooperation with each other. Competition and cooperation can in fact coexist as they do not lead to mutually exclusive strategies, and they can both have an impact on hospital quality. The demand for wider health-care coverage requires that hospitals and other health-care organizations integrate their resources and expertise by creating inter-organizational linkages. Hospitals must coordinate their actions, operations, and plans to serve the public interest (Gittell and Weiss 2004) in several ways. First hospitals could share they resource in terms of facilities but also in terms of human resources. They could share the manager ability and moreover they can share the knowledge and skills of their physicians. Furthermore, hospitals can cooperate sharing their patients. Patient-sharing practices diffuse and grow in importance, but it remains unclear what drives these collaborations and if it leads to increase the healthcare quality. Iwashyna (2012) in their review of the literature on the transfer of critically ill patients, concludes that the destination of patients is not necessarily chosen on the basis of objective evidence about the performance and capabilities of the receiving hospital. Despite this, it is argued that stimulating hospital cooperation in terms of patient sharing relations means that appropriate patients could be transferred from lower to higher quality hospitals (Iwashyna et al. 2009). Nevertheless, while competition is extensively studied in the literature, cooperation and its impact on quality is relatively less investigated. In this thesis we investigate the relationship between cooperation and quality, and whether hospital cooperation increases the quality provided to the citizens.

1.2 Case study: Lombardy

In this thesis we investigate the Regional Healthcare System (RHS) of Lombardy (Italy), as an interesting case-study. The choice of Lombardy as case-study is due to the dimensions of the RHS as well as to the rules defining this healthcare system. Lombardy is the most populous region of Italy: it counts 10 million of citizens, and it is also one of the wealthiest, best educated and richest regions in Europe. Lombardy devotes each year roughly 18 billion euro to the healthcare system, and is composed by roughly 200 financed hospitals, 51% of which are public. Each year approximately 1,500,000 hospitalizations are delivered, and 10% of these concern patients living in a different region. This makes this region comparable to many European countries. The organization of the healthcare system in Lombardy is another reason to study this RHS. In Lombardy the RHS is based on competition, which means that public and private providers deliver hospital services and patients are free to choose where to be hospitalized. Furthermore, in Lombardy there is an effective P4P program, and hospitals cooperate with each other.
Before describing the details of the RHS in Lombardy, we start with a summary of the Italian NHS. At the beginning of the XX century and up to the fifties. the Italian NHS was based on a compulsory insurance system, where several workers' funds were usually financed jointly by employers and employees. The main problem was that only 90% of citizens had an healthcare coverage (i.e. free-lance workers were excluded) and different services were provided depending on the belonging to different workers' funds. We can say that equity was far from being a characteristic of the Italian healthcare system at that time. This situation radically changed in 1978 with the national reform n.833, when Italy adopted a *Beveridge* healthcare system. This is one of the framework existing in the world to organize a NHS, and is commonly adopted in most of the European countries. It takes the name from the UK liberal economist William Beveridge who was designated by Prime Minister Winston Churchill to design a welfare system for the UK after the Second World War (Beveridge et al. 1942). This model of healthcare system has two main pillars: is universalistic, namely healthcare is guarantee to all the citizens, and it is free for all, even if, in order to reach a balance in spending, some shrewdness is adopted. Following this organizational model, the Italian healthcare system is funded by taxation, and each citizen contributes in healthcare financing on the basis of its economic resources.
A second reform process was defined by two acts of the Italian government, in 1992 and in 1993 respectively. According to these two reforms the role of financing and programming the healthcare system have been separated from the providers of the healthcare services. The central government preserved the role of financing the entire healthcare system and defining the essential level of assistance, whereas each Italian region was deputed to define its own healthcare system. This was the first step in defining 21 different healthcare systems (one for each region) within the Italian healthcare system, a process that was completed with the constitutional reform in 2001.

In 1997 the Lombardy region exploited the national reform and designed a specific regional healthcare system. The regional law n.31 in 1997 transformed the Lombard RHS in a quasi–market, where the healthcare services are financed by the central government but not necessarily delivered by public providers (Colombo 2008). The main characteristics of this new healthcare system can be summarized in: 1) the separation between purchasers (Local Health Authorities - LHA) and providers (hospitals), 2) the competition between private and public providers and 3) freedom of choice of the hospital where to be admitted (Brenna 2011). Furthermore, Lombardy was one of the first regions in Italy who completely adopted a prospective payment system based on the Diagnosis Related Groups (DRGs). This type of payment system, invented and adopted in the US at the beginning of the eighties (Hsiao et al. 1986), has been implemented in Lombardy to increase the cost-efficiency of the hospital care (Barbetta et al. 2007). Instead of receiving a reimbursement according to the volume of hospitalizations provided, each hospital (private or public) receives a fixed reimbursement for each discharge according to a tariff established by regional law at the beginning of each year. The hospitals admitted to this form of public financing are those included in the healthcare system on the basis of certain requirements, established and monitored by the regional healthcare directorate (Vittadini et al. 2012).

In Lombardy a budget constrain limits the annual hospital reimbursement. The yearly hospital budget assignment is based on a bargaining between the hospitals and the regional officers. At the end of a year, there is an agreement between the managers of each hospital and the regional managers on the overall budget that each hospital can receive for the provided hospitalizations. This is a monetary cap and the hospital managers are free to choose how to allocate the patients' discharges in terms of DRG in order to not overcome this cap. Hospital managers may decide to allocate hospital's resources in the different wards according to the different remuneration levels provided by the DRG-tariffs scheme (Martini et al. 2014).

Starting from 2012, a new extra-budget is devoted to each hospital based on their performance evaluated on hospital outcomes. According to this effectiveness evaluation, the budget of each hospital is increased or decreased up to +2% or -2% respectively. In Italy, Lombardy is the only region adopting this type of policy, using quality evaluation in order to remodulate the hospital budget.

As anticipated, in Lombardy each year roughly 1,500,000 of hospitalizations are delivered to citizens living in Lombardy. In Figure 1.1 the trend of the hospitalizations in the last six years is presented, distinguished by total amount, number of ordinary hospitalizations (excluded day-hospitals, palliative cares and rehabilitation), total of hospitalizations in day-hospital and day-surgery. The trend of the ordinary hospitalizations is decreasing in the last years in order to increase the appropriateness of the hospitalizations and the efficiency of the system, and this is balanced by the increasing in the admissions in day-surgery, while other hospitalizations are delivered in outpatient treatment. Hospitals in Lombardy are dislocated in the entire region, with a predominant amount in the metropolitan area of Milan. Figure 1.2 shows the hospitals' location in the Region, distinct between private and public providers, and we can observe that

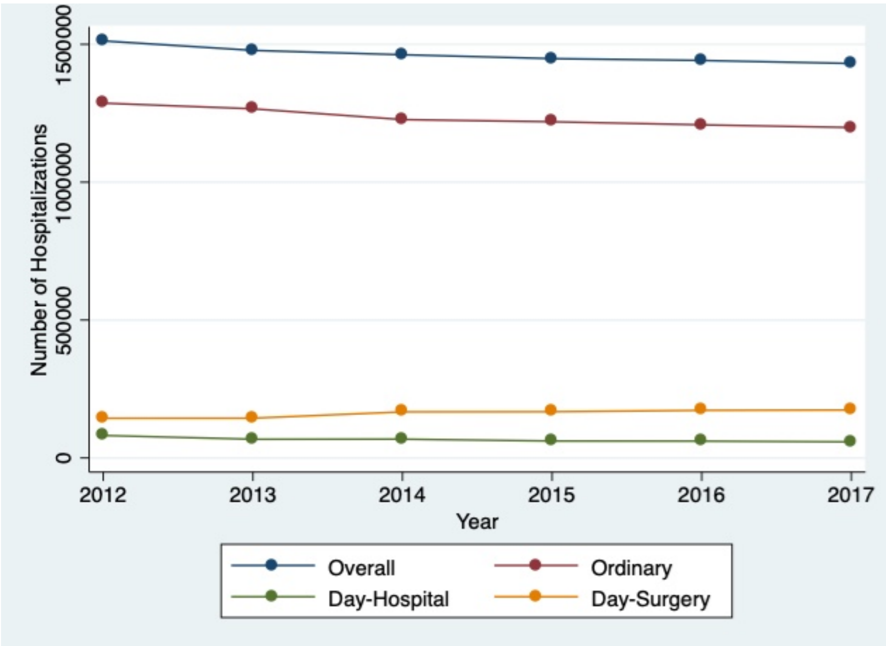


FIGURE 1.1: Number of hospitalizations in Lombardy from 2012 to 2017

private providers are mostly located in Milan, Brescia and Bergamo in the middle of the region. Furthermore, the map in Figure 1.2 shows that all the territory of the region presents a good coverage which is reduced in the mountain area in the north of the region. The location of the providers has an impact when we study hospital competition, where patients' choices is a driver of the competition and the distance between the patients' home and the hospital is the main covariate explaining these choices. The healthcare system in Lombardy is also characterized by 23 teaching hospitals and 21 monospecialized hospitals dedicated to cardiology, neurology and orthopaedics, defining a complete healthcare system where the supply for the citizens is qualified.

1.3 Data

In Lombardy healthcare system quality evaluation is based on the administrative data gathered from the regional healthcare information system, and it refers to the whole set of hospitalizations (Schede di Dimissione Ospedaliera, here-in-after SDO) provided in Lombardy. The SDO data source was defined by the Italian Ministry of Health in 1991. All the hospitals belonging to the national healthcare system must send the data concerning each provided hospitalization to the regional information system and each year the Italian regions send this set of data to the Ministry of Health. The information collected in the SDOs include patient's demographic data (e.g. age, sex, residence, education), characteristics of the hospitalization (e.g. hospital of admission, discharge ward) and clinical features (e.g. principal diagnosis, concomitant diagnosis, interventions). Each SDO record is distinct by an anonymous code that identifies the

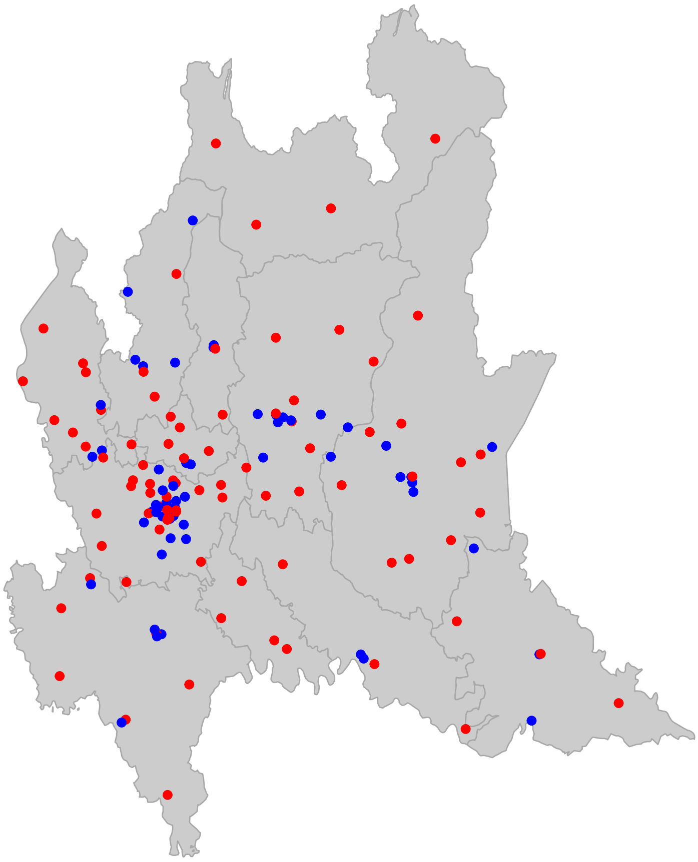


FIGURE 1.2: Number of hospitalizations in Lombardy from 2012 to 2017

document that cannot be used to discover the identity of the patient. In order to manage the data it is important to define a unique key for the hospitalization and this can be obtained concatenating the information related to the hospital of admission, the identifier of the hospitalization and the date of admission. In each SDO the clinical information of the hospitalization is collected in several items. First of all in each hospitalization a principal diagnosis is indicated, and then, up to five co-diagnosis which can complicate the severity of the patients can be indicated. For the surgical admission the main procedure delivered to the patient is indicated and up to five secondary procedures can be included. The information in terms of diagnosis and surgical procedures allows to obtain the DRG associated to the hospitalization. Furthermore, there are several algorithms which can be used to combine diagnosis and DRG in order to obtain a set of comorbidities affecting the patients. In particular, the Charlson (Charlson et al. 1987) and the Elixhauser (Elixhauser et al. 1998) algorithms are widely adopted. Finally, the information related to the reimbursement can be used in order to perform cost-effectiveness analysis. Originally conceived for economic- management purposes, the SDO are currently used for clinical-epidemiological studies and to support health planning activities, as well as for monitoring the clinical hospital risk. Furthermore, the quality of this information is a characteristic that allows researchers to use SDO for hospital quality evaluation. In Lombardy, the SDO data source is recognized for having a good quality and a high level of internal coherence. A wide activity of control of the way the hospitals collect data in the SDO data source has been performed during the last fifteen years. In specific, in each Lombard LHA a team of selected professionals check the consistency of the SDO in comparison with the clinical documents stored in the hospitals, for about 10% of the annual discharges (roughly 150,000 hospitalizations).

1.4 Multilevel models for predicting quality

As mentioned before, a quite extensive literature suggests that multilevel models offer solutions for studying relationships between health outcomes and covariates in complex hierarchical data structures, considering both individual and aggregate levels of analysis (Christiansen and Morris 1997, Goldstein and Spiegelhalter 1996, Leyland and Boddy 1998, Leyland and Goldstein 2001, Normand et al. 1997, Rice and Leyland 1996). Healthcare data are usually characterized by a hierarchical structure. Patients hospitalized in different hospitals define a hierarchy in the data, and this implies that both patients and hospital membership influence each others. Ignoring this relationship risks invalidating the use of ordinary statistical tools for data analysis A first issue depending on ignoring the hierarchal structure relates with the assumptions behind the underlying statistical tests, that treat the data as an independent random sample. If we accept that patients admitted in a hospital are more alike to patients admitted in a different hospital, the assumption of independence is violated, and, therefore, the validity of the statistical tests is disputable. A second order of issues relates with the level of aggregation of the data. If we deal with a hierarchy in the data, we can read the results of our analysis at the aggregated data level (hospitals in our framework) or individual (patients). In this case we can face a problem of ecological or atomistic fallacy, which consist respectively on analysing data at the individual level and drawing erroneous conclusions at the aggregated one or *viceversa*. In order to overcome the challanges related to the hierarchical structure of the data, the statistical literature has introduced multilevel models. An extensive introduction to multilevel models can be found in Hox (1995), Goldstein (2010), and Snijders and Bosker (2012).

Let $(\boldsymbol{X}, Y)$ be defined in some finite space $\Omega \subseteq \mathbb{R}^d \times \mathbb{R}$, where Y is the response variable and $\boldsymbol{X}$ the vector of level-1 covariates. We view the data as having a two-level structure with lower level observations (level-1), nested within higher level observations (level-2).

To indicate the level-2 unit that individual i belongs to, we add a second subscript j so that y_{ij} is the value of Y for the i-th individual in the j-th second level unit, and x_{ij} is the obervation for one covariate X for the i-th individual in the j-th second level unit. Let us suppose that there are J level 2 observations with n_j level-1 observations in the j-th second level unit, so that the total sample size is $n = n_1 + \cdots + n_J$. Let us assume that the mean of Y is ν and $\upsilon = f(\nu)$ is the function that links the expected values of Y to the predicted values given by the regression model. Then multilevel models, following the notation in Hox and Roberts (2011), are defined as:

$$\boldsymbol{\upsilon}_j = \boldsymbol{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\boldsymbol{U}_j, \tag{1.1}$$

where $\boldsymbol{\upsilon}_j$ is the $n_j \times 1$ vector of expected values of Y for each j level-2 unit, $\boldsymbol{X}_j$ is the $n_j \times d$ matrix of the covariates, $\boldsymbol{\beta}$ is the $d \times 1$ vector of the fixed parameters. $\boldsymbol{Z}_j$ is the $n_j \times r$ matrix for the r subset of covariates included in the model as random effects, and $\boldsymbol{U}_j$ is the $r \times 1$ vector of the unknown random effects, distributed as a $\mathcal{N}_r(0, \boldsymbol{\tau}^2)$. Considering for simplicity a model with only one random covariate the formulation becomes:

$$\boldsymbol{\upsilon}_j = \beta_{0j} + \beta_{1j}X_j \tag{1.2}$$

where $\beta_{0j} = \beta_0 + u_{0j}$ and $\beta_{1j} = \beta_1 + u_{1j}X_j$, with

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}\right).$$

When a Gaussian dependent variable is considered, the model includes an individual error term ε_{ij}, which is assumed to be distributed as a $\mathcal{N}(0, \sigma^2)$.
Multilevel models account for the correlations among observations in the same cluster. In the case of a Gaussian random intercept model the intraclass correlation coefficient (ICC) is given by

$$\text{ICC} = \frac{\tau_0^2}{\tau_0^2 + \sigma^2}, \tag{1.3}$$

where σ^2 is the variance of the individual error term in the model (Snijders and Bosker 2012).
When a Gaussian random slope model is adopted, and considering for simplicity a multilevel model including one covariate X, the ICC depends on the values of X, and assumes this formulation:

$$\text{ICC}(y_{ij}, y_{i'j}) = \frac{\tau_0^2 + \tau_{01}(x_{ij} + x_{i'j}) + \tau_1^2 x_{ij}x_{i'j}}{\sqrt{\tau_0^2 + 2\tau_{01}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2}\sqrt{\tau_0^2 + 2\tau_{01}x_{i'j} + \tau_1^2 x_{i'j}^2 + \sigma^2}}, \tag{1.4}$$

where x_{ij} and $x_{i'j}$ are the values of X for two observations belonging to the same group j (Goldstein et al. 2002).
The Restricted Maximum Likehlihood (REML) approach is adopted in order to estimate the parameters of multilevel models (Corbeil and Searle 1976, Patterson and Thompson 1971). In a multilevel framework, maximum likelihood estimation produces biased estimates for the parameters because it does not consider the heterogeneity induced by the clusters observed at the second level. REML allows to obtain unbiased estimates, separating the estimation of the fixed part from the random part of the model.

1.5 Contribution of the thesis

Chapter 2 in this thesis relates to the statistical evaluation of quality in healthcare. The results of an outcomes' evaluation based on classical statistical tools such as multilevel models, can be affected by unobserved heterogeneity. This is typically captured by finite mixture models which, however, do not capture observed heterogeneity. In chapter 2, we develop an extension of the cluster-weighted mixture model (Ingrassia et al. 2012; 2015) to the multilevel framework, in order to account for both known and latent structures of the data. Furthermore, we develop an Expectation-Maximization algorithm for parameter estimation and a parametric bootstrap approach for assessing the variability of the estimators.
The first part of this chapter is dedicated to the methodological presentation of the extension of the cluster weighted models to the multilevel logistic framework. A simulation study follow this methodological part and then an application to real data is provided. The main variables included in the application concern the mortality, used as dependent variable of the models, and some covariates. Mortality is calculated from the SDO where the date of death for the patients is included and this allows us to calculate if the patient dies within 30 days after the discharge or not. The comorbidity index is calculate using the Elixhauser algorithm (Elixhauser et al. 1998), whereas the DRG weight is a characteristic of the DRG. The inclusion of sex and age complete the risk adjustment process.
In Chapter 3, we evaluate the effectiveness of the P4P program adopted in our case-study, on the basis of five health outcomes and across a wide range of medical conditions. The policy evaluation is based on a difference-in-differences approach (Abadie 2005). The main variables involved are the five outcomes, which are calculated using the SDO. Mortality is calculated as describe before, while readmission is calculated using the anonymous identifier of the patient: once a patient is discharged from an hospital, we search in the following SDO if the same person is admitted for the same MDC within 45 days. Return in the surgery room is calculated within the same hospitalization and only for surgical admissions, whereas transfer is identified as a patient discharged from an hospital and admitted to another one in the same day or within 24 hours. Voluntary discharges is declared by the hospital in a variable included in the SDO. The risk adjustment process is completed including in the model the following covariates: sex, age, DRG weight and comorbidities. The model is also controlled by hospital characteristics such as ownership, teaching status, level of technology and if the hospital is monospecialized or general.
In Chapter 4 we study the effect of competition on adverse hospital health outcomes in Lombardy, a pro-competitive context where information about hospital quality is not publicly available. Although risk-adjusted hospital rankings are estimated yearly in this region, such rankings are provided only to hospital managers and are not available to general practitioners or citizens. Hence, patients may choose the hospital where to be admitted on the basis of different criteria. n this chapter we include the same variables discussed before in order to study the effect of the

competition on the health outcome, but what diversifies this chapter is a first stage where we study the patients' choices. In this part of the chapter we are able to include the distance from the place where the patient live to the hospital location. This variable is the main driver of the patients' choices. Predicted patients' choices give us the opportunity to calculate an index of competition which is used in the model relating competition and quality. This index is extensively presented in the chapter.

Finally, in Chapter 5 we analyse the role of cooperation in stimulating quality in healthcare in Lombardy. To this aim, we exploit the statistical models developed in the growing field of network science. First, we use a social relations model to identify the determinants of hospital transfers, and then we use the predicted transfers to test whether there exists a relationship between these flows and the quality of origin and destination hospitals. In this chapter we use the patients transfers calculated as in Chapter 3, and they are adjusted using age and DRG weight together with the number of hospital discharges and an index indicating the degree of centrality of each origin and destination hospital in the network. In the step of the analysis where we study the effect of the cooperation on quality we also include the hospital ownership in order to study if there is a different behavior according the public or private status of the hospitals.

1.6 Publications

Chapter 2

A reduced version of this chapter is published on the proceeding of the conference CLADAG 2017, organized by the Italian Society of Statistics, with this reference:

Berta P, Pennoni F, Vinciotti V (2017). Outcome evaluation in healthcare: the multilevel logistic cluster weighted model. In: Book of Short Papers CLADAG 2017. p. 1-6, ISBN: 978-88-99459-71-0, University of Milano-Bicocca Milan (Italy), 2017.

Furthermore, this chapter is under review for publication in Statistical Analysis and Data Mining as a joint work with my supervisor Veronica Vinciotti.

Chapter 3

This chapter is based on a joint paper written with my colleague Alina Peluso and my supervisor Veronica Vinciotti, which is published in Empirical Economics with this reference:

Peluso A, Berta P, Vinciotti V (2018). Do pay-for-performance incentives lead to a better health outcome?, DOI: 10.1007/s00181-018-1425-8, Forthcoming.

Chapter 4 This chapter is based on a paper published with my second supervisor Francesco Moscone on the Journal of Royal Statistical Society (Series-A) whit this reference:

Berta P, Martini G, Moscone F, Vittadini G (2016). The association between asymmetric information, hospital competition and quality of healthcare: evidence from Italy. Journal of the Royal Statistical Society. Series-A. Statistics in Society, vol. 179, p. 907-926, ISSN: 0964-1998.

Chapter 5 This chapter is a joint work with my supervisors Veronica Vinciotti and Francesco Moscone.

Chapter 2

Quality evaluation in health using cluster weighted models

As discussed in Chapter 1, the performance of hospitals is usually evaluated by multilevel models. Performance evaluation based on administrative data deals with individual information about patients which are nested within the hospitals where they are admitted. This is a hierarchical data structure and we assume that patients admitted to a hospital are more alike that patients admitted to another one. This assumption implies that the independence between observations is violated and linear models cannot be considered. In these cases, multilevel models must be adopted, assessing in this way the variability within and between hospitals. Despite the multilevel model allows us to consider the heterogeneity for observed groups, this is not suitable for data with unobserved heterogeneity. Even if risk-adjustment and observed heterogeneity can be asessed by multilevel models, as stated by Lilford et al. (2004), there is a source of variation influencing hospital quality evaluation which cannot be measured or observed. This latent variability can systematically influence the outcomes and depends on factors such as lifestyle (smoking, physical activity, drinking, etc.) or previous clinical conditions. These factors affect patient severity but we cannot include them in a risk-adjustment model because they cannot be measured or observed. The effect of this latent source of unobserved heterogeneity leads to the variation of the regression coefficients between groups of individuals sharing similar but unobserved characteristics. To overcome such drawbacks, in this chapter we propose a multilevel cluster-weighted model for binary dependent variables, a new mixture model approach for handling hierarchical data. This approach is applied to the healthcare context where often the outcomes are dichotomous, and allows us to evaluate the relative effectiveness of the hospitals through a family of multilevel models, each of them describing a different sub-population of patients.

2.1 Finite mixture models

The application of multilevel models requires that data have a natural hierarchical structure, where patients are nested into wards and hospitals. In cases when a natural grouping in the data is unknown but one expects latent groups, finite mixture models allow to account for this heterogeneity in the response distribution by splitting the population into a finite number of relatively homogeneous classes (McLachlan and Peel 2000). The finite mixture framework is then characterized by the idea that sometimes data can be observed as a composition of different groups in which the observations can be clustered to better represent their heterogeneity. Assuming that a given population Ω can be partitioned into C groups, say $\Omega_1, \dots, \Omega_C$, we refer to finite mixture distributions as the distribution of a random variable Y described by a weighted sum of the C components, that is

$$p(y) = \sum_{c=1}^{C} w_c p(y; \Omega_c), \tag{2.1}$$

where $p(y; \Omega_c)$ is the density of Y within group c and w_c are non-negatives constants varying from 0 to 1 and the sum of which is equal to 1. In a mixture framework, the w_c are called the mixing weights and $p(y; \Omega_c)$ are the component densities of the mixture (McLachlan and Peel 2000).

In a regression framework, Y is conditioned on a d-dimensional vector of covariates $\boldsymbol{X}$, and the finite mixture model can be expressed in the following way:

$$p(y|\boldsymbol{x}) = \sum_{c=1}^{C} w_c p(y|\boldsymbol{x}; \Omega_c). \tag{2.2}$$

2.2 Cluster weighted models

Ingrassia et al. (2012) have generalized the framework described in the previous section by introducing the so-called Cluster-Weighted Model (CWM). Here, the joint density of the response and the covariates is clustered into groups. This results in a mixture of local models, which are represented by the conditional densities of the response given the covariates within a group, weighted both by the local densities of the covariates, which are typically not considered within standard mixture regression models, and the usual mixing weights.

Thus, CWMs are formulated as:

$$p(\boldsymbol{x}, y; \boldsymbol{\theta}) = \sum_{c=1}^{C} p(y|\boldsymbol{x}; \boldsymbol{\xi}_c)\phi(\boldsymbol{x}; \boldsymbol{\psi}_c)w_c, \tag{2.3}$$

where $p(y|\boldsymbol{x}; \boldsymbol{\xi}_c)$ is the conditional density of Y given $\boldsymbol{X}$ in Ω_c, which depends on a vector of parameters $\boldsymbol{\xi}_c$, $\phi(\boldsymbol{x}; \boldsymbol{\psi}_c)$ is the marginal density of $\boldsymbol{x}$ in Ω_c, which depends on a vector of parameters ψ_c, and w_c is the weight of Ω_c in the mixture (with $w_c > 0$ and $\sum_{c=1}^{C} w_c = 1$). The set of all parameters of the model is denoted by $\boldsymbol{\theta}$, and is given by

$$\boldsymbol{\theta} = (w_1 \dots w_{C-1}, \boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_C, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_C).$$

Once the model is fitted to data, the units of the population are classified to the clusters according to their posterior probability:

$$P(\Omega_c|\boldsymbol{x}_i, y_i) = \frac{p(y_i|\boldsymbol{x}_i; \boldsymbol{\xi}_c)\phi(\boldsymbol{x}_i; \boldsymbol{\psi}_c)w_c}{\sum_{k=1}^{C} p(y_i|\boldsymbol{x}_i; \boldsymbol{\xi}_k)\phi(\boldsymbol{x}_i; \boldsymbol{\psi}_k)w_k}. \tag{2.4}$$

The relationship between CWMs and standard finite mixture models, as defined in Section 2.1, is explained in Ingrassia et al. (2012). In particular, the authors have theoretically illustrated that, when the marginal part of the model, which relates to the distribution of $\boldsymbol{X}$, does not depend on the latent groups, CWM contains the finite mixture of Gaussian (Everitt and Hand 1981, McLachlan and Peel 2000) and the finite mixture regression models (McLachlan and Peel 2000, Wedel and DeSarbo 1995) as special cases.
In general, accounting for the distribution of X within each latent group is what distinguishes CWM from standard finite mixture models.

2.3 Multilevel regression mixture models

When both known and latent groups have to be modelled, a strand of research has proposed an extension of the standard mixture models of Section 2.1 to the multilevel setting in order to disentangle latent classes within the natural groups in the data (Asparouhov and Muthén 2008, Muthén and Asparouhov 2009, Vermunt 2005).

Let us consider a multilevel model as defined in Chapter 1, but we now let the regression of Y on the vector of covariates $\boldsymbol{X}$ for the patient i in the hospital j vary across a patient-level latent class c. In this case, we estimate one multilevel for each latent group, and the multilevel regression mixture is defined as

$$\boldsymbol{v}_j = \boldsymbol{X}_j\boldsymbol{\beta}_c + \boldsymbol{Z}_j\boldsymbol{U}_{cj} \tag{2.5}$$

where the suffix c in the vectors of parameters identify the C multilevel models that will be estimated.
Considering for simplicity one random covariate, the model becomes:

$$\boldsymbol{v}_j = \beta_{0jc} + \beta_{1jc}X_j \tag{2.6}$$

where $\beta_{0jc} = \beta_{0c} + u_{0jc}$ and $\beta_{1jc} = \beta_{1c} + u_{1jc}$, with

$$\begin{bmatrix} u_{0jc} \\ u_{1jc} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{0c}^2 & \tau_{01c} \\ \tau_{01c} & \tau_{1c}^2 \end{bmatrix} \right).$$

Conditioning on the sub-population Ω_c, the distribution of $Y|X$ will be denoted by $p(y_{ij}|x_{ij};\boldsymbol{\lambda}_c)$, where $\boldsymbol{\lambda}_c = (\beta_{0c}, \beta_{1c}, \tau_{0c}^2, \tau_{1c}^2, \tau_{01c})$ is the vector of the parameters of the conditional distribution of $y|x$, for $c = 1, \dots, C$.
Here, the posterior probability of the c-th group for the multilevel mixture model is given by

$$P(\Omega_c|\boldsymbol{x}, y) = \frac{p(y|\boldsymbol{x};\boldsymbol{\lambda}_c)w_c}{\sum_{k=1}^{C} p(y|\boldsymbol{x};\boldsymbol{\lambda}_k)w_k}. \tag{2.7}$$

Thus, the multilevel regression mixture models is able to capture unobserved heterogeneity at the individual level, expressed in terms of random intercepts and slopes that vary between clusters, by modelling the probability of latent classes Ω_c.

2.4 The Multilevel Logistic Cluster-Weighted Model

Since finite mixture of regression models are a special case of cluster weighted models, and both cluster weighted models and multilevel models are very useful in the healthcare context, in this thesis we extend cluster weighted models to the multilevel framework, and thus we propose an extension to the finite mixture of regression models described in Section 2.3.
The general formulation of CWM follows that of Section 2.2, and the joint probability of the random vector of covariates $\boldsymbol{X}$ and the dependent variable Y is expressed as in Equation (2.3). In this case, the conditional part of the model depends on the distribution of Y, and $\boldsymbol{\xi}_c$ is the vector of the regression parameters. The marginal density of the vector of covariates is assumed to be distributed as a multivariate normal with mean $\boldsymbol{\mu}_c$ and covariance $\boldsymbol{\Sigma}_c$.
In the following we firstly introduce the multilevel CWM for Gaussian dependent variable, as developed by Berta et al. (2016), and then we define in detail the multilevel CWM (MCWM) for the cases of Bernoulli distributions of the variable Y conditioned of X, which are the most adopted in healthcare evaluation, since most health outcomes are dichotomous (e.g. mortality, readmissions). We further consider the simple case of one covariate only.

Gaussian Multilevel CWM

A multilevel cluster weighted model for a Gaussian dependent variable can be formulated as follows:

$$y_{ij} = \beta_{0c} + \beta_{1c}X_{cj} + u_{c0j} + u_{c1j}X_{cj} + \varepsilon_{ij}, \tag{2.8}$$

where $u_{c0j} \sim N(0, \tau_{0c}^2)$ are the random effects for the hospital j in the cluster c, whereas $u_{c1j} \sim N(0, \tau_{1c}^2)$ are the random effects related to the random covariate included in the model, and ε_{ij} is the error component.

Logistic Multilevel CWM

Considering the probability π of observing a binary health outcome Y, depending both on the hierarchical structure in the data (in our case patients within hospitals) and on possible latent groups, we consider a logit link and a random effect model, resulting in

$$\text{logit}(\pi_{jc}|\boldsymbol{X}, \Omega_c) = \beta_{0c} + \beta_{1c}X_{cj} + u_{c0j} + u_{c1j}X_{cj}, \tag{2.9}$$

where the assumptions related to the random effects are the same explained for the Gaussian case. It is interesting to notice that in healthcare u_{c0j}, the hospital random effects, can be interpreted as the relative effectiveness of hospitals with respect to the outcome y.

Cluster allocation

Regardless of the distribution of the dependent variable, each patient can be assigned to one of the C clusters according to the posterior probability

$$p(\Omega_c|\boldsymbol{x}, y) = \frac{w_c p(y|\boldsymbol{x}; \boldsymbol{\xi}_c)\phi(\boldsymbol{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)}{\sum_{k=1}^{C} w_k p(y|\boldsymbol{x}; \boldsymbol{\xi}_k)\phi(\boldsymbol{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}.$$

Typically, the unit is allocated to the cluster associated to the maximum posterior probability.

2.5 Inference for Multilevel Cluster-Weighted Model

2.5.1 The EM-algorithm

In the presence of latent groups, the parameters $\boldsymbol{\theta}$ are estimated by an Expectation-Maximization (EM) algorithm. Using the notation in Equation (2.3), the aim of the estimation process is to identify the vector of parameters $\boldsymbol{\theta}$ composed by

$$\boldsymbol{\theta} = (w_1 \dots w_{C-1}, \boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_C, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_C, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_C)'.$$

The log-likelihood for $\boldsymbol{\theta}$ can be expressed as:

$$\ell((\boldsymbol{x}, y)|\boldsymbol{\theta}) = \sum_{j=1}^{J}\sum_{i=1}^{n_j} \log\{\sum_{c=1}^{C} p(y_{ij}|\boldsymbol{x}_{ij}; \boldsymbol{\xi}_c)\phi(\boldsymbol{x}_{ij}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)w_c\}$$

where y_{ij} and $\boldsymbol{x}_{ij}$ are the observed values of Y and $\boldsymbol{X}$, respectively, for the i^{th} first level observation (patient) in the j^{th} second level unit (hospital), with $j = 1, \ldots, J$ and $i = 1, \ldots, n_j$, where n_j is the total number of patients admitted to the hospital j.
Following McLachlan and Peel (2000), the formulation of the CWM problem can be viewed as a situation of incomplete data and an EM algorithm can be applied in order to estimate the maximum likelihood and to identify the probability that the observation $(\boldsymbol{x}_{ij}, y_{ij})$ belongs to one of the identified clusters. Assuming a C-dimensional component-label vector $\boldsymbol{z}_{ij}$ where $z_{ijc} = 1$ if the observation $(\boldsymbol{x}_{ij}, y_{ij})$ belongs to the c^{th} cluster and 0 otherwise, the complete data log-likelihood function for the observation $(\boldsymbol{x}_{ij}, y_{ij})$ and the latent allocation z_{ijc} can be expressed as:

$$\begin{aligned}\ell_c((\boldsymbol{x}, y, z)|\boldsymbol{\theta}) = &\sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{c=1}^{C} z_{ijc}\log(w_c) + \\ &\sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{c=1}^{C} z_{ijc}\log[p(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\xi}_c)] + \\ &\sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{c=1}^{C} z_{ijc}\log[\phi(\boldsymbol{x}_{ij}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)].\end{aligned} \tag{2.10}$$

Equation (2.10) implies that the complete data log-likelihood is composed of the three parts of the model: the probability of belonging to one of the clusters, the regression part of the response given the covariates and the marginal multivariate Gaussian distribution of the covariates.

The EM algorithm follows an iterative process starting with an evaluation of the missing data based on the available data (E-step) and then a maximization of the expected log-likelihood (M-step). Considering the unknown vector $\boldsymbol{z}_{ij}$, the $(r+1)^{\text{th}}$ iteration of the EM-algorithm is based on the expectation with respect to z of the complete data log-likelihood $\ell_c((\boldsymbol{x}, y, z)|\boldsymbol{\theta})$ in Equation 2.10, with $\boldsymbol{\theta}$ estimated at the r^{th} iteration, i.e.

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(r)}) = E_{z|(\boldsymbol{x}, y);\theta^{(r)}}(\ell_c((\boldsymbol{x}, y, z)|\boldsymbol{\theta})).$$

This requires the calculation of the probability that the observation $(\boldsymbol{x}_{ij}, y_{ij})$ belongs to the c cluster, since

$$E(z_{ijc}|(\boldsymbol{x}_{ij}, y_{ij}), \boldsymbol{\theta}^{(r)}) = Pr\{z_{ijc} = 1|(\boldsymbol{x}_{ij}, y_{ij}), \boldsymbol{\theta}^{(r)}\} = \tau_c((\boldsymbol{x}_{ij}, y_{ij}), \boldsymbol{\theta}^{(r)}).$$

Given our proposed multilevel CWM,

$$\tau_c((\boldsymbol{x}_{ij}, \boldsymbol{y}_{ij}), \boldsymbol{\theta}^{(r)}) = \frac{p(y_{ij}|\boldsymbol{x}_{ij}; \boldsymbol{\xi}_c^{(r)})\phi(\boldsymbol{x}_{ij}; \boldsymbol{\mu}_c^{(r)}, \boldsymbol{\Sigma}_c^{(r)})w_c^{(r)}}{\sum_{k=1}^{C} p(y_{ij}|\boldsymbol{x}_{ij}; \boldsymbol{\xi}_k^{(r)})\phi(\boldsymbol{x}_{ij}; \boldsymbol{\mu}_k^{(r)}, \boldsymbol{\Sigma}_k^{(r)})w_k^{(r)}},$$

leading to

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(r)}) = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{c=1}^{C} \tau_c((\boldsymbol{x}_{ij}, y_{ij}), \boldsymbol{\theta}^{(r)}) \{\log(w_c) + \log[p(y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{\xi}_c)] + \log[\phi(\boldsymbol{x}_{ij}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)]\}.$$

At this point the M-Step performs the estimation of the maximum likelihood, obtaining the new parameters $\boldsymbol{\theta}$ for the next iteration of the E-Step. The iterative process continues until a pre-defined convergence criterion is met. The convergence is guaranteed when the Aitken acceleration index (Böhning et al. 1994, McLachlan and Peel 2000) is lower than a defined threshold, which is typically set to 1e-04.
In particular, under the assumption that the algorithm is concluded when at the iteration r the estimated log-likelihood $l^{(r)}$ is approaching some value l^*, we can observe that

$$l^{(r+1)} - l^* \approx a(l^{(r)} - l^*) \tag{2.11}$$

for all r, and a included between 0 and 1. Following Böhning et al. (1994), at the $(r+1)^{\text{th}}$ iteration, $a^{(r)}$ can be estimated as

$$a^{(r)} = \frac{l^{(r+1)} - l^{(r)}}{l^{(r)} - l^{(r-1)}} \tag{2.12}$$

and the Aitken accelerated estimate of the log-likelihood $l_A^{(r+1)}$ is

$$l_A^{(r+1)} = l^{(r)} + \frac{1}{1 - a^{(r)}}(l^{(r+1)} - l^{(r)}) \tag{2.13}$$

as demonstrated developing a geometric series in Böhning et al. (1994). The identification of an adequate number of clusters is a key issue in model-based clustering. This problem is strictly related with the model selection, and the EM algorithm is not able to estimate the optimal number of latent clusters in the data. Conversely, the number of latent groups must be specified as an initialization parameter of the EM algorithm. In this context Bayesian information criterion (BIC, Schwarz (1978)) is commonly adopted for choosing the number of latent clusters (McLachlan and Peel 2000), and the evidence state that penalized log-likelihood criteria, including Akaike's information criterion (AIC) and BIC, do not underestimate the optimal number of components (McLachlan and Rathnayake 2014). A commonly adopted criterion to select the optimal number of latent clusters requires to restart the estimation process several times initializing the EM algorithm with an increasing number of components. The optimal number of clusters is usually defined as the number which provides the lower value of BIC.

2.5.2 Standard errors via parametric bootstrap

The end of the EM algorithm provides two main results: the allocation of the observations to one of the identified clusters and the estimates of the parameters for both the fixed and the random effects of the regression part. Altough EM estimation process is widely adopted for MLE in model based-clustering, an assessment of the parameters variability is needed, due to the potential small dimension of the clusters or when the clusters are not well separated (Basford et al. 1997). Furhtermore, when the number of level 2 units is small the asymptotic MLE procedures may not accurately estimate the uncertainty associated with the parameter estimates from the EM algorithm (O'Hagan et al. 2015). For these reasons, we include in the estimation process a further step consisting of a bootstrap process, in order to assess the variability of the EM estimates. In particular, we implement a parametric bootstrap approach for mixed models, adapting the steps described in Carpenter et al. (2003) to the multilevel cluster weighted models framework. Considering as an example a multilevel random intercept logistic framework, and using the notation of Equation (2.9), the bootstrap approach follows these steps:

1. We denote by $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\tau}_0^2$ the fitted parameters in one of the identified clusters, where we drop the index c for simplicity;
2. We simulate the vector of the random effects $u_{0j}^* \sim N(0, \hat{\tau}_0^2)$, for $j = 1, \ldots, J$;
3. We simulate the bootstrap data y_{ij}^* from a Bernulli(π_{ij}^*) with
$$\boldsymbol{\pi}_j^* = \exp(\hat{\beta}_0 + \hat{\beta}_1 X_j + u_{0j}^*)/(1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_j + u_{0j}^*)),$$
maintaining the same sample size overall and within group;
4. We refit the model on y_{ij}^* and we obtain the set of bootstrap parameters $\hat{\beta}_0{}^*$, $\hat{\beta}_1{}^*$ and $\hat{\tau}_0^{2*}$;
5. We repeat the steps 2-4 B times, where B is the number of bootstrap iterations.

2.6 Simulation study

In this section, we present numerical studies based on simulated data, assuming a Bernoulli distribution for the dependent variable, conditional on one independent variable. We consider 150 units at level 1 (patients) and 20 units at level 2 (hospitals), so the data are in the form: $\{(x_{ij}, y_{ij})\,;\, i = 1, \ldots, 150 \text{ and } j = 1, \ldots, 20\}$. We simulate $n = 3,000$ observations, generated from the model in Equation (2.9) with $C = 3$ equally-sized latent classes ($w_1 = w_2 = w_3 = 1/3$), so that in this multilevel structure each hospital for each cluster discharge 50 patients. The covariate X is simulated according to a Gaussian distribution with mean specified

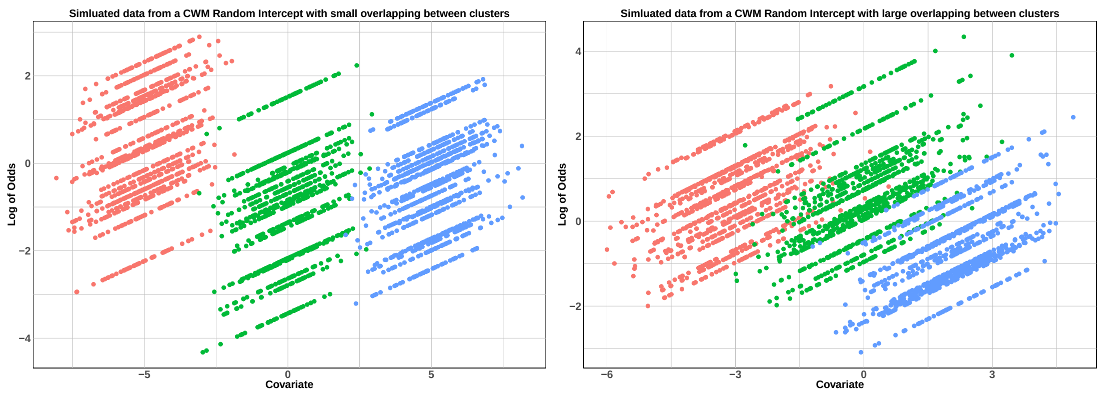
below, whereas all the variances in the model are set to 1.
The aim of these simulations is to compare our model (MCWM) with the multilevel logistic random mixture model (MRMM), which is presented in Section 2.3. All models have been fitted according to the maximum likelihood approach. The MCWM uses the EM algorithm described in Section 2.5.1 and coded in R. MRMM has been fitted using the MPlus software (Geiser 2012).
Models are compared on the basis of the estimated parameters, the accuracy of the classification of the binary dependent variable, and the analysis of the model's ability in allocating the observations to the right latent class. The estimation is compared by the mean over the number of latent clusters of the distance between the parameters fixed for the simulations and the estimated parameters. This comparison is provided separately for the intercepts β_{0c} and the slopes β_{1c} using the following index:

$$\Delta = \frac{\sum_{c=1}^{C}(\gamma_c - \hat{\gamma_c})^2}{C},$$

where γ_c are the true parameters and $\hat{\gamma_c}$ are their estimates. In order to analyse the accuracy of the model in predicting the values of the binary dependent variable, we generated 100 new datasets using the same parameters setted for the simulated data and then we apply the parameters estimated using both MCWM and MRMM for the training data. Adopting a threshold of 0.5 to the predicted probability of each model, we predict the values of the binary dependent variable. Hence we create the confusion matrix comparing the predicted y with the observed and we calculate the accuracy index as follows:

$$\text{Accuracy} = \frac{\sum_{i=1}^{2} \text{diag}_i(\text{Confusion matrix})}{\text{Number of observations}},$$

where diag_i are the two values in the diagonal of the confusion matrix. The comparison is provided by the mean of the 100 replicated tests. We, therefore, test the accuracy of the predictions using the ROC curve. The ROC curve is a plot comparing the true-positive rate (sensitivity) with the corresponding false-positive rate (1-specificity), which is helpful in evaluating a classification model. In our case, the sensitivity is the ability of the model to correctly classify each observation in the two values of the dependent variable (true positive rate), whereas the specificity is the ability of the model to correctly identify that the predicted value of an observation does not match the observed one (true negative rate). Both the sensitivity and specificity are based on a threshold which is used to define the confusion matrix. If, for example, the threshold equals to 0.5, all the observations with a predicted value grater than 0.5 are assigned to the value 1, otherwise to 0. In the confusion matrix this allocation is compared with the original value of the dependent variable. The results of both specificity and sensitivity, calculated for several thresholds, provide the points needed to define the ROC curve. The comparison between different models is based on the area under the curve: the higher is the area the better is the performance of the model (Zou et al. 2011).



(a) First Simulation *(b)* Second Simulation

FIGURE 2.1: Simulations from a random intercept CWM

The last comparison which we provide measures the ability of the models in allocating the observations. In this sense, we compare the allocation provided by the models with the allocation defined when we have simulated the data. At the end of the estimation process, we compare the number of observations allocated to the right cluster with the overall number of observations, obtaining the percentage of exact allocation.
Finally, in this section, we provide evidence about the performance of the bootstrap method for building confidence intervals for the parameters.

2.6.1 Comparison of models on data generated from a two-level random intercept CWM

In this study, we consider two simulations with different level of separation between clusters. The motivation is that the higher the level of proximity among clusters, the higher the expected difficulty for the estimation processes in allocating each observation to the right cluster. In particular, in the first simulation (Figures 2.1a) the observations in the clusters show a low level of overlapping, with only few observations that can be confused at the time of the allocation among clusters. In the second simulation (Figure 2.1b) the clusters are more overlapped and this should affect the ability of both the MRMM and the MCWM to correctly identify the latent groups and then to estimate the model parameters. Each color in the graph indicates one of the three simulated clusters. The parameter setting for the two simulations is presented in Table 2.1, following the same notation as in Equation (2.9) for regression parameters, where μ_c represents the mean of the density for the covariate X, depending on the C latent clusters. Because MCWM models the distribution of X, we expect that it should obtain a better allocation than MRMM, particularly when the level of overlapping between the clusters increases.
Table 2.2 shows the accuracy of the estimated parameters of the fixed part of the model for both MRMM and MCWM, using the measures described before. For the first simulation, we observe that both MRMM and MCWM estimate a set of parameters which is quite similar to

	Latent classes	First simulation	Second simulation
β_{0c}	$c = 1$	3	2
	$c = 2$	-1	0.5
	$c = 3$	-3	-2
β_{1c}	$c = 1$	0.3	0.5
	$c = 2$	0.3	0.5
	$c = 3$	0.3	0.5
μ_c	$c = 1$	-5	-3
	$c = 2$	0	0
	$c = 3$	5	2

TABLE 2.1: Parameters used for data generation in the two simulations from a random intercept CWM.

Model framework		Fist simulation	Second simulation
MRMM	$\Delta_{intercept}$	1.3250	5.3560
	Δ_{slope}	0.0001	0.5630
MCWM	$\Delta_{intercept}$	0.1273	1.3708
	Δ_{slope}	0.0001	0.4906

TABLE 2.2: Accuracy of the parameters' estimates for MRMM and MCWM for the two simulations from a random intercept CWM.

	Overall Accuracy
MRMM	0.6839
MCWM	0.7750

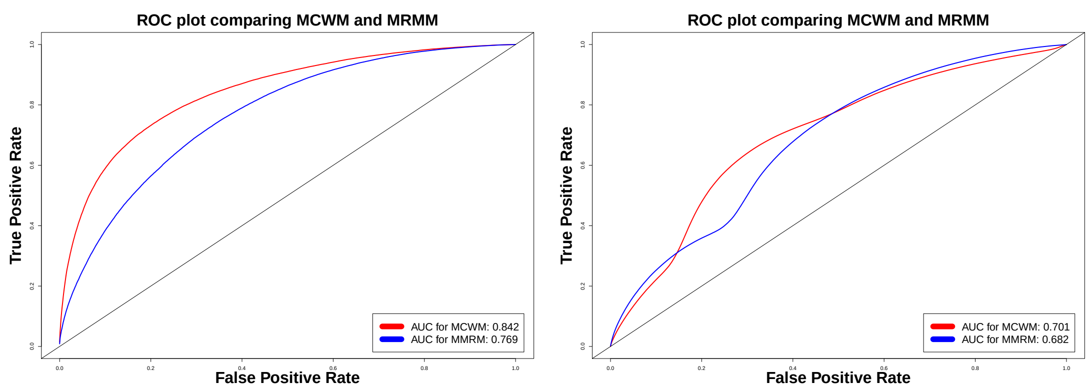
(a) First simulation

	Overall Accuracy
MRMM	0.6019
MCWM	0.6787

(b) Second simulation

TABLE 2.3: Performance comparison for MRMM and MCWM based on classification accuracy

the parameters used to simulate the data. Conversely, in the second simulation, the MCWM performs better moreover in terms of model intercept. Table 2.3 shows the comparison in terms of classification accuracy. Even in this case, the level of overlapping has an impact on the ability of our model in predicting the dependent variable. Nevertheless, in both simulations the accuracy of the MCWM is better than that of MRMM. Figure 2.2 shows the ROC curves for both the simulations. In the first simulation the area under the curve is good for both the compared model, and we can appreciate that the MCWM obtain a better performance. In the second simulation we observe a lower performance for both the models which reduce their ability in terms of both true and false positive rates. Finally, Table 2.5 shows the comparison in terms of the percentage of exact allocation of the observations into clusters. MRMM and MCWM perform both well in this first simulation, obtaining a performance close to 100% of exact allocation. This was an expected result as the clusters are quite separated and both MRMM and MCWM are able to allocate the units to their right cluster. In the second simulation, instead, the performance for both the models decreases, but MCWM performs better than MRMM.



(a) First Simulation *(b)* Second Simulation

FIGURE 2.2: ROC Curve comparing the accuracy of predicting the dependent variable in the two simulations from a random intercept CWM. AUC in the legend represents the Area Under the Curve.

	% of exact allocation
MRMM	98.7%
MCWM	99.5%

(a) First simulation

	% of exact allocation
MRMM	64.7%
MCWM	80.2%

(b) Second simulation

TABLE 2.5: Performance comparison for MRMM and MCWM based on percentage of exact allocation to the clusters

2.6.2 Bootstrap coverage

In Section 2.5.2 we described bootstrap procedure to derive the confidence intervals for the model parameters. Methods to construct confidence intervals are typically evaluated by the coverage. In our case the procedure described in Section 2.5.2 produces a confidence interval for each of the model parameters. In a simulation setting we construct 500 datasets from the same model after fixing the values of the true parameters. For each parameter, the coverage is defined by the percentage of times that the true parameter falls in the 500 generated confidence intervals. If the confidence level is set to the 95% one would expect the coverage close to this nominal value (Carpenter and Bithell 2000).

To verify the coverage, we perform a simulation study on a dataset with a hierarchical structure and a different number of units at level 1 and level 2 (see Table 2.7 for details). We consider a binary dependent variable and one covariate distributed as a standard normal. We consider the case of no clusters and we fix the parameters to $\beta_0 = -2$, $\beta_1 = 2$ and $\tau_0^2 = 4$, using the notation of Equation 2.9. Table 2.7 shows a relatively good performance in terms of coverage for the bootstrap algorithm, somewhat lower than 95% but in line with (Carpenter and Bithell 2000). In addition, the table shows that the performance increases with the number of observations

Parameters	Coverages (%) for the different sample sizes		
	40 level 2, 20 level 1	80 level 2, 40 level 1	160 level 2, 80 level 1
β_0	87.00%	89.10%	90.30%
β_1	90.00%	91.20%	91.80%
τ_u^2	87.30%	88.20%	89.10%

TABLE 2.7: Coverage of the parametric bootstrap approach for 95% confidence intervals from simulated data with varying sample sizes.

2.7 Mortality evaluation by a MCWM approach based on binary dependent variable

In this section we provide an application of the MCWM to real data in order to demonstrate its suitability for the evaluation in healthcare, and its ability in disentangle latent heterogeneity in the data. For this application we consider the Lombardy case-study described in Chapter 1, and we evaluate the hospital performance in terms of mortality, the most adopted healthcare outcome. We complete the outcome evaluation by providing the hospitals' league tables (Goldstein and Spiegelhalter 1996). In this way, we demonstrate that MCWM can be an optimal tool for policy makers interested in outcome evaluation. The outcome of interest is 30-day mortality, the most used proxy of quality in this research field. This outcome is measured by merging the hospitalization records with the registry of citizens conserved in Lombardy, where we can find the date of death for each patient. In this way we can identify whether a patient discharged alive has died within 30 day after the discharge.

Data concerning hospitalizations occurred in 2014 in 127 hospitals in Lombardy, provide information on a number of selected patients' characteristics, namely sex, age, the DRG weight, measuring the resources used by the hospital to treat each patient, and the Elixhauser index (Elixhauser et al. 1998), measuring the level of patients' comorbidities. Both the DRG weight and the Elixauser index are here considered as a proxy of patients' severity.

We analyse the MCWM on two different disciplines: cardiosurgery and medicine. Cardiosurgery is a highly specialized discipline admitting patients that need complex surgical intervention, but with a low risk of mortality, whereas medicine is a widespread general discipline, characterized by a high risk of mortality. We exploit the hierarchical structure of patients nested within hospitals using a multilevel model and we use the proposed MCWM to investigate whether there is evidence for further latent structures. We consider for simplicity models with two clusters. The descriptive statistics in Table 2.8 allow us to appreciate the different case-mix of patients admitted in the two considered wards and in the two identified clusters. Patients in cardiosurgery are on average younger than patients in medicine, while the risk of mortality in medicine is 10 times higher compared to cardiosurgery. The value of the DRG weight shows how cardiosurgery is a highly specialized discipline. The clustering composition for cardiosurgery indicates how the two latent groups mainly differ according to the age (in cluster 2 the patients are younger). Considering the clusters identified in medicine, we observe that patients in cluster 2 are younger than patients in cluster 1, which leads to a lower risk of mortality for this cluster compared to cluster 1. Moreover, higher levels of comorbidities are observed for the patients allocated to cluster 1.

We compare the results of the evaluation performed using MCWM with a typical evaluation

		DRG Weight	Comorbidity	Age	Female	Mortality
Cardiosurgery						
Cluster 1	Mean	5.4110	1.2235	70.5777	0.3305	0.0140
	Std Dev	2.8120	1.0693	8.4312	0.4704	0.1175
	Num of Obs	7,788	7,788	7,788	7,788	7,788
Cluster 2	Mean	5.5688	1.0551	42.6457	0.3363	0.0067
	Std Dev	3.3237	0.8868	8.4884	0.4727	0.0819
	Num of Obs	889	889	889	889	889
Medicine						
Cluster 1	Mean	1.0922	1.3463	79.0967	0.5179	0.1660
	Std Dev	0.7092	1.1230	9.1036	0.4997	0.3721
	Num of Obs	119,678	119,678	119,678	119,678	119,678
Cluster 2	Mean	0.9399	0.8611	44.0957	0.4851	0.0605
	Std Dev	0.5520	0.9337	10.3680	0.4998	0.2385
	Num of Obs	18,407	18,407	18,407	18,407	18,407

TABLE 2.8: Descriptive statistics of the two clusters identified by MCWM on the two separated wards of cardiosurgery and medicine.

based on multilevel model. This comparison is performed in terms of goodness of fit and parameters estimates of both the fixed and random effects.The goodness of fit is evaluated according to the BIC. Table 2.9 shows how the multilevel cluster-weighted model has a lower BIC compared to the classical multilevel model. Looking further at the parameter estimates, Table 2.9 shows how several effects of the covariates on the risk of mortality are different between the two latent groups, both in terms of magnitude and of direction of the relationship, and how these differences are not picked up by the standard multilevel model. The significance of the estimates is evaluated via the parametric bootstrap approach described before. For cardiosurgery, our proposed model finds a different direction and significance for the effect of age and DRG weight on mortality among the two clusters. Whereas in the standard multilevel model the coefficient related to the DRG weight is negative and not significant, the application of the MCWM shows how this relationship is negative and significant for the patients allocated in cluster 1, indicating that the higher the resources used by the hospital to treat patients in this group the lower their risk of mortality, while the relationship is positive and significant for the patients allocated in cluster 2, demonstrating that for these patients a higher severity implies a higher risk of death. Moreover, we observe that cluster 2 is characterized by a higher magnitude of coefficients related to sex and DRG weight and the estimated coefficient for comorbidities is slightly higher for cluster 1 compared to that of cluster 2. In contrast to this, in the ward of medicine, we do not observe any differences in the direction of coefficients in the compared models, but using the

MCWM we detect a different magnitude for the coefficients related to comorbidities, age and DRG weight. In particular in cluster 2 the effect of age and comorbidities is more than double compared to the coefficients in cluster 1. Therefore, in medicine we observe a gender effect in cluster 1, with female patients having a lower risk of morality in that cluster.

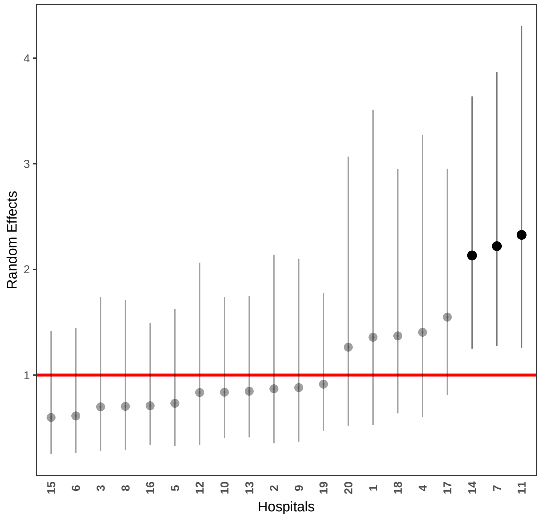
The effects detected in Table 2.9 have a significant impact on the final league tables, and show

Cardiosurgery	Estimates	Multilevel	MCWM	
			Cluster1	Cluster2
	(Intercept)	-8.3233***	-8.8560***	-10.2914***
	Female	0.1786	0.2278	0.6188
	Age	0.0517***	0.0601***	-0.0032
	DRG Weigth	-0.0038	-0.0849	0.4145***
	Comorbidity	0.2485***	0.3142***	0.1834
	BIC	-58295.35	-59399.93	
Medicine				
	(Intercept)	-5.7630***	-6.9378***	-12.0587***
	Female	-0.2683***	-0.3280***	-0.0745
	Age	0.0491***	0.0613***	0.1541***
	DRG Weigth	0.3679***	0.3773***	0.5489***
	Comorbidity	0.0677***	0.0328***	0.7610***
	BIC	-1128332	-1167850	

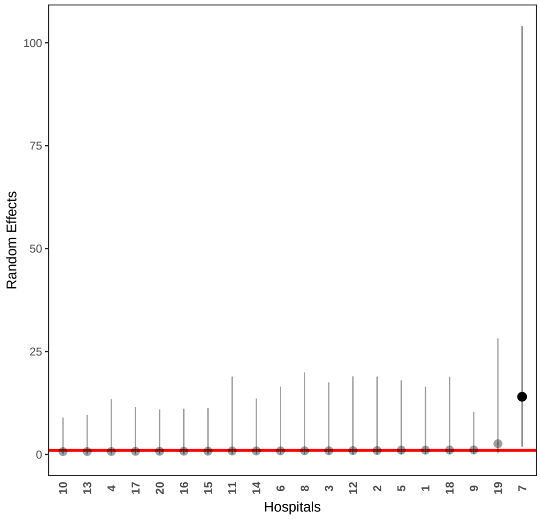
TABLE 2.9: Regression coefficients of the multilevel and the MCWM models fitted to the Lombardy healthcare data for the cases of cardiosurgery and medicine.

also here a difference between the results obtained by the proposed MCWM model and by a standard multilevel approach. Figure 2.3 shows the league tables for cardiosurgery for the MCWM (top) and multilevel (bottom) approaches. Figure 2.4 provides the same results for medicine. These figures are drawn based on the estimated random effects and on confidence intervals obtained using the same parametric bootstrap approach described before. This is implemented, for each cluster, in the function `plotREsim` in the R package `merTools` (Knowles and Frederick 2016). Hospital random effects different from the overall average (i.e. when the confidence interval does not cross the red line) are highlighted in bold. The figures show how, in cardiosurgery, the league tables of the multilevel model and of cluster 1 of MCWM are the same, but MCWM allows to detect a bad performance related to the hospital coded as `7` in cluster 2 . This is the only hospital presenting a bad performance in this cluster, and it is the only hospital with bad results both in cluster 1 and cluster 2. This means that the patients allocated in cluster 2 receive the same quality in all the hospitals except for hospital `7`.

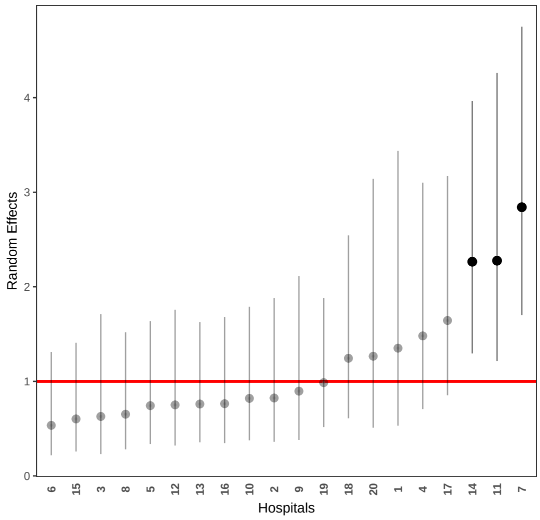
In medicine, a direct comparison of the league tables is complicated by the large number of hospitals involved in this analysis. However, we are able to compare the overall heterogeneity of Figure 2.4c with the cluster specific heterogeneity in Figure 2.4a and Figure 2.4b. As we observed for cardiosurgery, patients allocated in cluster 2 receive a more homogeneous quality,



(a) League Table for MCWM - Cluster 1



(b) League Table for MCWM - Cluster 2



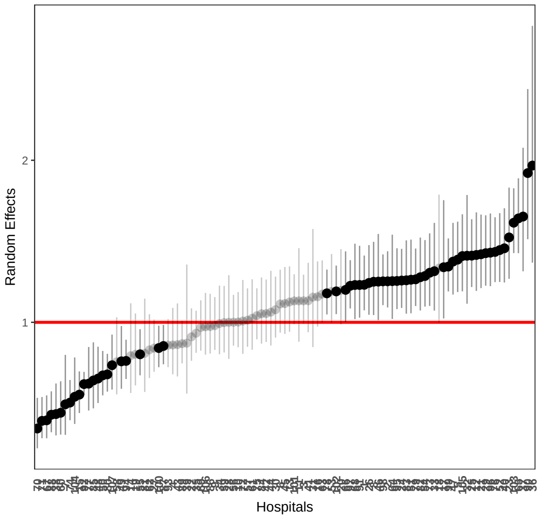
(c) League Table for Multilevel Model

FIGURE 2.3: League Tables for the MCWM (top) and Multilevel Model (bottom) in Cardiosurgery

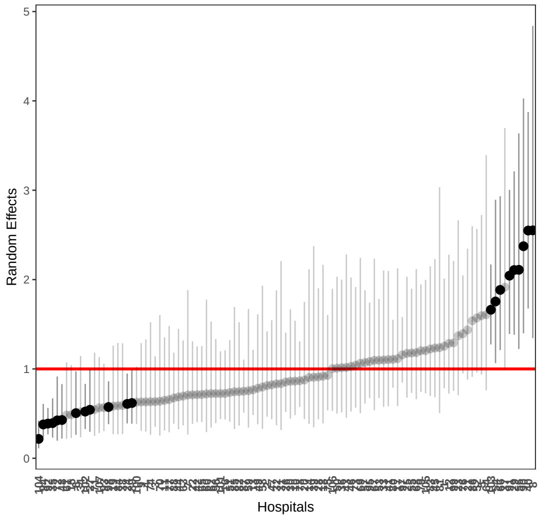
and this could be a useful cluster for policy makers to identify the hospitals that have a bad performance for a specific analysis.

2.8 Conclusions

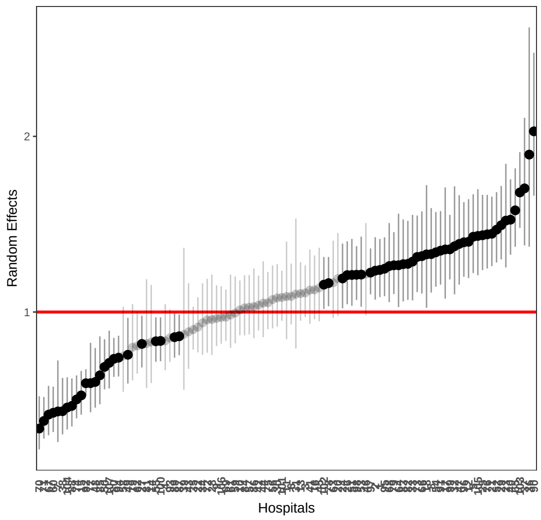
In this chapter we presented an extension of cluster-weighted models for hierarchical data and a binary dependent variable. The extension of the multilevel cluster-weighted models to the case of Bernoulli dependent variable, is particularly interesting in healthcare context, where often binary outcomes are considered. The proposed model allows to identify latent clusters in the data, related to both the outcome and the risk-adjustment variables, as well as to account for



(a) League Table for MCWM - Cluster 1



(b) League Table for MCWM - Cluster 2



(c) League Table for Multilevel Model

FIGURE 2.4: League Tables for the MCWM (top) and Multilevel Model (bottom) in Medicine

the hierarchical structure of the data which is typical in healthcare evaluations. We presented inference of the model, including an EM-algorithm for parameter estimation and a parametric bootstrap approach for building confidence intervals for the parameters.

Numerical studies compared the performance of the MCWM with the performance of the mixture of multilevel logistic models. The results showed that the MCWM achieves better (or at least not worse) results than the MRMM.

An empirical application is provided, using a rich dataset gathered from the Lombardy healthcare system. The application confirmed that the proposed MCWM detects two well-defined latent groups within the hierarchical structure of hospitals. Interestingly, the regression coefficients have different signs, magnitude and statistical significance for the two different groups, showing the advantage of this method compared to a standard multilevel model. The Bayesian information criterion supported this comparison. In addition to the fixed effects, the league tables of hospitals constructed from the random effects showed different patterns between the two latent groups.

Although the application shown here is mostly illustrative of the method, it may stimulate further investigations of the determinants of latent class membership in a similar context. Such investigations may focus on the patients' severity based on unobserved characteristics, and also considering different clinical disciplines or illnesses, such as AMI. Furthermore, this new model framework opens a new way in the healthcare evaluation, in particular it can help in handling with the issue underlined by Lilford et al. (2004), which specified that an overall evaluation does not consider latent patients characteristics, neither observable nor measured characteristics . The adoption of MCWM can reduce this issue providing new evidence in the effectiveness evaluation process.

In general, the adoption of the MCWM may have great implications for policy makers and healthcare managers. Indeed, the use of MCWM highlighted the effects that could be masked using a classical multilevel approach, and how these could impact the final rankings of hospitals. This is clearly expressed in the application where the league tables provided by MCWM make clear the quality provided in each cluster and how the quality by a hospital can be different focusing on different clusters.

However, the proposed model can be widely applied in all research fields where there is a binary outcome and a hierarchical structure of the data. For example, education is a typical field of research where data are characterized by a hierarchical structure and binary outcomes are often considered.

Chapter 3

Policy evaluation using a Difference-in-Difference approach

In this chapter, we evaluate the effectiveness of a pay-for-performance program on the basis of five health outcomes and across a wide range of medical conditions. In Lombardy a rewarding program was introduced in 2012, and we aim to evaluate this policy using a difference-in-differences approach. The model includes multiple dependent outcomes, that allow quantifying the joint effect of the program, and random effects, that account for the heterogeneity of the data at the ward and hospital level. Our results show that the policy had a positive effect on the hospitals' performance in terms of those outcomes that can be more influenced by a managerial activity, namely the number of readmissions, transfers and returns to surgery room.

3.1 Literature review on P4P programs

As anticipated in Chapter 1, the adoption of a Pay-for-Performance (P4P) approach aims to drive the hospitals in improving quality. The idea behind the implementation of a P4P approach is quite simple: in order to improve the overall quality delivered, healthcare providers are given the opportunity to have their reimbursements increased when they achieve specified quality benchmarks (Alshamsan et al. 2010, Eijkenaar et al. 2013). From an economics perspective, the hospital is considered as a profit maximizer agent which is encouraged to compete for quality in order to obtain a financial reward, rather than to attract more patients. Therefore, a P4P program is considered efficient when an improved quality of care is achieved with equal or lower costs (Emmert et al. 2012), and, clearly, the evaluation of the quality delivered is a crucial part to every P4P approach.

The aim of the current chapter is to contribute to the existing literature by providing a thorough evaluation of a P4P program and its effect on the overall quality of the healthcare system. As in

the evaluation of any policy, a choice needs to be made about which health outcome to use for quantifying the impact of the P4P program. In many studies, a single outcome is considered. For example, Sutton et al. (2012) quantify the impact of the P4P adoption in England by analysing the hospital overall mortality. In addition, the evaluation of P4P programmes is often confined to specific clinical conditions, such as acute myocardial infarction (AMI), coronary artery bypass graft surgery (CABG), heart failure, pneumonia, and hip/knee replacement (Glickman et al. 2007, Jha et al. 2012, Levin-Scherz et al. 2006, Shih et al. 2014, Sutton et al. 2012). In contrast to these studies, we analyse the P4P effect using five different health outcomes and based on the overall case-mix hospitalizations of the wards considered. Moreover, for the first time in a P4P study, we investigate the policy effect with regards to hospital ownership, by evaluating possible different reactions to the P4P program among the private (for-profit and not-for-profit) and public providers, and also with regards to the different wards, by evaluating whether surgical and medical wards reacted differently to the policy.
The chapter proceeds as follows: in Section 3.2 we describe the healthcare system in Lombardy and the adopted P4P program; in Section 3.4 we describe the chosen methodological approach; in Section 3.3 we present the data used in the analysis and in Section 3.5 we discuss the results of the policy evaluation. Section 3.6 concludes the chapter.

3.2 The P4P program in Lombardy

The definition of the healthcare system in Italy and, in specific, in Lombardy, has been discussed in Chapter 1. Here, we describe how Lombardy introduced in 2012 a new reform which included a P4P program, whereby the increment of the hospital annual budget, based on a weighted mean of the hospital's evaluated outcomes.The hospitals are ranked according to this measure: the first hospital in the ranking receives an increment of 2% of its annual budget, the worst one gets a penalty of 2%, whereas all the others receive an amount between the interval $[-2\%, +2\%]$, and proportional to the distance between their score and the score of the last hospital in the category's ranking (p.84 of Region (2011), Region (2012)).
In the first instance, the regional healthcare management decided to evaluate the weighted outcome measures only on 9 wards, i.e. cardiology, cardiosurgery, neurosurgery, neurology, oncology, general medicine, urology, orthopaedic, surgery. The wards were chosen according to the coverage within the hospitals, the inclusion of both medical and surgical disciplines as well as the level of specialization (cardiosurgery and neurosurgery). Further details on the policy introduction can be found in the regional resolution (Region (2011)).
Following these premises, the study discussed in this chapter, in line with the designs adopted by previous studies (Lindenauer et al. 2007, Rosenthal et al. 2005), analyses 9 hospital wards covering a wide range of medical conditions, exogenously selected for the treatment group, and

subjected to the P4P program, whereas the other hospital wards were not involved in the program, and they are considered for the control group. Data were collected both two years prior and two year post introduction of the policy for all hospitals in the Lombardy region. The aim of this chapter is then to evaluate the effect of the policy on the basis of the data collected.
It is interesting to note that the incentive is provided to the hospital as a whole, as typical of P4P programmes in healthcare (Cashin et al. 2014). The individual hospitals have then a large accountability on how the allocate the incentive payments. Typically, provider institutions allocate the financial resources to make general improvements in the service delivered, and in particular related to the performance measures. In the case of the Lombardy region, it is also possible that the physicians and/or nurses working in the treated wards received a direct bonus as a drive to performance improvement. This is however bound to vary across hospitals, so we do not expect to see the impact of this in our policy evaluation.

3.3 Data and descriptive statistics

In this section, we give the specific of the data used for the policy evaluation.
The database was gathered from the Lombardy healthcare information system. Data were collected on patients admitted to 142 hospitals during the four years 2010-2013 (two before and two in the policy-on period). In this period the hospitals provided 3,581,389 hospitalisations, coded in the available hospital discharge chart. In our analysis, we included patients admitted for acute care and we excluded patients living outside the region, patients younger than two years old or patients hospitalized in day-hospital, rehabilitation or palliative treatments.
Table 3.1 provides details for the variables considered in the study and the five outcomes. We used variables both at the patient and ward/hospital level. At the patient level, there is information on their gender, age, number of transit to the intensive care unit during hospitalization, the weight of the financial reimbursement corresponding to the patient's disease, and the comorbidity index. The latter is measured as in Elixhauser et al. (1998) and indicates the presence of one or more additional diseases or disorders co-occurring with a primary disease or disorder. At the hospital level, we know whether the hospital is affiliated to a medical school in which medical students receive practical training, whether the hospital is mono-specialistic or general, and whether there is presence of high-technology instrumentation in the ward. Finally, we include the hospitals' ownership, which categorizes the hospital as private for profit, private not-for-profit or public, and we distinguish wards whose prevalent activity is surgical from the medical ones. The effectiveness of the policy is evaluated over the five health outcomes, namely mortality, readmissions, transfers, returns, and voluntary discharges. We should clarify that the outcome return to the surgery room can be evaluated only for the surgical wards.

	Untreated				Treated			
	Pre-policy		Post-policy		Pre-policy		Post-policy	
	2010	2011	2012	2013	2010	2011	2012	2013
Patient								
MALE	0.2589	0.2613	0.2646	0.2673	0.5399	0.5413	0.5397	0.5383
	(0.43)	(0.43)	(0.44)	(0.44)	(0.49)	(0.49)	(0.49)	(0.49)
AGE	46.076	46.585	46.973	47.212	64.526	64.877	65.054	65.384
	(21.1)	(21.1)	(21.2)	(21.3)	(18.7)	(18.5)	(18.6)	(18.5)
DRGWEIGHT	0.892	0.9127	0.9139	0.919	1.2974	1.3252	1.3167	1.3277
	(0.81)	(0.84)	(0.83)	(0.85)	(1.12)	(1.15)	(1.12)	(1.13)
COMORBIDITY	0.2379	0.2128	0.2156	0.2099	0.4082	0.3303	0.325	0.3121
	(0.58)	(0.55)	(0.56)	(0.55)	(0.72)	(0.66)	(0.65)	(0.64)
INTCARE	0.015	0.0164	0.017	0.0174	0.0644	0.0676	0.0677	0.0687
	(0.12)	(0.12)	(0.12)	(0.13)	(0.24)	(0.25)	(0.25)	(0.25)
Ward/Hospital								
TECHNOLOGY	0.8585	0.8588	0.8614	0.8683	0.8079	0.807	0.8111	0.8119
	(0.34)	(0.34)	(0.34)	(0.33)	(0.39)	(0.39)	(0.39)	(0.39)
TEACHING	0.2684	0.2708	0.2754	0.2734	0.2455	0.2456	0.2471	0.2456
	(0.44)	(0.44)	(0.44)	(0.44)	(0.43)	(0.43)	(0.43)	(0.43)
SPECIALISED	0.052	0.0474	0.0482	0.049	0.0387	0.0386	0.0406	0.0393
	(0.22)	(0.21)	(0.21)	(0.21)	(0.19)	(0.19)	(0.19)	(0.19)
SURGICAL	0.5637	0.5535	0.5646	0.562	0.5088	0.4884	0.4942	0.487
	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.5))	(0.49)
OWN:NOPROFIT	0.0758	0.0765	0.077	0.0793	0.0947	0.0948	0.0975	0.096
	(0.26)	(0.26)	(0.26)	(0.27)	(0.29)	(0.29)	(0.29)	(0.29)
OWN:PROFIT	0.1376	0.1373	0.1346	0.1264	0.2314	0.2354	0.2308	0.2327
	(0.34)	(0.34)	(0.34)	(0.33)	(0.42)	(0.42)	(0.42)	(0.42)
OWN:PUBLIC	0.7866	0.7862	0.7884	0.7943	0.6739	0.6698	0.6717	0.6713
	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.49)	(0.5))	(0.49)
Outcomes								
TRANSFERS	0.0056	0.0052	0.0036	0.0035	0.0127	0.0127	0.0053	0.0051
	(0.07)	(0.07)	(0.06)	(0.05)	(0.11)	(0.11)	(0.07)	(0.07)
RETURN	0.0592	0.0632	0.0099	0.0108	0.0431	0.0443	0.0154	0.0161
	(0.23)	(0.24)	(0.09)	(0.10)	(0.20)	(0.20)	(0.12)	(0.12)
MORTALITY	0.0268	0.0276	0.029	0.0273	0.0593	0.0608	0.0611	0.0601
	(0.16)	(0.16)	(0.16)	(0.16)	(0.23)	(0.23)	(0.23)	(0.23)
READMISSIONS	0.1216	0.1149	0.1117	0.1091	0.1335	0.1277	0.1211	0.1111
	(0.32)	(0.31)	(0.31)	(0.31)	(0.34)	(0.33)	(0.32)	(0.31)
VOLDISCH	0.0084	0.0085	0.0082	0.0084	0.0088	0.0081	0.0076	0.007
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)

TABLE 3.1: Sample means and standard deviations in brackets for the covariates in the study from the Lombardy hospital inpatient stays for each year before and after the policy introduction.

Table 3.1 reports the average (and the standard deviations in brackets) of the variables in the dataset by treatment and across the four years of the study (two pre and two post policy). It appears that the mix of patients within the treated and untreated wards is relatively stable over time, but that there are differences between the two groups. In particular, patients that are admitted to the treated wards are on average older than those admitted to the untreated ward. In addition, the treated wards consider higher risk patients than the untreated wards in terms of DRGs weight, number of comorbidities and intensive treatment. The percentage of comorbidities (roughly 30%) is however still relatively small compared to other countries e.g. 0.69% in Northern Ireland in 2011/2012 (Reilly et al. 2015). This is justified by the coding rules that affect the healthcare system in Lombardy, whereby only the comorbidities directly connected with the treated DRGs are registered. Considering the variables related to the hospitals and the wards, we observe that the overall composition of the hospitals has not changed during the policy period, with surgical wards covering around 51% of the overall admissions. Moreover, 71% of the hospitalizations are provided by the public hospitals, whereas 30% of the patients are admitted to a private provider (20% in the for profit hospitals and 9% in the not-for-profit). With regards to the health outcome measures, three out of the five outcomes, namely transfers, return to the surgery room and readmissions, show a reduction after the introduction of the P4P program. The aim is to assess the significance of this finding after adjusting for the patient-level covariates identified in Table 3.1.

3.4 The Econometric Approach

We test the effect of the policy using a difference-in-differences (DID) approach (Abadie 2005, Blundell et al. 2004), using the data described in Section 3.3. To justify the suitability of this approach, the following considerations are needed:

1. The wards are split into a *treatment* group - the 9 wards that are used for the hospital evaluation - and a *control* group - the remaining wards. The allocation of the wards in one of these groups was made exogenously prior to the introduction of the policy (Region (2011)). There is an underlying assumption here that, although the incentive is provided to the hospital as a whole, the incentive is dictated only by the performance of the wards *treated*. Combined with the fact that the individual wards operate autonomously, the *untreated* wards can be considered as an independent group. A similar analysis was conducted by Sutton et al. (2012), where the treatment and control groups are defined within each hospital on the basis of selected diagnoses.

2. Units do not switch between the control and the treatment group: improvements in performance of the control group do not affect the financial incentives gained by the hospital.

We will however test whether there is evidence of a distortion of the hospital behaviour aimed at inflating the performance evaluation, such as the lift of resources in favour of the treated wards.

3. Any macro changes affect both groups equally and differences between the treatment and the control group remain constant in the absence of treatment, i.e. a parallel trend prior to treatment. The check of this assumption is going to be discussed later in the results section. Of notice is also the fact that the regional resolution was formally announced in December 2011 (Region (2011)), and applied from early January 2012 (Region (2012)). Thus, hospitals had no possibility to anticipate changes.

As discussed in Section 3.3, this policy evaluation is based on five health outcomes. Given the mix of patients in the different wards, the outcomes are first adjusted by patients characteristics via the use of a multilevel logistic mixed effect model (Goldstein 2010, Snijders and Bosker 2012). This model allows to account for the hierarchical structure of the data whereby patients are clustered into wards and wards are nested into hospitals. In addition, the longitudinal structure of the data means that a time effect is also to be expected. Following the notation introduced in Section 1.4, let Y_{iwjt} represent a binary health outcome for patient i (with $i = 1, \ldots, n_{wjt}$) in the ward w (with $w = 1, \ldots, W_{jt}$), belonging to the hospital j (with $j = 1, \ldots, J_t$), hospitalized at time t (in years, $t = 2010, \ldots, 2013$). Let π_{iwjt} be the conditional probability of Y_{iwjt} being equal to 1. We consider the model

$$\log\left(\frac{\pi_{iwjt}}{1-\pi_{iwjt}}\right) = \alpha + \boldsymbol{\eta} X_{iwjt} + \mu_{wjt} + u_{jt}, \tag{3.1}$$

where $\boldsymbol{\eta}$ is a vector of coefficients for the X_{iwjt} patient-level covariates, μ_{wjt} is a random effect of the ward w nested within hospital j at time t, capturing the latent heterogeneity of the wards, whereas u_{jt} captures the latent heterogeneity of the hospital j at time t. μ_{wjt} and u_{jt} are independent and identically distributed, $N(0, \tau_\mu^2)$ and $N(0, \tau_u^2)$, respectively, and are assumed to be uncorrelated with the regressors. Compared to the model described in Section 1.4 a time effect t and a ward effect w are included.

The model in equation 3.1 returns the patients' predicted probabilities

$$\hat{\pi}_{iwjt} = \frac{\exp\left(\hat{\alpha} + \hat{\eta}\, X_{iwjt} + \hat{\mu}_{wjt} + \hat{u}_{jt}\right)}{1 + \exp\left(\hat{\alpha} + \hat{\eta}\, X_{iwjt} + \hat{\mu}_{wjt} + \hat{u}_{jt}\right)}, \tag{3.2}$$

which we collapse at the ward level over time in order to obtain the average predicted health outcome

$$\mathrm{HO}_{wjt_m} = \frac{\sum_{i \in P_{wjt_m}} \hat{\pi}_{iwjt}}{|P_{wjt_m}|}, \tag{3.3}$$

where P_{wjt_m} is the set of patients admitted in the ward w of the hospital j in the month m ($m = 1, \ldots, 12$) of the year t and $|P_{wjt_m}|$ is the cardinality of this set.

The aim is now to quantify the policy effect on the basis of the five (adjusted) health outcomes. As we anticipate a correlation between the five health outcomes, we consider a multivariate DID model, rather than a separate model for each outcome. Adopting a multivariate model is not common in healthcare literature, but in this way, we are able to quantify the overall effect of the policy across all health outcomes, as well as at the individual level. Let then $\text{HO}^{(\nu)}_{wht_m}$ denote the health outcome ν, namely readmissions ($\nu = 1$), mortality ($\nu = 2$), return to the surgical room ($\nu = 3$), transfers ($\nu = 4$) and voluntary discharges ($\nu = 5$), at month m of year t ($t = 2010, \ldots, 2013$) of ward w ($w = 1, \ldots, W_j$) belonging to hospital j (with $j = 1, \ldots, J$). We consider the following multivariate mixed model:

$$\begin{aligned}\text{HO}^{(\nu)}_{wjt_m} \quad &= \alpha^{(\nu)}_j \; + \; \beta^{(\nu)} \, \text{TREATED}_{wj} \; + \; \textstyle\sum_{h=2011}^{2013} \gamma^{(\nu)}_h \, I(h = t) \; + \\ &\textstyle\sum_{h=2011}^{2013} \delta^{(\nu)}_h \, \left(I(h = t) \cdot \text{TREATED}_{wj}\right) \; + \; \upsilon^{(\nu)} \, \text{MONTH}_{t_m} \; + \; \epsilon^{(\nu)}_{wjt_m}, \end{aligned} \tag{3.4}$$

where the dummy variable TREATED_{wj} indicates whether the ward w is in the treatment group or not, the indicator variable $I(h = t)$ indexes the four years of the study (two pre and two post policy), with 2010 set as reference category, MONTH is a continuous variable, taking values 1 to 48 and added to correct for a possible seasonality effect, $\alpha^{(\nu)}_j$ is the random hospital effect for outcome ν, and the error $\epsilon^{(\nu)}_{wjt_m} = (\epsilon^{(1)}_{wjt_m}, \ldots, \epsilon^{(5)}_{wjt_m})$ has a multivariate distribution $\epsilon_{wjt_m} \sim N_5(0, \Sigma)$, with the covariance Σ accounting for possible dependencies between the different outcomes. The parameter $\delta^{(\nu)}_j$ is of interest in this model. Under the assumption of a parallel trend pre-policy, we expect $\delta^{(\nu)}_{2011} = 0$ for all outcomes, whereas the parameters $\delta^{(\nu)}_{2012}$ and $\delta^{(\nu)}_{2013}$ represent the DID of average outcomes between the treated and control wards from the pre to the post-policy years. The two different parameters for the post-policy period let us detect whether the impact of the policy was immediate in the first year of its introduction or whether it was delayed in the second year (Ayyagari and Shane 2015). This model allows us to detect the effect of the policy across all wards.

A second objective of the study is to detect whether the reaction to the P4P adoption is different depending on the ward's type. In particular, we group all wards into two types: surgical and medical, and extend the model in equation 3.4 to:

$$
\begin{aligned}
\mathrm{HO}^{(\nu)}_{wjt_m} = \alpha^{(\nu)}_j &+ \beta^{(\nu)}\,\mathrm{TREATED}_{wj} + \sum_{h=2011}^{2013} \gamma^{(\nu)}_h\, I(h=t) \\
&+ \sum_{k=1}^{2} \lambda^{(\nu)}_k I(k=\mathrm{SURGICAL}_{wj}) \\
&+ \sum_{h=2011}^{2013} \left(\delta^{(\nu)}_j\, I(h=t)\,\cdot\,\mathrm{TREATED}_{wj}\right) \\
&+ \sum_{h=2011}^{2013}\sum_{k=1}^{2} \left(\mu^{(\nu)}_{hk}\, I(j=t)\,\cdot\, I(k=\mathrm{SURGICAL}_{wj})\right) \\
&+ \sum_{k=1}^{2} \left(\varphi^{(\nu)}_k I(k=\mathrm{SURGICAL}_{wj})\cdot\mathrm{TREATED}_{wj}\right) \\
&+ \sum_{h=2011}^{2013}\sum_{k=1}^{2} \left(\psi^{(\nu)}_{hk} I(h=t)\cdot I(k=\mathrm{SURGICAL}_{wj})\cdot\mathrm{TREATED}_{wj}\right) \\
&+ \upsilon^{(\nu)}\,\mathrm{MONTH}_{t_m} + \epsilon^{(\nu)}_{wjt_m},
\end{aligned}
\tag{3.5}
$$

with the variable SURGICAL defined as 1 if the prevalent activity of the ward is surgical and 0 otherwise. In this model, the DID parameters $\psi^{(\nu)}_{hk}$, $j = (2012, 2013)$, are of interest as they represent the differences in average outcomes between the surgical treated wards and the surgical control wards, from the pre to the post policy period and with respect to the medical wards which are taken as the reference category. For this model, we do not consider the health outcome returns to the surgery room as this is observed only for the surgical wards.

Finally, in the results section, we also consider a similar model for the detection of possible differences in the reaction to the P4P adoption depending on the type of hospital ownership. In particular, we compare private for-profit, private not-for-profit and public hospitals. Due to the more strict budget constrains for private hospitals, these hospitals may react more actively to the policy than public ones. Furthermore, private for-profit hospitals are more oriented towards profit than the other hospitals and may therefore be more driven to increase their outcome measures in order to obtain a financial reward.

3.5 Results of the policy evaluation

In order to assess whether there has been an improvement in the healthcare quality following the introduction of the P4P policy, we use the multivariate DID approach discussed in Section 3.3. Table 3.2 reports the fixed effects estimates of the model in equation 3.4. As all outcomes are constrained to be between 0 and 1, the parameter estimates and the p-values are computed

	MORTALITY	READMISSIONS	RETURN	TRANSFERS	VOL. DISCH.
MONTHS	0.001	-0.001	0.001	-0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
TREATED	0.02***	0.004***	-0.037***	0.006***	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
YEAR_{2010}	0.044***	0.13***	0.084***	0.009***	0.009***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
YEAR_{2011}	0.044***	0.125***	0.082***	0.008***	0.008***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
YEAR_{2012}	0.045***	0.122***	0.021***	0.006*	0.008**
	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)
YEAR_{2013}	0.041***	0.118***	0.022***	0.005	0.008**
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)
$\text{TREATED}\cdot\text{YEAR}_{2011}$ (δ_{2011})	0.002	0.001	0.002	0.001	-0.001
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
$\text{TREATED}\cdot\text{YEAR}_{2012}$ (δ_{2012})	0.001	-0.005***	0.026***	-0.005***	-0.001
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
$\text{TREATED}\cdot\text{YEAR}_{2013}$ (δ_{2013})	0.005***	-0.011***	0.025***	-0.005***	-0.001
	[0.001]	[0.001]	[0.003]	[0.001]	[0.001]

The coefficients and standard errors (in brackets) are reported. *** represents significance at the 1% level, ** represents significance at the 5% level and * represents significance at the 10% level.

TABLE 3.2: Estimates for the fixed effects for the model in equation (3.4).

by a non-parametric bootstrap approach. For this, we use a method specifically developed for multilevel modelling (Carpenter et al. 2003), and which has been described in Section 2.5.2.

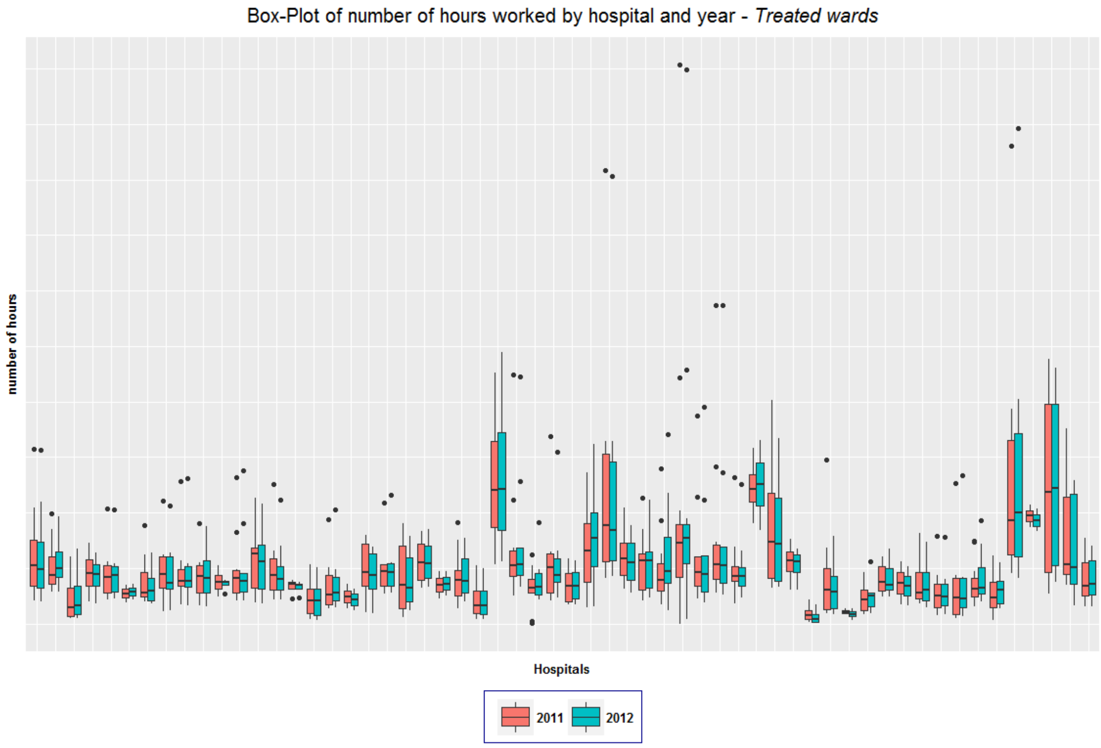
3.5.1 Testing the assumptions of a DID approach for policy evaluation

Table 3.2 shows how the parameters δ^{ν}_{2011} of the interaction between TREATED and YEAR_{2011} are not significantly different from zero. This provides evidence in favour of the parallel trend assumption for each individual health outcome, i.e. the differences between the average outcome of the treatment and control group are constant prior to the introduction of the policy. This assumption is needed in order to evaluate the impact of the policy using a DID approach. As we require a parallel trend to be satisfied for all health outcomes simultaneously, we use a multivariate analysis of variance test (MANOVA) to test the null hypothesis $H_0 : \delta^{(1)}_{2011} = \ldots \delta^{(5)}_{2011} = 0$ under the model in equation 3.4. The Wilks' lambda statistics returns a p-value of 0.2676, which provides further evidence in support of the parallel trend assumption across all health outcomes. Given that the incentive is provided to the hospital as a whole, it is also necessary to test whether the introduction of the P4P may have had a negative spillover effect between the treated and the untreated wards. This would violate the assumption of independence between the two groups and thus bias the policy evaluation. Although within each ward the physicians and nurses detain managerial freedom on whether and how to treat the patients, spillover effects could take the form of hospitals lifting resources in favour of the treated wards to the expense of the untreated wards. To this aim, we assess whether there has been a difference in the total number of hours worked by physicians and nurses within each hospital between the treated and the untreated wards from the year 2011 (pre-policy) to 2012 (post-policy). We consider 58 hospitals which have a balanced proportion of treated/untreated wards. Figure 3.1 shows the box-plot of the number of hours worked by hospital and year. The figure shows how, within each hospital, the number of hours worked is stable across the two groups and between the pre and post-policy period, suggesting that no shift of resources occurred, at least at the level of labour. This is supported by a non-significant p-value for the year-treatment interaction term (0.812) from a negative binomial generalised linear model which includes also fixed effects for hospitals. In addition to the allocation of resources, another possible spillover effect could result from the sharing of technological resources between the different wards. This may have an impact on surgical outcomes, such as the return to the surgery room in our case. We have no data to evaluate this, but we will take this into consideration when interpreting the results of the policy evaluation analysis.

Together with the spillover effects mentioned above between wards within the same hospital, the different providers may have also reacted to the policy by avoiding to treat high risk patients (Levaggi and Montefiori 2013). In order to check for this potential distortion, we have analysed whether the cream skimming index, calculated as in Berta et al. (2010), changed significantly between the pre and the post policy period. As above, we restrict the analysis to the hospitals which have a balanced proportion of treated/untreated wards and we perform the pre-post analysis separately for the treated and untreated groups. Using a multiple regression model, we find only four hospitals (out of 58) with a significant negative interaction with the post-policy term, two for the treated wards (p-values: $4.54E-08, 0.0025$) and two for the untreated ones (p-values: $0.02, 0.0314$). Thus, we conclude that overall the hospitals show no evidence of a gaming behaviour in selecting the mix of patients in the post-policy period.

3.5.2 Do the hospitals react positively to the policy?

We are now in a position to evaluate the impact of the P4P policy by considering the estimates of the coefficients of the interaction between the treatment variable and the post-policy years in Table 3.2, i.e. δ^{ν}_{2012} and δ^{ν}_{2013}. As all health outcomes are improved if they are reduced, a significant and negative coefficient for these interactions would mean that the P4P introduction had a positive effect on quality. This result is confirmed for readmissions ($\delta_{2012} = -0.0051$, $\delta_{2013} = -0.0112$) and transfers ($\delta_{2012} = -0.0046$, $\delta_{2013} = -0.0047$). This is a clear signal that the hospital activity was modified as a result of the P4P introduction, as both readmissions and transfers are directly affected by the hospital organization. In particular, the results show that the P4P program may have reduced the hospital attitude of readmitting patients in order to increase the number of the DRGs provided (Berta et al. 2010). The reduction in the transfers of the patients between hospitals in the treated wards is also particularly encouraging, considering that transfers are directly linked to the patient safety and continuity of care.



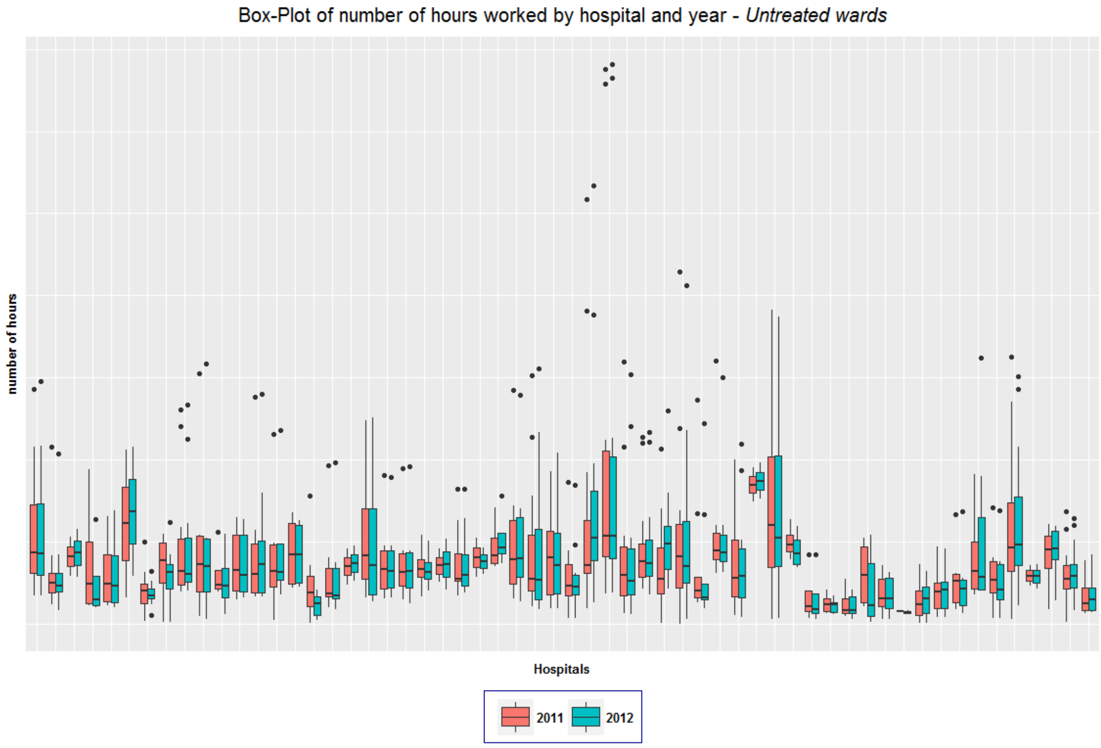


FIGURE 3.1: Box-plot of the number of hours worked by hospital and year for the *Treated* (top) and *Untreated* (bottom) wards.

In order to further quantify the impact of the policy and to confirm the significance of the results on the health outcomes in absolute terms, Figure 3.2 plots the marginal effects of each health outcome in equation 3.4 for treated and untreated wards and over the observation period (Ai and Norton 2003, Karaca-Mandic et al. 2012). As well as verifying the parallel trend in the pre-policy period, the plots show a clear improvement for readmissions and transfers. In particular, there is an absolute difference of 0.91% and 1.52% in the average number of readmissions between the treated and untreated wards in the year 2012 and 2013, respectively, and of 0.31% in the year 2011, whereas there is a difference of 0.19% and 0.18% in the average number of transfers between the treated and untreated wards in the year 2012 and 2013, respectively, and of 0.72% in the year 2011. This leads to DID reductions of 0.60% (readmissions) and 0.53% (transfers) in 2012 compared to 2011 and a further reduction of 0.61% (readmissions) and 0.01% (transfers) in 2013. The predicted percentages of reduction correspond to a P4P-related saving of 4,324 readmissions and 4,295 transfers in the treated wards in 2012 and a further reduction of 4,871 readmissions and 157 transfers in 2013. The picture for the other three health outcomes is more complex than for transfers and readmissions. The average number of returns to the surgery room seems to increase in the treated wards more than in the untreated after the introduction of the policy, as δ_{2012} and δ_{2013} are positive and significant. This is shown in Figure 3.2, which, on the other hand, shows also how the P4P incentives improve the performance for both the treated and untreated wards. This is an interesting result, suggesting that the managerial impact in the hospital organization caused by the adoption of the P4P program has changed the overall hospital performance with regards to the surgical activity. A possible explanation to this could be given by a spillover effect between the treated and the untreated wards, as all wards may be benefiting from potentially improved technology in the surgery room.

For the other two health outcomes, voluntary discharges and mortality, the coefficients of δ_{2012} and δ_{2013} are not significantly different from zero. Figure 3.2 shows how the number of voluntary discharges decreases already before the P4P introduction. With regards to mortality, it is reasonable to believe that, when hospitals are checked for effectiveness on more than one output, they will focus on those outcomes that are easily measurable. This is observed by Propper et al. (2008) in the context of a competition analysis. From this point of view, readmissions, transfers and return to the surgery room represent well-measured outcomes. Hence it is possible that hospitals have focussed their efforts on those easily measured and better observable activities in order to increase their performance and then gain financial rewards.

3.5.3 Do surgical and medical wards react differently to the policy?

We fit the model in equation 3.5 to the data in order to answer this question. The results, omitted in full for brevity, show evidence of a differential impact of the P4P introduction for the two health outcomes that were significant in the global analysis above. In particular, there is

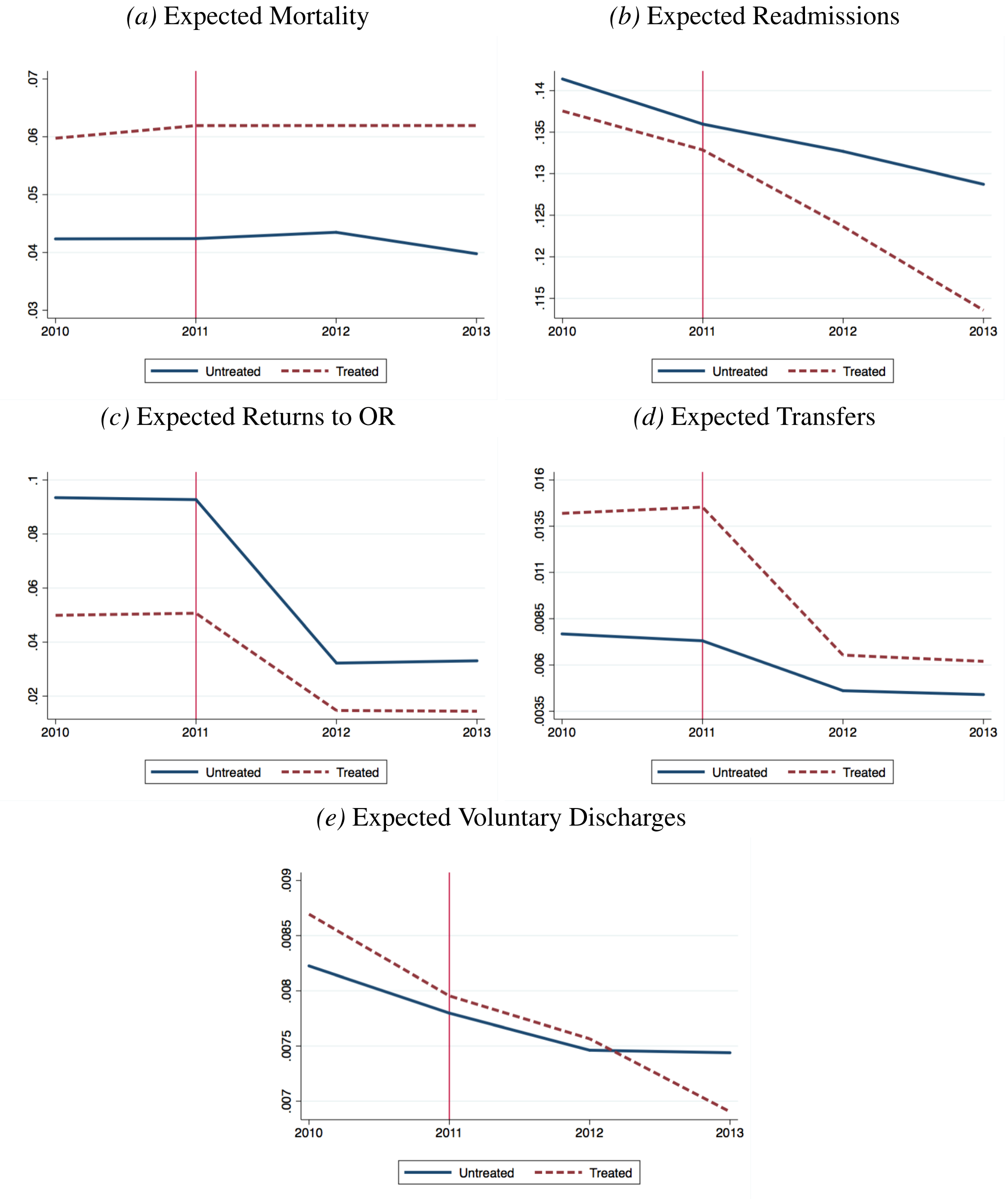


FIGURE 3.2: Marginal effects of all health outcomes per year and treatment for the model in equation 3.4.

evidence that the P4P program impacted more on the medical wards than on the surgical ones in terms of number of readmissions ($\psi_{2012} = 0.008$, p-value $= 0.0102$; $\psi_{2013} = 0.0307$, p-value $=< .0001$) and number of transfers ($\psi_{2012} = 0.0117$, p-value $= 0.0002$, $\psi_{2013} = 0.012$, p-value $= 0.0001$). This is shown visually also by the marginal effects in Figure 3.3. This finding can be explained by the fact that the surgical healthcare pathways are more rigorous and more linked to fixed guidelines than those on medical hospitalizations, which instead tend to be more flexible and more dependent on managerial actions and hospital organization.

(a) Expected Readmissions *(b)* Expected Transfers

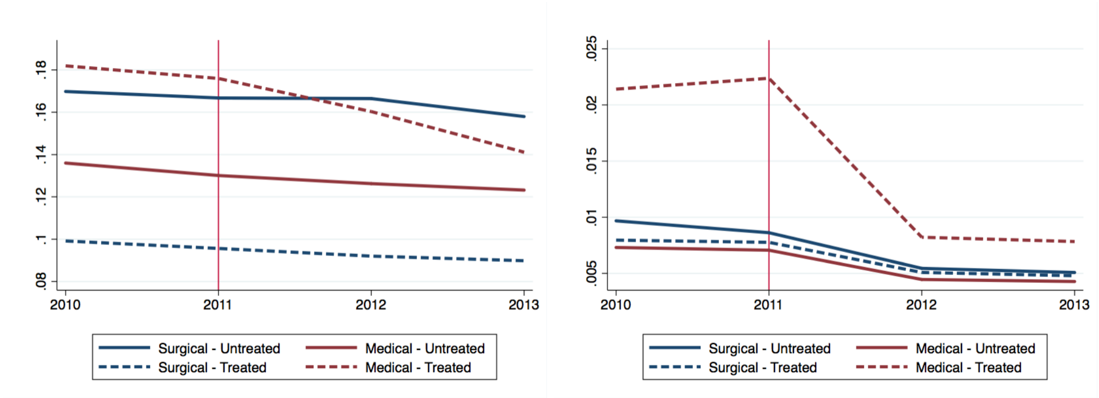


FIGURE 3.3: Marginal effects of readmissions and transfers per type of ward, year and treatment for the model in equation 3.5.

3.5.4 Do private and public hospitals react differently to the policy?

Previous studies have found no dependency between hospital ownership and efficiency (Barbetta et al. 2007) or hospital ownership and competition (Berta et al. 2016), suggesting that the long term adoption of a quasi-market system in Lombardy has reduced the expected differences between the hospital types.

In this section, we test whether the hospitals reacted differently to the introduction of the P4P policy, depending on their ownership. In order to answer this question, we use a model like equation 3.5, but with SURGICAL replaced by a variable representing the ownership type (OWN), where public is taken as the reference category. Once again, the interactions $\psi_{jk}^{(\nu)}$ are of interest in this model. In line with the existing literature, the results show only limited evidence in support to a hypothesis of a different reaction: apart from readmissions in 2012 ($\psi_{2012,\text{not-for-profit}} = -0.01964$, p-value $= 0.0004$; $\psi_{2012,\text{private}} = -0.0096$, p-value $= 0.0062$), the interaction for readmissions in 2013 and all interactions for transfers, for both the private for profit and not-for-profit categories, are not statistically significant. This is an interesting result meaning that the monetary incentive is an interesting motivation to improve the quality of care for all types of ownership and not only for the profit-maximizer providers (profit hospitals).

3.6 Conclusions and future work

The P4P approach has been adopted in many countries in order to encourage improvements in the quality of healthcare by supplying financial incentives to healthcare providers. In this study, we evaluate the impact of a specific P4P program adopted in the Lombardy region (Italy) in 2012. Differently to previous studies, we perform the analysis considering the whole healthcare system, evaluating multiple health outcomes over a number of clinical areas. We analyse data over four years, two before (2010/2011) and two after (2012/2013) the implementation of the program. The policy was applied to all hospitals in the Lombardy region, but the incentive was calculated only on the basis of the performance of 9 wards. The fact that the selection of these wards was made exogenously, combined with the fact that we observe a parallel trend pre-introduction of the policy and that we have found no evidence of spillover effects between the treated and untreated wards in terms of allocation of resources, have led us to use a multivariate DID approach for the evaluation of the impact of the policy.

Our study shows that two out of the five health outcomes considered i.e. readmissions and transfers, support the hypothesis that the P4P introduction had a positive effect on quality. The picture for the other three health outcomes is more complex than for transfers and readmissions. Considering the returns to the surgery room, our results show that the P4P incentives improve the performance for both the treated and untreated wards. We speculate that this may be the result of improved technology in the surgery room which all the wards have benefited from. The last two health outcomes, voluntary discharges and mortality, did not show changes that can be attributed to the P4P adoption. This can be explained by considering the fact that when hospitals are checked for effectiveness on more than one output, they will focus on those outcomes which are more easily driven by a managerial intervention in order to improve their performance and to obtain the financial incentives.

Moreover, our study shows that the medical wards have reacted to the P4P program more strongly than the surgical wards, whereas only limited evidence is found to suggest that the policy reaction was different across different types of hospital ownership. Overall, the results show that the healthcare system in Lombardy was positively impacted by the P4P implementation, as anticipated by Castaldi et al. (2011): there is evidence of a reduction in some adverse health outcomes and of a general change in the hospital organization in order to improve the healthcare services provided to the citizens. Lastly, the evaluation study found no evidence of a distortion of the hospital behaviour aimed at inflating the performance evaluation, such as cream skimming behaviour.

The analysis assume linearity between the dependent variable and the covariates. This assumption is supported by the economic literature in casual inference. In particular, Angrist and Pischke (2008) discuss the theoretical and practical advantages of the linear models over generalized linear models in terms of linking inference with causality, and causal interpretation of regression coefficients. Furthermore, Hellevik (2009) demonstrates that if the purpose of the analysis is not prediction but causal decomposition, the problem of fitted values falling out the range 0-1 is no longer relevant. Despite this, when the dependent variable is a rate, statistical literature suggests to use beta regression models or logistic regressions for proportions or linear models with log-transformation of the dependent variable. Given this, our choice of a linear link can be seen as a limitation and future progress of this work could be the application of one of the suggested approaches (beta regression, etc.).

This study has some implications. Firstly, Lombardy should extend the adoption of the P4P program across the whole regional healthcare system in order to improve the overall hospital activity. Secondly, given the positive impact of the P4P program in Lombardy, the adoption of a similar strategy is suggested to the other regional healthcare systems in Italy. This would stimulate improvements in quality for the regions that already perform relatively well, but, in particular, this would be an important incentive for these regions with a lower qualified healthcare system.

Future work on the evaluation of P4P programs could explore additional aspects, for which data were currently not available. Firstly, it would be interesting to test the impact of the P4P program in terms of the number of intra-hospital infections and complications, or other outcomes directly related to the performance of the hospitals' physicians and the improvement of technology. Secondly, it would be useful to conduct a comparative analysis between the Lombardy region and neighbouring regions which are not subjected to P4P programmes. This would help also in controlling for spillover effects between the treated and the untreated wards within the same hospital, such as those resulting from the sharing of common technology and resources. Thirdly, our analysis has focussed solely on the impact of the P4P programs on the hospital effectiveness. It would be interesting to extend the current analysis to understand whether the monetary incentive had an impact also on the hospital efficiency. Finally, we believe that further research is needed to assess the impact of P4P programs over a long time frame, as encouraged by Werner et al. (2011).

Chapter 4

The association between asymmetric information, hospital competition and quality of healthcare

In this chapter, we study the effect of competition on adverse hospital health outcomes. In particular, we consider a context, the healthcare system in Lombardy, in which information about hospital quality is not publicly available. Using patient-predicted choice probabilities, we construct a set of competition indices and measure their effect on hospital quality, evaluated by a composite index of mortality and readmission rates.

4.1 How competition between hospitals may impact quality in healthcare

As introduced in Chapter 1, competition in healthcare is a key issue among the recent reforms implemented in several Western Countries. These interventions originate from the belief that to stimulate hospitals' competition leads to increase NHS quality (Beckert et al. 2012) reducing unnecessary costs (Kessler and McClellan 2000). Despite this, the wide literature on hospital competition found mixed results about the effect of competition on quality. Four key factors exist that may shape how competition between hospitals impacts quality:

1. Institutional settings of the hospital-market supply side (Gaynor et al. 2012, Kessler and McClellan 2000, Moscone et al. 2012, Propper et al. 2004, Tay 2003)

2. The degree of patient freedom of choice (Beckert et al. 2012, Cooper et al. 2011, Moscone et al. 2012, Tay 2003, Varkevisser et al. 2012)

3. The hospital competitive strategy (Cooper et al. 2011, Kessler and McClellan 2000, Tay 2003)

4. The degree of information regarding hospital quality (Dranove et al. 2003, Dranove and Sfekas 2008).

In order to study the effect of competition on hospital quality, several issues need to be considered. First of all, in many settings, patients have imperfect information about hospital quality. Second, measuring hospital quality is difficult, since it involves multiple dimensions (e.g., quality of life, better treatments, customer satisfaction, etc.). This leaves a certain degree of uncertainty regarding the quality level provided by a hospital. Third, patients may collect different signals of quality, such as hospital rankings based on some specific indicators (e.g., mortality rates, readmission rates, etc.), suggestions of the general practitioners (GPs), word-of-mouth within their social networks. These signals may influence their hospital choice, but it is still unclear whether they are a determinant of the real choices made by patients. However, patient hospital choice is the result of a trade–off. This involves the patient's beliefs regarding hospital quality and the distance between the patient's residence and the hospital. Since the distance is a cost for the patient, the higher is the distance the lower is the chance that a hospital is chosen, unless the quality difference (between a close and a far hospital) justifies travelling more in order to be hospitalized.

Most of the previous contributions on hospital competition are based on the seminal paper by Kessler and McClellan (2000). Their approach is based on modelling patients' hospital choices. Predicted patients' choices are used to build measures of competition, and then these measures are adopted to model the relationship between competition and hospital quality. However, although recent years have witnessed a growing interest on the competition in the healthcare sector, the debate about the effect of competition on quality is still alive in the scientific community as empirical evidence from different countries reports mixed results. While some studies suggest that competition amongst hospitals has positive effects on health outcomes (Gaynor 2016), others support the argument that more competition may even harm the health of patients (Propper et al. 2004). Our analysis considers a setting in which hospital competition is quality based and depends on the number of hospitals that a patient can reach in a reasonable time as well as fixed prices and asymmetric information regarding the quality of providers. In this context, it is important to understand the possible effects of hospital competition under different degrees of information regarding quality that is available to patients. We are working under the assumption that patients are rational agents who will maximize their utility if properly informed regarding quality and that they will choose the hospital that provides the best combination between quality and geographical distance (or travel time) from their residence. It is reasonable also to assume that, for more complicated treatments, patients will be willing to travel more for a high quality hospital and that, even for non-complicated treatments, patients also will select a high quality hospital that is relatively close to their residence and not simply the closest hospital.

4.2 Literature review on competition in healthcare

The literature on hospital competition has focused mainly on the US and UK markets. Although patients are free to choose in both markets, hospitals in the USA may set both prices and quality. In contrast, in the UK, hospitals can only move quality since prices are regulated. Most studies have investigated the effect of hospital competition on health outcomes as measured by the Herfindahl-Hirschman index (HHI)(Cooper et al. 2011, Gaynor 2006, Gaynor and Haas-Wilson 1998, Gaynor et al. 2012, Kessler and McClellan 2000, Propper et al. 2008; 2004, Tay 2003). Recent work, following the approach by Kessler and McClellan (2000), has used predicted flows based on (exogenous) patient characteristics and patient-to-hospital distance when computing the HHI, rather than actual patient flows. This allows to avoid endogeneity problems when studying the effect of the HHI on healthcare quality as well as distortions in defining the geographical area representing the potential hospital market (Kessler and McClellan 2000). The geographical area, if defined by using observed choices, may be influenced by hospital quality, which leads to larger areas for high quality hospitals, and remains unobserved by the researchers. Several studies have investigated the effects of hospital competition on hospital quality in the US market. Kessler and McClellan (2000) used individual data on non-rural elderly Medicare patients hospitalized for heart attack treatment in 1985, 1988, 1991 and 1994. They provided ordinary least squares estimates of the effect of hospital competition and showed that competition leads to better health outcomes, lowering 30-day mortality hospital rates, reducing treatment costs. Tay (2003) used data from 1994 for patients with acute myocardial infarction (AMI). She estimated a mixed logit model and showed the importance of quality in patient choice and provided evidence that they are willing to travel more if the quality of treatment is higher.

Various studies have analysed the influence of competition on healthcare quality in the UK. Propper et al. (2008; 2004) studied hospital mortality rates for AMI and found a negative effect of competition. They used aggregated hospital level measures and tried to avoid endogeneity problems in the HHI by estimating potential demand rather than observed choice. Cooper et al. (2011) implemented a difference-in-differences econometric model to study the effect of recent UK pro-competition reforms, founding a positive effect on mortality rate. In a more recent study, Gaynor et al. (2012) adopted patient level data for a coronary artery bypass graft procedure and investigated the effect of patients' freedom of choice on mortality rates. They found that giving patients the possibility of selecting their hospitals when they know the quality of the hospitals significantly reduces mortality rates. Gaynor et al. (2012) tackled the issue of freedom of choice and information on hospital quality and the results are very close in spirit to our contribution.

Little empirical work exists on the effect of competition on the healthcare sector in Italy. Moscone et al. (2012) studied the effect of patient hospital choice of an imperfect measure of hospital quality (the effect of word-of-mouth social interaction given by the percentage of patients living in the same area who have previously made the same treatment choice). They studied the choices of patients suffering from heart disease who were receiving treatment in one Italian region (Lombardy). Using administrative data that included the whole population, they showed that the informal neighbourhood effect does not influence the health outcomes and even may lead patients to make suboptimal selections.

Lastly, only few references have explored the possible effects of asymmetric information in healthcare despite the relevant insights that are achieved by some very famous early contributions (Akerlof et al. 1970, Spence 1973) and the large number of subsequent references (an excellent review is in Mas-Colell et al. (1995)). Dranove and Sfekas (2008) analysed the effect of disclosing hospital report cards in the USA and showed that it may induce selection of patients, which means that hospitals may not admit patients with bad health statuses because they do not want to worsen their rankings. Dranove et al. (2003) showed that spreading information on hospital quality does not necessarily improve the performance of top ranking hospitals, probably because the rankings confirm patients' informal perceptions on the different quality levels. Although Varkevisser et al. (2012) provided evidence that patients tend to choose better quality hospitals in the Netherlands, their study did not show whether this choice produces a market premium for top quality hospitals.

Another key factor in the relationship between hospital quality and competition is the choice of variables representing health outcomes. Several works in the literature use mortality rate as the quality indicator (see, among others, Kessler and McClellan (2000), Tay (2003), Propper et al. (2008; 2004), Beckert et al. (2012) and Cooper et al. (2011)). Although most of these references focused on treatments for AMI, some researchers (see, among others, Goldstein and Spiegelhalter (1996), Iezzoni et al. (1996) and Lilford et al. (2004)) have criticized the use of mortality in treatments different from AMI as many diseases (e.g. chronic illness) have very low mortality risks associated to them. In the USA, there is growing evidence (Neuman et al. 2014) that mortality rates alone cannot capture hospital differences in treatment provided to patients.

This chapter sheds light on why the empirical literature often rejects the theoretical result that more competition leads to better health outcomes in a fixed-price setting. The key factor in our analysis is the effect of the asymmetric information that does not provide to the citizens the proper tools to make an informed choice of the hospital where to be admitted. This fact has an impact on the hospital quality, reducing the effect of the competition.

The remainder of the chapter is organized as follows. Section 4.3 describe the data, whereas Section 4.4 presents the econometric strategy. Section 4.5 discusses estimation results, and Section 4.6 concludes with some suggestions for future research.

4.3 The data and descriptive statistics

As introduced in Chapter 1, Lombardy is an interesting case-study for hospital competition due to the presence of public and private providers that compete with each other in attracting patients.

Moreover, Lombardy is also a very interesting environment to study the effects of asymmetric information on hospital competition. In fact, even if the regional government evaluates the hospital quality every year, the results of this evaluation are available only for the hospitals, whereas they are hidden for both patients and their GPs. In this way the patients' choice of the hospital where to be admitted cannot be based on the hospital quality.
We gathered administrative data on all patients who were admitted to the cardiac surgery, cardiology and general medicine wards in any public or private hospital in the Lombardy region in 2012. Data on each patient, extracted from the hospital discharge charts, include demographic characteristics such as age, gender and place of residence (the municipality), clinical information such as principal diagnosis and codiagnosis, main and secondary procedures, comorbidity, length of stay, type of admission (planned or via the emergency room), the ward of admission and type of discharge (e.g. death), financial information such as the DRG and hospital discharge chart reimbursement and the information on the GP referral. Such data were matched with information on hospitals (ownership and teaching status, technology, etc.), and on the travel distance in minutes from the patient's residence (the municipality) to the hospital. Information on travel distance, expressed in units of time, was computed using *Google Maps*. The algorithm computes the fastest route from an individual's residence to the hospital by car and the distance is set to 0 if an individual's street address and the hospital location are identical. Furthermore, in order to account for the influence of social relations in the choice of the hospital, in which to be admitted, we added a network variable in the model. This is measured by the percentage of citizens living in the same municipality as the patient, who have chosen the same hospital to be admitted in the 12 months preceding the analysis. The assumption related to this variable is that in the Lombardy region, like in the rest of Italy, most of the population is concentrated in small-to-medium sized municipalities that are characterized by a strong historical and cultural identity as well as autonomy guaranteed by the Italian legislative structure. Family members usually live within the same municipality and meeting with friends and relatives is encouraged through local associations, cultural events, activities of the local parishes and so forth. Within the same municipality, historical, political, social and religious forces may encourage interaction between people, which is the main reason for using this as a reference area for building the network variable. We include such information, which we call a network effect, with aim to detect a mechanism shaping the preferences of individuals and ultimately influencing their decisions about the provider where to be hospitalized.
In order to evaluate the hospital quality we use an adverse outcome measured as the presence or not oh either mortality or readmission. This outcome is equal to 1 whether the patient dies in hospital or if the patient dies or is readmitted within 30-days after the discharge.
We removed from the data set any patient whose source of admission was other than elective. After this cleaning procedure, we were left with a total of 194,020 patients of whom 9,121 were admitted to cardiac surgery, 71,499 to cardiology and 113,400 to general medicine. These patients were admitted to the cardiac surgery ward of 20 hospitals, the cardiology ward of 76

hospitals and to the general medicine ward of 124 hospitals in the Lombardy region.

Table 4.1 defines the variables which were selected for the empirical analysis. Table 4.2 pro-

Variables	
Individual-Specific Variables	
Choice_{ij}	1 if patient i is admitted to hospital j.
Distance_{ij}	Time distance from residence of patient i and hospital j (in minutes).
Network_{ij}	% of residents living in the same municipality as patient i admitted in ward w of hospital j in the 12 months previous to the analysis.
Age_i	Patient i age in years.
Male_i	1 if patient i is male,
GP_i	100 x n. patients in the zip code sharing the GP with i / n. patients in the zip code.
Ward- and Hospital-Specific Variables	
Adverseout_{wj}	Hospital j composite index of adverse health outcomes (i.e., 30-day mortality or readmission) in ward w.
Male_{wj}	% of males in ward w of hospital j.
Age65_{wj}	% of patients over 65 in ward w of hospital j.
ICU_{wj}	% of transits in intensive care unit for patients in ward w of hospital j.
DRGWEI_{wj}	Average DRG weight in ward w of hospital j.
Private_j	1 if hospital j's ownership is for-profit.
NFP_j	1 if hospital j's ownership is not-for-profit.
Mono_j	1 if hospital j is mono specialized.
Teaching_j	1 if hospital j is a teaching hospital.
Technology_j	1 if the hospital j has a high technology assessment (i.e., an intensive care unit department) and 0 otherwise.
Heart_{wj}	1 if ward w in hospital$j$j is cardiac surgery.
Cardio_{wj}	1 if ward w in hospital j is cardiology.
Medicine_{wj}	1 if ward w in hospital j is general medicine.
Beds_{wj}	Number of beds in ward w of hospital j.
Ranking_{wj}	Percentile rank of ward w in hospital j in the league table of the Lombardy quality evaluation program.

TABLE 4.1: List of variable definitions

vides descriptive statistics for the set of patient-specific characteristics, whereas Table 4.3 for the hospital-specific variables. The statistics show that cardiac pathology affects more males than females and that patients in cardiology and general medicine are older than patients in cardiac surgery. Looking at the distance variables, we note that patients who are admitted to cardiac surgery are more willing to travel longer distances and that their network size is smaller compared with cardiology and general medicine patients. Focusing on the ward and hospital level variables, we note that general medicine has high mortality rates. As expected, patients who are admitted to cardiac surgery are (relatively) young and are undergoing highly specialized expensive treatment and interventions. Conversely, patients who are admitted to cardiology and general medicine are older, often affected by a number of comorbidities, and admitted for a variety of treatments and interventions. Table 4.3 shows also that the three wards have different

Variables	**Cardiac Surgery**		**Cardiology**		**General Medicine**	
	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
Distance$_{ij}$	29.18	20.72	19.1	15.72	15.87	13.02
Network$_{ij}$	38.01	29.77	66.1	33.59	77.86	29.15
Age$_i$	64.78	17.11	69.34	13.82	72.91	15.68
Male$_i$ (%)	65.52	47.5	64.78	47.77	49.56	50
GP$_i$	14.11	17.85	14.23	18.02	18.27	20.54

TABLE 4.2: Descriptive statistics of individual-specific variables.

Variables	**Cardiac Surgery**		**Cardiology**		**General Medicine**	
	Average	St. Dev.	Average	St. Dev.	Average	St. Dev.
Adverseout$_{wj}$ (%)	5.31	1.62	7.25	2.54	17.12	5.6
Male$_{wj}$ (%)	66.71	4.74	63.08	6.76	48.32	6.45
Age65$_{wj}$ (%)	64.51	7.68	70.76	8.13	77.05	8.39
ICU$_{wj}$ (%)	76.38	42.47	20.77	40.57	1.55	12.36
DRGWEI$_{wj}$	5.1	2.86	1.67	1.07	1.07	0.71
NFP$_j$ (%)	10	30.77	10.52	30.89	12.1	32.61
Private$_j$ (%)	40	50.26	26.31	44.37	13.59	34.26
Technology$_j$ (%)	95	22	92.1	27.14	74.57	43.55
Teaching$_j$ (%)	40	50.26	15.71	36.72	17.81	38.26
Mono$_j$ (%)	5	22.36	1.31	11.47	0.91	9.48
Beds$_{wj}$	21.12	14.83	22.5	16.44	43.87	31.39
Ranking$_{wj}$	49.45	29.46	53.13	28.21	49.18	27.54
Num of hospitals	20		76		124	

TABLE 4.3: Descriptive statistics of ward- and hospital-specific variables.

compositions in terms of ownership and teaching status. Private and teaching hospitals often have cardiac surgery wards, whereas public, non-teaching hospitals often have cardiology and general medicine wards. However, as expected, cardiac surgery wards have high technology equipment. Finally, Table 4.4 offers a set of descriptive statistics on patient-to-hospital distance,

Ward	**Min**	**25$^{\text{th}}$ Percentile**	**Average**	**Median**	**75$^{\text{th}}$ Percentile**	**Max**
Cardiac Surgery	0	16	29.18	24	38	152
Cardiology	0	10	19.1	16	24	188
General Medicine	0	7	15.87	13	20	197

TABLE 4.4: Descriptive statistics of patient-to-hospital time distance by ward.

measured in minutes of time, for the three wards. It is interesting to observe that the average distance to the cardiology ward is much shorter than for the cardiac surgery ward: about 19 min for the former versus 29 min for the latter. This may be explained by the fact that patients who are admitted for cardiac surgery may face more complex interventions and thus are more willing to travel further to receive high quality treatment. Table 4.4 shows that, overall, in a context of asymmetric information regarding hospital quality in Lombardy, patients tend to choose nearby hospitals, showing little propensity to travel far for hospital treatments.

4.4 The econometric strategy

To study the effect of competition between hospitals on health outcomes we adopt a two-stage approach. In the first stage we study patient hospital choices as a function of a set of patient characteristics, the hospital travel distance and the network effect. In the second stage, we compute a set of HHIs (one for each ward and for each hospital) by using the predicted choice probabilities estimated in the first stage and we then analyse their effect on hospital quality. In order to evaluate quality, we focus on a composite index of mortality and readmission rates.

4.4.1 Modelling patients' hospital choices

In the first stage, we investigate patient choices using the conditional discrete choice model prosed by McFadden (1973). In our context we aim at describing the choice of a patient (individual) of the hospital where to be admitted (alternative). Patients choose the hospital where to be admitted on the basis of their characteristics, the hospital travel distance, and the network effect described above. Suppose that, for each ward considered in this study (cardiosurgery, cardiology and medicine), the observable choice of individual i of being admitted to hospital j is modelled as a conditional logit:

$$\pi_{ik} = \frac{\exp(\rho \text{Distance}_{ik} + \boldsymbol{\delta}_k \text{Network}_{ik} + \boldsymbol{\gamma}_k \text{GP}_i + \boldsymbol{\beta}_k \boldsymbol{x}_i)}{\sum_{j=1}^{J} \exp(\rho \text{Distance}_{ij} + \boldsymbol{\delta}_j \text{Network}_{ij} + \boldsymbol{\gamma}_j \text{GP}_i + \boldsymbol{\beta}_j \boldsymbol{x}_i)} \tag{4.1}$$

where π_{ik} is the probability that the patient i chooses the hospital k, depending on the patient's characteristics and the characteristics of the chosen and of the alternative hospitals. In order to have the identifiability of the parameters, without lost of generality, we set the parameters corresponding to the last hospital to 0., i.e. $\delta_J = 0$, $\gamma_J = 0$ and $\beta_J = 0$. In Equation (4.1), the covariates Distance$_{ij}$, GP$_i$ and Network$_{ij}$ are defined as in Table 4.1, and the the vector **x** of patient characteristics includes the variables Age$_i$ and Male$_i$

A positive and significant coefficient attached to the network effect (δ_j) means that a subset of the population, sharing informal information on the quality of the j^{th} hospital, increases the conditional probability of choosing it for each member of this subset. A negative and significant coefficient implies that, *ceteris paribus*, the patients will make a choice that is different from that of them neighbours. In addiction, by including the variable GP$_i$, which measures the fraction of patients in the postal code area of patient i sharing a GP with patient i, we aim to capture the potential impact that the GP's advice may have on the choice of the hospital at which to be treated.

We estimate Equation (4.1) for each ward separately by maximum likelihood using the *asclogit* procedure in the statistical software Stata 13.

4.4.2 Constructing a competition index

Once the predicted patients' choices are modelled, we move to analying the relationship between competition and quality. In order to achieve this aim, we need to convert the predicted choices obtained from Equation (4.1) into a competition index. Using the predicted probabilities at this stage avoids potential endogeneity in patient flows and in the definition of hospital catchment areas as underlined by Kessler and McClellan (2000). In fact, real patient flows can be influenced by variables such as the teaching status or the size of a hospital, which are connected with hospital quality. The endogeneity problem may also arise because hospitals with higher quality could obtain higher market shares and thus the index of competition may be affected by the dependent variable. Such endogeneity may bias results when regressing the HHI on health outcomes. As indicated by Kessler and McClellan (2000), building theoretical patients based on exogenous factors may overcome these problems.

Previous literature defines the potential markets of hospital-specific HHIs as the area surrounding each hospital by using an array of arbitrary lengths (e.g. 30 km) (Bloom et al. 2015, Siciliani and Martin 2007). To avoid the possible biases in computing the HHI by using these ad hoc methods, we follow Kessler and McClellan (2000) and use the predicted flows that are estimated by Equation (4.1) to compute HHI indices by exogenously assigning each patient to a given geographic area identified by the local healthcare zone, called the LHA. In Lombardy, there are 15 LHAs and each patient is exogenously assigned to one. Let N_q be the number of patients i living in the LHA q. For each patient i in this area, let $\widehat{\pi}_{iwj}$ be the predicted probability that for a ward w the patient i chooses hospital j. The share of patients living in the LHA area q who are predicted to choose hospital k over the predicted flow of patients living in LHA q to all the hospitals is

$$\alpha_{qwk} = \sum_{i=1}^{N_q} \widehat{\pi}_{iwk} \bigg/ \sum_{j=1}^{J} \sum_{i=1}^{N_q} \widehat{\pi}_{iwj}, \tag{4.2}$$

where J is the number of hospitals operating within a given ward (e.g. cardiology) in Lombardy. Finally, we can compute the competition index HHI for a given LHA q by:

$$\mathrm{HHI}_{wq} = \sum_{j=1}^{J} \alpha_{qwj}^2. \tag{4.3}$$

The next step consists of defining the weight for hospital j of the LHA area q relative to all LHA areas in Lombardy:

$$\widehat{\beta}_{qwj} = \sum_{i=1}^{N_q} \widehat{\pi}_{iwj} \bigg/ \sum_{i=1}^{N_{wj}} \widehat{\pi}_{iwj}, \tag{4.4}$$

where N_{wj} is the total number of patients admitted in ward w of hospital j in Lombardy. The last step is computing the HHI for ward w in hospital j, given by

$$\mathrm{HHI}_{wj} = 10,000 \sum_{q=1}^{Q} \widehat{\beta}_{qwj} \mathrm{HHI}_{wq}, \tag{4.5}$$

where the value 10,000 is arbitrarily chosen to rescale the index.

Hence, each ward–hospital has an HHI competition index that is a weighted average (using each hospital patient share in LHA q) of the exogenously defined LHA q competition index. HHI_{wj} varies between $10,000 * 1/J$ (competition) and $10,000$ (monopoly), with larger values indicating a decrease in the degree of competition.

4.4.3 Modelling the effect of competition on the hospital quality

The second stage of our econometric approach is designed to verify the influence of competition on hospital adverse health outcomes. Let $\mathrm{Adverseout}_{wj}$ be the adverse health outcome rate for ward w of hospital j, as defined in Table 4.1. In our second stage we consider the following multilevel model, based on the assumptions described in Section 1.4

$$\mathrm{Adverseout}_{wj} = \alpha + \boldsymbol{\beta} \boldsymbol{x}_{wj} + \boldsymbol{\gamma} \boldsymbol{z}_j + \theta \mathrm{HHI}_{wj} + u_j + \varepsilon_{wj}, \tag{4.6}$$

where $\boldsymbol{x}_{wj}$ is a set of ward and hospital-specific characteristics $\boldsymbol{z}_j$ is a set of hospital-specific attributes and u_j is a hospital-specific random effect. As regressors, in addition to the key variable HHI, we control also for other covariates. We include the following variables, defined in Table 4.1: $\mathrm{Age65}_{wj}$ in order to account for patient health status, which is highly correlated with chronic conditions, DRGWEI_{wj} and ICU_{wj}, with the aim of identifying the treatment complexity. Further, we have included the dummies Mono_j, $\mathrm{Teaching}_j$, NFP_j, $\mathrm{Private}_j$, and $\mathrm{Technology}_j$. Since we do not have information on specific technological equipment, we adopt the presence in hospital i of an intensive care unit as a proxy for hospital technology classification. Although this feature fits well in the case of general medicine and cardiology, for heart surgery we identify a set of additional treatments that require high technology equipment such as

1. repair of atrial and ventricular septa with prostheses,
2. total repair of certain congenital cardiac anomalies and
3. heart replacement procedures.

Finally, we include ward dummies, Heart_{wj} and Cardio_{wj}, and an interaction term between the HHI and the hospital ownership.

The main focus is on the magnitude and significance of the parameter related to the HHI covariate. A positive and significant coefficient indicates that more competition increases the level of quality that is offered by hospitals, whereas a coefficient that is statistically insignificant would point to no effect of competition.

To conclude, we test whether there is a relationship between the hospital quality ranking evaluated by the regional healthcare directorate, and the patient choices predicted by Equation (4.1). Specifically, we estimate the following multilevel model by using data at the patient level:

$$\widehat{\pi}_{iwj} = \alpha + \beta \text{Ranking}_{wj} + \gamma \text{Beds}_{wj} + \eta \text{Heart}_{wj} + \lambda \text{Cardio}_{wj} + u_j + \varepsilon_{wj}, \tag{4.7}$$

where $\widehat{\pi}_{iwj}$ is patient i's maximum probability among all predicted probabilities (obtained from Equation (4.1)) of selecting each hospital in the region with ward w, whereas Ranking_{wj} is the hospital level ranking calculated by the Lombardy region within its quality evaluation programme. Beds_{wj} is the number of beds for each ward and Heart_{wj} and Cardio_{wj} are ward fixed effects to control for hospital-specific characteristics. An insignificant coefficient for the ranking variable indicates that the levels of hospital quality do not drive patients' choices and this would be a proof of the effect of the asymmetric information.

4.5 Estimation results

Table 4.5 summarizes results for the model in Equation (4.1). It reports a set of statistics on the regression coefficients estimated by maximum likelihood. For brevity we provide synthetic statistics of the hospital-specific estimates. As expected, patient-to-hospital travel distance has a negative influence on choices, implying that patients are more likely to choose closer hospitals relative to similar alternatives at longer distances (Sivey 2012). The coefficient that is attached to GP_i is positive, though with a mild effect and weak evidence.

Looking at the results for the network variable, it is interesting to observe that the estimated coefficients are large and the average t-ratio is statistically significant in all models, thus indicating neighbourhood effects. For cardiac surgery, the average coefficient is higher than for other wards, indicating that patients, *ceteris paribus*, are strongly influenced by the choice of their neighbours. We observe that patients who are admitted to this ward often need complicated and risky interventions and may spend more time and effort on gathering information on the quality of the ward when compared with other patients.

Table 4.6 shows the distribution of the HHIs calculated for the three wards and computed by using theoretical patient flows obtained by estimating Equation (4.1). The average, median, 25^{th} and 75^{th} percentiles of the HHI are consistently higher for heart surgery, indicating a lower degree of competition compared with cardiology and general medicine. It is interesting that we obtain the largest HHIs for hospitals that are quasi-local monopolists –in rural areas or very

Independent Variables	*Cardiac Surgery* Coefficient	*Cardiac Surgery* Standard error	*Cardiology* Coefficient	*Cardiology* Standard error	*General Medicine* Coefficient	*General Medicine* Standard error
Distance$_{ij}$	-0.100***	0.001	-0.160***	0.001	-0.171***	0.001
Network$_{ij}$						
Mean	5.732	0.516	4.024	0.522	1.036	0.25
Standard deviation	3.701		11.556		3.786	
Age$_i$						
Mean	0.001	0.007	0.024	0.007	0.021	0.005
Standard deviation	0.028		0.032		0.025	
Male$_i$						
Mean	-0.094	0.188	-0.064	0.199	0.012	0.152
Standard deviation	0.489		0.385		0.296	
GP$_i$						
Mean	0.003	0.006	-0.001	0.006	-0.0002	0.005
Standard deviation	0.013		0.019		0.013	
Constant						
Mean	1,730	0.516	-5.931	0.721	-3.316	0.336
Standard deviation	2.717		11.888		4.061	
Observations	172,280		4,601,876		12,700,700	
BIC	29,003.20		228,815.30		370,129.80	

Note: *** 1% significance, ** 5% significance, * 10% significance.

TABLE 4.5: Estimated regression coefficients of the model in Equation (4.1) for each of the 3 wards considered in this study. Summary statistics are given for the estimated parameters across all hospitals.

small cities– whereas the lowest HHI values are attached to hospitals in the densely populated areas of Milan and Bergamo.

Ward	**Mean (St.Dev.)**	**Min**	**25th percentile**	**Median**	**75th percentile**	**Max**
Cardiac Surgery	2,323.27 (1,140.86)	1,438.92 *(Ca' Granda Niguarda)*	1,641.56	2,007.57	2,532.19	6,480.12 *(C. Poma Mantua)*
Cardiology	1,319.98 (630.86)	564.94 *(Fatebenefratelli Milan)*	875.74	1,201.26	1,538.25	3,739.27 *(Valcamonica Esine)*
General Medicine	1,041.85 (601.2)	550.5 *(San Pellegrino Terme)*	642.11	949.65	1,115.02	5,238.45 *(Valcamonica Esine)*

TABLE 4.6: Descriptive statistics of the HHI from equation (4.5), computed for each ward.

Table 4.7 shows the estimation results for Equation (4.6), in order to quantify the influence of competition on hospital quality, measured by the dependent variable Adverseout$_{wj}$. Model 1 in Table 4.7 includes only patients' covariates, whereas in Model 2 hospital covariates are included. In Model 3 we control for the interaction between the competition index HHI and the hospital ownership. In all specifications we control for the ward of discharge. In all the estimates in Table 4.7 the coefficient related to the HHI is statistically insignificant. The interaction term between the HHI and ownership status is also insignificant, indicating that there are no significant differences regarding the effect of competition on adverse outcomes for public, private and not-for-profit hospitals.

Independent variables	*Model 1* *Coeff. (St. err.)*	*Model 2* *Coeff. (St. err.)*	*Model 3* *Coeff. (St. err.)*
HHI_{wj}	0.031 (0.051)	0.013 (0.039)	0.018 (0.056)
Age65_{wj}	0.193 (0.047)***	0.212 (0.048)***	0.215 (0.055)***
Male_{wj}	-0.111 (0.090)	-0.046 (0.104)	-0.046 (0.105))
DRGWEI_{wj}	0.006 (0.018)	0.006 (0.014)	0.008 (0.014)
ICU_{wj}	-0.009 (0.026)	-0.037 (0.029)	-0.037 (0.031)
Heart_{wj}	-0.137 (0.072)*	-0.105 (0.064)	-0.109 (0.062)*
Cardio_{wj}	-0.106 (0.014)***	-0.095 (0.016)***	-0.096 (0.020)***
Technology_j		0.022 (0.008)***	0.022 (0.010)**
Mono_j		-0.005 (0.020)	-0.005 (0.022)
Teaching_j		-0.004 (0.008)	-0.002 (0.008)
NFP_j		-0.026 (0.013)**	-0.014 (0.038)
Private_j		-0.013 (0.007)*	-0.017 (0.017)
HHIj*NFP_j			-0.110 (0.274)
HHIj*Private_j			0.034 (0.134)
Constant	0.069 (0.050)	0.016 (0.043)	0.013 (0.056)
Num of Obs.	220	220	220
Loglikelihood	382.51	391.82	392.19

Note: *** 1% significance, ** 5% significance, * 10% significance.

TABLE 4.7: Estimated regression coefficients of the model in Equation (4.6) to detect the effect of competition. Model 1 includes the characteristics of the patients in each ward. In Model 2 hospitals' characteristics are added. In Model 3 we include also an interaction between cooperation and hospital ownership.

As for the remaining regressors, the dummy variable for heart surgery has a negative and (weakly) significant coefficient in Model 1, suggesting that the likelihood of adverse outcomes for patients in this ward is relatively lower than for patients in general medicine. The estimated coefficient attached to Age65_{wj} is positive and statistically significant in all models, indicating that hospitals with a higher share of patients who are older than 65 years tend to have more adverse health outcomes. High technology hospitals have more adverse outcomes than non-high-technology hospitals. There is weak statistically significant evidence in Model 2 that private hospitals have lower adverse outcome rates than public hospitals.

We explain the absence of evidence of a relationship between quality and competition by the presence of asymmetric information about the 'true' quality of hospitals, which was also suggested by Moscone et al. (2012). In fact, the presence of asymmetric information may act as a barrier for competition to work effectively, since it may reduce the possible returns from investing in hospital quality.

Finally, Table 4.8 reports results for the estimation of the effect of hospital quality ranking on patient-predicted choice probabilities obtained from Equation (4.1). It is interesting that the effect of hospital rankings on predicted probabilities is always statistically insignificant. This result reinforces the role that is played by the presence of asymmetric information, which means that patients are correctly informed about the hospital quality.

Independent Variables	*Coefficient*	*Standard Error*
Ranking$_{wj}$	-0.000001	0.00003
Beds$_{wj}$	0.001***	0.00002
Heart$_{wj}$	0.140***	0.0025
Cardio$_{wj}$	0.076***	0.0014
Constant	0.430***	0.021
Num of Obs.	171,616	
ICC	0.67	

Note: Results show estimates of multilevel model in equation 4.6.
*** 1% significance.

TABLE 4.8: Estimated regression coefficients from Equation 4.7, which study the impact of hospital quality ranking on patient-predicted choice probabilities. This analysis allows us to detect the effect of the asymmetric information.

4.6 Concluding remarks and future work

In this chapter we investigated how competition affects the health of patients in regional quasi-market. We found that more competition does not seem to have a significant influence on the quality of hospitals. One explanation for this result is the lack of publicly available information on the quality of hospitals. The presence of such asymmetric information may exacerbate the influence of information that is gathered locally. It may also result in a reduced freedom of choice for patients, a lower degree of competition between hospitals and a lack of market premium for top quality hospitals. Our results point to the network effect as a barrier for competition to work effectively and indicate that patient choice is likely to be not affected by the true quality of hospitals. Our analysis may shed light on why empirical literature often rejects the theoretical result that more competition should lead to better health when prices are fixed.

Our contribution has two important policy implications. First, the results show that it is necessary and urgent to disclose information regarding hospital quality ranking computed within the regional quality evaluation programme, to GPs, patients and the wider public. As shown by Austin et al. (2015), a set of indicators delivered to the public must remain fixed for a sufficiently long period of time to avoid misunderstandings and confusion. As such, the presence of asymmetric information will be reduced and patients will tend to choose high quality hospitals and to enjoy the benefits of having invested in better healthcare. Although publicly available hospital rankings may certainly support patient choice and encourage providers to improve their quality, this may not be enough to encourage low quality hospitals to improve their quality of care.

Hence, our second policy implication is that the regional government should make a special intervention on behalf of these hospitals. For instance, such an intervention may give hospitals with only one or two wards, which is significantly below the regional average (i.e. indicated as belonging to group 3 within the quality evaluation programme), a time period–say, 1-2 years–within which they must make improvements. If a ward is still ranked in the bottom quality group after this period of time, it would be closed or receive a monetary penalty. These interventions in the regional hospital structure are essential to form a competitive hospital market.

Our results are open to further new research developments. In this chapter, following Moscone et al. (2012), we have used hospital network effects as a proxy for patient sensitivity to local information or social interaction. However, we observe that social interaction may be the result of other forces such as contextual or correlated effects (Brock and Durlauf 2001, Manski 1993). Future work will consider strategies for disentangling social interaction from the effect of other factors. A limitation of our work is that the study focuses on only a single cross-section. Future work will explore whether our results are consistent when using panel data. Another interesting extension is the analysis of healthcare quality at the surgical or team level or using the average surgical quality within the ward weighted by the number of surgeries. In fact, patients could choose their provider depending on the national and international reputation of a particular surgeon or medical team and average surgical quality is a more accurate measure of quality.

Finally, we remark that the indicators that are usually used in the literature are not sufficiently sensitive to detect variations in ward level quality. Although in this chapter we have mitigated this issue by using a composite index of adverse health outcomes, future work should include other indicators for hospital quality–e.g. clinical indicators describing the quality of the treatment that is used in various pathological conditions (Damberg et al. 1998, Iezzoni et al. 1996), process measures such as the frequency of using best practices in the treatment of a pathology (Joint Commission on Accreditation of Healthcare Organization (Davies 1994)), sentinel events representing unexpected occurrences (e.g. death or severe physical or psychological injury) (Joint Commission on Accreditation of Healthcare Organization (Davies 1994)) and quality-of-life outcomes indicating the general health condition of the patient (Damberg et al. 1998).

Chapter 5

The effect of cooperation on quality in the healthcare sector

Regardless of the specifics of a healthcare system, hospitals can be considered as altruistic economic agents which cooperate in order to improve the quality provided to the citizens.
In this chapter, we evaluate whether cooperation between hospitals has an impact on the hospitals' quality. We analyse the effect of cooperation on quality, by taking the network of patients' transfers between hospitals as a proxy of their level of cooperation. Firstly, we exploit data at patient and hospital level, and we identify the determinants of patients' flows with the use of a social relations model that accounts, among other things, for a potential correlation among the group specific effects of origin and destination hospitals. Secondly, we move beyond the discussion on the determinants of the network of patients' flows to assess whether this network between hospitals has a positive or negative effect on the overall mortality, hence on the quality provided to both the origin and destination hospitals.

5.1 Cooperation between hospitals in healthcare

This chapter is stimulated by a lack of existing literature on the effects of cooperation in healthcare. While focussing mostly on boosting competition in healthcare systems, policy makers have disregarded collaboration among providers as a force that can be shaped to improve the quality and efficiency of healthcare systems. In this work we analyse how hospitals cooperate in a competitive environment and we measure the determinants and outcomes of such cooperation. In Chapter 4 we presented a growing literature that studies how pro-competition reforms, implemented in a number of Western countries, can trigger interaction and reaction processes between hospitals and can have an impact on hospital decisions and hospital quality. According to this view, hospitals may set their own quality of services, adjusting their decisions by looking at the quality offered by neighbouring hospitals, often taken as those located within the same geographical area.

Aside from the aspect of competition, it is likely that collaboration *per se* plays an important role in boosting the quality of healthcare. Hospitals, motivated by reasons such as convenience or altruism, may decide to engage in mutually beneficial cooperation with each other, leading to improvements in overall health indicators. Collaboration among healthcare providers can take different forms. The most common form of formal collaboration is the merging process. Hospital merging is often defined as the consolidation with local competitors, which can take two form: in local multi-hospital systems, two or more hospitals in the same geographic market have common ownership, but maintain separate physical facilities and financial activities. In local mergers two or more hospitals in the same local market have common ownership and unify financial records, and may or not consolidate physical facilities (Dranove and Lindrooth 2003). The literature on mergers between private hospitals in US seems to lead little benefits in terms of prices and costs (Vogt and Town 2006). A recent work by Gaynor et al. (2012), contributed to this literature studying the effect of the Labour Party reform of healthcare in 1997, which imposed a radical programme of hospital closure, merging many hospitals geographically co-located. Extending the set of outcomes including financial performance, productivity, waiting times and clinical quality, the authors find little evidence that mergers increase the quality of the NHS. Furthermore, consolidation seems to reduce the role of the competition. Different common forms of cooperation concern clinical network information sharing, joint treatment or diagnostic centers, new shared assets and joint construction of new facilities. The small results produced by formal (or imposed) types of cooperation are not surprising. As argued by Westra et al. (2017), cooperation can influence outcomes overcoming formal agreements to constitute factual collaboration. Indeed, one the main feature of the informal types of cooperation is that there exists neither formal agreements between the two parties nor regional guidelines that regulate the flow of patients, but only the decision to cooperate. In the existing literature two forms of informal cooperation have been studied: patient transfers and professionals' affiliations. The cooperation in terms of professionals' affiliations cen be observed within hospital boundaries or in terms of professional sharing their knowledge. This form of cooperation create channels for transferring best-practices, and creating learning opportunities which should be related with outcomes. Evidence in the literature are mixed. While some studies indicate positive effect of inter-organizational learning (Westra et al. 2016), other studies indicate that physicians are unable to duplicate their performance from one organization to the next (Huckman and Pisano 2006, Westert et al. 1993). The second form of cooperation relates the inter-hospital patient transfers (also known as inter-facility or secondary transfers), namely the need for a hospital to transfer a patient when the diagnostic and therapeutic facilities required for that patient are not available at the given hospital, or when the complex case has been resolved and the patient can be transferred in a facility that is less technology intensive.

The aim of this chapter is to identify a positive effect of the hospital cooperation and hospital

quality. This aim is also based on the hypothesis that hospitals are altruistic agents, and they decide to transfer their patients to other hospitals when the benefits of transfer outweigh the risks. Altruism in healthcare is a limited topic of research, but growing evidence in the literature support this concept. In Arrow (1963) physician are differentiated by typical profit-maximizing economic agents, considering the patient benefits as a fundamental reason the be altruistic. Similar arguments can be found in the following literature, where altruism in healthcare has been considered as a measurable behavior, using discrete choice experiments (Ellis and McGuire 1986). Finally, a general overview is collected in Galizzi et al. (2015), where it is recognized that evidence in supporting the hypothesis of altruism in healthcare are rapidly growing. While the decision to transfer a patient is usually driven by matters such as infrastructure and availability of specialized care, the choice of the destination hospital by the manager of the referring (or origin) hospital is driven, among other things, by its proximity, availability and quality of care. However, one complication is that hospitals may not know the distribution of quality across the other hospitals in the healthcare sector, thus their choice will be driven by a measure of perceived quality. This source of asymmetric information, as discussed in Chapter 4, may produce two effects: if the perceived quality is reflecting the "true quality", we should expect that cooperation will improve health outcomes. However, if the perceived quality is negatively associated with the "true quality", cooperation may even harm patients. Hence, in this chapter we will study the impact that an informal mechanism of cooperation via inter-hospital patients' transfers has on the overall hospital quality of both origin and destination hospitals. We will first investigate the determinants of patients' transfers between hospitals, with the aim of finding some exogenous to quality, and we will then quantify their impact on hospital quality.
The rest of the chapter is organized as follows: in Section 5.2 we review the existing literature; in Section 5.3 we discuss the healthcare setting and the data, in Section 5.4 we study the determinants of patients' transfers; in Section5.5 we analyse the impact of cooperation on the quality of the healthcare system, and Section 5.6 concludes with some final remarks.

5.2 Literature background

Healthcare providers are increasing their level of cooperation and policy makers are called to strengthen interactions among healthcare stakeholders to improve both efficiency and quality of care (Mascia et al. 2012). Cooperation between economic agents in the healthcare sector can be represented by a network and studied using network models. Within the several applications of network analysis, we are focusing on the networks defined by patients' flows. These can be expressed as patients' mobility between regions or Local Health Authorities (LHA), or as patient's transfers. The first type of these networks is used to study the policy implications of inter-regional mobility. The majority of these studies adopt gravity models (Silva and Tenreyro 2006), where the flow of patients between two regions (or LHAs) is modelled as a function of the characteristics of the origin and destination region (or LHA), as well as the geographical distance amongst them. Within these studies, Levaggi and Zanola (2004) estimate a regression model for net patients' migration during the period 1994-1997. Empirical results show a south to north pattern of patients' mobility. The authors find that in wealthier regions there is a higher quality of services with a low outflow mobility. Similar results have been found for Spain (Cantarero 2006) and Japan (Shinjo and Aramaki 2012). In the same way, Fabbri and Robone (2010) focus on bilateral patient flows occurred in 2001, adopting spatial analysis. They show that most advanced LHAs are more engaged in containing patients outflow and exporting hospital services. Balia et al. (2018) study patients' flows among Italian regions for the period 2001-2010 using a dynamic spatially correlated random effect gravity model. They find that income, hospital capacity, and an indicator of regional technology level are the main determinants of flows.

The second way to study patients' flows concerns the sharing of patients. A specific literature focuses on collaboration in healthcare by studying networks of physicians (i.e GPs), but only few works look directly at the hospital collaboration, in terms of transfers of patients between hospitals (Barnett et al. 2011, Landon et al. 2012, Pollack et al. 2012). In this chapter we take this network as a proxy for the hospital cooperation, and we are interested in explaining these flows, and in quantifying their impact on quality.

The paper by Mackenzie et al. (1997) is considered one of the first attempts to analyse patients transfers between healthcare providers. Using data gathered from a survey performed on 278 intensive care units in 1994 in UK, the authors provide a descriptive analysis of patients transfers, and show how such an analysis can be used to drive a targeted allocation of patients transfers. Iwashyna et al. (2009), adopting an exploratory data analysis on Medicare patients in Connecticut, find that more critical patients tend to be transferred to high technology hospitals. Iwashyna et al. (2009), using statistical network analysis on nationwide US Medicare data find that patients transfer is not randomly distributed and that the centrality of a hospital in the network is associated with increased capability in delivery of services, suggesting that transfers direct patients toward better resourced hospitals.

A group of papers, on the other hand, has studied patients transfers by analyzing the main drivers of hospital cooperation. Lee et al. (2011), using data from hospital discharges in California, use a regression framework and a set of centrality indices and hospital characteristics, find that cancer specialized hospitals are more likely to receive transfers as well as the hospitals with higher patients' volume. However, the authors do not find a clear association between geographical distance between hospitals and transfers. Lomi and Pallotti (2012) and Caimo et al. (2017) adopt Exponential Random Graph Models (ERGM) to describe the factors related with the presence of cooperation among hospitals. They control for indices measuring the presence of an edge between hospitals, the presence of reciprocity in the edges, as well as the impact of the distance between hospitals as an effect that can affects costs and risks of a transfer. Using administrative data for 91 hospitals located in the Lazio region (Italy), the authors find that hospitals' proximity and sharing an administrative membership facilitate cooperation. Furthermore, they find the presence of local networks, with the tendency to reciprocity among hospitals.

The works reviewed above have contributed to identifying the determinants of cooperation defined by patients' transfers. There is a lack of knowledge about the relationship between the network of patients' transfers and the hospital quality. In this perspective, Mascia et al. (2012) use a panel negative binomial regression on administrative Italian data for 35 hospitals. The authors consider the number of transfers as dependent variable, and include as quality-related covariate the 45-day readmission rate. This paper concludes that patients' transfers are positively correlated with volume of hospitalizations and the membership to the same LHA, and negatively correlated with geographical distance. However, the hospitals with a better performance in terms of readmission are less involved in patients' transfers. Along the same lines, Lomi et al. (2014), using Italian data on 35 hospitals, analyse the relationship between patients transfers and risk-adjusted readmission rate. They adopt a multiplicative Cox function for empirical relational events (the transfers), finding that patients tend to be moved to hospitals providing better quality (i.e. lower risk-adjusted readmission rate). To the best of our knowledge, Mascia et al. (2015) is the only paper which studies the effect of patients' transfers on hospital quality. The authors analyse the effect of cooperation on hospital quality, using Italian data on 31 hospitals. They adopt a multilevel model approach to describe the impact of measures of centrality and ego-network density on readmissions within 45-days after the discharge. In this paper, the authors find that a dense network is more likely to reduce the quality provided.

In this chapter we propose a global approach considering both the factors influencing the hospital cooperation as well as the effect of this cooperation on quality. Consistently with the literature, we measure inter-hospital collaboration using the patients' transfers. We describe the characteristics of the network of transfers by applying, for the first time to the healthcare sector, the social relations model in the form proposed by Hoff (2005), where both the effects of the sender and receiver hospitals, their correlation and their reciprocity is considered. This analysis allows the identification of the exogenous determinants of transfers which can be used in a subsequent analysis to test the effect of inter-hospital collaboration on hospital quality.

5.3 Data

The context of this study has been introduced in Chapter 1, where we described the Lombardy healthcare system in detail, and, in particular, the reimbursement system based on DRGs and the presence of private and public providers competing with each other for patients.

In this chapter, we analyse data on patients discharged from 145 hospitals accredited with the regional healthcare system in Lombardy in 2014. In this year the hospitalizations were 1,541,996, of which 84% were ordinary and 16% were in day hospital or day surgery. Furthermore, hospitalizations of patients living outside the Lombardy region accounted for 10% of all admissions. The hospital discharge data contains demographic information such as age and gender, information on hospitalization (length of stay, special-care unit use, transfers within the same hospital or through other facilities, and within-hospital mortality), and a total of 6 diagnosis codes and surgical procedures defined according to the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9- CM). Only ordinary hospitalizations for patients aged more than 2 years were retained in the sample. Both planned and unplanned admissions are considered in line with the criteria adopted in similar studies (Lomi et al. 2014, Mascia et al. 2015; 2012). In the analysis of hospital cooperation in terms of patients transfers, there is no evidence to consider that the planning of the hospitalization may affect the hospital cooperation. Given this, unfortunately, the available data do not allow us to identify cases like urgent patients treated in the emergency department of one hospital and then directly transferred to another hospital more experienced in treating those specific patients. In this case, when the admission to the ED is followed by a transfer to another hospital, the patient is not recorded as hospitalized in the first hospital. To the best of our knowledge, this is in line with the rest of the literature about patients transfers and therefore considered as a common limitation among similar studies. Table 5.1 provides a set of descriptive statistics split by hospital ownership and

	Private Hospitals		Public Hospitals	
	mean	sd	mean	sd
		Outcomes		
Transfers	0.01	0.11	0.02	0.14
Mortality	0.05	0.21	0.06	0.24
		Patients Characteristics		
Age	62.27	18.56	60.29	20.26
Female	0.51	0.50	0.55	0.50
Comorbidities	0.59	0.89	0.65	0.94
DRG Weight	1.34	1.25	1.21	1.21
		Hospitals		
Num of Hospitals	65		80	
Num of Transfers	3,024		12,492	
Num of Hospitalizations	256,909		643,242	

TABLE 5.1: Descriptive statistics about outcomes, patients' characteristics and hospitals considered in the analysis. Information is split by hospital ownership.

grouped by outcomes, patients characteristics and hospitals. Around 45% of the hospitals are private, although they only cover 28% of the hospitalizations. It is interesting to observe that, while patient's demographic characteristics (age and gender) are similar for private and public hospitals, their case-mix is quite different, with private hospitals having on average patients with less comorbidities and higher DRG weight. This may be explained, at least in part, by the higher cream-skimming policy adopted by private hospitals (Berta et al. 2010). Total mortality rates are similar for both public and private healthcare providers, whereas the level of transfers is double in public hospitals.

5.4 Finding the determinants of network of transfers

Measuring hospital cooperation using the patients' transfers leads us to the definition of a network among hospitals in Lombardy. In this network each hospital is a node and the edges are

	Private Hospitals		Public Hospitals		Total	
	mean	sd	mean	sd	mean	sd
In–centrality	0.38	0.30	1.00	0.64	0.74	0.61
In–strength	54.30	42.79	144.00	91.55	106.88	87.20
Out–centrality	0.45	0.73	0.95	0.99	0.74	0.93
Out–strength	64.32	105.50	136.93	143.27	106.88	133.50
Num of Hospitals	65		80		145	

TABLE 5.2: Network description split by hospital ownership. In-centrality refers to the number of hospitals from which a hospital receives transferred patients. In-strength is the total number of patients received by a hospital. Out-centrality is the number of hospitals where the patients are transferred from a hospital. Out-strength is the number of patients transferred by a hospital.

defined by the connections between two hospitals with a weight defined by the number of the patients transfers. In Table 5.2 we describe this network using strength indices and centrality indices which are normalized for the total number of hospitals, i.e. the total number of nodes in the network. The in-centrality measures the number of hospitals from which a hospital receives transferred patients, whereas the out-centrality indicates the number of hospitals where the patients are transferred from a hospital. On the other hand, strength indices define the total number of patients moved in the hospital network. The in-strength index refers to the number of patients received by a hospital, whereas the out-strength index measures the number of patients transferred by a hospital (Fernández-Gracia et al. 2017). Private hospitals show a higher in-centrality index compared to the out-centrality, whereas the inverse is observed for public ones. This confirms that the private hospitals are more involved in the network as sender hospitals than receiver. The same relationship is explained by the strength indices, evidencing that public hospitals are more engaged in a cooperative framework defined by the patients' transfers.
In order to identify partitions of the network where the nodes belonging to the same community strongly connected among them and sparsely connected with the nodes belonging to different communities, the hospitals' network has been analyzed applying the community detection method of Blondel et al. (2008). This community detection method, based on the improvement of the modularity of the identified communities, is one of the best performer methods in the comparison provided by Yang et al. (2016). The goodness of the community detection has been evaluated on the basis of the modularity index M of a network with N nodes and m links. Let

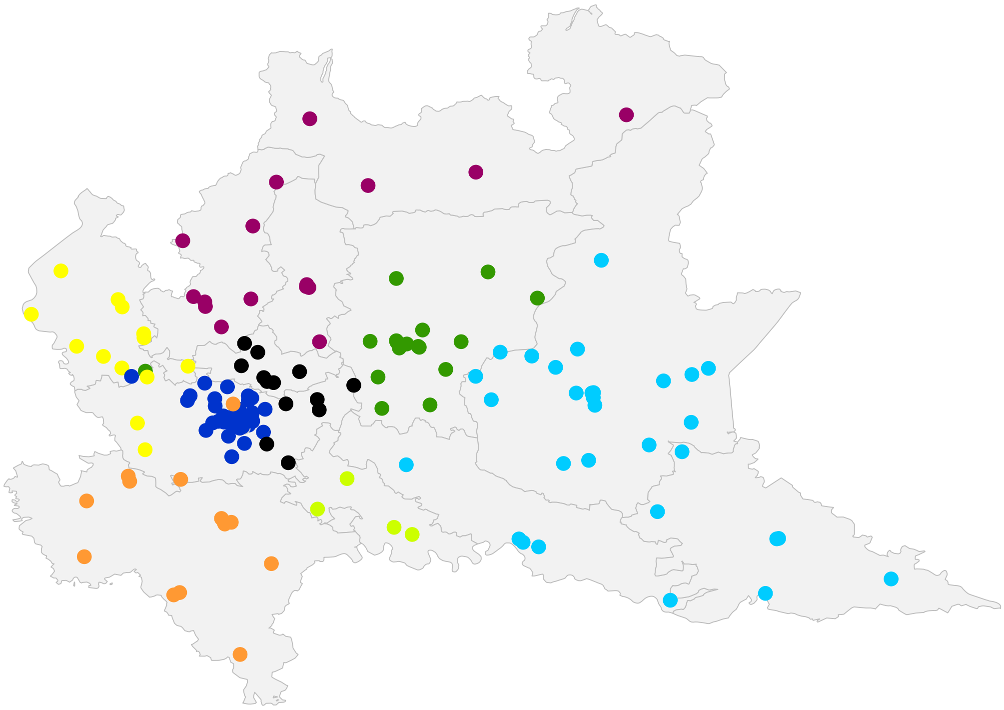


FIGURE 5.1: Maps of the hospitals and their belonging to the communities detected in the network of patients' transfers using the multilevel community detection method

$C_1, ..., C_c$ be a given candidate grouping of the network in c groups, the modularity M is defined as follows:

$$M = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) C_i C_j \tag{5.1}$$

where A is the adjacency matrix of the network, k_i is the degree of the node i, $C_i = 1$ if vertex i. The modularity index M increases when the community detection identifies nontrivial groups compared with a random assignment of the nodes to the groups.

In our analysis the modularity shows a strong relationship between the communities identified in the network and suggest a relationship with the geographical location, as can be observed in Figure 5.1. Nine communities are detected and most of them correspond to specific municipalities in Lombardy (Pavia, Lodi and Bergamo, respectively identified by orange, red and green dots). The metropolitan area of Milan, where the most part of the hospitals are located, is characterized by three different communities. The eastern part of Milan shares the community with the municipality of Monza-Brianza (black dots), whereas the western part shares the hospitals with the community of Varese (yellow dots). The central part of this metropolitan area identifies one single community (blu dots). Finally two more communities are identified. The northern one is characterized by the presence of the Alps mountains and include the municipalities of Sondrio, Como and Lecco (purple dots). The last community detected is defined by the eastern part of the region and is shared by the municipality of Cremona, Mantua and Brescia (azure dots). The community detection process has shown how both the distance between hospitals and the administrative co-membership are substantial factors in defining the hospitals' network, consistently with the literature (Caimo et al. 2017, Lomi and Pallotti 2012, Mascia et al. 2012). On the basis of this analysis, we will include distance, which is also directly correlated with co-memberships on our model of patients' transfers.

In particular we aim to model patients' transfers using exogenous covariates not related to quality. We define a transfer between hospitals by a patient discharged from a hospital and then admitted in another hospital on the same day or the next one (Iwashyna et al. 2009). In order to exclude any patients' involvement in this process, we exclude the voluntary discharges decided by the patients. Let then T_{ij} define the number of transfers between hospital i and j. The transfers are modelled by

$$\begin{aligned} T_{ij} &= \alpha + \beta HD_i + \gamma HD_j + \delta A_i + \zeta A_j + \phi DW_i + \\ &+ \psi DW_j + \varrho C_i + \theta C_j + \eta D_{ij} + \varepsilon_{ij} \end{aligned} \tag{5.2}$$

$$\varepsilon_{ij} = a_i + b_j + \vartheta_{ij}$$

$$(a_i, b_i)' \sim MVN(0, \Sigma_{ab}), \quad \Sigma_{ab} = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix},$$

$$(\vartheta_{ij}, \vartheta_{ji})' \sim MVN(0, \Sigma_\vartheta), \quad \Sigma_\vartheta = \sigma_\vartheta^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

where a_i and b_j, with $i = 1...N$ and $j = 1...N$, are the random effects for the sender and receiver hospitals, respectively, and ϑ_{ij} the errors. This model includes a covariance among the ε_{ij} given by:

$$E(\varepsilon_{ij}^2) = \sigma_a^2 + \sigma_b^2 + \sigma_\vartheta^2, \quad E(\varepsilon_{ij}\varepsilon_{ik}) = \sigma_a^2,$$

$$E(\varepsilon_{ij}\varepsilon_{ji}) = \rho\sigma_\vartheta^2 + 2\sigma_{ab}, \quad E(\varepsilon_{ij}\varepsilon_{kj}) = \sigma_b^2,$$

$$E(\varepsilon_{ij}\varepsilon_{kl}) = 0, \quad E(\varepsilon_{ij}\varepsilon_{ki}) = \sigma_{ab}$$

where σ_a^2 represents the correlation of observations having a common hospital sender, whereas σ_b^2 defines the dependence of observations having a common hospital receiver, and ρ measure the "reciprocity" between sender and receiver hospitals, that is the dyadic correlation between the observation i, j and j, i.

This social relations model belongs to the class of mixed models where random effects are included in order to control for both sender and receiver hospitals. We estimate the model in Equation (5.2) using the *ame* function in the *AMEN* R-package (Hoff et al. 2017) that allows us to deal with dyadic variables and statistical dependencies in the data. This R-package does not support count dependent variable, thus we normalize the variable T_{ij}, using an Anscombe transformation (Anscombe 1948), and we assume a Gaussian distribution for the errors.

In Equation (5.2) we have included a number of covariates for both the origin (i) and destination

(j) hospital. In particular, we control for the number of discharges, in order to rescale the dependent variable. The variable DRG Weight (DW) is a measure of resources the hospital employs to treat patients, and patients' age (A) is measured as an average at the hospital level. Both these variables are included in the model as a proxy of patients' severity. In addition, we include in the model a dyadic geographical variable, D, indicating the distance between hospitals. We also control for the degree of centrality (C) of each origin and destination hospital in the network in a geographical sense. In particular, the degree of centrality index measures the number of edges for each hospital (vertex) in the network based on the distances among hospitals. This index is calculated starting from an adjacency matrix, that define an edge if two hospitals are distant less than 30 minutes of effective time travel. Including the origin and destination degree of centrality means adjusting the model for the hospitals' concentration in a pre-defined space. The hypothesis is that a higher value of this index for the origin hospital indicates a wider choice set for the hospital that needs to decide where to transfer a patient. Table 5.3 shows the estimated coefficients of the social relations model in Equation 5.2. The

	Estimate	Std. Error
Intercept	3.0321***	0.5481
Distance (D_{ij})	-0.0258***	0.0011
Origin		
Num of Hospital Discharges (HD_i)	0.0001***	0.0002
DRG Weight (DW_i)	0.3967**	0.1652
Age (A_i)	-0.0021	0.0052
Centrality Index (C_i)	-2.5273***	0.2641
Destination		
Num of Hospital Discharges (HD_j)	0.0001***	0.0003
DRG Weight (DW_j)	0.1423	0.1756
Age (A_j)	-0.0011	0.0057
Centrality Index (C_j)	-2.7512***	0.2712
σ_a^2	0.1341	0.0192
σ_{ab}	0.1131	0.0171
σ_b^2	0.1405	0.0191
ρ	0.6458	0.0064

Note: *** Sign at 0.01,** Sign at 0.05, *Sign at 0.1

TABLE 5.3: Modelling patients' transfers using Equation (5.2)

negative and significant relationship between the degree of centrality and the transfers explains the role of these indices in the model. As expressed above, the choice to control for the degree of centrality allows us to measure the set of choices available for the hospital as sender and also the set of potential cooperator hospitals as receiver. The negative effect indicates that when the set of opportunity where to transfer a patient increases the number of patients sent to a specific hospital reduces and, moreover, a destination hospital receives less patients when the number of hospitals in its network increases. In order to check the robustness of this result, we have estimated the model in Equation (5.2) including different degree of centrality indices based on several thresholds of the distances (between 20 and 40 minutes). We have observed a correlation over 0.90 among all the predicted values estimated including these different indices.

Considering the distance between two hospitals, as expected, the shorter is the distance the higher is the number of transfers, despite the high density of hospitals in Lombardy. This ensures that a patient can be transferred by reducing the cost for travelling and the risks for the patient associated to the transfer (Caimo et al. 2017, Landon et al. 2012, Mascia et al. 2012). The age and the DRG weight do not have a significant effect on the transfers, except for the DRG weight at origin, where the positive coefficient indicates that the patients that are transferred are the more complicated ones. This results supports the hypothesis of an altruistic behavior.

In addition, our social relations model includes also the measure of the reciprocity between the observations, represented by the ρ coefficient in Table 5.3. This parameter takes into account the dyadic correlation between the observations in a pair (Hoff 2005), and its high value suggests a high correlation between the two.

5.5 The impact of cooperation on the hospital quality

The baseline hypothesis of this study states that cooperation between hospitals improves the overall quality in a healthcare system. In order to evaluate this hypothesis, we adopt a statistical approach that measures the effect of the cooperation between hospitals on the hospital quality. Quality is measured using hospital mortality, defined by a variable assuming value 1 if the patient dies in hospital or within 30-days after the discharge, and 0 otherwise.

Considering that the observed mortality depends on the different case-mix within hospitals, we add to the empirical strategy an evaluation of the healthcare quality to obtain a risk-adjusted mortality. Given the binary outcome mortality, let π_{pj} be the probability that the patient p dies in the hospital j, with $p = 1, ..., P_j$ the number of patients discharged from the hospital j with $j = 1, ..., J$ and $P = P_1 + ... + P_J$ the total number of patients admitted in Lombardy. The hierarchical structure of the data leads to the adoption of a multilevel logistic model, as introduced in Section 1.4, and described by the following equation:

$$\log\left(\frac{\pi_{pj}}{1-\pi_{pj}}\right) = \alpha + \boldsymbol{\eta}\boldsymbol{X}_{pj} + u_{0j} \tag{5.3}$$

where $\boldsymbol{X}_{pj}$ represents a set of patients' characteristic and $u_{0j} \sim N(0, \tau_0^2)$ is the hospital-specific random effect. Table 5.4 shows the results of the quality evaluation based on mortality. All the patient's covariates included in the model significantly affect the risk of death. This is allowing us to obtain predicted mortality adjusted for the different case-mix hospitalized by each provider.

Covariate	Estimates	Std. Error	Covariate	Estimates	Std. Error
Female	-0.180	0.010	AIDS/HIV	1.039	0.091
Age	0.072	0.000	Lymphoma	1.095	0.037
DRG Weight	0.182	0.003	Metastatic Cancer	1.702	0.018
Congestive Heart Failure	0.744	0.014	Solid Tumor Without Metastasis	0.286	0.016
Valvular Disease	-0.375	0.032	Coagulopathy	1.401	0.050
Pulmonary Circulation Disorders	0.701	0.035	Weight Loss	3.193	0.029
Peripheral Vascular Disorders	0.136	0.030	Fluid and Electrolyte Disorders	0.837	0.025
Hypertension, Uncomplicated	-0.656	0.026	Blood Loss Anemia	-0.353	0.045
Paralysis	0.943	0.070	Deficiency Anemia	-0.539	0.052
Other Neurological Disorders	0.946	0.028	Alcohol Abuse	0.244	0.049
Chronic Pulmonary Disease	0.069	0.021	Depression	-0.925	0.099
Hypothyroidism	-0.574	0.074	Hypertension, Complicated	-0.693	0.037
Renal Failure	0.365	0.019	Constant	-8.752	0.063
Liver Disease	1.256	0.026			
Observations	900,151				
Number of groups	145				

Note: All variables are significant at 0.01

TABLE 5.4: Quality evaluation based on mortality

Hence, we sum these probabilities over all the patients admitted in every hospital to obtain

$$W_i = \sum_{p=1}^{P_i} \widehat{\pi_{pi}}$$

where $\widehat{\pi_{pi}}$ is the expected probability for patient p in hospital i given by the model in Equation (5.3).

At this point, we define the overall quality for the hospitals i and j in terms of mortality as

$$W_{ij} = W_i + W_j, \tag{5.4}$$

representing the number of patients death for each hospitals' pairs. We decided to adopt a measure of overall quality instead of a measure of mortality split by both sender and receiver hospitals because we are interested in estimating the impact of cooperation on the overall quality of the healthcare system. In fact, if a hospital sends a patient with a very high risk of mortality to another hospital and the patient dies, this increases the mortality of the receiver but does not impact on the mortality of the pair, since the patient would have most likely died in the hospital from which she was transferred. Considering the overall mortality allows us to take into account the effect of the cooperation on the mortality for both the sender and receiver hospitals.

From the model described in the previous section we obtain the predicted transfer, adjusted for a set of covariates not related to the quality, and this guarantees that the predicted transfers can be taken as an exogenous predictor of the overall mortality.

Let then $\widehat{T_{ij}}$ define the number of predicted transfers between hospital i and j estimated by the model in Equation (5.2). In order to avoid other problems of endogeneity, $\widehat{T_{ij}}$ is calculated excluding the hospital random effects because they can be related with characteristics such as the teaching status or the size of a hospital, which can affect hospital quality.

At this point, we relate the predicted cooperation ($\widehat{T_{ij}}$) with the hospital performance defined in Equation (5.4) (W_{ij}). This step can be formulated as a Poisson mixed effect model, and defined as follows:

$$E(W_{ij}|\widehat{T_{ij}}, HD_{ij}, OWN_{ij}) = \exp(\alpha + \xi\widehat{T_{ij}} + \beta HD_{ij} + \boldsymbol{\theta} OWN_{ij} + u_i + u_j) \tag{5.5}$$

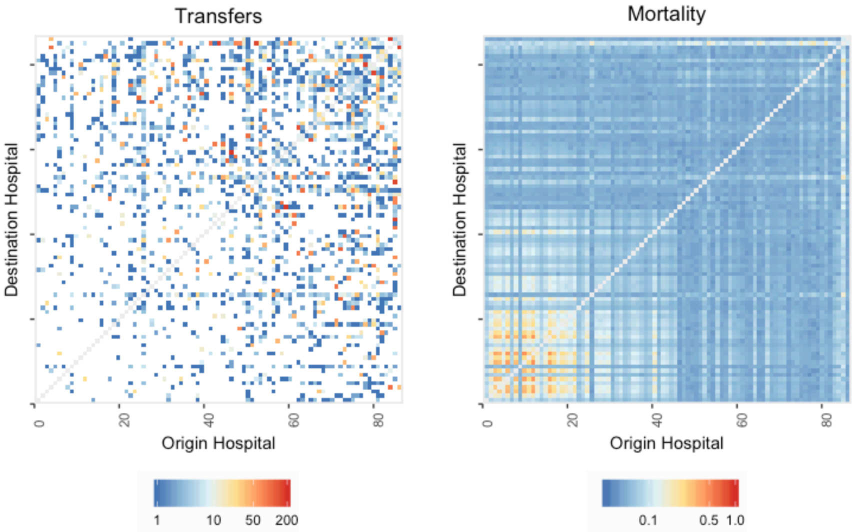
where $u_1, ..., u_{145} \sim N(0, \sigma_u^2)$. The coefficient ξ in Equation (5.5) is of interest in order to demonstrate the hypothesis that the cooperation $\widehat{T_{ij}}$, modelled as in Equation (5.2), increases the hospital quality. The model is also controlled for the discharges (HD) of the hospitals' pairs, and for their ownership (OWN_{ij}) that can be public and public or private and private, or from public to private and from private to public. Table 5.5 shows the results of the model

	Estimate	Std. Error
(Intercept)	5.7380***	0.0431
$\widehat{T_{ij}}$	-0.0203***	0.0030
Num of Hospital Discharges (HD_{ij})	0.0593***	0.0020
Private vs Private (OWN_{ij} (1))	-0.3906***	0.0513
Public vs Private (OWN_{ij} (2))	-0.1211***	0.0358
Private vs Public (OWN_{ij} (3))	-0.1214***	0.0358

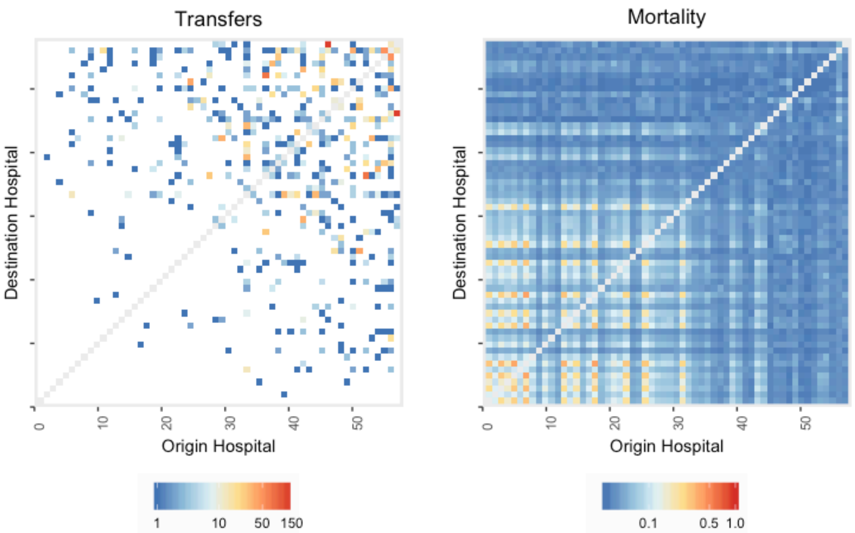
Note: *** Sign at 0.01,** Sign at 0.05, *Sign at 0.1

TABLE 5.5: Modelling overall mortality adjusted for predicted patients' transfers

described in Equation (5.5). This is the core of the chapter, where we analyze the effect of cooperation on quality (mortality). This analysis confirms the hypothesis behind the study, showing a negative and significant coefficient indicating that the higher is the cooperation between a pair of hospitals the lower is the overall risk adjusted mortality for the two hospitals. In an institutional setting affected by asymmetric information about the hospital quality, the hospitals follow their perceived quality and their ability in building an informal network. This positive effect on the quality means that this informal network has a positive effect for the patients, for the hospital performance and, at the same time, increases the overall quality of the healthcare system.
The analysis is concluded by a deepening based on the hospital ownership. In specific, we want to disentangle the differences by ownership concerning the relationship between cooperation and quality. To do this, we add to Equation (5.5) the interactions between predicted transfers at first stage and the hospitals' ownerships. The result of this analysis is presented in Figure 5.2. On the left, the heatmap represents the observed transfers between hospitals' pairs sharing the same ownership (public vs public on the top and private vs private at the bottom), whereas on the right, the heatmap shows the expected mortality produced by the model in Equation (5.5), including the aforementioned interaction, scaled by the number of discharges of the hospitals' pairs. We observe that public hospitals are more involved in cooperation than private, whereas in both Subfigures 5.2a and 5.2b the analysis confirms that regardless of the ownership, the cooperation is effective in reducing mortality.



(a) Heatmap for public hospitals



(b) Heatmap for private hospitals

FIGURE 5.2: The effect of the patients transfers on the predicted overall mortality. The hospitals in the heatmaps are sorted by the in-degree index and the colors of the points are defined in the log-scale, except for the null transfers where the points are white.

5.6 Discussion

This work has studied the informal cooperation between hospitals, analysing the determinants of this and the effects on the hospital performance.

The analysis of the determinants of patients' transfers is in line with the literature (Caimo et al. 2017, Lomi et al. 2014, Lomi and Pallotti 2012). Our findings show that geographical distance between hospitals plays an important role in explaining the flows of patients between them: the shorter the distance, the higher is the number of transfers. This result was confirmed by observing that the communities detected by the community detection method based on the network of transfers had a clear geographical nature.

Most relevant is the effect of the degree of centrality in the explanation of the mechanisms of patients' transfers. In our model we find that when the number of hospitals which transfer patients to a specific hospital increases, the number of patients transferred by this hospital decreases. This seems to say that a big receiver tends to be a small sender. We also show that the higher is the number of opportunities where to send the patients the lower is the number of patients transferred to a specific hospital. This support the hypothesis that the choice of the hospital where to transfer a patient is based on the appropriate solution for the patients' needs. But this could be also a critical point that undermines the effect we found in this work. Indeed, according to Mascia et al. (2015), when a network is sparse and the opportunities to cooperate decrease, the quality provided increases. In this sense a policy intervention, in a context where the number of hospitals where to transfer the patients is high, should artificially reduce the density of the network, i.e., addressing the transfers to a pre-identified group of hospitals. This should lead to a better identification of the hospitals where the patients should be transferred to reduce their risk of adverse outcomes.

The main contribution to the use of advanced statistical methods in healthcare relates to a novel global way to analyze the effect of the inter-hospital cooperation on hospital quality. First of all, differently from similar works (Caimo et al. 2017), we study the hospital network defined by patients transfers, considering the dimension of the edge and not only the presence or absence of a connection between the nodes (hospitals) in the network. Secondly, we distinguish a first stage where we model the determinants of patients transfers, and we take care of the potential endogeneity due to the quality. To this aim, we do not include among the covariates explaining the patients transfers any variable which can be related with the hospital quality. In this way we avoid a potential bias that could affect the analysis of the relationship between cooperation and quality.

To the best of our knowledge, before this study, only the work by Mascia et al. (2015) considered the hospital quality as a dependent variable in the analysis of this topic. In contrast to their work, we do not limit the study to the effect of the network characteristics on the quality of the transferring hospital, but we define a measure of quality that considers the benefits for the patients, for the hospitals and, as a consequence, for the overall healthcare system. The results in this chapter support the hypothesis that cooperation among hospitals leads to better health outcomes for both the origin and destination hospitals. The ownership analysis demonstrates that public hospitals are more engaged in hospital cooperation, but that cooperation is effective also for the private ones. In synthesis our findings encourage the hospital cooperation, suggesting that the wide heterogeneity in hospital quality observed in the overall healthcare system, could be reduced with a mechanism of inter-hospital cooperation (Iwashyna et al. 2009).

This chapter could have several prominent policy implications. First of all, the results show that it is necessary and urgent to disclose information regarding hospital quality ranking computed within the regional quality evaluation programme. The presence of asymmetric information must be reduced at the patients' level as well as at the hospital level, so that the informal networks can be supported by this information. Furthermore, our findings suggest that policy makers should support cooperation, involving in this process all the hospitals belonging to the healthcare system. Moreover, policy makers should support hospitals' networks based on geographical proximity, same administrative membership (municipality or Local Health Authority) and hospital characteristics and abilities.

Future works can be proposed in order to increase the knowledge about the impact of hospital cooperation. First of all, the study of the hospital cooperation should be crossed with the competition policy existing in the Lombardy region, which has been studied in Chapter 4. The relationship between competition and cooperation should be analysed with the aim of understanding whether, at least in our case-study, hospital competition limits a positive hospital networking (Lomi and Pallotti 2012, Mascia et al. 2012).

Second, we choose one measure of cooperation, but this does not imply that the hospitals could collaborate in several other ways. An interesting further analysis should be implemented including different cooperation measures and also different quality indicators. For example, it could be of interest to consider the scientific collaboration among professionals, which can be derived by the scientific works published jointly by physicians operating in different hospitals.

Finally, in order to generalize of our findings, we would like to use similar approaches to study hospital cooperation in different healthcare systems, where the effect of cooperation on the hospital quality has not been studied. In the US, for example, where the hospitals are free to choose where to transfer their patients, the access to a better quality hospital is not guaranteed (Iwashyna 2012, Lomi et al. 2014, Veinot et al. 2012). A further evolution of this study could analyse the effect of cooperation on quality at different magnitude levels of the transfers' distribution. An approach considering a non linear dependence between quality and transfers can be adopted in order to identify possibly different patterns for different level of transfers. Part of this distribution could suggest that increasing the intensity of the informal network increases the overall quality, but after some threshold the level of transfers could become negative for the patients and as a consequence for the quality.

Chapter 6

Conclusions

6.1 Summary

The general scope of this thesis was to shed light on the role of statistical techniques within a healthcare quality evaluation framework. To this aim, we studied several topics which are typical of the healthcare literature in the fields of statistics, economics and econometrics. We highlighted how statistical methods can be used in a variety of contexts, such as to test the impact of a market condition such as competition on the healthcare quality, to measure the causal effect of a policy on the quality improvement, to analyse the effect of cooperation on the hospital performance, and to measure quality in the first place. At the same time we implemented a model framework, which demonstrated that the usual methods adopted for quality evaluation can be improved by an advancing in statistical methodology.

In particular, in Chapter 1 we introduced the thesis, with a background on the quality in healthcare, the statistical methods used, most notably multilevel models, and the case-study considered in this thesis, which is the Lombardy region in Italy. The regional healthcare system in Lombardy, is characterized by a population of 10 millions of citizens, where public and private providers compete with each other, patients are free to choose where to be hospitalized, and a pay-for-performance program based on an annual effectiveness evaluation was recently implemented. For all these reasons we consider Lombardy region a relevant case-study.

In Chapter 2 a methodological extension of the cluster weighted models is presented. This statistical framework introduced in the literature by Ingrassia et al. (2012) is a generalization of the widespread finite mixture models. Although multilevel models are widely used for hospital evaluations, to disentangle observed heterogeneity at the hospital and ward level, they are not able to identify the presence of latent heterogeneity. This can be overcome by the adoption of a finite mixture of regression models, but in this case the observed heterogeneity due to the patients clustering within hospitals is avoided. Differently from these approaches, the cluster weighted multilevel models allow us to consider the presence of both latent and known heterogeneity. The exploration of this new modelling approach demonstrated the ability of cluster weighted models to disentangle the latent groups within the individual data in a healthcare context and a good performance in terms of model fitting both on simulated and real data. Finally, the hospital league tables performed using a cluster weighted approach were shown to improve the quality of the performance evaluation in healthcare.

Policy evaluation based on advanced statistical methods is a relevant topic for the healthcare sector. In Chapter 3 we exploited a recent reform in Lombardy which introduced a pay-for-performance program aimed at increasing the hospital quality. We evaluated the causal effect of this policy using a difference-in-differences approach. Difference-in-differences is one of the most used approach in the policy evaluation context in order to disentangle the causal impact of a policy, but it requires that the design of a quasi-experimental framework respects some assumptions. In particular the treated and untreated groups of observations must be independent from each other and the difference between treated and untreated units in the period before the policy implementation has to be constant (parallel trend). Using a logistic linear mixed model we evaluated the effect of the implementation of the pay-for-performance program on multiple outcomes while accounting for the heterogeneity of the data at the multiple nested levels. The results demonstrated the positive effect of the policy, in particular for those outcomes that can be more influenced by a managerial activity. This finding is relevant for an international audience, due to the limited adoption of the pay-for-performance programs and moreover to the limited evaluation of those existent.

In Chapter 4 we aimed at evaluating the effect of competition between hospitals on the hospital performance. We introduced this economic concept and we presented how this topic is defined within the healthcare sector. Furthermore, we summarized the huge literature on competition in healthcare. The main research question in this chapter was to understand the role of competition in the Lombardy healthcare quasi-market, and to contribute to the general literature by analyzing the effect of asymmetric information. In this context asymmetric information is due to the fact that the citizens and the GPs do not have accurate information about the hospital quality and they cannot use this information in order to choose the hospital where to be admitted. This condition can reduce the effect of pro-competition reforms on quality. The main result of this chapter confirms most of the evidence in the literature, namely that competition does not influence on the hospital quality. Our analysis may contribute to the empirical literature, explaining why often the result that more competition improves the quality in healthcare is rejected. We concluded that a public disclosure of the hospital quality could impact on the overall performance of the healthcare system.

Chapter 5 is dedicated to the cooperation in healthcare. Literature on the effects of cooperation in healthcare is limited, and researchers are mostly focused on competition between hospitals, disregarding the existence of collaboration among providers. Starting from the idea that the hospitals and the physicians act as an altruistic economic agent, in Chapter 5 we analyse hospital cooperation in a competitive environment and we measure if cooperation increases hospital quality. We contribute to the lack of the existing literature in several ways. Firstly, we used the patients' transfers as a measure of cooperation and we adopted a social relations model from the network modelling literature to detect the determinants of cooperation. Secondly, we estimated if the network between hospitals has an effect on the overall mortality, i.e. on the quality provided by both the origin and destination hospital. The positive findings of this analysis showed that hospital cooperation should be supported by policy makers in order to stimulate the improvement of the healthcare systems.

This thesis suffers from some relevant limitations, apart from those included in each chapter, many of which related to the data and design of the study. First of all, in this thesis we are considering a single case study (Lombardy) and this is a limit for the generalizations of our results. Furthermore, Lombardy is a particular case in the context of the healthcare. In fact, Lombardy is a region which can organize its proper healthcare system, but is included in a national healthcare system. This situation makes a study on Lombardy different from typical study in healthcare literature, where, usually, national healthcare systems are treated. In Lombardy for example the opportunity to attract extraregional patients can influence the cooperation within the region, and this is not a condition affecting national healthcare systems. Furthermore, in Lombardy the hospital density is not uniformly distributed over all the region. The metropolitan area of Milan is characterized for including the majority of the hospitals and the 30% of the population. In this sense, considering for example chapter 5, it could be that transfers are less risky within Milan compared to the rest of the region. Considering, instead, chapter 4, the higher level of hospital density in Milan increase the level of competition in this area, compared to the rest of the regional territory. One more limitation concerns the data availability. In fact, in all the analysis we must consider only hospitalizations for patients living in Lombardy, because, after they are discharged, we are not able to collect data for patients living in other regions but hospitalized in Lombardy. This implies for example that we are not able to detect if they die 30 days after the discharge or if they are re-admitted in another hospital. For this reason we are loosing the 10% of the information in our analysis, without loss of generality, but excluding a substantial part of hospitalizations. Other limitations, in particular for chapter 4 and 5, concerns that using observational data we are unable to make inferences about causation, therefore, effect estimates should be cautiously interpreted. Furthermore, as with other studies using administrative data, information was lacking on potential confounding variables such as lifestyle factors. Despite this, the use of administrative data has been widely adopted in healthcare literature because such data, compared to clinical registries, are easily accessible, relatively inexpensive to use, and enable information to be collected on the entire population. Other variables that are important to the analysis might have being omitted or unmeasured in this thesis such as information on patients that may die before reaching the hospital, which could provide indication on whether hospital patient populations differ across in terms of opportunity to be assisted in a limited time after the clinical event, which is one of the main driver of mortality in chapter 4.

6.2 Recommendation for future research

In this thesis we explored a wide range of topics in healthcare using several statistical techniques. The results in Chapter 2 should encourage policy makers and researchers who deal with quality evaluation to exploit model based clustering approaches. We introduced the multilevel logistic cluster weighted model as a tool for disentangling latent heterogeneity in the data, improving the quality evaluation. In order to enhance the opportunities of using this model framework, cluster weighted multilevel models should be developed for count data. At the same time, our model is limited in disentangling the latent heterogeneity at patient level. One of the main developments consists on a cluster weighted multilevel model able to detect clusters at hospital level.
Chapter 3 and Chapter 4 demonstrated that statistics is a resource which can be exploited in order to study respectively the causal effect of the policies and hospital competition on the quality of cares. Concerning Chapter 3, we observed that the pay-for-performance programs are limited and most of the times their effect on hospital quality is not evaluated on the basis of a causal inference framework. It would be interesting to test the impact of the P4P program in terms of outcomes directly related to the performance of the hospital physicians. Furthermore, It would be useful to compare the P4P program in Lombardy with neighbouring regions which do not implement a P4P program. Finally, different dimensions of the quality should be investigated, i.e. efficiency. Chapter 4 demonstrated that the asymmetric information should be considered when the effect of the hospital competition on quality is evaluated. Furthermore, we suggest the opportunity to study the effect of the hospital competition on the efficiency, which is a topic that received a reduced attention in the scientific literature. We also suggest to compare the effect of the hospital competition on quality in Lombardy with the effect of competition in other Italian regions where we do not expect the presence of competition due to the different healthcare system's rules. Finally, in Chapter 5 we summarized a novel contribution to the literature concerning hospital cooperation. There is an evident lack in the scientific literature in terms of papers analyzing the hospital cooperation. We hope that this work stimulate researchers in exploring this area of healthcare sector. Moreover, there is also a lack of the literature on statistical networks dedicated to the healthcare sector. Healthcare is a framework where stakeholders are naturally connected (i.e. hospitals with hospitals, GPs with hospitals, patients with patients) and we consider that the scientific literature on health economics and health statistics could benefit by a most extended adoption of statistical network analysis for study the healthcare sector. In

any case, future works should analyse the underlying mechanisms of the hospital cooperation, and different measure of competition should be tested and compared with the patients' transfers.

Bibliography

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies 72*(1), 1–19.

Ai, C. and E. C. Norton (2003). Interaction terms in logit and probit models. *Economics letters 80*(1), 123–129.

Aigner, D., C. K. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. *Journal of econometrics 6*(1), 21–37.

Akerlof, G. et al. (1970). The market for lemons. *Quarterly journal of Economics 84*(3), 488–500.

Alshamsan, R., A. Majeed, M. Ashworth, J. Car, and C. Millett (2010). Impact of pay for performance on inequalities in health care: systematic review. *Journal of health services research and policy 15*(3), 178–184.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Ansari, M. Z. and B. T. Collopy (1996). The risk of an unplanned return to the operating room in australian hospitals. *ANZ Journal of Surgery 66*(1), 10–13.

Anscombe, F. J. (1948). The transformation of poisson, binomial and negative-binomial data. *Biometrika 35*(3/4), 246–254.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American economic review 53*(5), 941–973.

Ash, A. S., S. F. Fienberg, T. A. Louis, S.-L. T. Normand, T. A. Stukel, and J. Utts (2012). Statistical issues in assessing hospital performance. *Commissioned by the Committee of Presidents of Statistical Societies The COPSS-CMS White Paper Committee*.

Asparouhov, T. and B. Muthén (2008). Advances in latent variable mixture models. In G. Hancock and K. Samuelson (Eds.), *Advances in Latent Variable Mixture Models*, Charlotte, pp. 27–51. Information Age Publishing.

Austin, J. M., A. K. Jha, P. S. Romano, S. J. Singer, T. J. Vogus, R. M. Wachter, and P. J. Pronovost (2015). National hospital ratings systems share few common scores and may generate confusion instead of clarity. *Health Affairs 34*(3), 423–430.

Ayyagari, P. and D. M. Shane (2015). Does prescription drug coverage improve mental health? evidence from medicare part d. *Journal of health economics 41*, 46–58.

Balia, S., R. Brau, and E. Marrocu (2018). Interregional patient mobility in a decentralized healthcare system. *Regional Studies 52*(3), 388–402.

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

Barnett, M. L., B. E. Landon, A. J. O'malley, N. L. Keating, and N. A. Christakis (2011). Mapping physician networks with self-reported and administrative data. *Health Services Research 46*(5), 1592–1609.

Basford, K., D. Greenway, G. McLachlan, and D. Peel (1997). Standard errors of fitted component means of normal mixtures. *Computational Statistics 12*(1), 1–18.

Beckert, W., M. Christensen, and K. Collyer (2012). Choice of nhs-funded hospital services in england. *The Economic Journal 122*(560), 400–417.

Berta, P., G. Callea, G. Martini, and G. Vittadini (2010). The effects of upcoding, cream skimming and readmissions on the italian hospitals efficiency: a population-based investigation. *Economic Modelling 27*(4), 812–821.

Berta, P., S. Ingrassia, A. Punzo, and G. Vittadini (2016). Multilevel cluster-weighted models for the evaluation of hospitals. *Metron 74*(3), 275–292.

Berta, P., G. Martini, F. Moscone, and G. Vittadini (2016). The association between asymmetric information, hospital competition and quality of healthcare: evidence from italy. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 179*(4), 907–926.

Beveridge, W. H. et al. (1942). *Social insurance and allied services*. HM Stationery Office London.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment 2008*(10), P10008.

Bloom, N., C. Propper, S. Seiler, and J. Van Reenen (2015). The impact of competition on management quality: evidence from public hospitals. *The Review of Economic Studies 82*(2), 457–489.

Blundell, R., C. Meghir, M. C. Dias, and J. V. Reenen (2004). Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association*, 569–606.

Böhning, D., E. Dietz, R. Schaub, P. Schlattmann, and B. Lindsay (1994). The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. *Annals of the Institute of Statistical Mathematics 46*(2), 373–388.

Brekke, K. R., L. Siciliani, and O. R. Straume (2008). Competition and waiting times in hospital markets. *Journal of Public Economics 92*(7), 1607–1628.

Brenna, E. (2011). Quasi-market and cost-containment in beveridge systems: the Lombardy model of Italy. *Health Policy 103*(2), 209–218.

Brock, W. A. and S. N. Durlauf (2001). Interactions-based models. *Handbook of econometrics 5*, 3297–3380.

Caimo, A., F. Pallotti, and A. Lomi (2017). Bayesian exponential random graph modelling of interhospital patient referral networks. *Statistics in Medicine 36*(18), 2902–2920.

Cantarero, D. (2006). Health care and patients' migration across Spanish regions. *The European Journal of Health Economics 7*(2), 114–116.

Carpenter, J. and J. Bithell (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in medicine 19*(9), 1141–1164.

Carpenter, J., H. Goldstein, and J. Rasbash (2003). A novel bootstrap procedure for assessing the relationship between class size and achievement. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 52*(4), 431–443.

Cashin, C., Y.-L. Chi, P. Smith, et al. (2014). *Paying for performance in health care: implications for health system performance and accountability*. Open University Press.

Castaldi, S., A. Bodina, L. Bevilacqua, E. Parravicini, M. Bertuzzi, and F. Auxilia (2011). Payment for performance (p4p): any future in italy? *BMC public health 11*(1), 1.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases 40*(5), 373–383.

Christiansen, C. L. and C. N. Morris (1997). Improving the statistical approach to health care provider profiling. *Annals of internal medicine 127*(8_Part_2), 764–768.

Colombo, A. (2008). The 'lombardy model': Subsidiarity-informed regional governance. *Social Policy & Administration 42*(2), 177–196.

Cooper, Z., S. Gibbons, S. Jones, and A. McGuire (2011). Does hospital competition save lives? evidence from the english nhs patient choice reforms. *The Economic Journal 121*(554).

Corbeil, R. R. and S. R. Searle (1976). Restricted maximum likelihood (reml) estimation of variance components in the mixed model. *Technometrics 18*(1), 31–38.

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

Damberg, C., E. A. Kerr, and E. A. McGlynn (1998). Description of data sources and related issues. *Health Information Systems. Design Issues and Analytical Application*, 43–76.

Davies, A. R. (1994). *A guide to establishing programs for assessing outcomes in clinical settings*. Joint Commission on Accreditation of Healthcare Organizations.

Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite (2003). Is more information better? the effects of "report cards" on health care providers. *Journal of Political Economy 111*(3), 555–588.

Dranove, D. and R. Lindrooth (2003). Hospital consolidation and costs: another look at the evidence. *Journal of health economics 22*(6), 983–997.

Dranove, D. and A. Sfekas (2008). Start spreading the news: a structural estimate of the effects of new york hospital report cards. *Journal of health economics 27*(5), 1201–1207.

Dubois, R., R. Brook, and W. Rogers (1987). Adjusted hospital death rates: Potential screen for quality of medical care. *American Journal of Public Health 77*, 1162–1167.

Eijkenaar, F. (2012). Pay for performance in health care: An international overview of initiatives. *Medical Care Research and Review 69*(3), 251–276.

Eijkenaar, F., M. Emmert, M. Scheppach, and O. Schöffski (2013). Effects of pay for performance in health care: a systematic review of systematic reviews. *Health policy 110*(2), 115–130.

Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey (1998). Comorbidity measures for use with administrative data. *Medical care 36*(1), 8–27.

Ellis, R. P. and T. G. McGuire (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of health economics 5*(2), 129–151.

Emmert, M., F. Eijkenaar, H. Kemter, A. S. Esslinger, and O. Schöffski (2012). Economic evaluation of pay-for-performance in health care: a systematic review. *The European Journal of Health Economics 13*(6), 755–767.

Everitt, B. and D. Hand (1981). *Finite mixture distributions*. Chapman and Hall.

Fabbri, D. and S. Robone (2010). The geography of hospital admission in a national health service with patient choice. *Health Economics 19*(9), 1029–1047.

Fernández-Gracia, J., J.-P. Onnela, M. L. Barnett, V. M. Eguíluz, and N. A. Christakis (2017). Influence of a patient transfer network of us inpatient facilities on the incidence of nosocomial infections. *Scientific reports 7*(1), 2930.

Fillmore, H., C. Dubard, G. Ritter, and C. Jackson (2014). Health care savings with the patient-centered medical home: Community care of North Carolina's experience. *Population Health Management 17*(3), 141–148.

Folland, S., A. C. Goodman, M. Stano, et al. (2007). *The economics of health and health care*, Volume 6. Pearson Prentice Hall Upper Saddle River, NJ.

Galizzi, M. M., T. Tammi, G. Godager, I. Linnosmaa, and D. Wiesen (2015). Provider altruism in health economics. *Työpaperi: 2015_004*.

Gaynor, M. (2006). What do we know about competition and quality in health care markets?[electronic resource]. *URL: http://www. bristol. ac. uk/medialibrary/sites/cmpo/migrated/documents/wp151pdf*.

Gaynor, M. (2016). Consolidation and competition in us health care. *Health Affairs Blog*.

Gaynor, M. and D. Haas-Wilson (1998). Change, consolidation, and competition in health care markets. Technical report, National bureau of economic research.

Gaynor, M., M. Laudicella, and C. Propper (2012). Can governments do it better? merger mania and hospital outcomes in the english nhs. *Journal of health economics 31*(3), 528–543.

Gaynor, M., C. Propper, and S. Seiler (2012). Free to choose? reform and demand response in the english national health service. Technical report, National Bureau of Economic Research.

Geiser, C. (2012). *Data analysis with Mplus*. Guilford Press.

Gittell, J. H. and L. Weiss (2004). Coordination networks within and across organizations: A multi-level framework. *Journal of Management Studies 41*(1), 127–153.

Glickman, S. W., F.-S. Ou, E. R. DeLong, M. T. Roe, B. L. Lytle, J. Mulgund, J. S. Rumsfeld, W. B. Gibler, E. M. Ohman, K. A. Schulman, et al. (2007). Pay for performance, quality of care, and outcomes in acute myocardial infarction. *Jama 297*(21), 2373–2380.

Goldstein, H. (2010). *Multilevel Statistical Models* (4th ed.). London: Wiley.

Goldstein, H., W. Browne, and J. Rasbash (2002). Partitioning variation in multilevel models. *Understanding Statistics 1*(4), 223–231.

Goldstein, H. and D. J. Spiegelhalter (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 385–443.

Grillo Ruggieri, T., P. Berta, A. M. Murante, and S. Nuti (2018). Patient satisfaction, patients leaving hospital against medical advice and mortality in italian university hospitals: a cross-sectional analysis. *BMC health services research 18*(1), 51.

Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity 43*(1), 59–74.

Hoff, P., B. Fosdick, A. Volfovsky, and Y. He (2017). *AMEN: Additive and Multiplicative Effects Models for Networks and Relational Data*. R package version 1.3.

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the american Statistical association 100*(469), 286–295.

Hox, J. and J. K. Roberts (2011). *Handbook of advanced multilevel analysis*. Psychology Press.

Hox, J. J. (1995). *Applied multilevel analysis*. TT-publikaties.

Hsiao, W. C., H. M. Sapolsky, D. L. Dunn, and S. L. Weiner (1986). Lessons of the new jersey drg payment system. *Health Affairs 5*(2), 32–43.

Huckman, R. S. and G. P. Pisano (2006). The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science 52*(4), 473–488.

Hwang, S. W., J. Li, R. Gupta, V. Chien, and R. E. Martin (2003). What happens to patients who leave hospital against medical advice? *Canadian Medical Association Journal 168*(4), 417–420.

Iezzoni, L. I. (1997). The risks of risk adjustment. *Jama 278*(19), 1600–1607.

Iezzoni, L. I., A. S. Ash, M. Shwartz, J. Daley, J. S. Hughes, and Y. D. Mackiernan (1996). Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *American Journal of Public Health 86*(10), 1379–1387.

Ingrassia, S., S. C. Minotti, and G. Vittadini (2012). Local statistical modeling via a cluster-weighted approach with elliptical distributions. *Journal of classification 29*(3), 363–401.

Ingrassia, S., A. Punzo, G. Vittadini, and S. Minotti (2015). The generalized linear mixed cluster-weighted model. *Journal of Classification 32*(1), 85–113.

Iwashyna, T., J. Christie, J. Kahn, and D. Asch (2009). Uncharted paths: hospital networks in critical care. *Chest 135*(3), 827–833.

Iwashyna, T., J. Christie, J. Moody, J. Kahn, and D. Asch (2009). The structure of critical care transfer networks. *Medical care 47*(7), 787–793.

Iwashyna, T. J. (2012). The incomplete infrastructure for interhospital patient transfer. *Critical Care Medicine 40*(8), 2470–2478.

Jha, A. K., K. E. Joynt, E. J. Orav, and A. M. Epstein (2012). The long-term effect of premier pay for performance on patient outcomes. *New England Journal of Medicine 366*(17), 1606–1615.

Karaca-Mandic, P., E. C. Norton, and B. Dowd (2012). Interaction terms in nonlinear models. *Health services research 47*(1pt1), 255–274.

Kessler, D. P. and M. B. McClellan (2000). Is hospital competition socially wasteful? *The Quarterly Journal of Economics 115*(2), 577–615.

Knowles, J. and C. Frederick (2016). *merTools: Tools for Analyzing Mixed Effect Regression Models*. R package version 0.3.0.

Krumholz, H. M., Z. Lin, P. S. Keenan, J. Chen, J. S. Ross, E. E. Drye, S. M. Bernheim, Y. Wang, E. H. Bradley, L. F. Han, et al. (2013). Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. *Jama 309*(6), 587–593.

Landon, B. E., N. L. Keating, M. L. Barnett, J.-P. Onnela, S. Paul, A. J. O'Malley, T. Keegan, and N. A. Christakis (2012). Variation in patient-sharing networks of physicians across the united states. *Jama 308*(3), 265–273.

Laudicella, M., P. L. Donni, and P. C. Smith (2013). Hospital readmission rates: signal of failure or success? *Journal of health economics 32*(5), 909–921.

Leape, L. L., T. A. Brennan, N. Laird, A. G. Lawthers, A. R. Localio, B. A. Barnes, L. Hebert, J. P. Newhouse, P. C. Weiler, and H. Hiatt (1991). The nature of adverse events in hospitalized patients: results of the harvard medical practice study ii. *New England journal of medicine 324*(6), 377–384.

Lee, B. Y., S. M. McGlone, Y. Song, T. R. Avery, S. Eubank, C.-C. Chang, R. R. Bailey, D. K. Wagener, D. S. Burke, R. Platt, et al. (2011). Social network analysis of patient sharing among hospitals in Orange County, California. *American journal of Public Health 101*(4), 707–713.

Levaggi, R. and M. Montefiori (2003). Horizontal and vertical cream skimming in the health care market.

Levaggi, R. and M. Montefiori (2013). Patient selection in a mixed oligopoly market for health care: the role of the soft budget constraint. *International Review of Economics*.

Levaggi, R. and R. Zanola (2004). Patients' migration across regions: the case of Italy. *Applied Economics 36*(16), 1751–1757.

Levin-Scherz, J., N. DeVita, and J. Timbie (2006). Impact of pay-for-performance contracts and network registry on diabetes and asthma hedis® measures in an integrated delivery network. *Medical Care Research and Review 63*(1 suppl), 14S–28S.

Leyland, A. H. and F. A. Boddy (1998). League tables and acute myocardial infarction. *The Lancet 351*(9102), 555–558.

Leyland, A. H. and H. Goldstein (2001). *Multilevel modelling of health statistics*. Wiley.

Lilford, R., M. A. Mohammed, D. Spiegelhalter, and R. Thomson (2004). Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *The Lancet 363*(9415), 1147–1154.

Lindenauer, P. K., D. Remus, S. Roman, M. B. Rothberg, E. M. Benjamin, A. Ma, and D. W. Bratzler (2007). Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medicine 356*(5), 486–496.

Lomi, A., D. Mascia, D. Q. Vu, F. Pallotti, G. Conaldi, and T. J. Iwashyna (2014). Quality of care and interhospital collaboration: a study of patient transfers in Italy. *Medical Care 52*(5), 407.

Lomi, A. and F. Pallotti (2012). Relational collaboration among spatial multipoint competitors. *Social Networks 34*(1), 101–111.

Mackenzie, P. A., E. A. Smith, and P. G. Wallace (1997). Transfer of adults between intensive care units in the United Kingdom: postal survey. *BMJ 314*(7092), 1455.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

Martini, G., P. Berta, J. Mullahy, and G. Vittadini (2014). The effectiveness-efficiency trade-off in health care: The case of hospitals in Lombardy, Italy. *Regional Science and Urban Economics 49*, 217–231.

Mas-Colell, A., M. D. Whinston, J. R. Green, et al. (1995). *Microeconomic theory*, Volume 1. Oxford university press New York.

Mascia, D., F. Angeli, and F. Di Vincenzo (2015). Effect of hospital referral networks on patient readmissions. *Social Science & Medicine 132*, 113–121.

Mascia, D., F. Di Vincenzo, and A. Cicchetti (2012). Dynamic analysis of interhospital collaboration and competition: empirical evidence from an Italian regional health system. *Health Policy 105*(2), 273–281.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

McLachlan, G. J. and D. Peel (2000). *Finite Mixture Models*. New York: John Wiley & Sons.

McLachlan, G. J. and S. Rathnayake (2014). On the number of components in a gaussian mixture model. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 4*(5), 341–355.

Moscone, F., E. Tosetti, and G. Vittadini (2012). Social interaction in patients' hospital choice: evidence from italy. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 175*(2), 453–472.

Muthén, B. and T. Asparouhov (2009). Multilevel regression mixture analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 172*(3), 639–657.

Neuman, M. D., C. Wirtalla, and R. M. Werner (2014). Association between skilled nursing facility quality indicators and hospital readmissions. *Jama 312*(15), 1542–1551.

Normand, S.-L. T., M. E. Glickman, and C. A. Gatsonis (1997). Statistical methods for profiling providers of medical care: issues and applications. *Journal of the American Statistical Association 92*(439), 803–814.

Normand, S.-L. T. and D. M. Shahian (2007). Statistical and clinical aspects of hospital outcomes profiling. *Statistical Science*, 206–226.

O'Hagan, A., T. B. Murphy, L. Scrucca, and I. C. Gormley (2015). Investigation of parameter uncertainty in clustering using a gaussian mixture model via jackknife, bootstrap and weighted likelihood bootstrap. *arXiv preprint arXiv:1510.00551*.

Patterson, D. and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika 58*(3), 545–554.

Pink, G. H., A. D. Brown, M. L. Studer, K. L. Reiter, and P. Leatt (2005). Pay-for-performance in publicly financed healthcare: some international experience and considerations for Canada. *HealthcarePapers 6*(4), 8–26.

Pollack, C. E., G. Weissman, J. Bekelman, K. Liao, and K. Armstrong (2012). Physician social networks and variation in prostate cancer treatment in three cities. *Health Services Research 47*(1pt2), 380–403.

Propper, C., S. Burgess, and D. Gossage (2008). Competition and quality: evidence from the nhs internal market 1991–9. *The Economic Journal 118*(525), 138–170.

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

Region, L. (2011, 6 December). Act 2633: Determinazioni in ordine alla gestione del servizio socio sanitario regionale per l'esercizio 2012.

Region, L. (2012, 23 January). Act 349: Approvazione del metodo per l'individuazione dell'indice sintetico di performance per le strutture di ricovero.

Reilly, S., I. Olier, C. Planner, T. Doran, D. Reeves, D. M. Ashcroft, L. Gask, and E. Kontopantelis (2015). Inequalities in physical comorbidity: a longitudinal comparative cohort study of people with severe mental illness in the uk. *BMJ Open 5*(12).

Rice, N. and A. Leyland (1996). Multilevel models: applications to health data. *Journal of Health Services Research 1*(3), 154–164.

Rosenthal, M. B., R. G. Frank, Z. Li, and A. M. Epstein (2005). Early experience with pay-for-performance: from concept to practice. *Jama 294*(14), 1788–1793.

Rosko, M. D. and R. L. Mutter (2008). Stochastic frontier analysis of hospital inefficiency: a review of empirical issues and an assessment of robustness. *Medical Care Research and Review 65*(2), 131–166.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics 6*(2), 461–464.

Shih, T., L. H. Nicholas, J. R. Thumma, J. D. Birkmeyer, and J. B. Dimick (2014). Does pay-for-performance improve surgical outcomes? an evaluation of phase 2 of the premier hospital quality incentive demonstration. *Annals of surgery 259*(4), 677.

Shinjo, D. and T. Aramaki (2012). Geographic distribution of healthcare resources, healthcare service provision, and patient flow in Japan: a cross sectional study. *Social Science & Medicine 75*(11), 1954–1963.

Siciliani, L. and J. Hurst (2005). Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 oecd countries. *Health policy 72*(2), 201–215.

Siciliani, L. and S. Martin (2007). An empirical analysis of the impact of choice on waiting times. *Health Economics 16*(8), 763–779.

Silva, J. S. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and statistics 88*(4), 641–658.

Sivey, P. (2012). The effect of waiting time and distance on hospital choice for english cataract patients. *Health Economics 21*(4), 444–456.

Snijders, T. A. and R. J. Bosker (2012). *Multilevel analysis* (2nd ed.). London: SAGE Publications.

Spence, M. (1973). Job market signaling. *The quarterly journal of Economics 87*(3), 355–374.

Sutton, M., S. Nikolova, R. Boaden, H. Lester, R. McDonald, and M. Roland (2012). Reduced mortality with hospital pay for performance in England. *New England Journal of Medicine 367*(19), 1821–1828.

Tay, A. (2003). Assessing competition in hospital care markets: the importance of accounting for quality differentiation. *RAND Journal of Economics*, 786–814.

Varkevisser, M., S. A. van der Geest, and F. T. Schut (2012). Do patients choose hospitals with high quality ratings? empirical evidence from the market for angioplasty in the netherlands. *Journal of Health Economics 31*(2), 371–378.

Veinot, T. C., E. A. Bosk, K. Unnikrishnan, and T. J. Iwashyna (2012). Revenue, relationships and routines: The social organization of acute myocardial infarction patient transfers in the United States. *Social Science & mMdicine 75*(10), 1800–1810.

Vermunt, J. K. (2005). Mixed-effects logistic regression models for indirectly observed discrete outcome variables. *Multivariate Behavioral Research 40*(3), 281–301.

Vittadini, G., P. Berta, G. Martini, and G. Callea (2012). The effect of a law limiting upcoding on hospital admissions: evidence from italy. *Empirical economics 42*(2), 563–582.

Vogt, W. B. and R. Town (2006). How has hospital consolidation affected the price and quality of hospital care?

Wedel, M. and W. S. DeSarbo (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification 12*(1), 21–55.

Werner, R. M., J. T. Kolstad, E. A. Stuart, and D. Polsky (2011). The effect of pay-for-performance in hospitals: lessons for quality improvement. *Health Affairs 30*(4), 690–698.

Westert, G. P., A. P. Nieboer, and P. P. Groenewegen (1993). Variation in duration of hospital stay between hospitals and between doctors within hospitals. *Social Science & Medicine 37*(6), 833–839.

Westra, D., F. Angeli, M. Carree, and D. Ruwaard (2017). Understanding competition between healthcare providers: Introducing an intermediary inter-organizational perspective. *Health Policy 121*(2), 149–157.

Westra, D., F. Angeli, E. Jatautaitė, M. Carree, and D. Ruwaard (2016). Understanding specialist sharing: A mixed-method exploration in an increasingly price-competitive hospital market. *Social Science & Medicine 162*, 133–142.

WHO, W. H. O. (2000). *The world health report 2000: health systems: improving performance*. World Health Organization.

Yang, Z., R. Algesheimer, and C. J. Tessone (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports 6*, 30750.

Zou, K. H., A. Liu, A. I. Bandos, L. Ohno-Machado, and H. E. Rockette (2011). *Statistical evaluation of diagnostic performance: topics in ROC analysis*. CRC Press.