Università degli studi di Cagliari

DOTTORATO DI RICERCA

SCIENZE ECONOMICHE ED AZIENDALI

Ciclo XXIX

TITOLO TESI

ADMINISTRATIVE DATA IN HEALTH CARE: EMPIRICAL APPLICATIONS IN THE ITALIAN CONTEXT

Settore/i scientifico disciplinari di afferenza
SECS-P/01, SECS-P/03

Presentata da:	Moro Daniela
Coordinatore Dottorato	Prof. Andrea Melis
Tutor	Prof. Rinaldo Brau

Esame finale anno accademico 2015 – 2016
Tesi discussa nella sessione d'esame marzo – aprile 2017

Abstract

Scientific and economic research based on administrative data has grown considerably over the past two decades. The use of administrative data, in particular hospital discharge data, offered researchers the opportunity to develop new scientific evidence of the determinants of health, as well as evaluate government policies aimed at improving people's health.
This thesis is divided into three chapters. Chapter 1 offers a review of the literature on administrative data, discussing advantages and disadvantages, issues of data quality and the data's research potential. Administrative data from the SDO database are used in the second chapter to examine the determinants affecting patient choice of distant destinations. Discrete choice models reveal that the factors that are most influential to patient hospital choice are distance, number of beds and hospital case mix. In chapter 3, difference-in-difference models are used to evaluate the effect of an Italian regulatory policy that prohibited the night sale of alcohol in motorway service areas. Combining data provided by the Ministry of Health for the years 2007-2013 on admissions for alcohol intoxication and accidents with information about hospital location and socio-economic characteristics of the area, our results suggest the effectiveness of the policy ban in reducing the number of alcohol related hospitalizations when aggregating data at patient municipality of residence.

Nel corso degli ultimi vent'anni l'attività di ricerca scientifica ed economica basata sull'uso di dati amministrativi è cresciuta notevolmente. L'uso dei dati amministrativi, ed in particolare l'uso dei dati ospedalieri, ha offerto ai ricercatori l'opportunità di sviluppare nuove conoscenze ed evidenze scientifiche non solo sulle determinanti della salute, ma anche sulla valutazione dell'efficacia di politiche governative volte al miglioramento della salute degli individui. Il lavoro di ricerca si articola in tre capitoli. Nel primo capitolo si effettua una ricognizione della letteratura scientifica che si è sviluppata a seguito della diffusione dei dati amministrativi. Nel dettaglio, si evidenziano vantaggi e svantaggi associati all'uso dei dati amministrativi, si discutono le questioni concernenti la qualità del dato nonché il potenziale di ricerca derivante dall'incrocio delle informazioni amministrative con quelle provenienti da altre banche dati (ad esempio survey, census etc). Nel secondo capitolo, sfruttando le informazioni amministrative provenienti dalla banca dati SDO, si approfondiscono le determinanti che incidono sulla scelta del paziente verso destinazioni lontane. Attraverso l'uso di modelli econometrici che consentono di valutare le differenze nei comportamenti di scelta dei pazienti, si evince che i fattori che incidono maggiormente sulla scelta ospedaliera sono la distanza, il numero dei posti letto e la complessità dei casi trattati nelle strutture ospedaliere. Infine, nel terzo capitolo, la valutazione dell'efficacia della politica italiana di regolazione sulla vendita di alcolici nelle aree di servizio autostradale nelle ore notturne avviene attraverso la stima di modelli *difference-in-differences*. Unendo i dati sui ricoveri per intossicazioni alcoliche e incidenti stradali, forniti dal Ministero della Salute per gli anni 2007-2013, con informazioni relative alla localizzazione degli ospedali e le caratteristiche socio-economiche del territorio, i risultati delle stime mostrano che esiste un'evidenza circa l'efficacia della politica di divieto nella riduzione del numero di ricoveri ospedalieri associati all'uso di alcol quando si considerano i dati a livello di comune di residenza del paziente.

Contents

List of tables

List of Figures

To my family

Introduction

Over the last three decades increased use of administrative data through improved access and data linkage have become one of the most cost-effective way to support empirical research. Initially recorded for financial purposes, for both government and commercial payers, these electronic records, which contains a huge amount of information on patients (i.e. age, gender, zip code) and hospitalizations (main diagnoses, procedures, date of admission and discharge, etc.), have started to be used in a broad range of applications.

Despite administrative data are collected almost by all states and other level of government worldwide, the quality and the quantity of information recorded vary among countries with differences in the completeness of the data, the ways administrative data are disseminated, and data disclosure policies. Administrative data offer to government agencies, provider associations, policymakers and researchers new opportunities to produce official statistics and to provide new insights into social inequalities, evaluation of policy reforms and human behaviours that would have never otherwise been possible.

In light of these considerations, this thesis analyses the usefulness of hospital administrative data to perform economic research. Specifically, this dissertation contributes to the empirical literature that examines the impact of patients' behaviours on hospital choice. It also exploits the potential of administrative data to evaluate a policy change on alcohol night sale and assess its impact on population health.

Chapter 1 provides an overview of the characteristics of administrative data along with the opportunities they may offer to understand human behaviour and inform policy choices. First, it provides a discussion on the evolution of administrative data from financial to empirical tool for research along with a focus on their advantages and disadvantages. Coverage and availability of data are included among the strengths, while incomplete samples and inaccuracies are described among limitation. Second, this chapter investigates the potential and the research opportunities associated to data matching: linking data have been found to be useful to conduct a wide range of study design, such as interrupted time series, natural experiments, long term follows up of surveys and research studies, costing and economic evaluation. Third, it proposes a brief outline of the main hospital administrative data available in OECD countries together with a description of their differences and similarities. Finally, evidence from the economic literature will also be presented.

In Chapter 2, data from the Italian hospital discharge records and data related to the characteristics and outcomes of the Italian hospitals are used to analyse patient hospital choices in the particularly case of Sardinia.

Extensive research has used patient level data to analyse the determinants of patient hospital choices. The phenomenon of free patient choice, also intended as a manifestation of individual freedom of choice of a hospital structure within the national territory, constitutes a central argument on the management and organisation of healthcare systems. The relevance of patient choice may be traced in several factors. On the one hand, free patient choice gives an incentive to provider to compete in quality, especially in a context of fixed prices. On the other hand, the argument is of interest because of the financial implication associated with patient extra-regional choices for regions with net passive mobility, since they are charged of the compensation of hospital treatment outside the region while incurring in the fixed costs needed for ensuring health-care services to the residents under their RHS. In this framework, chapter 2 focuses on the Sardinia region, and specifically investigates the determinants of hospital attractiveness once the distance hurdle has been overcome by patients. Thus, patient hospital choices among alternative hospitals services are modelled in the study by means of conditional and mixed logit models.

In Chapter 3, the effectiveness of a policy ban which has prohibited the on-trade sale of spirits on motorway services areas has been investigated.

It is well known that alcohol abuse affect people well being in several ways. For instance, the World Health organization has repeatedly documented that alcohol abuse is responsible for people' disabilities and deaths. Besides the economic values related to the passing away of young lives, alcohol consumption is also a source of negative externalities contributing to the deterioration of working performance, absenteeism, crimes, motor vehicle accidents and hospitalizations. As a way to reduce the negative consequences linked to alcohol abuse, in Italy, several regulation policies have been adopted. In this chapter, linking patient level data from the Italian Ministry of Health for the period 2007-2013 with information regarding hospitals locations and the socio-economic characteristics of the surrounding areas, the effectiveness of the above mentioned policy ban, which has prohibited the on-trade sale of spirits on motorway services areas between 10 pm and 6 am since 10th July 2010, has been investigated. Exploiting the heterogeneity of the location of hospitals (municipalities), which are partly contiguous to Italian motorways, and partly located in geographical areas without motorways (which can be viewed as the control group), the effect of the policy ban on hospitalizations related to car accidents and alcohol intoxication has been analysed by means of difference-in-differences methods and "*exposure*" approaches.

.

Chapter 1

Administrative data in health economic research

According to The American Statistical Association (1977), administrative data are "*collected and maintained for the purpose of taking action on or controlling actions of an individual person or other entity*." Similarly, administrative datasets in health care are commonly designed to collect and store patient information through the work of healthcare providers. These data, which are mainly used for financial purposes and for the reimbursement of health services, have become the subject of interest of policymakers, healthcare insurers, and public health and economics researchers. They represent a powerful source of information. In fact, they may provide information on inpatients hospitalizations over time, major patient diagnoses and procedures, trauma, outpatient care, rehabilitation services, etc.

The nature of these data allows researchers to elaborate more accurate and up-to-date health economics statistics, generate opportunities for experimental studies, test hypotheses and to measure economic effects and outcomes (Einav and Levin 2014). When combined with other types of health data (e.g. survey data, medical records), administrative data may offer new research opportunities in economics.

This chapter's analysis will concentrate on the usefulness of administrative health data to empirical research, focusing on key aspects of assembling different types of health records that are appropriate for empirical research. More specifically, the chapter provides information on administrative data and a description of the main data sources of empirical research. The next two sections review empirical economics literature, which incorporates administrative health data, to analyse not only health outcomes but also the effects of governmental policies on patient health.

1.1 Overview of administrative data

1.1.1 Administrative data: from a financial to an empirical tool

The term administrative data is traditionally used to describe records that depend on the operation of administrative systems, especially in public agency sectors (Elias, 2014). More specifically, administrative data have been defined by Woollard (2014) as those information gathered with the objective of registration, transaction and record keeping. "*Health data*", "*administrative health records*" and "*administrative claims data*" are generally used as synonyms of "*health care administrative data*" (Cadarette and Wong 2015).

Within several OECD countries, administrative data have been collected since the adoption of Diagnosis Related Groups (DRG) in the 1980s (Moise 2001).[1] The adoption of the DRGs system made it possible to classify a set of patients with similar diseases into groups based on common required resources. Since then, medical records have been used to monitor hospital activities, allowing administrators to evaluate the volume of hospital production and also estimate hospital costs. Although administrative data contain limited clinical information (i.e. demographic characteristics, main patients diagnoses and procedures codes), these data are more often used as starting point to derive *structure* (health equipment and facilities), *process* (the set of mechanisms/steps within a process who lead to a specific outcome) and *outcome* (improvement or worsening of patient health status) indicators. Usually employed to evaluate hospital care provision, these indicators may constitute a useful framework to perform *ad hoc* studies once specific critical issues have been defined. Several

[1] Countries as Scandinavia and United Kingdom even though do not use DRG systems, have large national registries or large national administrative databases, which record diagnoses and treatments for the entire population.

report cards, which consist on annual reports to provide consumers information on hospital quality, and physician practice profiles, which permit comparison among physicians in terms of resource consumption, are derived from administrative data. However, according to Iezzoni et al. (1997) the lack of exhaustive clinical information and the financial nature of these data, might compromise the opportunity to derive accurate quality assessment from administrative data.

Considering the amount of information that administrative health data can provide, researchers took note of the data source's potential and began using them to conduct robust clinical, policy and economics analyses.

1.1.2 Discussion of the advantages and disadvantages of administrative data

There are advantages and disadvantages associated with the use of administrative data. A particularly strong point of hospital data is the completeness of inpatient information, which makes those records more accurate and reliable than other sources of data. Providers directly collect information about diagnoses and procedures, details on individual characteristics and the types of services received, which leads many to consider this data source to be more accurate than others. Another advantage of health data is the ability to represent the entire population of interest. In fact, current administrative datasets contain millions of records and, as in the case of hospital data, can often enclose the entire universe of patients. It can be argued that such a large sample size makes it possible to analyse clinical events in small subsamples as well as to implement novel economics methodologies (e.g. experimental approaches) that allow for precise predictions and estimate results. Because discharge data are regularly updated, they are usually accessible for a long period of time. Another key aspect of hospital discharge records is the longitudinal structure of observational data. This allows researchers to follow patients over time and to exploit changes in outcomes (Card et al. 2010) or the effectiveness of state programs. In addition, health data allow avoidable medical costs to be evaded, reducing the burden of data processing (e.g. discharge data are more precise) and improving administrative efficiencies. Last, but certainly not least, administrative data help to reduce sample errors and to enhance external data validity. Indeed, large sample sizes can help reduce selection bias that might affect estimates' validity (Nguyen and Barshes 2010, Mazzali and Duca 2015).

However, as mentioned before, administrative data also have limitations, for example, incomplete samples and inaccuracies in medical record reporting. The quality of data is compromised if the data are underreported or fail to correctly identify procedures codes or a diagnosis. Moreover, because discharge data only report those events occurring within a hospital ward, another disadvantage of administrative health data is that they do not include records of those patients who pass away before being admitted to a hospital. In this case, especially in clinical studies, the possibility of underestimating the phenomenon object of analysis present a limitation. Finally, because of data sensitivity and privacy protection, data access is strictly regulated by governmental institution, thus limiting the range of variables accessible to researchers.

1.2 Traditional surveys records and administrative data

Administrative data began to be collected in the 1980s for financial purposes and only recently have become the object of interest of researchers conducting empirical studies. Together with survey data, administrative data sources have had an important effect on both the quality and quantity of empirical research conducted in health economics in recent years. Although administrative data are the most important tool, the combination of administrative and survey data offers researchers the greatest potential for empirical analysis.

A social survey is defined as "*an investigation about the characteristics of a given population by means of collecting data from a sample of that population and estimating their characteristics through the systematic use of statistical methodology*" (OECD, 2000). Commonly designed to answer at specific research questions, surveys are used to measure population characteristics and, in the case of the health sector, to answer questions related to patient health risks, health equity and health system performance.

Different methods may be used to collect survey data. In-person, telephone and computer-assisted interviews, along with online and mailed questionnaires, are common examples of survey research methods. In-person and telephone interviews are usually employed to obtain information from small groups of people on a wide range of topics. The advantages of gathering information through these two methods are screening accuracy (i.e. respondents cannot give incorrect information about themselves regarding age, sex or race), the capture of respondent discomfort with sensitive matters, maintenance of interviewee focus and avoidance of errors that may arise from distractions. However, administered interviews are usually costly and time consuming and require all interviewers to be trained on the object of the study. On the contrary, self-administered interviews, such as mailed and online questionnaires, are generally employed when a large group of subjects in different geographical areas must be reached. These methods allow researchers to obtain a broad range of information about the population and to ensure the privacy of interviewees. Moreover, the low cost of data collection allows researchers to increase the probability of obtaining questionnaire responses relative to others interview methods. Unfortunately, self-administered surveys are less accurate than administered interviews, as they do not allow for probing questions or clarifications.

Despite the amount of information provided by survey data, administrative health data have been proven to be more reliable and accurate than those obtained through social surveys

(Jones 2007). To understand the reasons for the success of administrative data in empirical researches, it is useful to compare, albeit briefly, administrative and survey data. Table 1.1 below summarizes the differences and similarities of the two data sources.

Table 1.1 Characteristics of survey and administrative data

Social Surveys	**Administrative data**
Data are collected to answer different research questions	Data are gathered for financial purposes
Data may be very large and complex	Data may be large and complex
Limited time span	Long time span
Highly systematic	Semi-systematic
Known sample/population	Data may be cluttered (i.e. it may require data management to clean the dataset from missing or wrong data)
	Multidimensional (i.e. it may entail several piece of data which have to be brought together through data linkage)
	Generally and known sample/population

Source: Connelly et al. (2016)

The first difference relates to purpose: while surveys are generally conducted to answer at particular set of research questions and target specific population and factors of interest, administrative data are mainly collected to fulfil a financial need rather than for research purposes. For these reasons, administrative data are less expensive and require fewer trained staff to handle health information. Second, surveys data may suffer from problems related to the sample size (e.g. limited time span). In fact, survey data are not intended to follow trends over time. It is well known that survey data are not helpful in analysing changes in the population over time unless two or more surveys are performed in different years. On the contrary, one reason administrative data are appealing is because of the ability to provide information over a long period of time, which allows the researcher to observe changes in outcomes over time. Third, claims data generally provide more accurate coverage of the target populations by eliminating the typical non-response rate of survey data, which depend on the respondents' decisions to answer. Similarly, administrative data may suffer from measurement error problems caused by missing data or coding errors. Fourth, while administrative data are generally used to offer researchers a logical framework to monitor specific areas of problems, survey data usually complement these data offering more in-depth information on the phenomenon being studied.

Although these differences, survey and administrative data linkage may offer new research perspectives and opportunities for enhancing the availability of information about the clinical and socio-economic characteristics of the population.

1.3 Linked databases. What potential?

1.3.1 Basics of linked data

The research value of claims data can be further enriched by matching data from multiple sources concerning the same entity. For example, linking detailed information from social surveys with reliable administrative datasets gives academics the opportunity to address new research questions that otherwise may not have been possible. Specifically, record linkage offers the advantage of increasing the number of explanatory factors and supplementing the information available from administrative data with those provided by surveys, allowing researchers to offer better descriptions of economics phenomena (Kunn 2015). Considering the importance that data linkage is acquiring in the research sector, it is useful to explain what the data linkage technique consists of and its potential for research and policy evaluation.

According to Winglee et al. (2005), record linkage is "a *process of pairing records from two files and trying to select the pairs that belong to the same entity*". Every time a patient comes into contact with certain health services, for example, when they visit a GP, receive medical treatments in an emergency department or are admitted to a hospital, new information is collected. Taken alone, those data do not permit a patient's continuous health pathway to be evaluated. However, when a personal identifier is available, it is feasible to add together information from different data sources, thereby improving data quality, re-employing existing information for new studies and reducing the costs associated with data collection. In the health sector, for example, linked data may allow researchers to explore a wide range of effects, such as the impact of birth weight on future education scores, that would not have been possible in the past due to a lack of data linkage procedures.

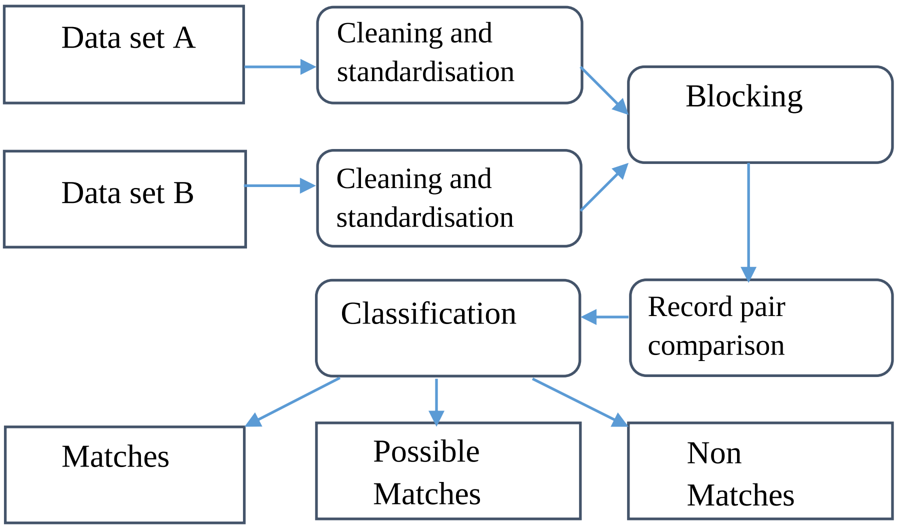
Administrative data are well suited for linkages to different datasets. For examples, claims data can be matched to data from other administrative sources (e.g. information on hospital admissions and data from death registries) or survey data or combined with aggregate data (e.g. hospital records to unemployment data). However, it is not a simple procedure to link records from different data sources belonging to the same person. Thus, data linkage requires both knowledge of the datasets to be linked and thorough expertise in statistics and programming in order to define the appropriate methods for combining records and reducing data errors (Bradley et al. 2010).

Data linkage is easier when all data sources employ a unique personal identifier number. A personal identifier is considered perfect when it is unique and permanent. In countries such as Denmark, Finland and Norway a unique personal identifier is given to each person at birth (Lunde 1975). Since 1947, the National Tax Board of Sweden has assigned a personal identity number (PIN) to each individual. In this country, the PIN is commonly used as a key variable that follows a patient and their medical records over time and matches patient data with different registers such as “*The patient registers*”, the “*Cancer Register*” and the “*Cause of Death register* “(Ludvigsson et al. 2009).

Several methods can be used to link administrative records to other sources of data and generate more comprehensive and reliable datasets for analysis. Figure 1.1 below gives an overview of a record linkage process. Administrative data may suffer from missing and incorrect information. Therefore, when two datasets (A and B) are to be linked, data cleaning and standardization procedures are the first step of an appropriate data linkage. When linking datasets with thousands of records, the number of possible required matches increases. Thus, in order to limit the number of pairs comparisons, "*traditional data linkage techniques employ a blocking variable (one or a combination of record characteristics) to partition the dataset into different blocks*" (Christen and Goiser 2007). All data with the same blocking variable value are placed into the same block, allowing for comparison among only the records within the same block (Cristen and Goiser 2007). Different iterations within the blocking variable are performed in order to avoid possible errors caused by incorrect coding procedures. Each recorded pair resulting from the blocking process is then compared using different field values. These fields are compared to return the matching weight, which is then used to classify records as matches, possible matches and non-matches.

There are two main types of data linkage methods: deterministic and probabilistic. Deterministic linkage entails a perfect match between unique variables (e.g. personal identity number). This means that in the presence of errors (i.e. incorrect coding of personal identifier), changes in names (e.g. the presence in a database of a maiden name and of a married name in the other), or other discrepancies in key data fields, it will not be possible to obtain perfect matches. Because unique variables are not present in all datasets, probabilistic data linkage (also called *fuzzy matching*) uses multiple variables (i.e. last name/first name, birth data, demographic information, medical information, etc) to properly defined a match or a non-match. Thus, according to Christen and Goiser (2007) "*pairs of records are considered as matches if their common attributes predominantly agree, or as non-matches if they predominantly disagree.*"

Figure 1.1 General linkage process



Source: Christen and Goiser (2007)

1.3.2 The benefits of data linkage

Data linkage represents an inestimable tool for providing new insights into population health. It can help researchers improve the value of existing data at a low cost without the need for new data collection. Once the data linkages have been realized, in fact, there is no need to again incur the linkage cost for each new research project (Chamberlayne et al. 1998). In a time of economic hardship, when research money is limited, relying on existing data to realise scientific research is important for ensuring population health and the quality of care while also keeping costs low. This means that re-using existing information could be a more cost-effective way to inform policy evaluation.

Linking data from different sources also provides other benefits. For example, data linkage helps researchers to reduce the duration of projects and policy research and to improve data quality and integrity (Scottish Government 2010). In their paper, Kemm et al. (2010) discuss the invaluable benefits of health service and social record linkage to not only support policy evaluation, but also health knowledge and data quality. Importantly, data linkage allows researchers to use a wide range of study designs when conducting their research: interrupted time series, comparison of exposed and unexposed, long-term follow up of surveys and research studies, cost and economic evaluation.

1.3.3 Issues with data linkage

As underlined above, linking data from several sources has enabled researchers to examine a wide range of phenomena. However, data linkage is not a panacea: matching records via personal identifiers may raise problems of data confidentiality, privacy and data security. In fact, according to Flowers and Ferguson (2010, p.276) "*data linkage is thought to increase the likelihood of disclosure of individuals…and there are considerable concerns over confidentiality issues and data ownership. Paranoia about security lapses has not helped and there is bureaucracy around data access, such that it is becoming increasingly difficult to access any patient data without prior consent*".

Several options have been identified in order to protect confidentiality in health research. Informed consent is one of them. Mason and Laurie (2010) claim that only through the use of informed consent is it possible to legitimate medical research and limit the problem of data confidentiality. However, for researchers carrying out longitudinal studies using health data, it is difficult or sometimes impossible to obtain consent from each patient (Brett and Deary, 2014). In the case of hospital discharge data, for example, obtaining informed consent from a large number of persons would be too costly and time consuming (Regidor 2004). Other techniques may be employed when it becomes impossible to obtain patients' informed consents. Data anonymisation is a common technique used in large datasets to protect patient information, avoid information disclosure and allow governmental institutions to share data without patient permission. Data anonymisation may be obtained thought encryption, *k*-anonymity, generalization and perturbation. Data encryption ensures data quality and integrity, which means the dataset's structure is not modified before the analysis. Sweeney (2002) first addressed the problem of data protection through the concept of *k*-anonymity. A dataset has the *k*-anonymity property when "*every combination of values of quasi-identifiers can be indistinctly matched to at least k respondents*" (Ciriani et al. 2007). One of the way *k*-anonymity may be achieved is through the generalization technique, which removes data specificity and protect individual information by creating summary data.

In summary, problems related to data confidentiality, privacy and data security can be overcome through the use of coded information and unlinked data. Linkage processes create only a low and easily manageable risk of data disclosure.

1.4 Administrative datasets in OECD countries

In most OECD countries, claims data are gathered primarily for financial purposes rather than for research. Table 1.2 provides an overview of the main health administrative data sources used to perform empirical studies. It should be noted that, in the interest of selectivity, we consider only the main sources of information used in economics research in OECD countries. Thus, the following will be a description of the characteristics of the most relevant database along with their strength and limitations.

Medicare claims data

One of the most important inpatients source of information in the USA is the Medicare dataset. Medicare is a federal health insurance program that provides medical services to all Americans citizens age 65 or older. That said, Medicare dataset provides administrative health records for all aspects of care among all patient beneficiaries of health insurance across the United States. Specifically, Medicare dataset collect information on patient main diagnosis, medical procedures, length of stay, age, sex, DRGs from all hospitals in which patients' beneficiary of Medicare insurance have been admitted. Mortality rates and hospital utilizations are also recorded in this dataset. There are several strengths associated to the use of this source of data. For instance, it provides an almost complete coverage of hospitalizations of adults age 65 or older (98%), making it one of the most important source of health information in US. It allows researchers to conduct epidemiological and empirical studies at low costs and it permits data linking with external data sources such as US census, cancer registries (e.g. SEER/Medicare) and surveys (e.g. Health and Retirement Study, Panel study of Income Dynamics, National Health Interview Surey). However, there are also some limitations. As example, Medicare dataset gives information only on acute and diseases care omitting both long term and home care information (National Institute on Aging, National Institutes of health). In addition, diseases like dementia and diabetes are often problematic to be recognized and for these reasons often under-diagnosed.

Hospital Episode Statistics

The Hospital Episode Statistics (HES) may be dated back to 1987, when administrative health data began to be recorded with the aim to monitor hospital activity. Nowadays, HES constitutes a rich source of information about all patients' admissions, outpatient appointments and Accident and Emergency attendances (A&E) in all NHS hospitals in England. The primary goal of HES is to allow hospital financial remuneration. However, as other administrative data sources, HES are commonly used by researchers, providers' organizations and patients as a tool to assess the quality of clinical services and to provide healthcare analyses. In fact, HES contains information on patients (e.g. age, sex, NHS number, zip code), hospitalization type (e.g. NHS trust, admission and discharge data, method and source of admission, discharge method) and medical aspects (e.g. diagnoses and main procedures).

The main strength of HES is the possibility to be linked with other datasets through a unique identifier (i.e. the NHS number, year of birth, etc.) so that, as examples, readmission and fatality rates as well as chronic diseases can be traced. On the contrary, weaknesses of the current dataset are represented by coding errors, the lack of data on primary care and by the limitation caused by the absence of data coming from other sources than hospital trusts in England (e.g. data of Walsh patient are collected in Patient Episode Database Walsh) (Garrat et al. 2010). Because of the above mentioned limitations, other datasets have been generated for clinical, research and quality purposes.

The Danish National Patient Register

Conceived in 1977, the Danish National Patient Register is considered to be one of the oldest and finest datasets of its type (Lynge et al. 2011). Actually, the NPR recorded somatic and psychiatric outpatient and inpatients admission in Danish hospital departments. Each record contains personal characteristic of patient (e.g. the unique personal identification number, patient age, gender, place of residence), administrative (e.g. date of admission and discharge, length of stay) and clinical information (e.g. diagnoses and main procedures).

Commonly used to monitor hospital and health service utilization, NPR constitute a useful tool to monitor chronic disease, to assess quality and to perform scientific research (Schmidt et al. 2015).

The main strength of these dataset is the possibility to be linked through unique person identification number to other source of information in order to produce long term follow up spanning birth to death. However, some limitations may rise with respect to errors in coding diagnoses (e.g. using wrong ICD-10 code to describe a disease) or by typing, which lead to a decrease in the completeness, quality and validity of the dataset information.

Analogies and differences

Considered the main OECD data sources used by researchers to conduct public health and economic analyses, Table 1.2 reveals that almost all datasets exhibit information about patients' demographic characteristics of (e.g., age, sex, place of residence), administrative aspects (e.g., date of admission and discharge) and clinical information (e.g., main diagnosis, secondary diagnosis and procedures). Looking at the table it seems clear that despite the amount of information administrative data can provide, the quality and quantity of database information may vary among countries. For example, while hospital administrative data in Sweden and Denmark may include information on the entire population, in the USA, the collection of claims data is limited to specific age groups (i.e. Medicare beneficiaries aged 65 years or older), income groups (i.e. patients with limited resources - Medicaid program) or members of private health insurances (Lichtman et al. 2015, Schmidt et al 2015). In other countries, data collection may be limited to public hospitals or to specific provinces. One example is provided by Canada, where only several provinces submit hospital data through their health ministries.

A patient personal identifier is recorded in almost all data sources.[2] In France a personal medical record is assigned to all patient, who is given the possibility to hide the information he/she does not want to disclose (Quantin et al. 2008). Because of the sensitivity nature of health data, in France the chance to link patient level data with information contained in others database is still not exploited.

[2] Patient identifier consists on an alphanumeric code, which allow researchers and clinicians to link administrative data sets with information coming from medical datasets or other sources of data.

In many Nordic countries, such as Denmark, Finland, Belgium, the Netherlands and the United Kingdom, people are given a homogeneous personal identification number that it may be used in other fields than the purely health care sector. For instance, in Denmark, administrative health data are usually matched with information contained in other administrative datasets (e.g. the Cause of Death Registry, the Danish National Registry of Births), clinical registries (e.g. Danish registries for studies of medical genetic diseases), population surveys and epidemiologic field studies (Schmith et al. 2015). In Netherland, a personal record is assigned to everyone and it is commonly used to help government to perform public tasks. Example of data linkage, in this country, are the combination of the population register (e.g. the Basisregistratie Personen) and employer register (e.g. Werknemersbestand); the Netherlands Twin Register (NTR) with the Achmea Health Database (AHD) and the Koala cohort with information on pharmacies contained in the Stichting Farmaceutische Kengetallen (SFK) (van Grootheest et al. 2015). However, in Netherland, as in the case of the SFK dataset, sometimes personal records are anonymised to protect patient identity, thus compromising the process of data linkage.

In other states, such as in Australia it is a common practice link administrative data with information from Medicare Claims Database, while in Quebec, researcher usually link these data with information coming from RAMQ physician claims database. Using patient postal code, in Canada, administrative data may also be linked to census data.

It is clear that a common issue among OECD countries is related to data confidentiality and the potential inappropriate use of personal information that may affect patients or their families. National laws along with international agreements protect patient privacy which in turn make it difficult sharing information and have access to patient level information.

In light of all these considerations, the use of administrative data to perform both epidemiological and economics analysis requires caution.

Table 1.2 Main hospital administrative data sources

Country	Administrative data source	Years	Coverage	Linkage capacity	Individual and socio-economic characteristics
Australia	National Hospital morbidity Database (NHMD)	1993-94 to 2012-13	Hospitalizations in both public and private acute care hospitals	Personal id number is collected	Demographic data (i.e. sex, date of birth, age group, indigenous status, area of residence of patient), Length of stay (admission and leave day), Clinical and related data (e.g. principal and secondary diagnosis, MDC, admission mode, external causes of injury)
Australia	Hospital Morbidity Data system	1970-2016	All inpatient discharge summary data from all public and private hospitals in Western Australia.	Database has been linked to the Medicare Claims Database	Demographic data (i.e. indigenous status, marital status), payment classification
Canada	Discharge Abstract Database (DAD)	1994-95 to 2015-16	All acute inpatient discharges recorded in all provinces with the exception of Quebec	It is possible linked the database with others at national level	Demographic information and clinical data
Canada	MED-ECHO	1987-onward	All acute care hospitalization in Quebec	It can be linked with the RAMQ physicians claims data	Demographic data (e.g. age, sex, area of residence), Length of stay (e.g. admissions and leave day), clinical information (e.g. principal and secondary diagnose). Additional information may be obtained by linking through the use od postal code MED-ECHO data with census data
Canada	Manitoba Health Research Data Repository	1985-1999	Hospital discharge data in Manitoba	It may be linked with Office of vital statistics, Manitoba cancer treatments and Research foundation	Demographic (e.g. gender, postal code) length of stay, clinical information (e.g. procedure codes, main diagnoses)
USA	Medicare claim data	1984-onward	All acute care discharges for people aged 65 or more		Demographic characteristics (e.g. age, sex, postal code), Length of stay, clinical diagnosis

Country	Administrative data source	Years	Coverage	Linkage capacity	Individual and socio-economic characteristics
Finland	Hospital Discharge Register	1969-1993	Data for all patients discharged in Finland	Yes	Patient characteristics (e.g. identity number, municipality of residence), Waiting list entry date and date of admission, Length of stay , Clinical information (e.g. diagnosis)
Finland	Care Register for Health care	1994-	Data for all patients discharged in Finland	Yes	Patient characteristics (e.g. identity number, municipality of residence), Waiting list entry date and date of admission, Length of stay , Clinical information (e.g. diagnosis)
France	Programme de médicalisation des systèmes d'information (PMSI)	1994 onward	All discharges in both public and private hospitals	Yes. Possible at patient level with SNIR-AM (ambulatory care claims data) and ESPS survey	Patient characteristics (e.g. birth data, sex, zip code), Length of stay, Medical information (e.g. diagnoses)
Denmark	National Patient Register	1977- onward	Includes all hospitalization from Danish public hospital	Yes. Patient identifier may be linked with information contained in the Cause of Death Register	Patient characteristics (e.g. age, sex, place of residence), Clinical information
GBR	Hospital Episode Statistics (HES)	1989-90 onward	Includes information on admissions, outpatient visits and A&E attendances		Patient characteristics (e.g. age, sex, place of residence), Clinical information
Sweden	National Patient Register (NPR)	1987-	Includes all patients discharged from public hospitals in Sweden with the exception of private facilities	Yes. Patient identifier may be linked with population registers	Patient characteristics (e.g. personal registration number, gender, age, place of residence), administrative data (e.g. data of admission and discharge), Medical data (e.g. main diagnosis, secondary diagnosis and procedures)

Note Moise (2001)

1.4.1 Experience from Italy

As in other countries (i.e., the UK, Denmark, Sweden), Italy makes extensive use of administrative health records not only for financial purposes but also for a wide range of scientific studies. The earliest and perhaps most used administrative health database in Italy is the SDO ("*Scheda di dimissione ospedaliera*").

The SDO was defined by the Decree of the Ministry (D.M.) of Health on December 28, 1991, as an instrument to collect patient discharge information recorded within national territory on a yearly basis, both in public and in private hospitals. It represents an important administrative and clinical tool for not only monitoring hospital activities but also for enabling a wide range of clinical analyses.[3] For these reasons, the SDO has become an integral part of patient medical records and has both clinical and forensic relevance. On July 17, 1992, new Ministerial guidelines were defined in order to facilitate the SDO filing process.[4] The document more clearly defined the function of the SDO, which consists of "*the systematic collection for each case treated, of personal data and clinical information, of the assistance outcomes and the resources used for this purpose, must guarantee hospital network organization at central and regional level as well as the management of all activities at the healthcare institutions level. In addition, it must provide the activation of specific programs to evaluate the use of resources and to deliver quality of care".*

In an institutional framework characterized by changes in the National health system (NHS), to better regulate the procedures for the transmission of the SDO, the Italian Ministry of Health enacted the DM on July 26, 1993, contributing to the evolution of the SDO. The

[3] According to the D.M., the SDO must contain information on the following: the patient's full name, sex, date and municipality of birth, marital status, place and region of residence, citizenship, individual health code, the patient Local Health Authority, type (whether it is ordinary or day hospital) and reason for hospitalization, date/hour and department of admission, type of activity, presence of any traumatism or poisoning, day/department/mode of discharge, the main diagnosis at the moment of patient discharge, concomitant diseases or complications, whether or not the patient has had a surgery, and other interventions and procedures.

[4] Linee guida 17 giugno 1992 "la compilazione, la codifica e la gestione della scheda di dimissione ospedaliera istituita ex DM 28/12/1991"

ministerial decree more clearly regulated the systematic collection of administrative and clinical health information, inviting all hospitals, both public and private, to send at least quarterly patient discharge information to Regional or Autonomous Provinces (in the specific case of Bozen and Trento).[5] Moreover, following organizational changes and the promulgation of Law 724 on December 23, 1994, the SDO became a suitable instrument for obtaining economic remuneration within the system.[6] In fact, the information contained in the SDO provides the basis for the Diagnosis Related Groups (DRG) system, which was created to define hospital treatments (e.g., the type of surgery) and is used to quantify hospital remunerations. Each DRG is associated with a specific tariff calculated by the resources used for the treatment, considering not only direct and general costs but also the length of hospitalization. Each admission is then encoded through software called DRG GROUPER, which processes the information contained in the SDO.

On October 27, 2000, DM 380 reaffirmed the role of the SDO as an instrument for collecting patient information and defined it as "*a concise and faithful representation of the medical records, aimed at allowing the systematic collection of economic and quality information contained in patients' clinical files*." The Ministerial Decree updated the information contained in the SDO, adopting the classification ICD-9-CM (International Classification of Diseases and Clinical Modification), which organizes diseases and traumas into 17 groups based on specific criteria.[7] Each groups is in turn coded with five digits, thus totalling approximately 16,000 codes and permitting greater detail in the disease descriptions. This structured classification of diseases coupled with defined therapeutic and diagnostic procedures makes it possible to uniformly describe, throughout the entire national territory, the conditions determining the use of the health care facilities.

5 Hospital institutions, such as nursing homes, mental health hospitals, and gated communities, are not obligated to send patient discharge information to the Regions or Autonomous Province.
6 Misure di razionalizzazione della finanza pubblica
7 The ICD-9-CM is derived from the ICD-9 Classification developed by the World Health Organization.

In order to reduce regional variability in the coding of diagnoses, surgical procedures and therapeutic procedures, the Conference of State and Regions (June 6, 2002) defined general rules for the use of the ICD-9-CM classification. However, considering the complexity of the information collected through the SDO, several control checks were also required to validate the information provided. It is clear that, while important for the evaluation of the effectiveness, appropriateness, equity and efficiency of health care, quality information can only be obtained through a continuous process of coding procedure improvement. These evaluations activities, which are carried out using administrative data collected through information systems, become essential to the governance of regional health systems and clinical hospital departments.

Although the collection of administrative health data in Italy has increased in the last twenty years, concerns related to inaccuracies and data quality are still present. Moreover, not all hospital data are included among collected data items (e.g., outpatient cases). Thus, knowledge about diagnoses and treatments related to an individual patient or a group of patients is limited. When viewed critically, Italy appears to be behind other European countries like Norway, Sweden and Finland in the use of administrative data for empirical research. While data linkage appears to be a widespread and positive practice in Scandinavian countries, in Italy, as in France, the possibility of linking patient-level data with information about chronic disease or socio-economic characteristics is still difficult or impossible. Linking administrative data to information from other data sources requires the use of a common unique personal identifier. However, in Italy, there is no common key code able to identify individuals and follow them across education, employment and tax payment. For these reasons, analyses may suffer from difficulties in capturing full patient pathway information. Encouraging expanded use of administrative data along with the spread of record linkage practices will enable researchers to conduct empirical studies that would not otherwise be possible.

1.5 Administrative data: evidence from the literature

The volume of collected data is growing rapidly as a consequence of the spread of advanced technological tools. Notably, the availability of observational data has impacted the type of research conducted by professionals. Hamermesh (2013) reviews all papers published in the three most important economics journals over the period 1960-2010, finding that most articles published before the 1980s were theoretical or based on "ready-made data." During that time, datasets were scarce and those available only provided information about the same variables for a limited time span, making it impossible for researchers to correctly assess the phenomena of interest. Since then, the ability to observe the same factor over time has contributed to changes in the methodologies adopted by researchers and led to an increase in empirical analyses.

According to Einav and Levin (2014), analytical data sets offer the opportunity to evaluate variations across the population, health, education, wages, productivity and regulation. For example, Piketty and Saez (2014) analysed changes in income and wealth in economically developed countries using tax data. This analysis would not have been possible through the use of survey data, which are not able to properly measure income levels due to the small sample size. On the contrary, tax data provide information for over a century and permit a more accurate analysis of the evolution of inequalities.

Administrative data also play a role in answering questions related to differences in wages and firms' productivity. During the 1980s, according to Abowd et al. (1999, p.251), labour economists "*have lamented the lack of microeconomic data relating characteristics of the firms to characteristics of their workers because such data would permit researchers to begin* *to disentangle the effect of firm-level decisions from the effects of choices made by workers*." However, thanks to the spread of micro data in this area, it is now possible to handle diversity in wages and productivity within firms (Abowd et al 1999, Syverson 2011) as well to evaluate the effect of job displacement on specific socio-economic outcomes such as income and firm earnings. An example is shown in a paper published by VonWatcher et al (2009). The researchers employ a large database of earnings and employment histories of the USA inhabitants over 30 years, to evaluate the long-term effects of job separation on male earnings. When comparing earnings from workers who left their jobs in the 1980s with those who did not experience job loss (control group), they find important losses in annual income following mass layoffs (30%).

The longitudinal structure of administrative data has also been useful to researchers hoping to explain outcomes using data from previous years. An example is shown by Black et al. (2007). Using data on births from the Medical Birth Registry of Norway between 1967 and 1997, researchers combined birth information with those obtained from administrative data on the Norwegian population between the ages of 16 and 74 during the years 1986-2002. They examined the relationship between low birth weight and adult outcomes to find that birth weight plays a role in patients' future outcomes. Other researchers, instead, have used administrative data to perform randomized experiment to monitor outcomes (Einav and Levin 2014). Chetty et al. (2014) analysed the influence of teacher quality on students' achievements. The authors match information on administrative school district data with those on federal income tax data to perform quasi-experimental estimates in order to study the impact of teacher turnover on students' future wages.

Even though their primary goal is financial, administrative data are often useful for documenting quality of care and measuring hospital performance and health outcomes. These indicators, moreover, serve the purpose of supporting patient choice of hospital providers.

For instance, Aakvik and Holmas (2006) have used administrative data from Norwegian municipalities, during the period 1986-2001, to explore the link between the presence of GPs and subsequent health outcomes.

Moreover, claims data serve at the important purpose of supporting patient hospital choice and informing government policies.

1.5.1 Health data, patient choice and quality indicators

The use of administrative health data is fairly widespread phenomenon among researchers today compared to thirty years ago. But then, what makes administrative data so interesting? On the one hand, the use of patient-level data can be explained by their ease of accessibility and lower cost. On the other hand, these data also allow researchers to compare hospitals based on performance and health outcomes. In fact, the particular structure of health data may be used to determine levels of preventable admissions, the complexity of treated health case, the average length of stays, waiting times and readmissions and to evaluate the quality of service offered to consumers. Iezzoni et al. (1994) were the first to assess the value of administrative data in health quality and patient research. In their study, the authors demonstrate how administrative data could be used to inspect health complications that may potentially be avoided through improved care. Considering their level of detail in patient characteristics (i.e. patient lieu of residence and Local Health Authority), those data are used to explain and also analyse the factors driving patient hospital choice. However, describing patient’s choices with only administrative data is not an easy task.

Information and patient choice

Progress in digitizing health data have allowed, over the years, to aggregate an increasing amount of data in electronic databases and making them accessible to professionals of the health-care sectors, to researchers and potential health-care users. The release of this information have generated the question about the ability of patients to use health-care information to make informed hospital choices.

In different European health care systems, patients are enabled to choose their favourite in-country health-care provider under the idea that patients represent themselves best agents to make this type of decision (Wahlstedt and Ekman, 2016). If free patient choice was introduced with the goal of increasing competition between health care providers, it was also introduced to guarantee equity in the provision of health care by allowing patients to choose among different health care providers, both public and private. Administrative data, in this context, constitute an important source of health information, since it is obvious that a patient needs complete information in order to make a wise decision.

According to the neoclassical model, patients in health care markets act as free agents seeking to maximize their own well-being and utility. Perfect competition requires a large number of buyers (e.g., patients) and sellers (e.g., health providers) so that identical products (e.g., health services) are sold at the lowest price. Perfect information about the products is also required in this framework. The idea is that in the presence of a large number of suppliers with no barriers to entry or exit, more competition will generate efficient outcomes. Unfortunately, these conditions rarely exist together in the real world, and products in the market usually have some differences. Consider hip replacement surgery: even though the same service is offered by different health providers, it is plausible to think that the quality of service differs among providers. Thus, each product in a hospital market is unique and has a monopoly power. It should be stressed that obtaining perfect information in health markets is difficult, which suggests that the health sector has informational problems.

Healthcare markets, as with other markets in the economic system, are characterized by high levels of uncertainty. Arrow (1963, p. 951) was the first to discuss the effects of uncertainty and asymmetric information, arguing "*uncertainty as to the quality of the product is perhaps more intense here than in any other important commodity… Further, the amount of uncertainty, measured in terms of utility variability, is certainly much greater for medical care in severe cases than for, say, houses or automobiles.*" On the one hand, product uncertainty depends on providers' inability to predict incidence and patient hospitalizations for a particular disease. On the other hand, uncertainty is also generated by the existence of information asymmetry: one party (usually the seller) has more information than the other party (the buyer). The lack of information and degree of uncertainty lead to the "agency problem". The principal-agent relationship requires the presence of a contract under which one person, the agent (i.e., doctor), is able to make decisions affecting another person, the principal (i.e., patient) (Smith et al. 1997). Furthermore, because of the presence of information asymmetry, the agent has more information than the principal. This allows agents to misreport effort and maximize their own objectives at the expense of patients' interests. Common examples of this relationship include patients and physicians, managers and physicians, and purchasers and providers (Goddard et al. 2000).

Information asymmetry is known to lead to market failure. In healthcare, the market failure is caused by asymmetrical information between providers and consumers. Unlike other markets in which consumers can usually test the product before buying it, it is difficult in health markets to assess the quality of care services beforehand. Sometimes consumers are not conscious of the service's value even after purchasing it. In these circumstances, consumers base their choices on a provider's reputation rather than the product quality (Dixon et al. 2010). When patients are faced with the decision of where to be treated, they often find it difficult to inform their choices with quality information and hospital performance indicators (Boyce et al. 2010). For all of these reasons, patients seem to prefer the advice of friends and parents over public information when making such decisions (Gibbs et al 1996).

The importance of information in decision making is underlined by Lansley (2010) in his first speech as UK Secretary of State for Health: *"patients, they need to know who is providing quality, safe, effective, accessible services. Information will drive higher standards. It's not just about choice, although patients value choice, even if the choice they make continues to be their local practice and their local hospital. But information and choice hold people to account. So our vision will be of an information revolution across the NHS... Putting the information out there – accessible to everyone – is a catalyst. It drives comparison and performance"*(in Boyce et al. 2010). It seems to be quite clear that in a context characterized by uncertainty, patients need accurate information and health literacy to understand information and make appropriate choices. It is possible to mitigate market failure problems by increasing the amount of information available. Hence, through a policy of information transparency, patients may have greater amounts of more easily accessible information that will lead them to make informed hospital choices. Economists and policy-makers believe that patients make optimal decisions when they are sufficiently informed (Boyce et al. 2010), but is more information always better? On the one hand, if the quality of information available to consumers' increases, it may facilitate patients' choice of health providers and encourage hospitals to compete in service quality. On the other hand, the amount of information may negatively affect patients' choices. Because patients in the health market do not act rationally with respect to agents in other markets, there is evidence that, in these circumstances, less information is better. The results from a study conducted by Peter et al. (2007) confirmed the idea that "*less is more*" when patients have to compare performance indicators to make hospital choices. According to Schwartz (2004), having many alternatives to choose from is important for well-being but may have negative effects such as discontent and even choice deferral. In addition, Hibbard et al. (1997) show that consumers must be able to process and use only a small number of variables, yet they always prefer to have more, rather than less, information to sustain their decisions. More knowledge of available alternatives is commonly associated with freedom of choice, *autonomy, self-control and intrinsic motivation* (Zuckerman et al. 1978, Ryan and Deci 2000). For others, more information can lead to quality improvement if and only if performance measures are not publicly available.

Overall, we note that information plays an important role in reducing market failures caused by asymmetric information. Thus, health care institutions should undergo some important changes in order to make information relating to indicators of both process and outcomes, obtained through the actively use of administrative health data, more reliable, accessible and understandable to patients. Hence, understanding how consumers make decisions and enabling them with high quality information are vital to accurate choices.

Competition, quality and patient choice

In many European countries (e.g. Italy, the UK, Sweden, France), patients are free to choose their healthcare provider within their country. The importance of promoting patient free choice rests upon the idea that, in a fixed price regime, greater competition fosters provider competition to attract more patients, which in turns lead to improvements in the quality of health services (Beckert et al. 2012, Brekke et al. 2014, Dixton et al. 2010).

A large number of studies have focused on the impact of hospital competition on quality. Gaynor et al. (2013) evaluate the effect on healthcare quality of a pro-competition policy that enhanced patient choice and promoted hospital competition. The authors explored changes in the hospital market structure and evaluated whether hospitals in less concentrated markets experienced higher quality before and after the introduction of the health system reform in 2006. Using hospital discharge data from the Hospital Episode Statistics (HES), researchers estimated a difference-in-difference model and found strong evidence that under fixed prices, competition leads to improvements in the quality of health services. Similar results were found by Cooper et al in a paper published in 2011. The authors focus on changes in mortality trends for acute myocardial infarction (AMI) to test whether mortality rates in hospitals located in competitive markets declined more than those in hospitals located in less competitive markets after the introduction of patient choice in 2006. The results indicate that AMI mortality decreased more in hospitals located in competitive markets than those in less competitive markets. These results were confirmed in a review of the literature on competition and quality in healthcare markets published by Gaynor and Town in 2012. They found the effect of competition on quality to be positive when prices are regulated. Thus, hospital competition within a market characterized by regulated prices and health indicators information may lead to improvements in the quality of healthcare services.

Related literature has attempted to use administrative data to identify the determinants affecting patient choice of healthcare provider. Empirical studies based on US data have found that quality and geographic distance are good predictors of hospital choice. For instance, Luft et al. (1990) analysed the effects of direct and indirect measures of quality and distance on patient hospital choice in California for seven admission categories during 1983. They showed that hospital choice is negatively affected by patient distance and positively affected by the quality of hospital services. Since Luft et al., many studies have analysed the impact of quality measures and distance on patient choice. Burns and Wholey (1992) analysed the effect of physician characteristics on patient hospital choice. Linking patient discharge data from the Arizona Department of Health Services for the year 1989, which provides information on the admitting physician and hospital characteristics (i.e. diagnoses, outcomes, age, sex, zip code of residence), with information on hospital characteristics from the AHA,[8] the authors found that patient choice is positively affected by hospital quality and

[8] American Hospital Association (AHA 1990).

that patients prefer local hospitals over those that are farther away (i.e. distance has a negative effect on choice). More evidence on the effect of distance and quality on patient choice is provided by Tay (2003). Using Medicare claims data for all patients who suffered a heart attacks, she found that hospital demand is influenced by both quality and distance and that "*increasing different aspects of quality substantially raises the predicted demand of the average hospital.*"

The role of distance and quality as determinants of hospital demand is also confirmed by recent European studies. For instance, Gaynor et al (2012) use data on patient hospital choice for coronary artery bypass graft surgery (CABG) to examine responsiveness of choice to quality in the English NHS before and after the reform of 2006. They find patient hospital choice to be influenced by distance, mortality rates and waiting times. Another study published by Beckert et al. (2012) analysed patients' hospital choice for elective hip replacements in England. The researchers employ a demand model, which allows for patient heterogeneity, to measure how patients evaluate hospital characteristics in making their choices. Combining patient-level data from the Hospital Episode Statistics with information on hospital characteristics for the period 2008-2009, they find that patients are responsiveness to distance and quality. In particular, they show that hospital demand decreases with distance, and that patient hospital choice increases with quality. Santos et al (2016) examined whether the choice of the family doctor in the English National Health Service (NHS) is influenced by differences in the quality of health service provided. Using a dataset on patients and practices obtained by linking NHS administrative datasets with census and socio-economic data, the authors found that patients prefer practices which show higher QOF (Quality and Outcome Framework) scores. In addition, they found that the valuation of practices declines as the distance between patients' homes and the medical practice increases. Similar findings were found in Italy in a study by Moscone et al. (2012) that investigated the effect of social interaction on patient choice. They employed hospital discharge data for all patients admitted with a diagnosis of heart disease in Lombardy over the period 2004-2007, finding a positive correlation between individual choice and people living in the neighbourhood. Their results suggest that distance and clinical quality do matter in hospital choices.

In a highly competitive market such as the the healthcare sector, hospital reputation and amenities matter deeply to consumers who face the decision of which hospital to choose. Varkevisser et al. (2012) used data on readmission rates after heart failure treatment and information on hospitals quality rankings to reveal that Dutch patients are more willing to choose hospitals with lower readmissions rates and good reputations. However, other researchers found amenities to be key factors of patient decisions. As in the airlines sector, hospital amenities such as food, staff care and the surrounding environment are acquiring greater importance (Newhouse 1994). Evidence suggests that several hospitals are investing in luxury amenities to boost patient satisfaction and increase hospital demand. Goldman and Romley (2008) were the first to examine the influence of hospital amenities on patient hospital choice. They used survey data for the area of Los Angeles from The National Research Corporation (NRC), a health marketing firm, to find that an increase of one standard deviation in amenities leads to an increase in hospital demand of approximately 38%. In summary, the empirical findings on the effects of quality and distance on patient choice show evidence that patients prefer to bypass the closest hospital in favour of one that is farther away only if the additional costs caused by distance are compensated by a higher quality of care.

1.5.2 Using administrative data to assess governmental policy

Administrative data constitute a valuable tool not only for informing policy decisions and practices but also for evaluating a wide range of health issues such as public safety, injury safety prevention, public health, disease surveillance, disease registries, health planning, community assessments, public reporting for purchasing and comparative reports, quality assessment, performance improvement, commercial applications and health services (Schoenman et al. 2005 in Andrews 2015). However, thanks to their longitudinal dimension, administrative data are also extensively used to evaluate the effectiveness of previously enacted governmental interventions.

Effectiveness of policy intervention on health outcomes has been discussed widely in the literature. For instance, Taubman et. al (2014) examine the effect of Oregon's health insurance experiment on hospital outcomes using patient-level data from the emergency departments of hospitals in the Portland area. In 2008, Oregon expanded the Medicaid coverage for low-income adults selecting 30,000 individuals by lottery from a waiting list of 90,000 people. Comparing outcome differences between patients in the treatment group with those in the control group, they found an overall increase in the use of emergency departments of about 0.41 visits per person. In the same line of investigation, Farrar et al. (2009) use English hospital episode statistics and Scottish morbidity records for the period 2002-2006 to assess the effect of the introduction of the system of 'payment by results' on changes in outcome variables of volume, cost, and quality of care. Difference-in-difference analysis along with analysis of patient level secondary data with fixed effects models have been used to compare outcomes between hospitals subject to the introduction of the payment by result scheme with a group of hospitals not implementing the intervention across England and Scotland. Their findings suggest a reduction in unit costs following the introduction of the tariff.

Indeed, other researchers have used administrative data to measure the impact of law changes on crimes, traffic accidents and hospital admissions. For instance, Wichi and Gmel (2011) use hospital-level data based on the Swiss Hospital Statistics (SHS) to analyse the effects of a ban on night alcohol sales in petrol stations and video libraries on the number of hospital admissions caused by alcohol misuse. Similarly, Marcus and Siedler (2015) evaluate the impact of a ban on alcohol sales in 2010 in the state of Baden Wuttemberg on alcohol-related hospital admissions in Germany over the period 2007-2011, using data from the German hospital diagnosis statistic to show that the ban results in a 7% reduction in hospitalizations among adolescents and young adults.

1.6 Conclusions

The number of studies using administrative data have significantly increased in the last thirty years. In fact, administrative health data constitutes a powerful source of information that may provide researchers with cheap, reliable and comprehensive information on patients. Data usually includes large sample size and are available for long time span, thus enabling researchers to measure trends and variations in health outcomes and monitor hospital performance over time. Moreover, claims data are well suited to allow researcher to evaluate policy changes, social problems and societal issues.

Unlike Nordic countries, the access to administrative dataset is still limited in many countries. In fact, while in Nordic countries the use of a homogenous personal identification number and data sharing have favoured the use of administrative data for research purposes, in countries like the USA, Canada, United Kingdom, France and Italy there is uncertainty over the extent to which researchers will be gain access to claims data (Connelly et al. 2016).

As administrative data are originally collected for administrative purposes, researchers have to be careful about the quality and accuracy of these data and consider ethical and legal issues that underpin its use. These shortcomings suggest that the optimal approach for research consist on the use of secondary data to answer specific question for which these data are well suited, but turns to primary data for other aspects of the key clinical questions.

Chapter 2

Modelling the individual patient choice: the Sardinian patient case

The Italian National Health System (NHS) was established in 1978 after the introduction of a general reform that ensured the universal coverage of care and replaced a previous system based on separated national health services organizations. Health care was ensured throughout the national territory by a network of Local Health Units ("*Unità Sanitarie Locali*"), Mountain Communities and Municipality organizations. However, the subsequent problems linked to the rising costs of healthcare have required further reforms, initiated in the early 1990s, introducing quasi markets, managerialism as well as free patient choice (Jommi et al. 2001). In that direction, the Legislative Decree (LD) no 517/1993, modifying the LD 502/1992 aiming at the "reorganization of the health care law", confirmed citizens' rights to universal health coverage (as previously defined by the law 833/1978) and introduced changes in the LHU status and organization.

These reforms were accompanied by the regionalization of the NHS, which was organized into twenty-one regional health services (RHS) with significant autonomy in choosing their own organizational model and in setting the mechanism to fund the Local Health Authorities (LHA). Free patient choice all over the country, coupled with regional budget responsibility, has been implemented by means of the Diagnosis Related Group (DRG) system. Each admission episode is figuratively reimbursed by the RHS where the patient is enrolled.

In case of extra-regional mobility, a bilateral compensation system is at work. This adds an “institutional competition” dimension to the “market” competition in which the elementary providers (i.e. hospitals) are involved. Thus, the economic relevance of extra-regional free patient choice is manifold. First, it generates an incentive to providers to foster quality, especially in a context of fixed prices (eg. Besley and Ghatak, 2003). Second, inter-regional mobility may help smaller regions to ensure all health-care services while exploiting economies of scale in the provision of specific health services (Levaggi and Menoncin 2008). Inter-regional patient mobility may represent also a potential source of financial loss for regions with net passive mobility, since they are charged of the compensation of hospital treatment outside the region while incurring in the fixed costs needed for ensuring health-care services to the residents under their RHS.

When the element object of analysis is the patient, the potential determinants of hospital choice and patient mobility refer both to the demand side (namely, patient characteristics such as education level, income and age), and to supply determinants (“structure factors” such as the availability and accessibility of the providers. “Process factors” like the availability of information, continuity of treatment, waiting time and quality of treatment. “Outcome factors”, often in the form of objective indicators such as mortality statistics (Balia et al. 2014)) (see Victoor et al. 2012, for a review).

To the eye of the researcher, most of hospital’s quality in the Italian NHS takes the form of a latent variable, not appearing in the data, but effectively shaping patient’s decisions. The unique dimensions that can be unveiled are constituted by a series of indicators directly monitored by hospital discharge records of the Ministry of Health (such as the case mix index), and a new series of clinical outcomes (such as thirty-day Acute Myocardial Infarction

(AMI)) monitored in the so-called “Programma Nazionale Esiti”(PNE), similarly to other international monitoring programs such as the “NHS Outcomes Framework Indicators” in the UK.

Hospital “choice” is mostly driven by distance and accessibility, although in Italy the choice of a very distant hospital represents an important part of overall inter-regional patient mobility. In this environment, the island of Sardinia represents an excellent case study. By exploiting its peculiar geographical location, which does not experience the typical bordering and neighbouring spillover effects of most European regions, we are able to assess whether patients’ mobility towards distant hospital care constitutes a distinct phenomenon with its own intrinsic characteristics. Moreover, given the above-mentioned geographical location and the related costs associated with the decision to move to a hospital outside the region of residence, we can better appraise how health-care quality affects patients’ decisions.

This chapter is structured as follows. In section 2.1 we review the literature on hospital choice. The following two sections describe respectively the data and key variables and the econometric model we are estimating in our analysis. Finally, section 2.5 contains summary conclusion of the results and a discussion.

2.1 Related Literature and approach

Patients in different European health-care systems are allowed to choose their favourite health-care provider within the country (e.g., Denmark, Netherlands, England, Sweden, Italy). The intuition behind patient choice is that, in a contest of fixed prices, providers of care will compete only in quality, and not in costs, in order to attract more patients (Gaynor et al. 2013, Gaynor and Town 2012).

Individual patient choice within European countries has been modelled in a number of studies, that analysed both structural and quality characteristics. The benefit of using “objective” indicators of healthcare quality is debated. For some scholars there is only evidence of a residual role (Rademakers, Delnoij and de Boer 2011). Selecting a good quality provider is not an easy practice for patients because of difficulties linked to data access and of the patients’ inability to understand quality rank measures and trust them. This suggests that patients are more willing to rely on their experience, that of friends or parents’ advice rather than considering public information. Evidence on this matter is provided by Schenider and Epstein (1998) and by Cutler et al. (2004), who found that only few patients seem to consider quality measures before choosing a hospital. By contrast, the role is important for some others. Studies of patient’s hospital choice determinants in the US have found that hospital demand increases with higher hospital quality. For instance, Luft et al. (1990) evaluate the effect of direct and indirect measures of hospital quality on patient choice in three geographic areas in California in 1983. Using a conditional logit model to analyse the influence of quality on patient choice, they found that patient flows are positively associated with high hospital quality. In the same line of investigation, Tay (2003) analyse the importance of quality differentiation when estimating the effect of competition in a spatially differentiated hospital care market. By means of US cross-section data, Tay exploits the way patients ponder differences in hospital quality when deciding where to be treated in, finding that quality is an important factor of hospital choice. Moreover, Pope (2009), exploiting Medicare data on all patients hospitalized in California and a sample of other bordering hospitals, shows that hospitals that improve quality rankings are able to attract more patients. Similar results have also been found for England, the Netherlands (Varkevisser et al. 2012) and Italy (Moscone et al. 2012). For instance, Beckert et al. (2012) analyse hospital choice using patient level data for hip replacement surgery in England, showing that hospital demand increases with quality. Likewise, Santos et al. (2016) examining whether the choices of family doctor practices in England are affected by differences in the quality of health services, show that patients are more likely to seek care in practices of a higher quality. Several studies have found the existence of a strong association between geographical distance (e.g. travel distance) and patients' hospital choices (Luft et al. 1990, Tay 2003, Varkevisser et al. 2012, Santos et al. 2016). This is also the case of Beukers et al. (2014), when investigating the determinants of patient choice for non-emergency hip-replacement over the years 2008-2010, and Moscone et al. (2012), when analysing the influence of social interaction on patients' choice. In detail, Moscone et al., using data on all patients admitted to hospitals in Lombardy, show that geographical factors, such as the distance from the hospital, are important predictors of hospital choice.

Research has also documented the effect of subjective measures on the choice of hospitals by patients. In a study based on data from the Netherland, for example, Varkevisser, van der Geest and Schut (2012), analysing the relationship between hospital quality and patient hospital choice for angioplasty, find that patients are more likely to choose hospitals with a good reputation. Similarly, in Italy, Messina et al. (2003) show that patients are more willing to travel to receive care when the nearby hospital has a negative reputation. In addition, according to Burge et al. (2004) "*when patient choice is offered, and information on reputation is available, all other things being equal, richer patients will migrate to hospitals with better reputation in greater numbers than poorer patients".*

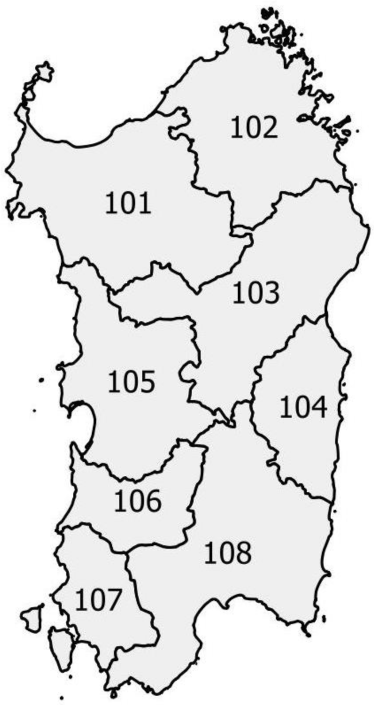
Another strand of literature has found evidence that patients seem to respond positively to specific hospital characteristics. For example, Roh et al. (2008) examining data from patients in Colorado, show that patients are more likely to prefer hospitals with more services and with a larger market share. Likewise, by examining data on Colorado rural patients, Roh and Moon (2005) find that hospital characteristics, such as the number of beds and hospital ownership, are able to affect patient hospital choice.

2.2 Description of the unit of analysis

In order to ensure the provision of the health care services on the whole territory, the health-care system in Sardinia is organized in eight different Local Health Authorities (LHA). As we can see from Figure 2.1 below, the main LHA is No 108 of *Cagliari*, followed by the No 101 of *Sassari*, No 103 of *Nuoro*, No 105 of *Oristano*, No 106 of *Sanluri*, No 107 of *Carbonia* and the smaller of *Lanusei* (No 104) and *Olbia* (No 102).

Health care is delivered in 46 hospitals, of which 71.1% are public, 27.3% are private accredited facilities and 2.3% are both hospital trusts (Aziende ospedaliere) and National Institutes for Scientific Research (IRCSS).

Figure 2.1 Distribution of the Local Health Authorities in Sardinia



Despite patients have the right to choose their favourite health care provider within the country, not only when specific services are not available in the region, the insularity of Sardinia and the distance from the mainland are some of the reasons explaining the scarce choices of extra-regional facilities by Sardinian patients.

2.3 Data and key variables

This analysis uses patient-level data provided by the Ministry of Health, which give us information on all patients enrolled and admitted to any hospital in Italy in the year 2013. Each inpatient hospitalization is classified using patient characteristics such as age, sex, comorbidities and the version 24 of the Diagnostic Related Group (DRG) system, which is defined by a "grouper" program through the international classification of diseases, 9th revision, clinical modification (ICD-9-CM).

In order to concentrate on actual patient choice, we focus on *"elective"* admission episodes by excluding from the database all patient admissions considered as emergencies and for which there is no doctor prescription.[9] A small share of admissions was dropped from the analysis when referring them to the hospital level or to the patient level. In the first case, we excluded those admissions with missing or invalid region codes (origin and destination), or lacking the indication of the ownership type. At the patient level, we dropped admissions without a major diagnostic code and those referred to patients with missing values for sex and age.

To provide an informative analysis of the reasons behind the patient choice of a distant hospital, we kept in the dataset only hospital discharges concerning Sardinian patients. In fact, Sardinia represents an interesting case study: it is isolated from the other Italian regions and not affected by health migration from neighbouring regions, which usually represent an important part of extra-regional health mobility at the national level. Considering the geographical position of the region, its insularity and the distance from the mainland, which make it difficult and expensive for patients to move to hospitals far from their place of residence, it is no coincidence that Sardinia is the Italian region with one of the lowest rates

[9] More specifically, the analysis excluded those hospitalizations falling under MDC 21 (Injuries, poison and toxic effect of drugs), MDC 22 (Burns), MDC 24 (Multiple Significant trauma) and MDC 25 (Human Immunodeficiency virus infections).

of patient mobility. Probably, when patients decide to leave their region of residence, to receive health care in a hospitals located in the mainland, it is only for reasons associated with specific needs and complex surgery.

The data provided in Table 2.1 confirmed what has been stated above: 94.71% of Sardinian patients choose to be hospitalized in facilities within their region, while extra-regional choices account only for roughly 5.29% of total cases. It is interesting to notice that, among the patients who benefitted from medical treatment outside their region of residence in 2013, the higher percentages of extra-regional choices are recorded for the LHA of Olbia (13.58%), followed by the LHA of Sassari and the LHA of Nuoro (5.91%). These numbers may be explained by the presence of important transport infrastructures at the local level. It is not by chance, in fact, that extra-regional mobility is most evident in north Sardinia, an area characterized by the highest number of port infrastructures and low-cost airlines offering patients daily connections to the mainland.

Table 2.1 Hospitalization of Sardinian residents, year 2013

	Total	LHA 101	LHA 102	LHA 103	LHA 104	LHA 105	LHA 106	LHA 107	LHA 108
Region	%	%	%	%	%	%	%	%	%
Continent	5.29	6.35	13.58	5.91	3.22	3.74	3.04	4.47	4.19
Sardinia	94.71	93.65	86.42	94.09	96.78	96.26	96.96	95.53	95.81

Thanks to the information provided by the SDO database, it is possible analyse the patient choices of distant hospital care more in detail. Table 2.2 shows the percentages of extra-regional choices across the other Italian regions. It emerges that about 36.48% of Sardinian patients chose to be hospitalized in Lombardy, 15.19% in hospitals located in Lazio, 14.93% in Emilia Romagna and 9.48% in Tuscany. However, some differences in patient choices appear when looking at hospitalizations by LHA of residence. In detail, we can see that around 44.28% of patients enrolled in the LHA of Cagliari, when leaving the region, prefer hospitals located in Lombardy, a much higher percentage when compared to other Sardinian LHA.

Analysing data on hospitalizations by age group, it is clear that extra-regional mobility is mostly determined, as we can see from Table 2.3, by those patients belonging to the age class 40-64 years (39.23%) and by the "*older adults*" (i.e. >64, 22.65%).

Table 2.2 Extra-regional hospitalizations of Sardinian residents, year 2013

	TOTAL	LHA 101	LHA 102	LHA 103	LHA 104	LHA 105	LHA 106	LHA 107	LHA 108
Region	%	%	%	%	%	%	%	%	%
Piedmont	4.73	6.48	4.21	4.60	6.25	3.79	7.05	5.06	3.54
Aosta Valley	0.18	0.21	0.35	0.17	0.00	0.11	0.00	0.33	0.07
Lombardy	36.48	31.67	31.75	32.51	33.33	35.71	41.73	38.99	44.28
Prov. Bozen	0.14	0.00	0.25	0.34	0.00	0.00	0.00	0.33	0.13
Prov. Trento	0.27	0.12	0.20	0.34	0.00	0.22	0.81	0.33	0.36
Veneto	6.23	5.02	6.22	4.94	7.99	7.14	9.21	8.97	6.33
Friuli Venezia Giulia	1.08	1.45	0.95	1.36	0.69	1.23	1.90	0.65	0.75
Liguria	6.43	10.63	5.02	6.13	6.60	6.47	8.94	6.36	3.84
Emilia Romagna	14.93	13.82	20.81	16.51	10.76	14.62	6.78	15.99	12.59
Tuscany	9.48	10.09	10.23	9.19	14.58	9.38	10.03	9.30	8.13
Umbria	0.46	0.46	0.85	0.60	0.69	0.11	0.00	0.82	0.23
Marche	0.70	0.66	0.45	0.60	0.00	1.34	0.27	0.49	0.92
Lazio	15.19	15.69	15.65	19.57	13.54	16.74	9.49	9.30	14.37
Abruzzo	0.37	0.58	0.60	0.26	0.69	0.11	0.27	0.16	0.20
Molise	0.11	0.29	0.05	0.09	0.00	0.22	0.00	0.00	0.03
Campania	1.38	1.20	1.40	0.77	2.08	1.34	0.54	1.14	1.84
Apulia	0.44	0.50	0.25	0.51	0.00	0.45	0.00	0.49	0.56
Basilicata	0.02	0.04	0.00	0.00	0.35	0.00	0.00	0.00	0.00
Calabria	0.06	0.12	0.00	0.00	0.00	0.22	0.27	0.00	0.00
Sicily	1.33	0.95	0.75	1.53	2.43	0.78	2.71	1.31	1.84

Table 2.3 Distribution of hospitalizations, by age groups (year 2013)

	Sardinia		Mainland Italy	
Age	Freq.	%	Freq.	%
5-14	6384	3.30	983	9.11
15-24	9546	4.94	740	6.86
25-39	28308	14.65	2391	22.15
40-64	67755	35.06	4234	39.23
> 64	81236	42.04	2445	22.65

As it has been stressed, Sardinian patients are more willing to choose hospitals in Lombardy, Lazio, Emilia Romagna and Liguria, namely Northern and Central regions of Italy. However, it should be noticed that patients' choices vary depending on age. Table 2.4 displays percentages of extra-regional mobility by age groups. With regard to children and adolescents (5-14 years old), Lazio and Liguria are the two regions more frequently chosen for hospitalizations outside Sardinia (39.06% and 24.42% respectively). These choices behaviours may be explained by the presence of two of the most important pediatric hospitals at the Italian level in these regions: the "*Bambin Gesù*" hospital, known to be one of the largest research centre for pediatric care in Europe (Rome, Lazio) and the "*Istituto Giannina Gaslini*" the largest children's hospitals in Northern Italy (Genoa, Liguria).

Table 2.4 Extra-regional hospitalizations of Sardinian residents, by age groups

	5-14	15-24	25-39	40-64	>64
Region	%	%	%	%	%
Piedmont	1.02	4.59	2.93	4.96	7.65
Aosta Valley	0.00	0.27	0.21	0.09	0.33
Lombardy	9.66	20.14	47.26	39.75	35.99
Prov. Bozen	0.00	0.14	0.25	0.09	0.16
Prov. Trento	0.00	0.41	0.25	0.33	0.25
Veneto	1.83	4.59	7.03	6.64	6.99
Friuli Venezia Giulia	0.81	1.22	0.67	1.11	1.51
Liguria	24.42	14.46	3.26	3.52	4.91
Emilia Romagna	10.78	16.35	13.63	16.56	14.60
Tuscany	11.39	15.54	7.74	9.28	8.92
Umbria	0.20	0.68	0.38	0.28	0.90
Marche	0.20	0.41	0.54	0.90	0.82
Lazio	39.06	19.19	10.92	12.09	13.91
Abruzzo	0.10	0.27	0.67	0.24	0.45
Molise	0.00	0.14	0.21	0.12	0.04
Campania	0.51	0.95	1.71	1.51	1.31
Apulia	0.00	0.14	0.29	0.52	0.70
Basilicata	0.00	0.00	0.00	0.05	0.00
Calabria	0.00	0.00	0.04	0.05	0.12
Sicily	0.00	0.54	2.01	1.91	0.45

Looking at Table 2.5, it is clear how extra-regional mobility is mostly due to disorders associated to the musculoskeletal system and connective tissue (17.62%), disorders of the female reproductive system (11.72%) and disorders of the circulatory system (8.05%). However, to be more precise it is good practice to show the most frequent DRG in passive mobility. Since DRGs allow to classify all hospitalizations into specific groups of costs/reimbursement, it is important to know the most frequent DRG in passive mobility as a way to limit financial losses associated with the reimbursement of hospital treatments outside the region of residence and, as well, to understand patient's need and improve services at local level.

Table 2.5 Extra-regional mobility for the five most common MDC, year 2013

MDC	Freq	%
8 - Diseases and disorders of the Musculoskeletal system and connective tissue	1902	17.62
13 - Diseases and disorders of the female reproductive system	1265	11.72
5 - Diseases and disorders of the circulatory system	920	8.52
1 - Diseases and disorders of the nervous system	869	8.05
17 - Myeloproliferative DDS (Poorly Differentiated Neoplasm)	718	6.65
6 - Diseases and disorders of the digestive system	625	5.79
9 - Diseases and disorders of the skin, subcutaneous tissue and breast	566	5.24
3 - Diseases and disorders of the ear, nose, mouth and throat	547	5.07
2 - Diseases and disorders of the eye	475	4.40
11 - Diseases and disorders of the kidney and urinary tract	425	3.94
10 - Diseases and disorders of the endocrine, nutritional and metabolic system	413	3.83
7 - Diseases and disorders of the hepatobiliary system and pancreas	402	3.72
4 - Diseases and disorders of the respiratory system	399	3.70
23 - Factors influencing health status and other contacts with health services	316	2.93
19 - Mental diseases and disorders	234	2.17
12 - Diseases and disorders of the female reproductive system	216	2.00
14 - Pregnancy, childbirth and puerperium	198	1.83
16 - Diseases and disorders of the blood and blood forming organs and immunological disorders	133	1.23
21 - Injury and poisoning	58	0.54
18 - Infectious and parasitic DDs (systemic or unspecified sites)	56	0.52
20 - Alcohol/Drug use or induced mental disorders	37	0.34
25 - HIV infections	16	0.15
24 - Multiple trauma	3	0.03

Table 2.6 reports the ten most frequent DRGs in passive mobility for the year 2013. Apparently, the descriptive statistics illustrate that patient extra-regional choices are determined by heterogeneous diseases, when actually the highest numbers of hospital

admissions are due to uterine interventions not linked to malignancies (5.89%). Note that interregional mobility, in this case, could be mainly driven by "privacy" motives.

It is interesting to underline that about 2.96% of the overall extra-regional choices are associated to DRG 410 (linked to chemotherapy). This finding is relevant because it may shed light on the reasons behind patients' mobility toward distant regions. Chemotherapy and radiotherapy (which consist respectively in the use of drugs and the administration of high doses of radiations to treat cancer) are treatments usually associated with long waiting lists in Sardinia, and this forces the patients to seek care outside the regional territory, mainly in Lombardy.

Thus, if it is true that extra-regional mobility represents an opportunity for those patients requiring highly specific medical services, usually offered by few highly specialized hospitals, on the other hand the choice of a distant hospital may be ascribed to regional health system inefficiencies, as it seems the case of cancer treatments. This means that the mobility of patients could be limited through an appropriate organization of services provision at the local level (AGENAS 2012).

Table 2.6 Extra-regional mobility for the ten most common DRGs, year 2013

DRG	Description	Freq	%
359	Uterine and adnexa proc for non malignancy W/CC	636	5.89
365	Other interventions on female reproductive system	486	4.50
410	Chemotherapy	319	2.96
544	Substitution of major joints/replanting of lower limbs	252	2.33
225	Foot intervention	195	1.81
256	Other diagnoses of muscholeskeletal system and connective tissue	157	1.45
503	Knee intervention without a primary diagnoses of infection	146	1.35
12	Degenerative diseases of the nervous system	143	1.32
467	Other factors influencing health status	136	1.26
55	Ear, nose, mouth and throat interventions	134	1.24

Starting from these considerations, two very different-in nature populations have been considered:

i. admissions for the five most common Major Diagnostic Categories (MDC) entailing extra-regional mobility (mdc 8-musculoskeletal system and connective tissue diseases; mdc 13- female reproductive system diseases; mdc 5-circulatory system disorders; mdc 1-nervous system disorders and mdc 17-myeloproliferative disorders)

ii. Cancer admissions, which show the highest percentages of mobility (icd-9-cm 150-159: malignant neoplasm of the digestive organs and peritoneum; icd-9-cm 211: benign neoplasm of other parts of digestive system, dpr 1974-1978: secondary malignant neoplasm of digestive system; dpr 2301-2309: carcinoma in situ of digestive tract; dpr 2352-2355: neoplasm of uncertain behaviour of digestive system and dpr 2390: neoplasm of unspecified nature of digestive system).[10]

In order to limit the analysis to a manageable set of alternatives, all Sardinian hospitals who provided 60% of patients' admissions and all extra-region hospitals with at least 5 yearly discharges (in the case of cancer treatment) and 10 discharges (relatively to MDC admissions) referable to Sardinian residents have been considered.

With respect to patient preferences, Table 2.7 presents detailed information on the hospitals structures chosen by Sardinian residents outside their region of residence. Altogether, we can notice that Teaching Hospitals (AOU) and Treatment and research institutes (IRCSS) are the first and second choice of patients who crossed their regional borders to seek care for diseases associated to the main 5 MDC listed above. [11,12]

[10] Cancer admissions have been defined using the ICD-9-CM international classification of diseases and procedures.

[11] Teaching hospitals are public health facilities providing health care services to citizens. In providing health care, teaching hospitals use advanced diagnostic and therapeutic equipment and innovative technologies.

[12] The treatment and research institutes (IRCCS) are defined by the Ministry of Health as hospitals of excellence with research purposes; they mainly focus on clinical and biomedical field of research and on the management of health services. The IRCCS admit patients for highly specialised treatment or to offer other high-quality according to the art. 13, paragraph 3, letter. d) of Legislative Decree 16 October 2003 n. 288.

These preferences, as it can be seen from Table 2.8, changes according to the types of diseases for which the patients move outside their region. However, focusing on cancer hospitalizations, we can see that Sardinian patients continue to choose IRCCS and AOU to receive their medical treatments.

Overall, we ended up with 45138 admission episodes with regard to the 5 main MDC diagnoses and 2575 admissions in the case of cancer, distributed among 30 and 22 hospitals respectively.

Table 2.7 Patients' choices of distant hospitals for the care of 5 Main MDC

Hospital code	Hospital name	Type
10909	AO CITTA' DELLA SALUTE E DELLA SCIENZA D	AOU
30097	ISTITUTO CLINICO CITTA' STUDI - MILANO	CCA
30147	POLICLINICO SAN MARCO S.R.L.-OSIO SOTTO	CCA
30905	OSPEDALI RIUNITI - BERGAMO	AO
30913	OSPEDALE CA' GRANDA-NIGUARDA - MILANO	AO
30924	OSPEDALE POLICLINICO S. MATTEO - PAVIA	IRCCS
30925	FONDAZ.IRCCS CA' GRANDA - OSPEDALE MAGGI	IRCCS
30934	CENTRO CARDIOLOGICO SPA "FOND. MONZINO"	IRCCS
30935	IRCCS S. RAFFAELE - MILANO	IRCCS
30943	IST. CLIN. HUMANITAS - ROZZANO	IRCCS
30947	I.R.C.C.S. POLICLINICO SAN DONATO	IRCCS
50017	OSPEDALE CLASSIFICATO SACRO CUORE - DON	OC
50912	AZ.OSP.UNIVERSITARIA INTEGRATA VERONA	AOU
80908	AZIENDA OSPEDALIERO-UNIVERSITARIA DI BOLOGNA	AOU
90901	AZIENDA OSPEDALIERO-UNIVERSITARIA PISANA	AOU
90903	AZ. OSPEDALIERO - UNIVERSITARIA CAREGGI	AOU
120905	POLICLINICO A. GEMELLI E C.I.C.	AOU
120906	POLICLINICO U. I	AOU
190923	A.O.U. POLICLINICO - VITTORIO EMANUELE	AOU

All these records have been considered as the outcome of a "hospital choice" process, in which the patient's choice set is composed of all potential 30 (22) hospitals included in the mdc (cancer) sample. By combining the information from the Hospital discharge dataset with the additional information from the database of the facilities in the Italian NHS, it is possible to condition the choice to the following wide set of hospital-specific characteristics.

Hospital capacity, measured by the number of beds. This is a commonly used attribute in hospital choice literature (Addams et al. 1991, Goodman et al. 1997, Tay 2003). As suggested by Porell and Adams (1995), patients seem to prefer not only highly qualified hospitals, but also hospitals with more beds and with more services. In general, hospital capacity is shown to affect the probability of choosing or not a hospital in a twofold manner. On the one hand, patients might be induced to choose hospital structures with high number of beds, to avoid being harmed in the care process because of overcrowded hospital wards. On the other hand, patients might be obliged to choose distant hospitals because of beds lack in hospitals close to their place of residence. In fact, the lower the number of hospital beds, the higher the waiting time and as a consequence the higher would be the probability of a patient to choose hospitals far away. However, this is also a controversial variable because a high number of beds in hospital may raise inefficiency issues.

Table 2.8 Patient's choices of distant hospitals for cancer treatment

Hospital code	Hospital name	Type
10908	OSPEDALE MAURIZIANO UMBERTO I - TORINO	AO
30913	OSPEDALE CA' GRANDA-NIGUARDA - MILANO	AO
30922	FOND.IRCCS "ISTIT.NAZ.LE TUMORI"MILANO	IRCCS
30935	IRCCS S. RAFFAELE - MILANO	IRCCS
30941	ISTITUTO EUROPEO DI ONCOLOGIA-MILANO	IRCCS
30943	IST. CLIN. HUMANITAS - ROZZANO	IRCCS
50022	CASA DI CURA PRIVATA POLISP. DOTT. PEDER	CCA
50901	AZIENDA OSPEDALIERA DI PADOVA	AO
50912	AZ.OSP.UNIVERSITARIA INTEGRATA VERONA	AOU
70901	IRCCS AOU S.MARTINO	IRCCS
80053	PRESIDIO OSPEDALIERO UNICO	OGD
80908	AZIENDA OSPEDALIERO-UNIVERSITARIA DI BOLOGNA	AOU
90903	AZ. OSPEDALIERO - UNIVERSITARIA CAREGGI	AOU
110905	A.O.U.OSPEDALI RIUNITI - ANCONA	AO
120905	POLICLINICO A. GEMELLI E C.I.C.	AOU
120906	POLICLINICO U. I	AOU

Hospital type. We constructed two dummy variables taking value 1 respectively when the hospital is a teaching or a private accredited institute. Hospital' teaching status is expected to impact positively on patient choice. Many factors can contribute to this result. First, teaching hospital are usually perceived as provider of higher quality of care, not only because they provide treatments of rare diseases and treat complex patients, but also because they offer their patients specialized services and advanced technologies (Neely and McInturff 1998, Ayanian and Weissman 2002). Secondly, a team of physicians (university professors, doctors and fellows) might care for a patient and this may be perceived as a signal of better medical care. Similarly, private accredited hospitals are expected to have a positive effect on patient choice. Private hospitals might be more attractive because they are supposed to provide personalized care and treatment to their patients.

Case mix index (CMI). This publicly monitored variable is calculated as the ratio between the average weight of admissions in a specific hospital and the average weight of admissions in the whole National Health System (NHS). More in detail, case mix index has been calculated in the following way:

$$CSI = \frac{\sum_{i=1}^{579} p_i \cdot N_{ih} \,/\, \sum_{i=1}^{579} N_{ih}}{\sum_{i=1}^{579} p_i \cdot N_{in} \,/\, \sum_{i=1}^{579} N_{in}}$$

where p_i is the specific relative weight for each DRG, N_{ih} represents the number of admissions for each Drg in each hospital and N_{in} is the number of admission in the whole Italian health system. The CMI reflects the clinical complexity (measured in terms of resource use intensity) associated with the need for more specialized hospital care (Buczko 1992, Tai et al. 2004). A value of the indicator higher than (or equal to) 1 indicates greater clinical severity compared to a set of reference hospitals. Then, we expect that severe patient will be more attracted from hospitals specialized in complex cases.

Market share. We used the market share measured by the ratio between the number of discharges in hospital *h* over the period and all discharges within the LHA over the same period of time. We looked at this measure to analyse the size of a hospital and its ability to attract patients compared to its competitors. The higher the value, the greater is the competitiveness and the patient perception of a hospital's ability to provide services and generate business.

Distance in kilometres of patients *i* LHA to hospital *h.* For the definition of kilometric distance, the following steps have been followed. Firstly, using information on hospital addresses provided by the Ministry of health, hospital locations have been geocoded. Second, using a shape file containing polygons for Sardinian LHA, centroid of each LHA have been defined. Then, the latitude and longitude coordinates of centroides have been calculated using a simple algorithm.[13] In this way, we can express *distance* as the Euclidean geographical distance from each patient LHA centroid and the centroid of each hospital. This variable is expected to capture the disincentive effect generated by the cost of mobility. Distance is supposed to negatively affect patient choice. Indeed, the greater the distance is, the higher the travel cost will be and the lower the number of patient willing to choose distant hospitals.

Hospital location (Rac). We finally considered whether a hospital is located in a regional administrative centre or not. Since Sardinian' patients have to take a plane or a boat to reach mainland distant hospitals, it can be argued that hospital locations matter for patients. Thus, we expect that hospitals located in regional administrative centres, and therefore more accessible, will be the ones more chosen by patients.

Hospital performance measured *in terms of Acute Myocardial Infarction – AMI - Mortality Rate (in the case of the MDC sample) and Surgery for malignant cancer of the Digestive system- SMCDS- Mortality Rate (as regards to cancer admissions)*. We exploited the

[13] The algorithm is based on a sum of triangle centroids weighted with their signed area.

information from the *Programma Nazionale Esiti* (PNE) as “objectives” indicators of hospital or district quality. An argument in support of the use of the AMI indicator is provided by Romano and Mutter (2004). They suggest that by using AMI mortality it is possible to bypass “*problems associated with unobservable differences in patient severity*”. Despite some performance indicators (i.e. hospital waiting times) might be affected by the presence of asymmetrical information and by hospital-manager incentives to “game” the system, another advantage of the use of the AMI mortality rate as a quality indicator is that this measure is not subject to hospital data manipulation. In addition, since AMI rates originate from emergency procedures in which patients cannot choose their favourite health care provider, emergency services would take patients to hospitals that are closer to their houses, and this restricts the possibility for hospitals to select healthier rather than more at-risk patients (Cooper et al. 2011). In the model estimates, we introduced an AMI risk adjusted for age, sex and patient comorbidities. In the case of cancer admissions, our indicator of hospital performance is whether or not a patient, operated for malignant neoplasm of the digestive system, died within 30 days of admission. In both cases, we expect that the higher the value of the indicators, the lower the demand and the probability of hospital to being chosen by patients.

As another indicator of hospital quality and performance we look at the *number of yearly treated cases under the categories of interest (Mdc/Cancer share)*. It is measured by the ratio between the number of patients' discharges for the "five main MDC" (cancer) categories in hospital *h* over the period and all discharges within the hospital over the same period of time. The idea is that the greater the number of procedures performed in a hospital and the better the medical treatment provided thanks to the learning-by-doing process. In fact, it is shared the belief that the more an individual performs a procedure the better they become in doing it (Arrow 1962). An important contribution is Avdic et al (2014), who evaluate the effect of learning-by-doing in the health care sector. Analysing the casual effect of production volumes on quality outcomes in cancer surgery, they give evidence of the positive effect of high surgical volumes on survivals rates and post-surgical complications. Thus, hospitals with higher values of the indicator are expected to attract more patients.

Doctor intensity, measured by the ratio between the number of doctors and the number of beds. We expected that a high number of doctors per beds would be able to influence positively the inflow of patients.

Table 2.9 shows the overall descriptive statistics for our study samples along with a brief description of variables. Moreover, as a way to give an informative picture of the characteristics of hospitals that have been chosen by patients, respectively in the case of MDC and cancer sample, tables 2.10 and 2.11 provide more details about hospitals attributes. It can be seen that, for the MDC sample, distant hospitals have in general better characteristics. For instance, looking at the Ami index, while the average value of the indicator for distant hospital is equal to 9.69, for closer hospitals that value is higher and equal to 11.71. Similar evidence can be drawn in the case of cancer sample.

Table 2.9 Descriptive statistics

		MDC				Cancer			
Variable name	Description	Mean	SD	Min	Max	Mean	SD	Min	Max
Patient characteristics									
Dist	Distance of patient *i* from hospial *h* (in Km)	393.19	248.38	3.01	728.58	437.18	237.65	3.01	740.42
Hospital attributes									
Case mix index (CMI)	ratio between the average weight of admissions in a specific hospital and the average weight of admissions in the whole NHS	1.09	0.19	0.80	1.68	1.12	0.11	0.90	1.28
Size (number of beds)	number of beds per hospital	746.60	561.92	112	2331	865.86	497.47	129	1570
Teaching	1 if hospital *h* is a teaching facility	0.50	0.50	0	1	0.50	0.50	0	1
Private accreditated hospitals	1 if hospital *h* is private accredited	0.27	0.44	0	1	-	-	-	-
Regional administrativ center (RAC)	1 if hospital *h* is located in a regional administrative center	0.43	0.50	0	1	0.68	0.47	0	1
Market share	ratio between the number of discharges in hospital *h* and discharges within the LHA	0.37	0.26	0.02	1	0.31	0.20	0.04	0.86
Doctor intensity	Number of doctors in hospital *h*/number of beds in hospital *h*	0.68	0.12	0.49	0.89	0.66	0.10	0.49	0.84
Mdc share	ratio between the number of discharges for main mdc disease in hospital *h* and discharges within the hospital *h*	46.67	13.46	28.10	95.15	-	-	-	-
AMI Risk adjusted (Radj)	number of death within 30 days/number of admissions for myocardial infarction X100 adjusted for counfounding factors	10.43	3.33	5.19	17.11	-	-	-	-
Cancer share	ratio between the number of discharges for cancer disease in hospital *h* and discharges within the hospital *h*	-	-	-	-	3.62	1.61	1.92	9.906
SMCDS Risk adjysted	number of death within 30 days/number of admissions for surgery caused by malignant cancer at the digestive system X100 adjusted for counfounding factors	-	-	-	-	5.56	3.63	1.15	16.66

Table 2.10 Descriptive statistics, by hospital structures (main MDC)

Hospital code	Hospital name	Beds	Doctors	ICM	Market share	AMI
10909	AO CITTA' DELLA SALUTE E DELLA SCIENZA D	2331	1597	1.08	0.63	7.79
30097	ISTITUTO CLINICO CITTA' STUDI - MILANO	270	239	1.25	0.03	11.14
30147	POLICLINICO SAN MARCO S.R.L.-OSIO SOTTO	293	158	1.14	0.08	9.82
30905	OSPEDALI RIUNITI - BERGAMO	1024	-	1.10	0.26	9.03
30913	OSPEDALE CA' GRANDA-NIGUARDA - MILANO	1066	-	1.17	0.09	7.04
30924	OSPEDALE POLICLINICO S. MATTEO - PAVIA	887	486	1.09	0.37	16.45
30925	FONDAZ.IRCCS CA' GRANDA - OSPEDALE MAGGI	974	732	0.91	0.11	7.30
30934	CENTRO CARDIOLOGICO SPA "FOND. MONZINO"	209	121	1.68	0.02	5.19
30935	IRCCS S. RAFFAELE - MILANO	1004	527	1.23	0.11	14.30
30943	IST. CLIN. HUMANITAS - ROZZANO	702	567	1.28	0.47	8.99
30947	I.R.C.C.S. POLICLINICO SAN DONATO	376	312	1.55	0.18	7.37
50017	OSPEDALE CLASSIFICATO SACRO CUORE - DON	310	188	1.04	0.41	6.94
50912	AZ.OSP.UNIVERSITARIA INTEGRATA VERONA	1433	833	1.16	0.71	11.14
80908	AZIENDA OSPEDALIERO-UNIVERSITARIA DI BOLOGNA	1570	827	1.12	0.35	6.05
90901	AZIENDA OSPEDALIERO-UNIVERSITARIA PISANA	1373	929	1.19	0.72	11.45
90903	AZ. OSPEDALIERO - UNIVERSITARIA CAREGGI	1495	1047	1.17	0.41	10.13
120905	POLICLINICO A. GEMELLI E C.I.C.	1558	766	1.02	0.39	6.94
120906	POLICLINICO U. I	1301	1041	1.00	0.28	12.15
190923	A.O.U. POLICLINICO - VITTORIO EMANUELE	957	736	1.05	0.27	14.97
200001	OSPEDALE CIVILE SASSARI	463	273	1.18	0.30	14.63
200002	OSPEDALE CIVILE ALGHERO	136	103	0.80	0.13	6.14
200012	P.O. GIOVANNI PAOLO II OLBIA	189	158	0.86	0.63	9.00
200017	P.O. SAN FRANCESCO	355	273	0.98	0.86	13.26
200019	P.O. 'NOSTRA SIGNORA DELLA MERCEDE'	112	94	0.85	0.69	11.88
200028	P.O.SIRAI	183	120	0.94	0.57	11.77
200031	P.O. SS. TRINITA'	313	248	0.98	0.15	9.39
200034	PRESIDIO OSPEDALIERO N.S. DI BONARIA	176	90	0.94	1.00	16.73
200052	P. OSPEDALIERO 'SAN MARTINO' - ORISTANO	247	177	0.99	0.56	17.11
200904	AZIENDA OSPEDALIERA G.BROTZU	550	417	1.16	0.20	7.77
200906	A.O.U. CAGLIARI	541	408	0.92	0.21	11.13

Table 2.11 Descriptive statistics, by hospital structures (cancer)

Hospital code	Hospital name	Beds	Doctors	ICM	Market share	SMCDS
10908	OSPEDALE MAURIZIANO UMBERTO I - TORINO	436	285	1.23	0.15	5.20
30913	OSPEDALE CA' GRANDA-NIGUARDA - MILANO	1066	-	1.17	0.09	5.59
30922	FOND.IRCCS "ISTIT.NAZ.LE TUMORI"MILANO	369	235	1.17	0.05	2.39
30935	IRCCS S. RAFFAELE - MILANO	1004	527	1.23	0.11	6.84
30941	ISTITUTO EUROPEO DI ONCOLOGIA-MILANO	312	262	1.26	0.04	4.16
30943	IST. CLIN. HUMANITAS - ROZZANO	702	567	1.28	0.47	5.55
50022	CASA DI CURA PRIVATA POLISP. DOTT. PEDER	260	140	1.15	0.29	6.64
50901	AZIENDA OSPEDALIERA DI PADOVA	1450	807	1.12	0.57	1.15
50912	AZ.OSP.UNIVERSITARIA INTEGRATA VERONA	1433	833	1.16	0.71	3.71
70901	IRCCS AOU S.MARTINO	1394	849	1.13	0.40	5.44
80053	PRESIDIO OSPEDALIERO UNICO	1379	841	0.98	0.33	5.38
80908	AZIENDA OSPEDALIERO-UNIVERSITARIA DI BOLOGNA	1570	827	1.12	0.35	4.22
90903	AZ. OSPEDALIERO - UNIVERSITARIA CAREGGI	1495	1047	1.17	0.41	2.75
110905	A.O.U.OSPEDALI RIUNITI - ANCONA	969	643	1.23	0.19	6.98
120905	POLICLINICO A. GEMELLI E C.I.C.	1558	766	1.02	0.39	4.97
120906	POLICLINICO U. I	1301	1041	1.00	0.28	2.68
200001	OSPEDALE CIVILE SASSARI	463	273	1.18	0.30	2.68
200017	P.O. SAN FRANCESCO	355	273	0.98	0.86	16.66
200029	P.O.SANTA BARBARA	129	82	0.90	0.31	14.70
200031	P.O. SS. TRINITA'	313	248	0.98	0.15	7.35
200904	AZIENDA OSPEDALIERA G.BROTZU	550	417	1.16	0.20	5.30
200906	A.O.U. CAGLIARI	541	408	0.92	0.21	1.87

2.4 Empirical analysis

2.4.1 Econometric approach

In principle, each admission episode can be seen as the result of a "choice" by a sample of $i = 1, 2, \ldots, N$ patients over a choice set of $h = 1, 2, \ldots, H$ mutually exclusive hospitals. This choice can be described by means of a random utility model such as the following:

$$U(\text{choice } h \text{ by patient } i) \equiv U_{ih} = V_{ih} + \varepsilon_{ih} = \beta' x_h + \varepsilon_{ih} \quad (2.1)$$

where the component of the vector $\mathbf{x_h}$ may either denote hospital attributes, or be individual-specific. Utility for each individual is obtained through the sum between an observable element V_h^i and a stochastic unobservable component ε_h^i. In this study, the elements included in the vector are the hospital attributes described in the previous section.

By assuming that the individual random elements $\boldsymbol{\varepsilon}_h^i$ are independently and identically distributed (IID), with an extreme value type 1 (Gumbel) distribution, we get the "*conditional logit*" model, in which the likelihood that patient i is admitted to hospital h is:

$$P[y_i = h|x_i] = \frac{\exp(\beta' x_{ih})}{\sum_{l=1}^{H} \exp(\beta' x_{il})} \quad (2.2)$$

The IID assumption leads to the also known independence of irrelevant alternatives (IIA) property, which states that the odds between two generic alternatives *k* and *l* are equal to:

$$\frac{P(i|k)}{P(i|l)} = \frac{e^{\beta' x_{ik}}}{e^{\beta' x_{il}}} = e^{\beta' x_{ik} - \beta' x_{il}} \quad (2.3)$$

so that the relative probabilities depend solely on the characteristics of those two *k* and *l* alternatives (Wooldridge 2011). The IIA property implies that the probability to choose between two alternatives is independent from the presence of additional alternatives other than k and l (McFadden, 1984). In presence of a subsets of similar alternatives, this independence condition may prove very strong. However, this assumption is hardly satisfied in models with similar alternatives. An example is provided by McFadden (1974). Consider a situation where commuters have to choose between travelling by *car* or by *red bus*. Thus, imagine that the commuter was equally likely to choose either of the traveling modes. Then, the probability of choosing each travel mode is equal to 0.50, so that the ratio in (2.3) is equal to 1. Now, suppose that a new alternative, *blue bus*, is introduced in the choice set. Imagining that people do not care about the bus colour, it follows that the individual will chose between them with equal probability. This means that, if the IIA holds, then the choice probability among travel mode must now be 0.3333, so that the commuter remains equally like to choose the car or a bus. But a more likely consequence would be for the probability to choose *car* to remain 0.50 and the probability to choose between *red bus* and *blue bus* to be 0.25 each. In this case, therefore, it can be noted how the properties of IIA lead to an overestimation of the probability of choosing the bus and to an underestimation of the probability of choosing the car so that the model results are not reliable. Hausman and McFadden (1984) propose a test of IIA assumption based on the idea that, if a subset of the choice set is not relevant with respect to the other alternatives, omitting those alternatives from the *conditional logit* model will not lead to inconsistent estimates. But when the null hypothesis of independence of irrelevant alternatives is violated, then, the estimate model is not adequate to describe the studied phenomenon.

Several models may be used in order to overcome the rigidity of the IIA assumption. Examples of models that relax the IIA assumption are the *multinomial probit* model, the *nested logit* model and the *mixed logit* model. In the next section the mixed logit model will be discussed in detail along with its advantages with respect to the model just analysed.

2.4.2 Mixed logit model

The Mixed Logit model (ML), also known as "*random parameter logit*" or "*mixed multinomial logit*", was originally formulated by Cardell et al. (1978) and subsequently reformulated, in the currently-known version, by Ben-Akiva and Bolduc (1996) and McFadden and Train (2000). Similar to the *conditional logit* model, the ML is used to analyse discrete choices. The ML allows to evaluate the importance of each attributes of a given alternative according to the characteristics of this alternative and of any other available. The advantage of this model is that "*it obviates the three limitations of standard logit by allowing for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time*" (Train 2003).

Considering the previous notation, suppose that a sample of individuals *i =1,.....,N* face a choice among *h* alternatives in each of *T* choice set.[14] Then, the utility associated to each set of alternatives *h* as evaluated by each individual *i* can be written as:

$$\begin{aligned} U_{hit} &= \sum_{k=1}^{K} \beta_{ik}\, x_{htik} + \varepsilon_{hti} \\ &= \beta_i' x_{hti} + \varepsilon_{hti} \end{aligned} \tag{2.4}$$

where x_{hti} represents the set of explanatory variables related to the alternatives and the socio-economic characteristics of the decision makers.[15] As we can notice, the parameter β_i is supposed to vary across decision makers. This means that, in the ML the parameters may vary from individual to individual and are not fixed in the whole sample. Although this specification is similar to the one of the *conditional logit* model, in the ML, β varies among individuals instead of being fixed. The factors β_i and ε_{ih} are not observed by the researchers and are handled as stochastic influences.

[14] With multiple choice set analyst usually mention a situation in which a person faces a choice on more than one occasion (e.g. in a longitudinal panel).

[15] The discussion on the Mixed Logit model is drawn part from chapter 15 of Applied Choice Analysis: a primer, by Hensher et al. (2005) and part from chapter 6 of Discrete Choice Methods with Simulations, by Train (2003, 2009) to which the reader is referred for more details.

That being said, the assumption that the errors have to be *iid* between alternative and time periods, may be overcome by introducing in the utility function stochastic parameters in the utility function, which may be both heteroskedastic and correlated through alternatives. Thus, the basic structure of random parameters is:

$$\beta_i = \beta + \Delta z_i \tag{2.5}$$

while the full set of random parameters can be written as:

$$\beta_i = \beta + \Delta z_i + \Gamma v_i = \beta + \Delta z_i + \eta_i \tag{2.6}$$

where β is the parameter distribution mean, $\boldsymbol{z}$ indicates the observable variables of the individual influencing the parameter, Γ is a lower triangular matrix allowing free variances and correlations of the parameters, and $\boldsymbol{v_i}$ is a vector of random unobserved preference variation. Consequently, η_i represents a random term that induces a correlation among parameters and whose distribution depends on the analysts' choice. Since the random term may vary across time and may generate correlation across choices, it may be modelled to take into account the correlation between alternatives and time periods so that the error term of the model is independently and identically distributed. Thus, for a specific value of η_i, the *conditional probability* on β_i is a simple *Multinomial Logit* (MNL):

$$L_{hi}(\beta_i | X_i, \eta_i) = \frac{\exp(\beta_i' x_{hi})}{\sum_h \exp(\beta_i' x_{hi})} \tag{2.7}$$

Hence, the probability calculated through a MNL is nothing but a simply probability of choice which for each individual, is conditioned to certain values of the random term η_i. Since, the analyst does not know β_i he is unable to condition on β. The ML, instead, allows estimating the unconditional choice probability and, therefore, the probability calculated by considering all the values assumed by the random parameters. The unconditional choice probability is:

$$P_{hi}(X_i, z_i, \Omega) = \int L_{hi}\,(\beta_i|X_i, \eta_i) f(\eta_i|z_i, \Omega) d\eta_i \tag{2.8}$$

Therefore, the likelihood that an individual *i* will select a specific alternative *h* is given by the integral of L_{hi} over the possible values of β_i. It is now clear why this model is named *mixed logit:* the choice probability P_{hi} is obtained by the combination of logits with *f*.

Overall the advantages of the ML are that it takes into consideration the random variation of the preferences, the fact that probabilities are not subjected to the IIA property and that they allow free variances and correlations of parameters.

2.4.3 Identification strategy of "distant care" attraction factors

Several approaches can be adopted in order to assess whether the main explanatory factors in the vector $\mathbf{x_h}$ play the same role independently of the hospital being near or far from the patients' place of residence. In this chapter, we exploit the fact that insularity creates a clear dichotomy between regional hospitals (accessible by car) and "mainland" hospitals, the different distance of which is in practice "normalised" by the necessity to take a flight. The control for the presence of differential effects is obtained, on the one hand, by interacting each hospital characteristic with a "Sardinian hospital dummy" and, on the other hand, imposing, in the utility function, different parameters for the two groups of alternatives represented by close Sardinian hospitals and Mainland distant hospitals. This entails estimating the coefficients in the upper part of the Table for distant hospitals located on mainland Italy. By doing this, we can better identify the reasons that drive patients to a particular hospital to get a specific treatment, once the distance hurdle has been overcome.

2.5 Results

The main results from the application of the Conditional Logit model to our dataset are presented in Table 2.12 and 2.13, which report five estimates for the sample of main MDCs (models 1a, 2a, 3a, 4a ,5a) and five for cancer (models 1b, 2b, 3b, 4b,5b).[16] Model 1a (1b) forms the baseline specifications that only differ from the exclusion/inclusion of interactions between hospital-level characteristics and a geographical dummy indicating whether the hospital is "near" (namely, located in Sardinia), whereas Models 3a, 4a and 5a (3b, 4b,5b) are enriched by the presence of a death indicator for Ami (Smcds) risk adjusted and other covariates.

Model 1a shows that all hospital attributes have and explanatory power in the choice of the place of treatment, given that the coefficients are statistically significant. The positive sign of the *case mix* indicator points out a direct relationship between specialization in relatively complex treatments and likelihood of choice of a given hospital.

The hospital capacity variable has allowed us to detect a significant and unforeseen behaviour. Indeed, it seems that a higher supply of beds is negatively perceived by patients.

As expected, geographical distance has a negative and significant effect on patient choices. This result suggests us that closer hospitals, with respect to far away health structures are more likely to be chosen by patients. Similar finding has been found in the literature, for example by Sivey (2012) and Moscone et al (2012).

The hospital type dummies must be evaluated with respect to the baseline category of public hospitals. Our regressions display positive signs for *Teaching* and *Private accredited* hospitals. The first result could be the by-product of the public reputation of teaching institutes. The result for private accredited facilities could be consistent whether with a demand-side or a supply-side interpretation. In fact, the positive sign may reveal that patients

[16] Conditional logit estimates have been obtained using the software STATA13

are more willing to choose a private facility for some intrinsic characteristics (e.g. shorter waiting lists, personalized care), or that private ownership makes hospitals more prone to carry out active policies aimed at attracting patients.

With regard to the market share, we find a unexpected negative and significant effect: as the market share goes up the likelihood of a hospital to being chosen goes down. It can also be seen that *Rac* is positive and significant. There are two possible explanations for this result. First, a higher choice probability for hospitals in regional administrative centres could reflect the geographical concentration of very specific treatments and doctors with high reputation. Secondly, access to health services in these areas may be easier for patients, especially when compared to distant hospitals located on the Mainland.

2.5.1 Focusing on distant hospital care

Several interesting differences emerge from the estimation of the models with interaction terms (2a, 3a, 4a, 5a), which represents the bulk of our analysis. Overall, the value of the pseudo R^2 indicates a better fit of the estimated model than without interactions:

i. case mix attribute shows two opposite signs: negative for distant hospitals, while positive for closer hospitals. In the linear effect, we have -3.21 for Mainland hospitals vs 7.10 for Sardinia;

ii. a different reasoning applies to the attraction role of *Rac*. In the case of distant Mainland hospitals, this attribute clearly reduces the probability of admission. Conversely, it is a positive drivers of patient mobility in the case of Sardinian hospitals.

iii. The role of teaching and private accredited hospitals *vis à vis* public hospital facilities appears to be different. In the case of distant hospital on the Mainland, being a specialised teaching structure seems to reduce the probability of admission while being a private accredited structure affect positively patient choice. Conversely, those attributes are significant determinants of closer hospital choices.

iv. The market share indicator negatively affects the choice probability of distant hospitals, while a positive effect is found for Sardinian hospitals.

In the specification 3a, the model with interactions is augmented by introducing the risk adjusted mortality for acute myocardial infarction occurring within 30 days since first admission[17]. It is interesting to note that this quality measure has a negative and significant effect on patient outflows, but a smaller positive effect in the case of "near" hospitals.

To better consider the effect of quality on hospital demand, in model 4a, the measure related to the *MDC share* has been introduced in the model. The above mentioned measure, which should capture hospital quality inherent to the process of learning by doing, positively affects the choice of distant hospitals, while a negative effect is found for Sardinian hospitals. This suggests that positive externalities are mainly operating in Mainland hospitals, where doctors working on these hospitals presumably treat more cases and, as a consequence, own more experience. Finally, in model 5a, Doctor's intensity appears as a positive choice determinant for close hospitals.

In Table 2.13, we present the results from the models related to cancer treatments. Model 1b shows that the choice probability is positively correlated to the hospital case mix. Apparently, specialization in cancer DRGs does entail a specialization in expensive treatments. In this case, the use of hospitals' CMI as an indicator for assessing the quality of healthcare specialization would nowadays constitute a "red herring". The effects of the remaining explanatory factors confirm substantially the findings for the main MDCs.

Model 2b, 3b, 4b and 5b report the estimated results for the regression with interactions. The estimates in the upper part of the Table detect some important results for distant

[17] AMI risk adjusted has been calculated considering the effects due to confounding factors (such as age, sex, comorbidities, et.).

hospitals. First, hospital size is a poor negative predictor of the probability of choice. By assuming that Sardinian patients are exploiting some shared information on the good quality of treatments that is not revealed from the data, this implies the lack of economies of scale in cancer treatments. Second, looking at the specification reported in model 3b we can see that the *SMCDS* risk adjusted does have a positive effect on the choice of distant hospitals, even though this result is not robust after the inclusion of other controls (see column 4b). However, these results must be analysed carefully since the selection of the dataset, due to the lack of information on SMCDS risk adjusted for different structures, may cause a problem of selection bias.

In an attempt to better evaluate whether patient choice of distant hospitals represents a particular phenomenon with its own characteristics, we have decided to re-estimate conditional logit models, imposing parameters in the utility function, to detect the main determinants of patient choice both in the case of close and distant hospitals.[18] Table 2.14 displays the results of the estimates both for the sample of main MDC (model 1a, 2a) and Cancer (model 1b, 2b). The upper part of the table reports the estimated coefficients of the determinants affecting distant hospital choice, while the lower part presents estimation results for those attributes influencing close hospital choices. Looking at model 1a, the results of the estimate suggest that almost all variables have a negative effect on Mainland hospital choices. It is interesting to notice that the estimated coefficient for *Distance* shows an unexpected positive sign, thus indicating that, probably, beyond a certain distance, the likelihood of hospital choice seems to be not negatively affected by travel distance. However, those results seem to be not robust after the inclusion of the *Mdc share* variable (model 2a).

Model 1b (Table 2.14) reports the results for the cancer sample. The effect of *Case mix* is completely different. As regard to distant hospitals, this attribute negatively affects the

[18] Conditional Logit models have been estimated using the software NLOGIT 5.

Table 2.12 Conditional logit model of patient choice for main admissions

	Model 1a			Model 2a			Model 3a			Model 4a			Model 5a		
	Coeff.		SE	Coeff.		SE	Coeff.		SE	Coeff.		SE	Coeff.		SE
Distance	-0.0331	***	0.0002	0.0022	***	0.0003	0.0021	***	0.0003	0.0031	***	0.0003	0.0039	***	0.0003
Case mix index	4.5160	***	0.0469	-3.2151	***	0.1471	-3.3227	***	0.1534	-7.3368	***	0.3163	-4.1506	***	0.5173
Size (number of beds)	-0.0027	***	0.0000	0.0009	***	0.0001	0.0009	***	0.0001	0.0015	***	0.0001	0.0016	***	0.0001
Teaching	1.4306	***	0.0176	-0.0968		0.0604	-0.0417		0.0640	0.2628	***	0.0730	-0.3104	**	0.1270
Private accredited hospitals	1.1376	***	0.0491	1.6142	***	0.0633	1.6057	***	0.0637	0.9242	***	0.0772	0.7941	***	0.0851
Rac	0.2549	***	0.0227	-1.0251	***	0.0691	-1.0606	***	0.0716	-0.9873	***	0.0706	-0.8710	***	0.0868
Market share	-0.5493	***	0.0299	-3.4101	***	0.1872	-3.5617	***	0.2002	-2.9083	***	0.2151	-2.5218	***	0.2812
Radj							-0.0152	*	0.0084	-0.0168	**	0.0085	0.0159		0.0113
Mdc share										0.0769	***	0.0045	0.0533	***	0.0063
Number of doctors per n. beds													-1.6729	***	0.2235
Dummy regio	Yes			No			No			No			No		
<u>Interactions</u>															
Distance*dsar				-0.0535	***	0.0004	-0.0534	***	0.0004	-0.0546	***	0.0004	-0.0548	***	0.0004
Case mix index*dsar				7.0927	***	0.1968	6.6491	***	0.2595	12.7298	***	0.4042	11.1158	***	0.5355
Size*dsar				-0.0028	***	0.0002	-0.0024	***	0.0003	-0.0038	***	0.0003	-0.0045	***	0.0004
teaching*dsar				1.5083	***	0.0807	1.2961	***	0.1103	1.4833	***	0.1337	2.3907	***	0.2164
Rac*dsar				2.2260	***	0.0788	2.2642	***	0.0811	1.8295	***	0.0880	1.4450	***	0.1003
Market share *dsar				3.5443	***	0.1904	3.6299	***	0.2022	3.0368	***	0.2197	2.7285	***	0.2823
Radj*dsar							0.0265	***	0.0100	-0.0033		0.0112	-0.0361	***	0.0146
Mdc share*dsar										-0.1014	***	0.0051	-0.0887	***	0.0067
N. doctors per beds*dsar													3.1526	***	0.2563
Log Likelihood	-79294.01			-64170.83			-64166.64			-63956.69			-62489.63		
Pseudo R^2	0.4835			0.5820			0.5820			0.5834			0.5736		
N. of admissions	45138			45138			45138			45138			44981		
N. hospitals	30			30			30			30			26		
Hausman-McFadden P-value	0.000			0.000			0.000			0.000			0.000		

Table 2. 13 Conditional Logit model of patient choice for cancer related admissions

	Cancer									
	Model 1b		Model 2b		Model 3b		Model 4b		Model 5b	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Distance	-0.0179 ***	0.0005	0.0098 ***	0.0010	0.0091 ***	0.0011	0.0097 ***	0.0013	0.0097 ***	0.0012
Case mix index	4.1583 ***	0.2729	-6.2465 ***	1.1988	-6.8299 ***	1.3096	-7.9836 ***	1.5056	-7.3991 ***	1.4079
Size (number of beds)	-0.0029 ***	0.0001	-0.0009 **	0.0003	-0.0009 **	0.0004	-0.0008 ***	0.0003	-0.0003	0.0002
Teaching	1.5060 ***	0.0693	1.1784 ***	0.1783	1.2593 ***	0.1947	1.5127 ***	0.2250	1.2481 ***	0.2034
Rac	0.7659 ***	0.0778	0.8054 *	0.4576	0.6176	0.4665	0.0645	0.1702	0.0597	0.1676
Market share	0.9641 ***	0.1449	2.0216 ***	1.2038	1.7716	1.2269				
Smcds Radj					0.0468	0.0453	-0.0105	0.0545	0.0617	0.0462
Cancer share							-0.0958	0.0677		
Number of doctors per n. beds									1.3106	0.7647
Dummy regio	Yes		No		No		No		No	
Interactions										
Distance*dsar			-0.0531 ***	0.0017	-0.0511 ***	0.0020	-0.0519 ***	0.0022	-0.0519 ***	0.0021
Case mix index*dsar			11.9951 ***	0.9737	11.8512 ***	0.9963	5.5289 **	2.5721	10.1435 ***	1.4995
Size*dsar			-0.0061 ***	0.0015	-0.0067 ***	0.0015	0.0038	0.0035	-0.0046 ***	0.0012
teaching*dsar			1.1542 **	0.5087	0.7045	0.5346	-1.8953 **	0.8741	0.2719	0.5184
Rac*dsar			1.6418 ***	0.4063	1.8488 ***	0.4265	1.1113 **	0.4546	1.3309 ***	0.3657
Market share *dsar			-1.3709	1.1287	0.0174	1.3497				
Smcds Radj*dsar					-0.1138 **	0.0521	0.0110	0.0547	-0.1139 **	0.0513
Cancer share*dsar							0.9756 **	0.4682		
N. doctors per beds*dsar									2.3187	1.6273
Log Likelihood	-4612.11		-3870.01		-3866.66		-3866.71		-3826.3101	
Pseudo R^2	0.4205		0.5138		0.5142		0.5142		0.5108	
N. of admissions	2575		2575		2575		2575		2569	
N. hospitals	22		22		22		22		21	
Hausman-McFadden P-value	0.000		0.000		0.000		0.000		0.000	

Table 2. 14 Determinants of distance vs close hospital choices

	MDC (8-13-5-1-17)				*Cancer*			
	Model 1a		Model 2a		Model 1b		Model 2b	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Distant hospitals								
Distance	0.0008 ***	0.0003	0.0043 ***	0.0003	0.0091 ***	0.0011	0.0097 ***	0.0013
Case mix index	-1.8858 ***	0.1245	-10.4782 ***	0.3618	-6.8299 ***	1.3096	-7.9836 ***	1.5056
Size (number of beds)	-0.0005 ***	0.0000	0.0023 ***	0.0001	-0.0009 **	0.0004	-0.0008 ***	0.0003
Teaching	-0.2056 ***	0.0557	0.0886	0.0717	1.2593 ***	0.1948	1.5127 ***	0.2250
Rac	-0.5595 ***	0.0642	-0.8521 ***	0.0708	0.6176	0.4665	0.0645	0.1702
Market share	-2.0965 ***	0.1872	-2.0078 ***	0.2143	1.7716	1.2269		
Ami Radj	-0.0102	0.0075	-0.0443 ***	0.0085				
Smdcs Radj					0.0468	0.0453	-0.0106	0.0545
MDC share			0.0700 ***	0.0056				
Cancer share							-0.0958	0.0677
Close hospitals								
Distance	-0.0511 ***	0.0003	-0.0516 ***	0.0003	-0.0420 ***	0.0013	-0.0423 ***	0.0014
Case mix index	2.3526 ***	0.2829	5.8363 ***	0.3804	5.0213 ***	1.7149	-2.4546	2.2304
Size (number of beds)	-0.0007 ***	0.0002	-0.0027 ***	0.0003	-0.0076 ***	0.0015	0.0030	0.0036
Teaching	0.9980 ***	0.0780	0.8936 ***	0.0895	1.9639 ***	0.4236	-0.3826	0.8847
Rac	1.1157 ***	0.0401	0.8936 ***	0.0550	2.4664 ***	0.2214	1.1758 **	0.4566
Market share	-0.0462	0.0484	0.1830 ***	0.0497	1.7889 ***	0.6495		
Ami Radj	0.0209 ***	0.0046	-0.0240 ***	0.0060				
Smdcs Radj					-0.0670 **	0.0292	0.0005	0.0223
Mdc share			-0.0246 ***	0.0025				
Cancer share							0.8798 *	0.4929
Log Likelihood	-64474.45		-63923.92		-3866.66		-3866.71	
Pseudo R^2	0.4141		0.4191		0.2698		0.2698	
N. of admissions	45138		45138		2575		2575	
N. hospitals	30		30		22		22	
Hausman-McFadden P-value	0.000		0.000		0.000		0.000	

probability of admission, while in the case of close hospitals it seems to be a much stronger positive determinant of patient choice. Though positive, the effect of teaching is smaller for distant hospitals (1.259 for Mainland hospitals vs 1.9639). Moreover, the quality indicator shows opposite but significant effects: positive in the case of distant hospital, while negative on patient choice in Sardinia.[19]

One of the main problems related to the use of the conditional logit model is represented by the violation of the assumption of independence of irrelevant alternatives (IIA). As we have mentioned before IIA assumption implies that the probability of choosing between hospital A over hospital B does not depend on the availability of another hospital alternative C (Train 2009). For instance, when the IIA is in force, the mechanism whereby people choose between watching a movie or attending a ballet is independent from whoever is performing a live concert the same day. It should be mentioned, according to Christiadi and Cushing (2007), that IIA assumption is restrictive, especially in the presence of multiple alternatives as in the case of patient hospital choices. Violation of the IIA assumption may lead to a wrong estimation of a specific alternative being chosen. This means that the model may overestimate the probability of choosing hospital x, and as a consequence underestimate the probability of choosing another hospital.

In the light of the complication associated with the violation of the IIA assumption, the test of Hausman-McFadden (1984) has been performed to investigate the validity of the above-mentioned assumption. Table 2.12 (2.13 and 2.14) shows the Hausman-McFadden test results. Since Prob>chi2 =0.0000, the model fitted on these data fails to meet the asymptotic assumptions of the Hausman test. Results were confirmed by the “suest” post estimation command. These findings lead us to think that other models, such as nested logit or mixed logit that relax the IIA assumption, can be implemented to explain patient choice of distant hospital care; thus, mixed logit models have been estimated in order to overcome the limitation related to the conditional logit model and to account for preference heterogeneity among decision makers.

[19] Taking a different viewpoint, the result suggests that improvements in Sardinian hospitals quality measure would effectively restrain extra-regional outflows.

Table 2.15 shows the main results obtained from the application of the ML model.[20] Specifically, the table reports estimates for the sample of main MDCs (model 1a) and Cancer (model 1b). Looking at the estimate results for model 1a, the upper part of the table shows that the overall fit of the ML is -62997.67, which represents an improvement over the *conditional logit* model fit of -79284.3 (Table I in Appendix I). There are, in addition, an extra number of degrees of freedom, which increase from 9 to 18 in the Mixed logit model. The table displays also the Chi-square statistics for the log-likelihood ratio test (LRT), which give us information on the significance of the model. In this case, it can be said that Model 1a is statistically significant since the very high value of the Chi-square statistic. The lower part of the table reports the estimate results with regards to the determinants of the model. Considering the results for the random parameters, the mean sample parameters are all statistically significant confirming the results previously obtained through the *conditional logit* model (with the exception of the *size* parameter that is showing, in this case, a positive sign).

Looking at the cancer sample, the estimation results for model 1b seem to suggest that patients prefer hospitals with high quality ratings. In addition, patient choice appears to be negatively affected by case mix and hospital capacity. Starting from this model, we allow the parameters to vary by age and sex. Hence, in Table 2.16, much information is provided for each explanatory variable. For instance:

- *Distance* relates to the mean of the random parameter of the distribution, not considering provider's attributes interacting with patient characteristics. Thus, the coefficient associated to *Distance* is the equivalent of the parameter β in the formula (2.6).
- *Distance:male (age)* refers to the coefficients obtained by interacting hospital attributes with patient characteristics, so that *male (age)* represents the individual characteristic $z_{i,}$, while the estimated coefficient is Δ.

[20] Mixed logit models have been estimated with the software NLOGIT5.

- *St.Dev.Distance* is the standard deviation of the estimated parameter of the distribution. So that. The coefficient associated to *St. Dev. Distance* gives the standard deviation of the random parameter η_i.[21]

The standard deviations of *Distance* and hospital capacity (*Size)* are highly significant (1%), suggesting that the coefficients vary among cancer patients. The effect of the distance on the marginal utility of each alternative and, therefore, on the choice probability of each patient can be understood by reconstructing the distribution of parameters associated with the variable. Consequently, the marginal utility associated with *distance* is:

$$\boldsymbol{Distance} = -\mathbf{0.0441} - \mathbf{0.0029} \cdot \boldsymbol{Male} + \mathbf{0.0929} \cdot (\mathbf{5/30}\boldsymbol{years}) - \mathbf{0.0135}(> \mathbf{64}\boldsymbol{years}) + \mathbf{0.0382} \cdot \boldsymbol{N} \tag{2.9}$$

where *N* indicates values randomly extracted from a normal distribution function. Basically, the effect of *Distance* on the probability of a hospital to being chosen differs by age groups. As expected, *Distance* appears to negatively affect hospital's choices made by older patient (>64 years). Indeed, it is well known that elderly people are less likely to travel to receive medical treatments with respect to younger patients. Similarly, hospital choices made by males are less influenced by *Distance*.

Finally, we re-estimate ML models distinguishing, for each alternative, the parameters associated to *distant* vs *close* hospitals. Table 2.17 illustrates the results of the ML models both for the MDC (model 1a) and Cancer sample (model 1b). In model 1a, the mean sample parameters are all statistically significant. Comparing ML estimates with those obtained using *conditional logit* models (Table 2.14), several differences can be found. However, the values of the log-likelihood and the pseudo-R^2 suggests a better fit of the ML model. Looking at the upper part of the table, *Distance* has a negative and significant effect (at 1%) on the probability of choosing a hospital in the Mainland. Similarly, the values of the estimated coefficient for *Case Mix* indicates a negative

---

[21] Note that if the standard deviation is not significant then, according to Henser et al. (2005), "*dispersion around the mean is statistically equal to zero, suggesting that all information in the distribution is captured within the mean*".

effect of specialization in complex care. In addition, patients with diseases associated to main MDC seem to not consider quality ranking indicators before to choose a distant hospital structure.

Interesting findings, in order to identify the peculiarities associated with distant hospital choices, can be drawn looking at the *Cancer* estimates in model 1b. As we can see, there is a clear dichotomy in the role played by *Distance* in influencing patient choice of close regional hospital *vs* distant hospitals: negative and significant in the case of Sardinian structures, while positive, though not significant, for Mainland hospitals. For distant choices, the *Case mix* shows a negative sign; thus specialization in complex cases is associated with reductions on the probability of hospital being chosen. On the contrary, for nearby hospitals, it seems that *Case Mix* encourage regional inflows (+5.28). An unexpected positive sign of the *Smcds Radj* indicator point out that patients seems to not consider hospital quality before to choose a medical structure located in the Mainland regions. A positive sign of *Rac* is found for close hospitals. This result can be explained by the fact that in Sardinia, hospitals with higher technological equipment and specialized physicians are concentrated in urban centres rather than in rural centres. Conversely, a symmetric negative effect it is shown for distant hospitals.

Table 2.15 Mixed logit model of hospital choices

	Model 1a		Model 1b	
Dependent variable	CHOICE		CHOICE	
Log-likelihood	-62997.67		-3903.25	
Restricted Log-likelihood	-153523.25		-7959.43	
Chi squared	181051.15		8112.37	
Degree of freedom	18		16	
P-value	0.000		0.000	
Pseudo R^2:	0.590		0.510	
Halton	100		100	
N. of admissions	45138		2575	
N. hospitals	30		22	
Distance	-0.0706	***	-0.0563	***
St. Dev. Distance	0.0411	***	0.0389	***
Case mix index	0.1408	**	-2.5938	***
St. Dev. Case mix index	0.0148		0.2616	
Size (number of beds)	0.0009	***	-0.0011	***
St. Dev. Size	0.0000		0.0025	***
Teaching	0.2155	***	0.1279	
St. Dev. Teaching	1.7467	***	0.3880	
Private accredited hospitals	1.5593	***		
St. Dev. Private accredited h.	0.0194			
Rac	0.7759	***	1.8681	***
St. Dev. Rac	0.1113		0.1209	
Market share	-0.7390	***	0.9387	**
St.Dev.Market share	0.0470		0.1186	
Ami Radj	0.0108	***		
St.Dev. Ami Radj	0.0447			
Smdcs Adj			-0.0956	***
St.Dev. Smdcs Adj			0.0037	
Dummy regio	Yes		Yes	

Table 2.16 Mixed logit model allowing for patient heterogeneity (Cancer admissions)

Dependent variable	CHOICE	
Log-likelihood	-3880.16	
Restricted Log-likelihood	-7959.43	
Chi squared	8158.55	
Degree of freedom	40	
P-value	0.000	
Pseudo R^2:	0.513	
Halton	100	
N. of admissions	2575	
N. hospitals	22	
Distance	-0.0441	***
Distance: Male	-0.0029	
Distance: Age 5-30 years	0.0929	**
Distance: Age >64 years	-0.0135	***
St. Dev. Distance	0.0382	***
Case mix index	-1.8576	**
Case mix index: Male	-0.4388	
Case mix index: Age 5-30 years	28.2396	
Case mix index: Age >64	-1.5187	*
St. Dev. Case mix index	1.7684	
Size (number of beds)	-0.0015	***
Size: Male	0.0008	**
Size: Age 5-30 years	0.0186	
Size: Age >64 years	-0.0002	
St. Dev. Size	0.0022	***
Teaching	0.3254	
Teaching: Male	-0.2133	
Teaching: Age 5-30 years	8.5406	
Teaching: Age >64 years	-0.3920	*
St. Dev. Teaching	0.2825	
Rac	2.0348	***
Rac: Male	-0.1540	
Rac: Age 5-30 years	5.5083	
Rac: Age >64 years	-0.1344	
St. Dev. Rac	0.1963	
Market share	1.0485	
Market share: Male	-0.3263	
Market share: Age 5-30 years	-24.6395	
Market share: Age >64 years	0.4161	
St. Dev. Market share	0.5641	
Smcds Radj	-0.0897	***
Smcds: Male	0.0077	
Smcds: Age 5-30 years	1.7356	
Smcds: Age >64 years	-0.0434	
St. Dev. Smcds Radj	0.0016	
Dummy regio	Yes	

Table 2.17 Mixed Logit estimates of patient hospital choice (distant vs far hospitals)

	Model 1a		Model 1b	
Dependent variable	CHOICE		CHOICE	
Log-likelihood	-62150.00		-3819.93	
Restricted Log-likelihood	-153523.25		-7959.43	
Chi squared	182746.49		8279.01	
Degree of freedom	28		28	
P-value	0.000		0.000	
Pseudo- R^2	0.595		0.520	
Halton	50		50	
N. of admissions	45138		2575	
N. hospitals	30		22	
<u>Distant hospitals</u>				
Distance	-0.0480	***	0.0021	
St. Dev. Distance	0.0247	***	0.0030	***
Case mix index	-0.4180	**	-3.4261	**
St. Dev. Case mix index	0.1201		0.7523	**
Size (number of beds)	-0.0004	***	-0.0028	***
St. Dev. Size	0.0003	*	0.0031	***
Teaching	-0.5170	***	0.9202	***
St. Dev. Teaching	0.0044		0.8473	**
Rac	-0.7423	***	-0.9549	
St. Dev. Rac	0.0434		2.1729	***
Market share	-2.4575	***	-1.5151	
St. Dev. Market share	0.9542	**	2.4711	***
Ami Radj	0.0285	***		
St.Dev. Ami Radj	0.0002			
Smcds Radj			0.0928	*
St. Dev. Smcds Radj			0.0012	
<u>Close hospitals</u>				
Distance	-0.0754	***	-0.0561	***
St. Dev. Distance	0.0347	***	0.0062	
Case mix index	-6.1583	***	5.2762	***
St. Dev. Case mix index	0.1837		0.2394	
Size (number of beds)	0.0063	***	-0.0078	***
St. Dev. Size	0.0012	***	0.0023	**
Teaching	-1.1705	***	2.1035	***
St. Dev. Teaching	0.0054		0.0966	
Rac	1.3965	***	3.3920	***
St. Dev. Rac	0.0422		1.4248	***
Market share	-0.4820	***	1.8067	*
St. Dev. Market share	0.0220		0.2254	
Ami Radj	0.0964	***		
St.Dev. Ami Radj	0.0105			
Smcds Radj			-0.0568	
St.Dev. Smcds Radj			0.0217	

2.6 Conclusions

In this paper, we investigate the effect of the main characteristics on patient hospital choice in case of cancer and five MDC treatments. Using data from all patients enrolled in a Sardinian LHA in the year 2013, we exploited the geographical location of Sardinia to model hospital admissions by means of Conditional and Mixed logit models. Our findings substantially confirmed what we expected on the basis of the literature related to hospital choice. The most effective pull factors are short distance, case mix, the number of beds and hospital type (teaching or private accredited).

However, several differences in the estimated effects can be drawn from some indicators when imposing parameters for close and distant hospitals. Focusing on the MDC sample, we found that the *Case mix* index and *Distance* affect negatively the odds of a hospital being chosen, even if the effect is smaller for distant hospitals. Moving on the econometric analysis of cancer treatments, we find that, for distant hospitals, the choice probability is inversely correlated to the hospital case mix. Probably, after receiving medical consultation outside the region, patients prefer receiving expensive treatment in a hospital near to their home. More interestingly, unless unexpected, we find that *Distance* has a positive effect on patient distant choices, thus indicating that beyond a certain distance, patients are less concerned by additional traveling distances. Furthermore, not in line with the literature, results for distant hospitals shows that teaching hospitals are not positive predictors of hospital choice.

Future research should include data on Southern Italian LHAs in order to better understand the characteristics driving patients' choice to distant hospital care, and whether the existence of bordering effects change the empirical evidence discussed in this work.

Chapter 3

Drinking and related hospitalizations: evidence from alcohol bans in Italy

Alcohol is a leading cause of death and disability. According to the Global Status Report on Alcohol and Health, alcohol consumption claims over 3 million of lives in 2012; 6% of all global deaths (WHO, 2014). In the same vein, the Italian National Institute for Health (ISS) estimates that excessive alcohol consumption is responsible for about 4% of all male deaths (2% of female) and roughly 37% of male (18% for female) deaths due to motor vehicle accidents.

Assessing the damaging effects generated by alcohol use is relevant not only from a pure public health perspective, but also from an economic and social point of view. Besides the economic value related to the passing away of young lives, it has been widely recognized that alcohol consumption is a source of negative externalities contributing to the deterioration of working performance, absenteeism and unemployment (Terza 2002, MacDonald et al. 2004, Johansson et al. 2007, Bacharach et al. 2010, Bockerman et al. 2015), to the reduction of family income and to the increase of crimes (Heaton et al. 2012). Furthermore, alcohol consumption is a contributory cause of injuries (Cherpitel et al. 2003), automobile and motor vehicle accidents (Levitt and Porter 2001, Taylor et al. 2010).

During the last twenty years, regulation policies have been adopted by several European countries with the aim to reduce heavy drinking and alcohol attributable injuries. For instance, in 2008, following the Intoxicating Liquor Act, Ireland introduced limitations on the off-premise sales of alcohol between 10.30 am and 10.00 pm during the week and from 12.30 pm to 10.00 pm on Sundays. In Lithuania, in 2009, the government revised the *Law on Alcohol control*, banning the sale of alcoholic beverage in retails stores between 10 in the night and 8 in the morning.[22]

In Italy, a national limitation of night alcohol sales was introduced with the article 14 of the law 30 March 2001 n.125, which banned the on-trade sale of hard liquors on motorway service areas from 10 p.m to 6 a.m.[23] A major strengthening of this ban, still limited to motorway services areas, was introduced on July 2010, by means of the article 53 of the law 29 July 2010 n.120, which starting form 13th August 2010 also prohibited the off-trade sale of hard-liquor between 10 in the night and 6 in the morning and the on-trade sale of "soft" alcohol beverage from 2 to 6 in the morning.[24] Through these legislative interventions, the main goal of the Government was to reduce the number of deaths and injuries caused by car and motor-vehicle accidents. As a by-product, making it more difficult the nightly consumption of alcohol is also expected to reduce the effect of alcohol on population health.

The purpose of this study is to assess whether the enforcement of the 2010 law has been a successful policy: a) does this stricter regulation contribute to reducing road accidents? b) do the policy achieve the indirect goal of reducing alcohol-related intoxications? Due to data availability (the administrative health data are available only from 2001 onwards), we will focus on the most recent legislative intervention (the 2010 law), though it would have been interesting to estimate either the effect of the 2001 law and the relative contribution of the 2010 law which made the alcohol ban much more restrictive.

Because the ban applies only to a subset of Italian roads (i.e. motorways) the implementation of the policy can be seen as a (quasi) natural experiments.[25] Using data from several administrative sources, we can exploit the heterogeneity between admissions occurring in hospitals located in

---

[22] With the exception of alcohol sold on international trains, boats, airplanes and in duty-free shops prohibiting also the sale of alcoholic drinks on tap as well as the consumption of alcoholic units in open packaging inside cars.

[23] The Italian roads are classified as follows: A) highways (characterized by dual or more carriageway.) B) Strade extraurbane principali; C) Strade extraurbane secondarie.

[24] According to the Italian legislation, alcohol beverages are termed "hard liquors" when the alcohol content in volume exceed 21%.

[25] A natural experiment "*examines outcome measures for observations in treatment groups and comparison groups that are not randomly assigned*" (Meyer, 1995).

municipalities sufficiently close to the motorways (these hospitals represent the treatment group), and admissions occurring in hospitals that are located in municipalities at a longer distance from motorways (which represent the control group).

We assume that the enforcement of the law has primarily affected those individuals who have been admitted in hospitals, or who live in municipalities that are closer to motorways (where the ban is at place). We will focus only on the second legislative intervention, though the two are clearly related. Difference-in-differences (DD) approach is used to identify the effect of the introduction of the ban on the number of car/motor vehicle accidents and alcohol-related hospitalizations. In the literature on alcohol bans, this approach has been used, inter alia, by Marcus and Siedler (2015) in Germany and by Green and Navarro Paniagua (2016) in England/Wales and Spain. We extend the analysis by estimating an alternative model that considers geographical distance as an explanatory variable. Because the alcohol ban in force on motorways should not affect individuals who do not drive in motorways as well as those who live, or work, sufficiently far from them, we expect that the smaller the distance between hospitals municipalities (patients' municipalities) and the motorways, the lower the number of alcohol-related hospitalizations, including car and motor vehicle accidents. For this reason, we expect that the impact of the alcohol ban vanishes as the distance between hospitals municipalities (patients' municipalities) and motorways increases.

We estimate basic DD models and DD models that include explanatory variables distinguishing between treated and control hospitals municipalities. We also present results for the same DD models estimated after aggregating admissions by patients' zip codes, thus distinguishing between treated and control patients municipalities.

In the following sections, we review the literature on alcohol control policies. Then in Sections 3.2-3.8 we describe the identification strategies and the data, while the econometric models we are estimating in our analysis are presented in section 3.9. Finally, Section 3.10 contains the results of the study and a discussion.

3.1 Related literature

Over the years, as a way to reduce social externalities due to alcohol use, many developed countries have introduced a number of regulatory policies. Example of those policies are provided in Table 3.1 and include, for instance, Pigovian taxes which compensate for the social costs inflicted by alcohol consumption; advertising regulation policies which target all consumption by preventing people from alcohol exposition in TV, magazines and outdoor billboards; drink-driving restrictions with the main goal to reduce injuries and road traffic fatalities; legal minimum age restrictions for consumption and / or purchase of alcohol.

These regulatory policies have aroused interest both in the epidemiological and public health literature as well as in the economic literature. For instance, Anderson et al. (2009) find that policy aimed at regulating alcohol availability and drink-driving policies have a positive effect in decreasing alcohol-related harm; while estimating the effect of educational programs, they find that there is no significant effect on alcohol harms reduction.

Table 3.1 Alcohol policies

Target	**Pricing policies**	**Regulation / enforcement**	**Education**	**Health care**
All consumption	Tax increase	Advertising regulation		
Heavy use / dependence	Minimum price		School-based programmes	Brief interventions & treatment of dependence Workplace interventions
Injuries		Drinking-and-driving restrictions Opening hours regulation		

Source: Sassi et al. (2015)

Similarly, Sassi et al. (2015) evaluate the impact of different alcohol policies in Canada, Czech Repubblic and Germany, finding that policies based on school education are less effective in reducing alcohol harm vis à vis brief intervention and pricing policies (i.e. tax increases). Several empirical analyses on the impact of alcohol control measures on people' behaviuor and health have also been provided by the economic literature (for a review see Carpenter and Dobkin 2011; Kenkel and Sindelar, 2011).

We next summarize the literature evidence on the effectiveness of alcohol policies in reducing negative externalities associated with alcohol consumption.

3.1.1 Pricing policies

Similar to the consumers in the market, alcohol drinkers may react strategically to higher alcohol prices by lowering their demand. Starting from this simple economic consideration, it appears obvious that excise taxes and minimum price represent good instruments to reduce social externalities associated to unhealthy behaviours. As a proof of this argument, Manning et al. (1995) using data on alcohol consumption from the National Health Interview Survey for the year 1983, show that the price-elasticity of demand influence alcohol consumption and that both categories of less and heavy drinkers are less sensitive to price chances. In the same line, Chaloupka et al. (2002) show how rises in alcohol costs cause a decline in drinking and heavy drinking. Moreover, through a review of the economics literature on alcoholism and its effects, Cook et al (2002) note that alcohol prices influence drinker behaviours and alcohol related harm. Another evidence is reported by Farrell et al. (2003), who evaluate the impact on addiction of changes in the cost of alcohol through the use of data on the United States. The researchers reveal that higher alcohol prices generate both reductions on alcohol consumption and decrease on addiction. However, the sensitivity of consumers to price changes is not uniform leading to different alcohol consumption patterns.[26] For instance, Nelson (2013) argue that a 1% increase in the price of alcohol is estimated to lead to a decrease in consumption of around 0.32 for beer, 0.57 for wine and 0.67 for spirits.

There is marked evidence that variation in tax excise between types of alcohol beverage products have the potential to affect alcohol consumption and related harm. For example, looking more closely at the effect of excise taxes on alcohol consumption, Daley et al. (2012), in a study

[26] The price-elasticity of demand is used in economics to define the percentage change in a good consumption following changes of 1% in its price.

conducted in the United States, estimate that a tax increase of 25 cents for alcoholic beverage would have the effect to decrease alcohol consumption for heavy drinkers of 58.6 units annually (-11.4 %). Similarly, Delcher et al. (2012) analysing data on the state of New York show that a tax increase on beer and on spirits is associated with a 7% reduction in mortality caused by alcohol-related diseases. Moreover, Ruhm et al. (1996), focusing on the effect of beer taxes on the number of motor vehicle fatalities in 48 contiguous states of America and using a fixed effect model, find that higher beer taxes seem to influence the number of road accident deaths.

3.1.2 School Education programs

Alcohol consumption, especially heavy drinking, affect socio-economic outcomes such as educational achievements and scholastic performance, crime, as well as drink-driving fatalities. As an example, Carrell et al. (2011) examine, through a regression discontinuity approach, the effect of alcohol consumption on student academic performances in the USA, finding that legal access to alcohol lead to a decrease in academic scores.[27] Similarly, using data from the University of Oregon between winter 1998 to winter 2007, Lindo et al. (2013) analyse the impact of legal drinking age on student performance. Their results show that alcohol consumption leads to a reduction of academic performance, although the magnitude of the effect is lower than the one detected in Carrell et al.

As a way to avoid or, at least, limit the negative effects generated by alcohol consumption, several prevention initiatives on adolescent have been performed. School alcohol education programmes are among those strategies used by government to modify young drinking behaviours and attitudes, to limit the use of alcohol, to prevent alcohol intoxications (e.g. binge drinking) as well as alcohol related harm (Babor et al. 2010). Three main phases can be detected in the evolution of school based intervention programmes on alcohol. The first phase, between 1960 to 1980, consisted on the provision of knowledge on the negative consequences associated to alcohol misuse through teaching sections and other information strategies. The second phase entailed *affective education* programmes based on "*broader issues of personal development such as decision making,*

[27] Following the National Minimum Drinking Age Act of 1984 In the USA, the legal drinking age is 21 years.

values clarification and stress management” (Cuijpers 2003). However, these approaches based on the spread of knowledge and values clarification have proven to be ineffective in changing adolescent behaviours (Hansen 1994, Botvin 1995, Babor et al. 2010).[28] Thus, in the third phase, from 1980 onward, social influence models have started to be used as effective programs to change adolescent drinking attitudes.

That said, although a bunch of school-education interventions on alcohol abuse have been used over the years to control adolescent drinking behaviour, there is mixed evidence on the effectiveness of these interventions in reducing alcohol misuse. On the one hand, review of the literature analysing the effectiveness of school-based intervention in reducing alcohol abuse on young people, shows the absence of a clear evidence of those interventions in preventing alcohol misuse (NICE 2007). On the other hand, school-based program associated to family therapy or brief intervention seem to be the most promising strategies to delay alcohol consumption and its negative effects (Babor et al. 2010).

3.1.3 Workplace intervention measures

Alcohol is a psychotropic substance able to influence individual behaviour and its socio-economic outcomes. Over the years, it has been proven that impairment in human cognitive abilities due to alcohol abuse may directly affect the labour productivity by way of unintentional injuries, absenteeism, fights and other inadequate behaviours. For instance, using data provided by the Alcohol Supplement of the National Health Interview Survey for the 1988 and applying an instrumental variable approach, which was used to control for the potential endogeneity of problem drinking, Mullahy and Sindelar (1996) find that problem drinking results in employment reduction and difficulties in the likelihood to obtain a job. Using the same survey data, Terza (2002) reviews the work of Mullay and Sindelar showing that problem drinking was associated with statistically

[28] For a complete review on educational programs and strategies on alcohol use see chapter 13 in Babor et al. (2010).

significant reductions in the probability of being employed. Similar effects were also found by MacDonald and Shields (2004) and Johansson et al. (2007).

Education, training programs, counselling and brief intervention fall among the successful strategies aimed at reducing negative effect of alcohol use on the workplace. For instance, a systematic review of the literature shows that brief intervention, psychological skills, training and peer referral may represent a successful approach to handle alcohol-related harm (Webb et al. 2009).

Several programs have been implemented, at different levels, in order to reduce negative consequences caused by alcohol abuse in the workplace. An interesting example is shown by the International Civil Aviation Organization (ICAO) program. The ICAO established specific drug and alcohol policies, which consist on screening programs, training and educational session, work re-integration for employee which have followed rehabilitation programs, to reduce alcohol misuse and alcohol related-problems in aircrews.

3.1.4 Regulation enforcement policies

Regulation policies aimed at reducing the number of alcohol sale licenses and their hours of activity are considered measures that may affect the negative externalities caused by alcohol abuse (Kenkel and Sindelar 2011). In the last twenty years, a large body of literature has documented the influence of these specific policies on health and socio-economic outcomes. For instance, several studies have focused on the effect of policies regulating days and/or hours of alcohol sale on consumption behaviour and related harms. Through a systematic review of the literature, Popova et al. (2009) find that policies aimed at regulating the availability of alcohol are effective in reducing not only consumption, but also its adverse consequences on the society. In the same line, reviewing the international literature on the effects associated to the increase in the hours of alcohol sale in UK, Australia, New Zealand and North-America, Stockwell and Chikritzhs (2009) highlight the relationship between opening hours and drinking-related problems. Among the main studies cited therein, Carpenter and Eisenberg (2009) evaluate the impact of repealing Sundays alcohol sales restriction using individual-level data on alcohol consumption in Ontario, Canada. They show an increase of 7 to 15% in Sunday alcohol sale following the aforementioned liberalization on alcohol sale ban. Additional evidence comes from a study conducted by Middleton et al. (2010), who examine the effect of the liberalization on alcohol sale restrictions during Saturday and/or Saturdays. Reviewing the literature focused on the effectiveness of restricting or preserving limits on days of alcohol sale, they observe an increase in drinking consumption and alcohol related problems, such as motor vehicle accidents and assaults, as a result of the increase in days of alcohol sale. However, the evidence on the effect of alcohol sale liberation is mixed. Vingilis et al. (2005) estimate the impact of the introduction of the Liquor License Act in Ontario, Canada, which increase alcohol trading hours from 1 am to 2 am. Using a quasi-experimental design, the authors find that one-hour increase in alcohol sales generate a modest increase in alcohol consumption and related harms.

Other studies examine the effect of restricting alcohol trading on consumers' behaviour with respect to crime as well as work absenteeism. In particular, implementing a difference-in-differences approach (DD) on data from the metropolitan area of São Paulo, Biderman et al. (2009) show how the adoption of dry laws cause a significant reduction in the rate of murders and deaths from motor vehicle accidents. In a similar way, again using a DD approach, researchers as Heaton (2012) and Gronqvist and Niknami (2014) document the impact of alcohol sales liberalization on crimes. Specifically, the strategy adopted by Heaton (2012) consists of analysing the effect of a policy repealing Sundays' alcohol sale in the State of Virginia on crime. They analyse differences between crimes occurred during workdays with those on Sunday before and after the policy change, and conclude that expanding alcohol sale on Sundays generates an increase on crime for about 5 to 10%. Gronqvist and Niknami (2014) assess the effect of expanding alcohol sales of one day on the number of crime committed by Swedish. Comparing Saturdays crimes with crimes occurred during weekdays, they conclude that the extension of alcohol availability increase crime. Exploiting the quasi-natural experiment provided by existing policies in England, Wales and Spain, a recent study by Green and Navarro-Paniagua (2016) shows also the existence of a positive relationship between local opening time on alcohol sale and work absenteeism.

Most closely related to our research, other studies have analysed the correlation between regulation on days and/or hour of alcohol sale and the number of alcohol-related hospitalizations as well as road accidents. Wichi and Gmel (2011) estimate the effects of a ban on the night alcohol sale at petrol stations and video libraries from nine in the evening and seven a.m. in the canton of Geneva in Switzerland. They find evidence of a 40% drop in hospital admissions due to the use of alcohol. Similarly, Marcus and Siedler (2015) evaluate the impact of a ban on alcohol sales, in force in 2010 in the state of Baden-Württemberg in Germany, on alcohol-related hospitalizations using data covering the period 2007-11. By using a DD and synthetic control approaches, the analysis shows a reduction of hospitalizations of about 7% among adolescents and young adults due to the alcohol ban. A similar result also emerges from the study by Green et al. (2014). They compare differences between road accidents in England and Wales (treatment states) with those that occurred in Scotland (control state) before and after the liberalization in the hours of alcohol sale in England and Wales, and show that the introduction of the new policy has had a positive effect on the reduction of road accidents in treated states. However, the evidence on the effect of policies in reducing social negative externalities linked to alcohol is mixed. For instance, analysing the effect of a restriction on alcohol sales on Sundays in 15 US states, Lovenheim and Steefel (2011) find no significant change in alcohol consumption habits and a weak effect on the number of vehicle fatalities.

3.2 Econometric method

The difference-in-differences design is commonly used by researchers to evaluate the effect of public interventions and/or policy changes. In this line, DD techniques measure changes in outcomes over time by looking at the treatment groups (those exposed to a treatment in the second period of time) and the comparison groups (those not exposed to the treatment during both periods of time) before and after the policy intervention.

DD approach has been widely used by researchers in the last thirty years, however the basic idea has a very long history. Snow (1855) was one of the firsts researchers to use a DD approach to study the effect of changes in water supply, in a specific district in London, on the transmission of cholera. Considering the quote written by Ralph R. Frerichs on the introduction of Snow's book, the basic idea of Snow consisted on performing "… *a natural experiment involving two London water companies, one polluted with cholera and the other not. He demonstrated that persons who received contaminated water from the main river in London had much higher death rates due to cholera…*" (in Lechner, 2010).

In economics, applications of the DD approach date back to the early twentieth century. Obenauer and von der Nienburg (1915) exploited the introduction of the minimum wage in the state of Oregon, USA, for a particular group of women working on retail stores, compared to the other states, to verify the effect of a legal minimum wage on several outcome variables (e.g. women efficiency, labour costs) before and after the state intervention. Another early example of application of DD method, in labour economics, is provided by Lester (1946). He employs a DD method comparing employment levels, of groups of firms with low average wages to groups of firms with higher wager, prior and after different minimum wage increases, to analyse employment effects on minimum wages.

To be more precise, the widespread diffusion of the DD technique occurred, when, according to Lechner (2010), economic researchers began to use more and more "*changes in state laws and regulations to define pre-treatment periods (prior to the introduction of the policy) and unaffected comparison groups (states having a different policy than the one of interest)*". An example is offered in a study published by Simon (1966), who estimate the arc elasticity by comparing state liquor sales before and after price rises, with other states not experiencing a liquor price change.

One of the most famous application of the DD methods is due to Ashenfelter and Card (1985). Using longitudinal data on trainees' earnings and a comparison group, they evaluate the effectiveness of training for participants in the 1976 CETA programs by means of a DD model.

Similarly, DD approaches have been used to analyse the impact of minimum wage on employment. An example is provided by Card and Krueger in 1994. The researchers used DD approach to evaluate the effect of an increase in the minimum wage, from \$4.25 to \$5.05 per hour, in New Jersey on employment. In details, comparing employment, wages and prices at fast food stores in both New Jersey and Pennsylvania (the comparison group) before and after the wage rose, the author were able to estimate the effect of the new government regulation.

Other researchers, instead, as we have seen before, have used DD approach to evaluate the effect of various types of alcohol ban on people's health and socio-economic outcomes.

3.2.1 Notation of the DD approach

Table 3.2 below shows the different components of the DD model. Let define the treatment variable, T, as a binary variable $d \in \{0, 1\}$, where T=1 indicates the treatment group. In practice, in the DD set-up, variables of interest are measured in two, or more, time periods, so that $t \in \{0, 1\}$. Thus, t equal to 0 indicates the time prior to the treatment, while t equal to 1 defines the period of time soon after the treatment is applied. Having said that, the first row of the table displays outcome variables for the treatment groups before, $E(Y_{t+1}^{0}|T_i = 1)$, and after, $E(Y_{t+1}^{1}|T_i = 1)$, the policy intervention. Similarly, the second row of the table presents outcomes for the control groups, equal to $E(Y_{t+1}^{0}|T_i = 0)$ before the intervention and $E(Y_{t+1}^{1}|T_i = 0)$ after the policy implementation.

Table 3.2 Component of the difference-in-differences model

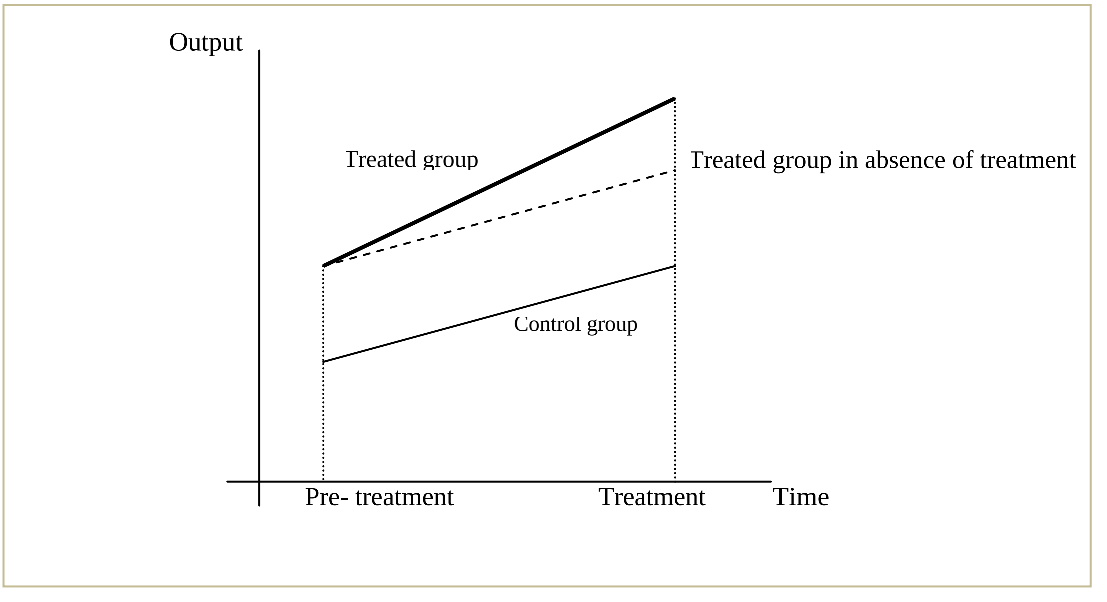
	Before the policy application	After the policy application	Difference
Treated group	$E(Y_{t-1}^{0}|T_i = 1)$	$E(Y_{t+1}^{1}|T_i = 1)$	$E(Y_{t+1}^{1}|T_i = 1) - E(Y_{t-1}^{0}|T_i = 1)$
Control group	$E(Y_{t-1}^{0}|T_i = 0)$	$E(Y_{t+1}^{1}|T_i = 0)$	$E(Y_{t+1}^{1}|T_i = 0) - E(Y_{t-1}^{0}|T_i = 0)$

The difference in outcomes among treatment groups before and after the policy intervention is given by the following difference: $E(Y_{t+1}^{1}|T_i = 1) - E(Y_{t-1}^{0}|T_i = 1)$, while the average changes among counterfactuals is obtained by calculating the difference in outcomes before and after the policy applications $E(Y_{t+1}^{1}|T_i = 0) - E(Y_{t-1}^{0}|T_i = 0)$. Then, the average effect identified by the difference-in-differences is:

$$DD = [E(Y_{t+1}^{1}|T_i = 1) - E(Y_{t-1}^{0}|T_i = 1)] - [E(Y_{t+1}^{1}|T_i = 0) - E(Y_{t-1}^{0}|T_i = 0)]$$

The crucial assumption for the DD to identify the effect of the policy is that in absence of the treatment, the difference between the treatment and comparison group is constant over time (Fig. 3.1). This condition is called by researchers "*Parallel paths assumption*".

Figure 3.1 The parallel-path assumption



It is well known that violation of parallel trend assumption will lead to biased estimation of the causal effect. But then, are there statistical tests to inspect the validity of the parallel trend assumption? At the moment, there are not statistical tests to verify whether or not the parallel trend assumption holds. However, to inspect the validity of the DD, several methods have been provided by researchers. Angrist and Krueguer (1999) argue that in order to validate the parallel trend assumption is important that trends are similar before the event of interest. Visually inspection is a good way to check the validity of the parallel trend assumption in presence of many time periods. Unfortunately, although pre-trends are similar, visual inspection is not able to consider the potential distortive effects caused by other policies changing in the same time span. For this reason, one way to “*test”* the validity of the parallel path consists on including, in the model estimate, leads and lags of the treatment. In order to understand how this “*test*” works, let consider a binary model:

$$Y_{ist} = \gamma_s + \lambda_t + \beta D_{st} + X_{ist}\delta + \varepsilon_{ist}$$

where γ_s denote the state fixed effect, λ_t the time fixed effect, D_{st} is a dummy indicating the treatment effect and ε_{ist} is the error term. The idea behind this test consists on including in the model *m* “leads” (as interactions between the treatment indicators for the pre-treatment periods and the time dummies) and q “lags (as interactions between the time dummies and the post-treatment periods) so that the model will be:

$$Y_{ist} = \gamma_s + \lambda_t + \sum_{j=-m}^{q} \beta_j D_{st}\,(t = k + j) + X_{ist}\delta + \varepsilon_{ist}$$

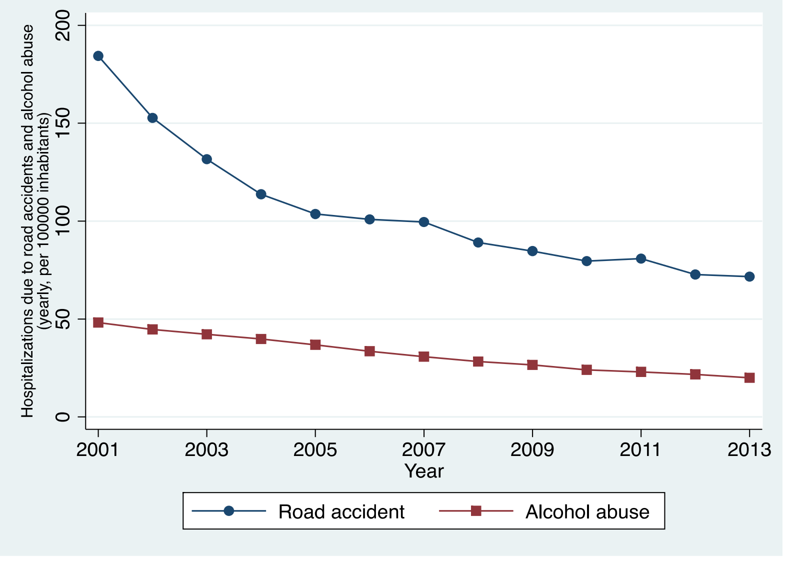
where *k* is the time at which the policy intervention come into effect and β_j the coefficient related to the *j*th lags or leads (Pischke 2005). If the outcome trends between treatment and control groups are the same, then the β_j coefficients for the pre-treatment period should not be statistically significant. An application of leads and lags to verify the parallel path assumption is provided by Autor (2003). He introduced both leads and lags in a DD model to investigate the effect of increased employment protection on the firm’ use of temporary help workers.

3.3 Identification strategy

As we have seen in the introduction, the main nationwide alcohol bans were introduced in Italy in the first decade of this century. Did these reforms obtain the expected effect? More in general, do night sales bans constitute an effective policy? A naïve interpretation of Figure 3.2 would lead to optimistic conclusions. Between 2001 and 2013, the accident-related hospitalization rate displays a sharp decrease moving from about 184 cases per 100,000 inhabitants to 72. A similar decreasing trend, though less strong, characterizes alcohol related hospitalizations, which have decreased of about 58% in 2013 with respect to 2001.

Many confounding factors could have played a role in the reduction of hospitalizations during the same period. As far as the road accidents are concerned, they include the improving of safety standards for cars and the penalty point system for driving licenses introduced in 2003.[29] In the case of alcohol related intoxications, the health promotion campaigns against alcohol abuse might have induced a change in individuals' drinking behaviour.[30]

Figure 3.2 Road accident and alcohol related hospitalization rates in Italy, 2001-2013.



Note: The figure shows road accident and alcohol related hospitalizations per 100000 inhabitants for the whole population over the periods 2001-2013 (Own elaboration on SDO data)

However, the peculiar nature of the legislative interventions, only referred to motorway service areas, can turn to be useful. Ceteris paribus, given that road accidents and alcohol consumption are certainly positively related, one should expect an additional negative effect on road accidents hospitalizations for those hospitals (municipalities) "near" the motorway network. Similarly, a negative additional effect on alcohol-related hospitalizations in the proximity of the motorway network could be in place, because the purchase of alcoholic drinks has become more difficult only in some geographical areas of the country. Henceforth, we discuss how the exploitation of this

[29] See De Paola and Scoppa (2013) for analysis of the effectiveness of this policy.
[30] See for example the National Alcohol and Health Plan (*Piano Nazionale Alcol e Salute*), Gaining Health 2007-2010.

geographical variation in the application of the ban enables us to adopt two different identification strategies. Due to the necessity of having the availability of ex ante and ex post information, we will focus on the 2010 major revision.

3.4 A diff-in-diff approach

If the policy has been effective, we expect a different evolution in the number of hospital admissions depending on the area in which they occurred (contiguous to motorways/ far from the motorways). Hence, the existence of an effect of the 2010 reform could be detected by analysing the differences between the number of discharges occurred in hospitals' municipalities located in areas close to the motorways (labelling them as "treatment group") and the remaining Italian hospitals' municipalities ("control group"); before and after the introduction of the policy by the Italian Parliament.

The concept of "area close to the motorway" could take the form of a geographical bandwidth. In this manner, the variation between admissions relating to hospitals within the bandwidth and those relating to hospitals located in distant geographic areas could be exploited. Then, as a way to reduce variability across data, admissions will be aggregated at hospitals' municipality level.

Let us define treated hospitals' municipalities those located within the bandwidth, and control ones those outside the bandwidth. In order to estimate an econometric model to assess the policy effectiveness, this dichotomous information can be easily used in the following regression model:

$$Y_{it} = \alpha_0 + \alpha_1 post_t + \alpha_2 treat_i + \beta(post_t * treat_i) + X'_{it}\eta + \varepsilon_{it} \tag{1}$$

where $post_t$ is a binary variable taking value 1 for the years after the implementation of the policy, 0 otherwise, $treat_i$ is a dummy variable equals to 1 for those hospitals' municipalities located within the bandwidth surrounding the motorway network and 0 for those located outside, and X'_{it} represents a vector of specific exogenous covariates.

A very difficult task is the definition of which geographical areas can actually be considered close or distant from the Italian motorway network. Using data on Italian hospital addresses we have geocoded hospitals' municipalities within a "*bandwidth"* arbitrarily set to 10 km.[31,32] This requires several steps. Firstly, we geocode motorway access points using data provided by POIGPS.[33] Then, we locate all motorway stations (i.e. tool booth) for which a fee is required for passage. However, as shown in Figure 3.3, since in some Italian regions as Calabria and Sicily different motorways stretches do not require a fee and are devoid of tool booth, we have also considered junctions as points of access to motorways. In this way, using both motorways tool booth and junctions we can identify access points for each Italian motorway (Figure 3.4). Secondly, by using hospital addresses, more than 1200 hospitals distributed in more than 560 municipalities have been located.[34] Finally, calculating a kilometric distance of 10 km from the motorway access points to the centroid of hospitals municipality, we have been able to define, not only a buffer distance of 10 km (our bandwidth) respectively to the right and to the left of the motorway access points (Figure 3.5), but also to detect hospitals admissions supposed to be mainly affected by the policy. In this way, it is possible to distinguish all hospitals' municipalities located close to the motorway bandwidth (treatment group), from those located far away (control group). Figure 3.6 illustrates the distribution of the hospitals municipalities in the Italian regions, which are concentrated in large part, as expected, along the motorways.

Starting from the idea that patients who are hospitalized in municipalities located within the bandwidth non necessarily live in those areas, to fully considered patients admissions affected by the policy ban, we have also decided to analyse the difference between the number of hospitalizations generated by patients living in municipalities close to the motorways ("treatment group") and those in distant municipality areas ("control ones") in both period of time (prior and

---

[31] Assuming close, an area that can be reach in 15 minutes, and considering a drive speed of 70 km/h, it has been possible to define the bandwidth of 10 km.

[32] Different geographical bandwidth, respectively at 14, 12, 8 and 6 km will be considered to check the sensitivity of ourestimates.

[33] POIGPS is an independent portal that provide point of interest related to the Italian roads, http://www.poigps.com/index.php. Data on motorway tool booth and junction are updated at February 2016.

[34] The software Quantum Gis has been used.

after the event of interest). Using data on patient zip codes, patients' municipalities of residence have been geocoded. Then, to identify patients' hospitalizations that might be influenced by the intervention, distances between the centroid of patient's municipalities and the closest motorways access point have been calculated. Fig. 3.7 and 3.8 below show the distribution of patient's municipalities along motorways respectively in the case of road traffic accidents and alcohol related hospitalizations.

3.5 An alternative approach: exposition to the policy ban

Applying the diff-in-diff approach strongly depends on the exact definition of what is close and what is far from the motorways. An alternative approach is possible, provided that one renounces to estimate the size of average treatment effects, and an exogenous (and possibly time-invariant) "exposure" variable is available.

For example, there is evidence documenting the effect of children exposition to chemical substances on specific health and socio-economic outcomes such as crime rates. If this is the case, then, the introduction of a regulation on atmosphere pollution emissions is expected to reduce the likelihood to commit future crimes specially for those children previously living in areas characterized by high lead exposition (Reyes, 2007). As another example, Acemoglu and Filkenstein (2008) investigate the effect of the introduction of the Medicare Prospective payment system (PPS) in the United States on October 1983 on firm input mix and technologies choices. In line with their theoretical expectations, they find a higher variation in hospitals characterized by a higher share of patients covered by Medicare, that is hospitals which were more exposed to Medicare change of reimbursement schemes.

Figure 3.3 Distribution of motorway toll booth stations

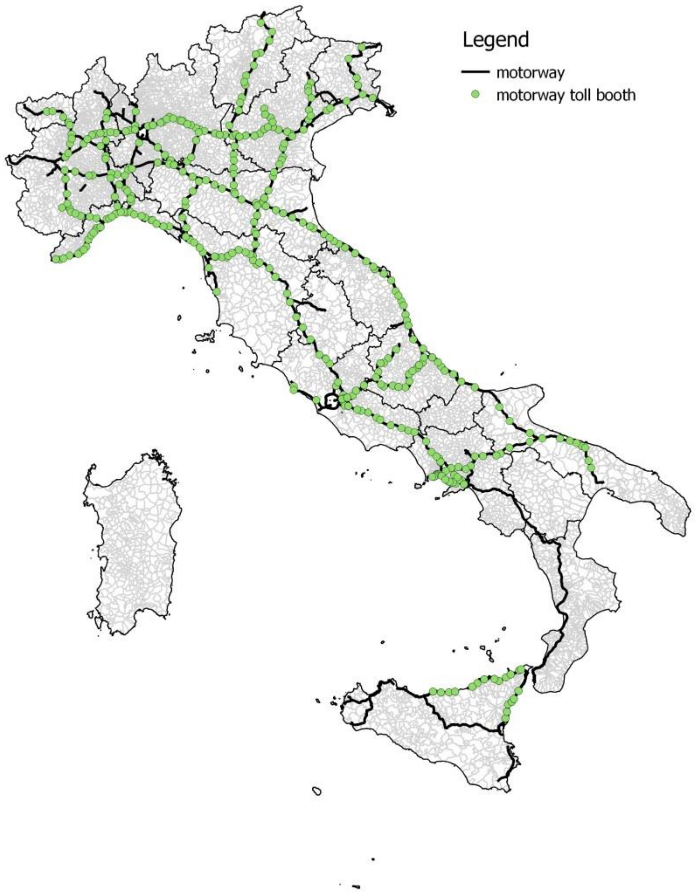


Figure 3.4 Location of motorway junctions

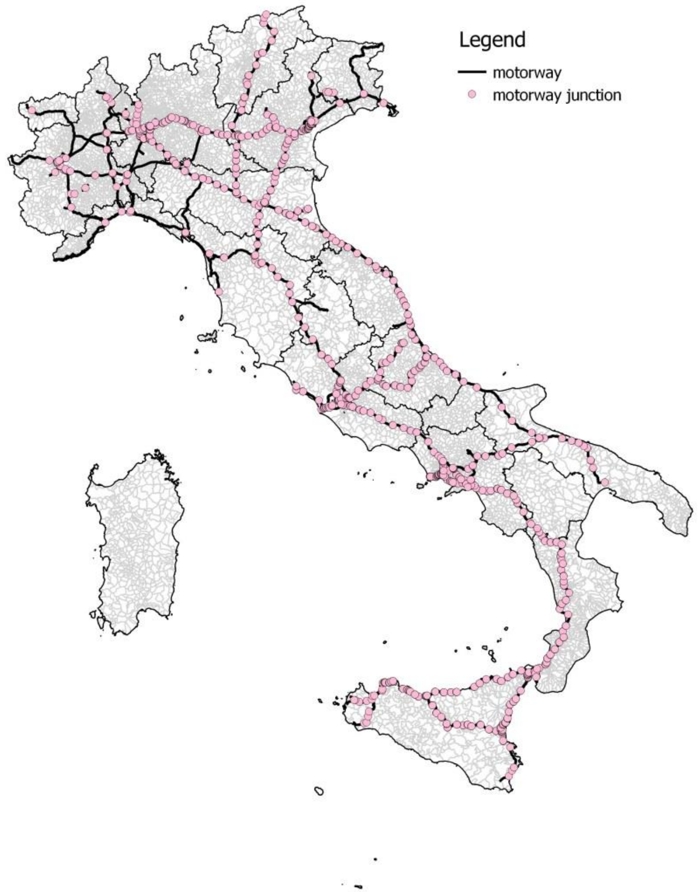


Figure 3.5 Motorways 10 km buffer distance

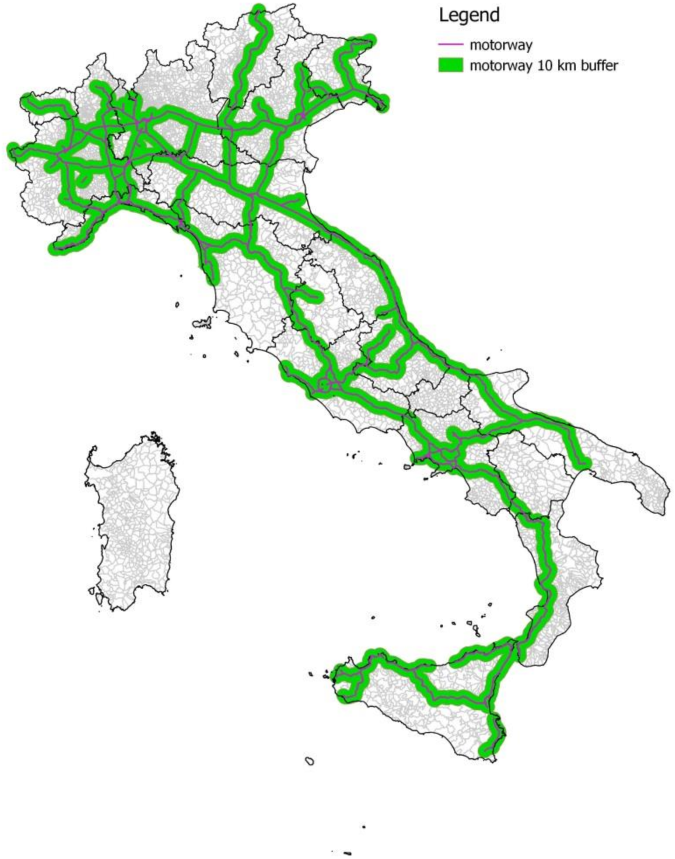


Figure 3.6 Distribution of hospital municipalities along motorways

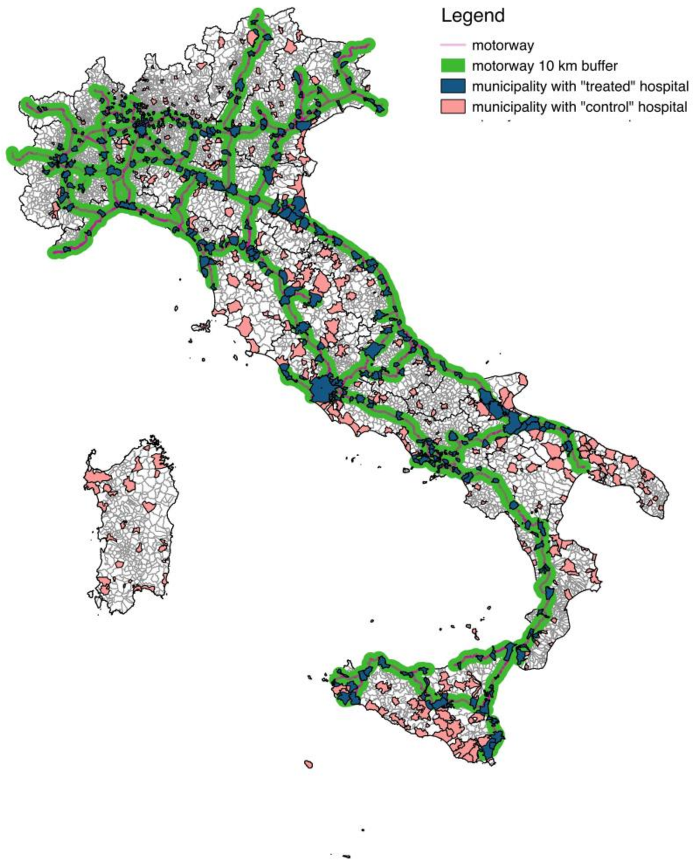


Figure 3.7 Distribution of patient municipalities of residence along motorways in the case of road traffic accidents hospitalizations

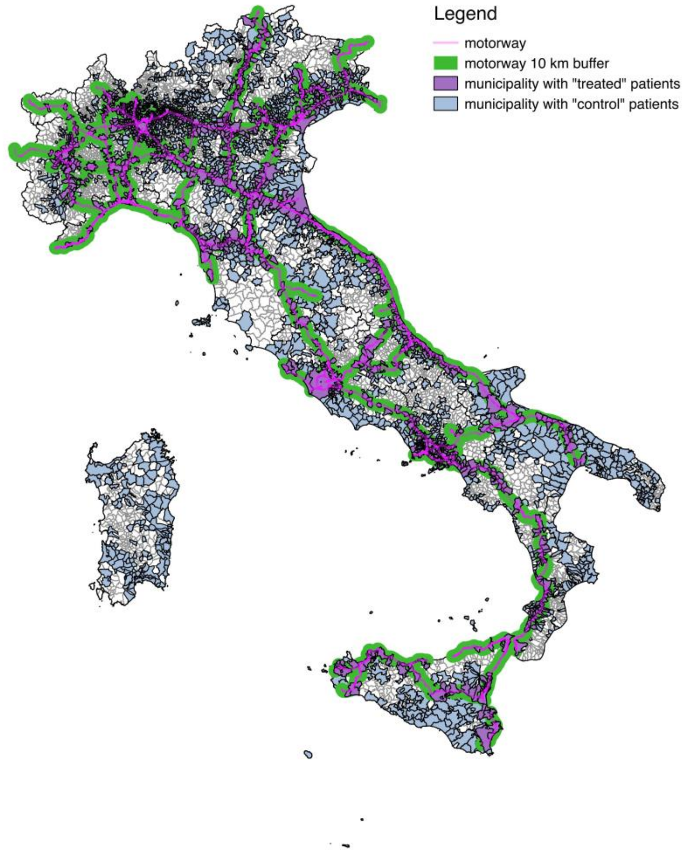
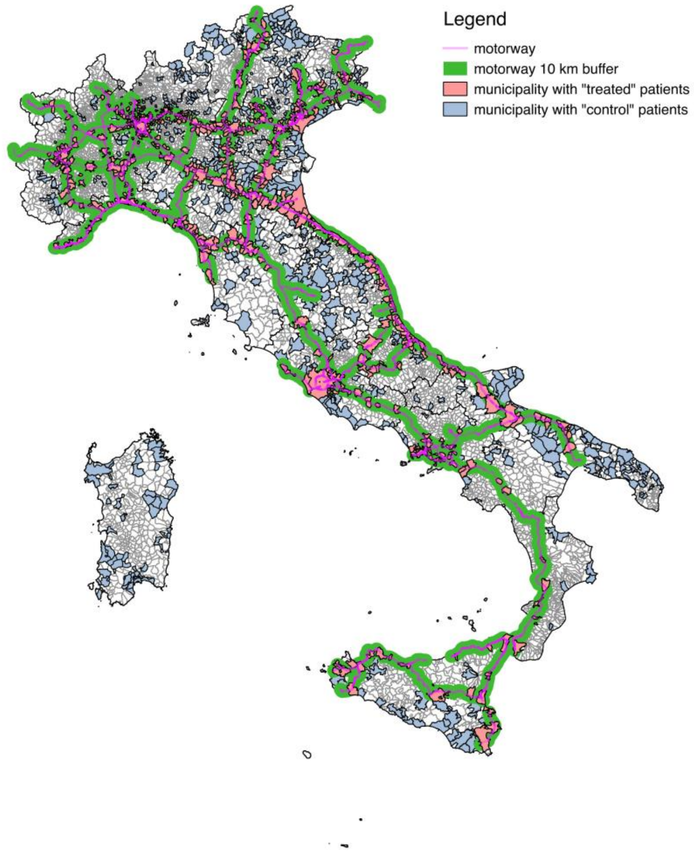


Figure 3.8 Distribution of patient municipalities of residence along motorways in the case of alcohol-related hospitalizations



Following a similar approach, we can choose as exposure variable a measure of the (exogenous and time invariant) closeness of hospitals, or municipalities, to the motorway network. We expect that the closer the hospital (municipality) to the motorways, the lower the number of hospitalizations due to alcohol intoxication and road accidents. Alternatively, the impact of the alcohol ban on motorway service areas should vanish as the distance between hospitals (municipality) and motorways increase. In practice, this implies carrying out a regression estimation with the inclusion of an interaction between the post-policy years and the inverse of the geographical distance between each hospital/municipality and the closest motorway access points:

$$y_{it} = \alpha_i + \gamma_t + \beta(POST_t \cdot c_i) + X'_{it}\eta + \varepsilon_{it}$$

where α_i are hospital or municipal fixed effects, γ_i are year dummies, X'_{it} represent a vector of specific covariates, and $POST_t$ is a dummy equal to 1 for the years following the policy change, and c_i is the inverse of distance.

Differently from our previous "10 km bandwidth approach", in this case we need a continuous indicator for distance. For these reasons, following a similar approach of that described in the previous paragraph, we firstly geocode motorway access points. Secondly, we continue by geocoding centroid of each hospital' municipality (patient's municipality). Figure 3.9 (3.11) and Figure 3.10 (3.12) below display the distribution of hospitals municipalities (patients' municipalities), with at least one hospital (or in the case of the analysis at patient levels, one discharge), along the national territory.

After having recovered the kilometric distance, the indicator ci can be constrained in the (0,1) interval. In particular, we set it equal to 1 if the hospital (municipality) i is located at less than 1 km from the closer motorway access point; and equal to 1/distance (in km) when the distances exceeding 1 km. We expect that the closer the hospital (municipality) to the motorways, the lower the number of hospitalizations due to road accidents and alcohol intoxication

Figure 3.9 Distribution of motorway access points and hospitals locations

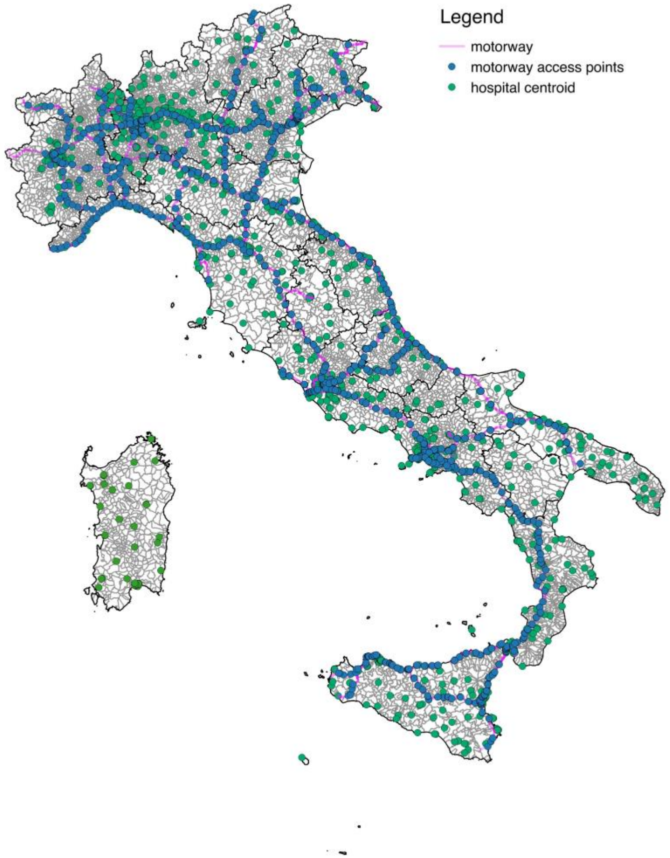


Figure 3.10 Distribution of motorway access points and municipality locations

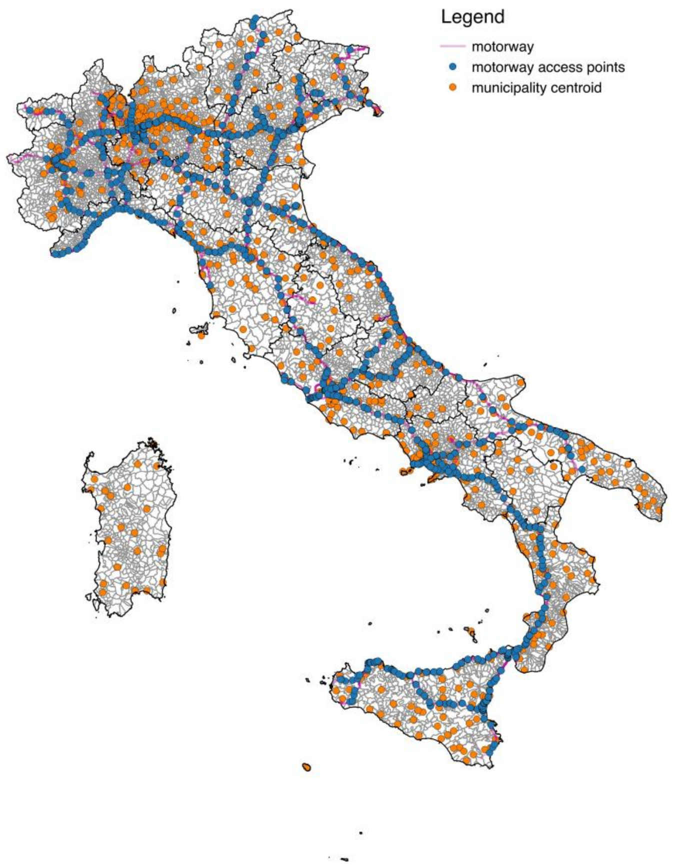


Figure 3.11 Distribution of motorway access points and patients' municipalities of residence in the case of road traffic hospitalizations

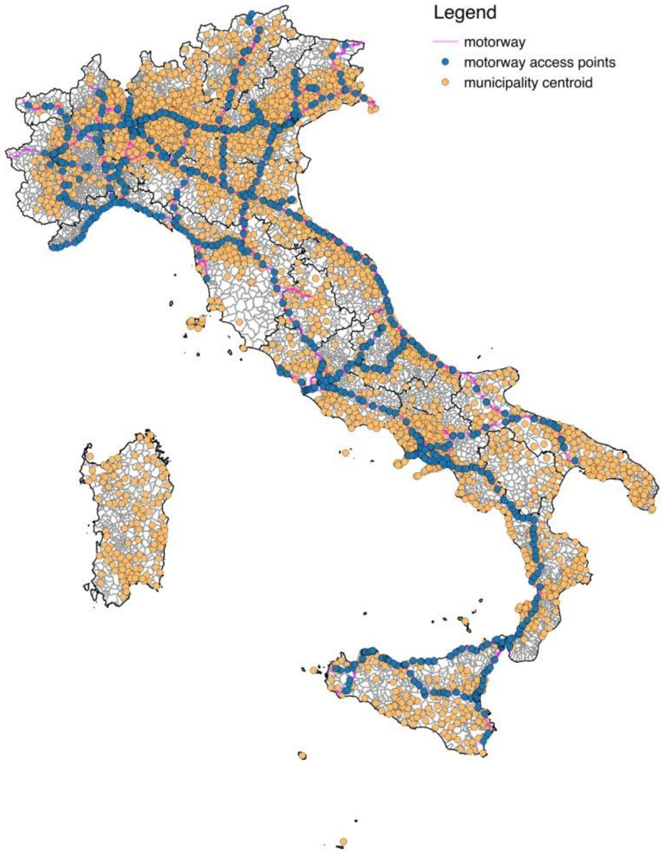
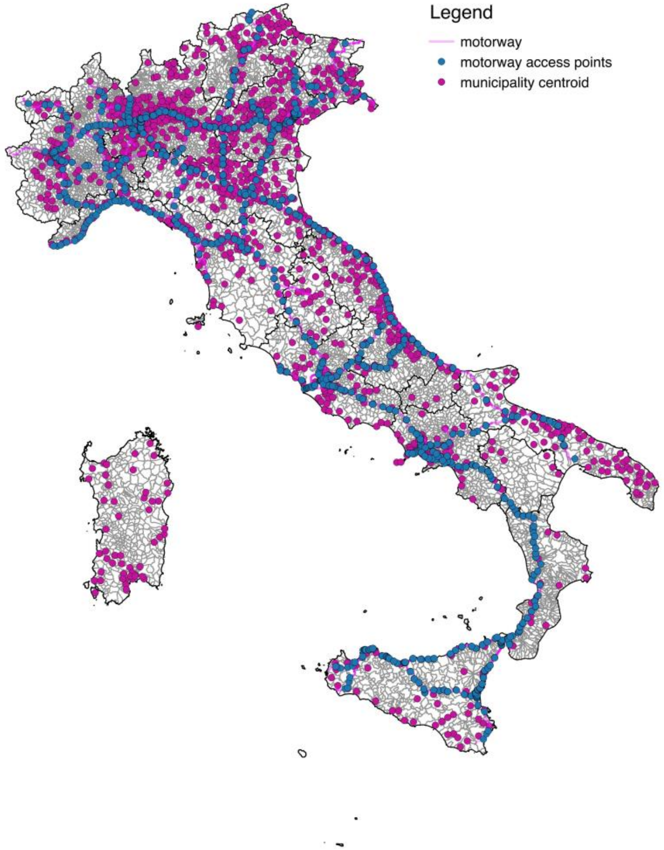


Figure 3.12 Distribution of motorway access points and patients' municipalities of residence in the case of alcohol related hospitalizations



3.6 Data and descriptive statistics

The 2010 legislative intervention on alcohol sales restrictions took place in a date for which full ex ante and ex post statistical information on hospital discharges is available. These research use administrative data provided by the Italian Ministry of Health and collected by all regions through the use of the "*Scheda di Dimissione Ospedaliera*" (SDO). The information collected by means of the SDO gives us detailed information on the socio demographic characteristics (e.g. sex, age classes, gender and province of residence), hospitalization features (e.g. day of admission and discharge), type of hospital (e.g. public or private) and clinical information (e.g. main diagnosis, diagnostic and therapeutic procedures) of all inpatients enrolled in the Italian hospitals over the period 2001-2013. It is important to notice that, although the SDO can provide information on secondary diagnosis, our dataset contains only information on the main patient's diagnosis. This means that the number of hospital admissions may be underestimated.

Notwithstanding the National Database of Hospital discharges enclose a rich set of information, it contains some inaccuracy mainly due to problems in the classification system, which vary over the years, as well as in the degree of correctness and completeness of some variable. For these reasons, a selection of the samples and specific procedures to clean data have been applied to our database in order to protect the quality of our study. At first, we proceed by eliminating all hospitalization records whose regions and Local Health Authorities (LHA) codes are incorrect or missing and all admissions made by foreigners in Italian Hospitals. Hence, we drop those records with an invalid code for sex, type of activity and hospitalization. Moreover, considering the variable related to the days spent by each patient in the hospital, we do not consider those hospitalizations which present a length of stay longer than 365 days.

To better analyse the effect of the motorway night alcohol sale bans, we have used these hospitalization data on a twofold manner:

- first, to evaluate whether or not the policy change has been a successful policy in reducing the number of road traffic accidents and fatalities;

- second, to inspect the effect of the policy ban on the number of hospitalizations caused by the abuse of alcohol.

We will produce appropriate descriptive statistics for the period 2001-2013. However, the econometric analysis will be uniquely based on data for the period 2007-2013, three years before and after the policy implementation.

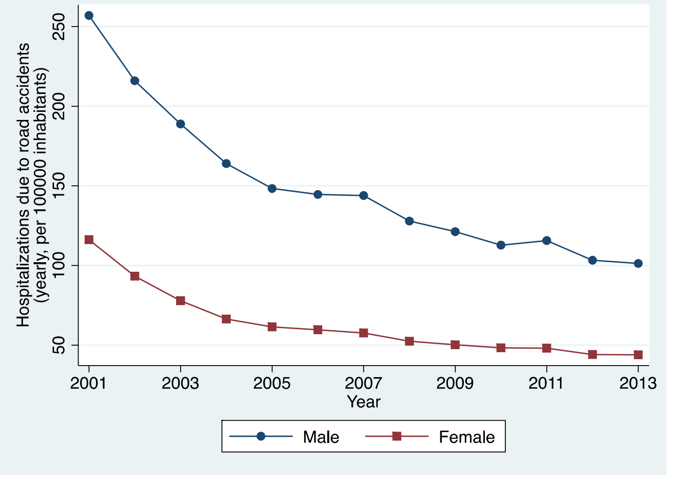
3.6.1 Descriptive statistics of general interest

<u>Road traffic accident hospitalizations</u>

Data on road traffic accident hospitalizations have been selected using information on traumatism admissions. In details, for each year considered in the analysis, we have chosen all hospital discharges which take value equal to three on the variable related to traumatism.[35]

Figure 3.13 plots trends of road traffic accidents grouped by sex. As we can see, the decrease is far greater for male when compared with female. In particular, while for female there has been a steady decline in the number of road accidents during all the thirteen years, for males, we assist to a different trend characterized in the year 2006/2007 and 2010/2011 by an increase in the number of accidents.

Figure 3.13 Road accident hospitalization rates, 2001-2013 (by sex)



[35] The variable "trauma" provided by the National Database of Hospital Discharges give us information about the presence or absence of traumatism for each hospitalization. It takes value equal to 1 when hospitalizations are caused by work accidents, equal to 2 in case of home accidents, equal to 3 whether hospital discharge are due to car accidents, a value equal to 4 in case of discharges related to violence versus other people, value of 5 when an attempt of suicide is recorded and it takes value equals to 9 in case of other accidents or poisoning hospitalizations.

From a more complete analysis, it appears that the reductions of road accident hospitalizations are larger for those aged 15-20, 20-24 and 25-29 (Figure 3.14). In addition, it is interesting to notice that, although there has been a decrease in the number of hospitalizations due to road accidents over the years, starting from 2010 we can see an increase on its number for the population belonging to the older age groups.[36]

Figure 3.14 Road accident hospitalization rates, 2001-2013 (by age groups)

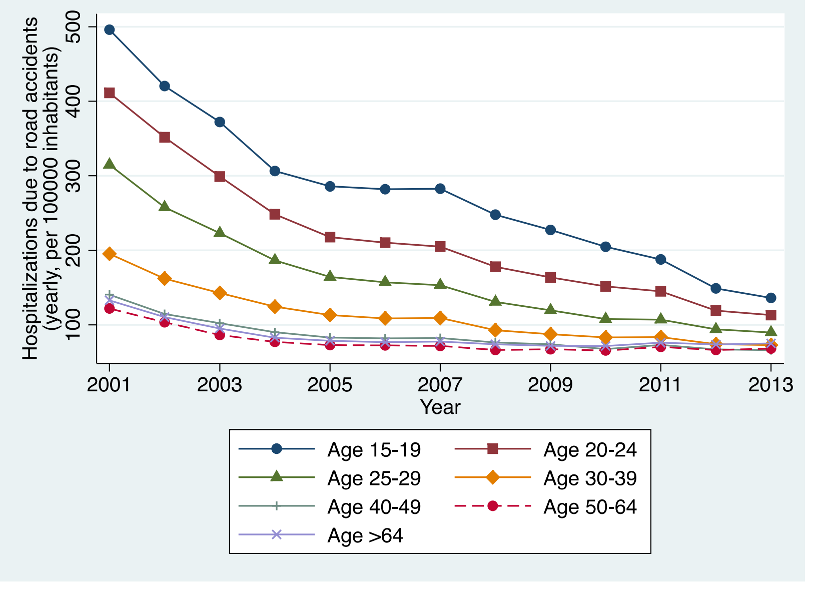


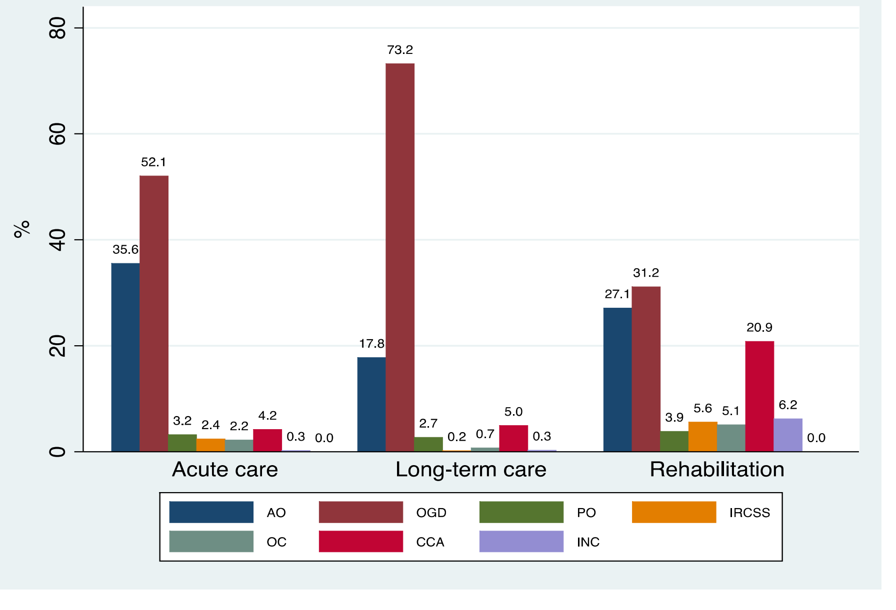
Table 3.3 shows the distribution of patients' admissions along the Italian regions and type of hospitalizations over the period 2001-2013. Of the 798,951 admissions due to car accidents, 98% of them are classified as "acute" admissions.[37] Furthermore**,** considering "acute" care admissions, 99% of hospitalizations are "ordinary" admissions: the highest percentage of car accident hospitalizations in "ordinary" regime is shown in Lombardy (19.5%). More in detail, Fig 3.15 displays the distribution of car accident hospitalizations by type of discharges and hospital activities. As we can see from the figure, most hospitalizations in "acute" care are recorded both in Hospital managed by LHA (OGD) and Hospital Trust (AO).

[36] See Tables I and III in appendix for more details on the distribution of road accidents hospitalizations.
[37] Acute care hospitalization provides residential service to patients who experience acute forms of disease. Long-term care consists on hospital services offered to those patients suffering from severe conditions and an impaired functional situation requiring specific treatment to get health improvement.

Table 3.3 Number of hospitalizations due to road traffic accidents, by regions and discharge type [38]

	Acute care				Rehabilitation				Long-term care			
	Ordinary admission		Day Hospital		Ordinary admission		Day Hospital		Ordinary admission		Day Hospital	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Piedmont	41054	5.32	65	0.72	3011	22.50	450	29.70	305	9.51	-	-
Aosta Valley	1551	0.20	-	-	-	-	-	-	-	-	-	-
Lombardy	150539	19.50	2827	31.29	5072	37.90	460	30.36	149	4.65	2	66.67
Prov. Bozen	8816	1.14	31	0.34	284	2.12	-	-	20	0.62	-	-
Prov. Trento	7048	0.91	352	3.90	586	4.38	118	7.79	46	1.43	-	-
Veneto	41565	5.39	772	8.54	354	2.65	66	4.36	109	3.40	-	-
Friuli V.G.	18460	2.39	-	-	1396	10.43	-	-	91	2.84	-	-
Liguria	12806	1.66	96	1.06	61	0.46	-	-			-	-
E. Romagna	60033	7.78	231	2.56	1502	11.22	296	19.54	2419	75.45	-	-
Tuscany	37470	4.85	786	8.70	217	1.62	45	2.97	11	0.34	-	-
Umbria	6344	0.82	34	0.38	43	0.32	-	-	3	0.09	-	-
Marche	17799	2.31	302	3.34	1	0.01	1	0.07	10	0.31	1	33.33
Lazio	55424	7.18	1780	19.70	309	2.31	41	2.71	13	0.41	-	-
Abruzzo	21909	2.84	209	2.31	56	0.42	-	-	8	0.25	-	-
Molise	5559	0.72	-	-	5	0.04	-	-	-	-	-	-
Campania	120722	15.64	-	-	150	1.12	-	-	3	0.09	-	-
Apulia	57650	7.47	-	-	42	0.31	-	-	1	0.03	-	-
Basilicata	4598	0.60	-	-	3	0.02	-	-	1	0.03	-	-
Calabria	26039	3.37	98	1.08	-	-	-	-	-	-	-	-
Sicily	54495	7.06	1385	15.33	158	1.18	30	1.98	10	0.31	-	-
Sardinia	21928	2.84	67	0.74	133	0.99	8	0.53	7	0.22	-	-

Figure 3.15 Distribution of road traffic accidents hospitalizations, by type of discharge and hospital activity



[38] For simplicity, information on hospitalizations related to “Nido” have been excluded from this table. Overall, among the period 2001-2013, 44 discharges belonged to the class Nido.

Alcohol related hospitalizations

We use administrative health data provided by the Ministry of Health to select alcohol related hospitalizations associated to a number of pathologies recurrent in the existing literature (Marcus and Siedler, 2015; Wichi and Gmel 2011). Namely, we include in our analysis hospitalization for "Mental disorders due to alcohol use" (code 291), "Acute alcohol syndrome" and "Acute alcohol intoxication" (303), "Alcohol abuse" (305)[39], "Toxic effect of ethyl alcohol" (980) of the international classification of diseases, clinical modification (ICD-9-CM), developed by the WHO. Table 3.4 shows the number of alcohol-related hospitalizations among the sub-mentioned ICD-9-CM categories.[40], It can be seen that the majority of hospitalizations are caused by acute intoxications (66.1%) and alcohol abuse (20.6). However, it is interesting to notice that there has been a substantial reduction in the number of alcohol-related admissions over the years. For instance, major declines are recorded in the case of hospitalizations related to the toxic effect of alcohol (-60.3%) and acute alcohol syndrome (-58.0%).

Table 3.4 Number of hospitalizations due to alcohol related intoxication, by ICD-9-CM codes (considering only patient with more than 9 years)

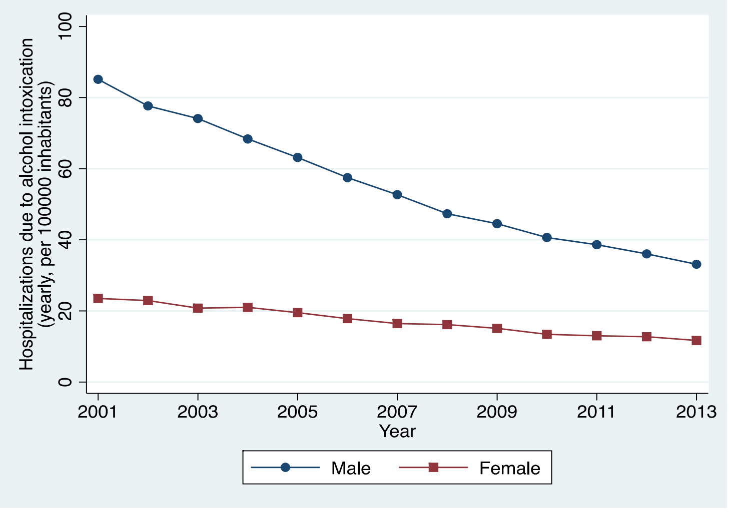
ICD 9 CM codes	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
291	3201	2920	2875	2680	2560	2622	2328	2242	2264	1909	1869	1709	1544
303	18597	17489	16389	15535	14307	12645	11796	10846	10248	9409	9026	8342	7810
305	5422	4861	4749	4679	4523	4297	3962	3632	3310	3108	2968	2727	2472
980	224	195	157	207	137	139	115	138	126	92	77	118	89
Total	27444	25465	24170	23033	21527	19703	18201	16858	15948	14518	13940	12896	11915

The proportion of elective admissions due to alcohol abuse are higher for men (74.8%) than for females (25,2%). However, Figure 3.16 shows decreasing trends in the number of those hospitalizations for both male and females during the whole period analysed in this study.

[39] More in detail, we consider dpr 30500, 30501, 30502, 30503

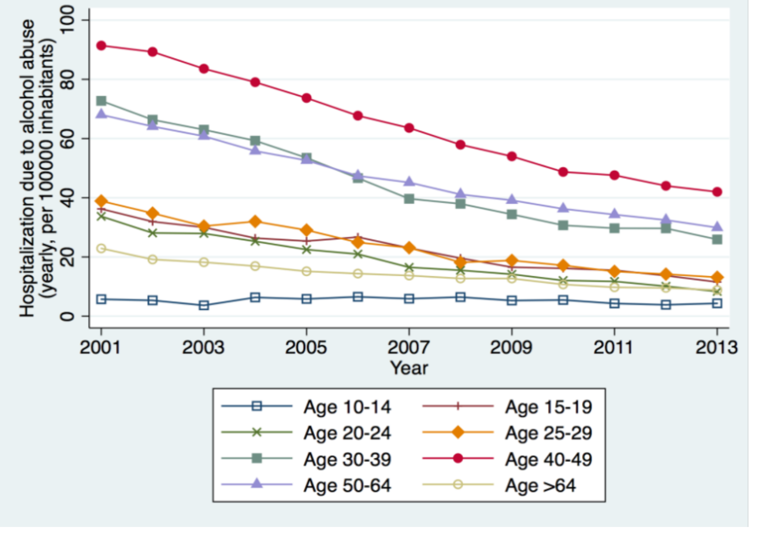
[40] We decide to restrict our analysis only on patient aged 10 and older.

Figure 3.16 Alcohol related hospitalization rates, 2001-2013 (by sex)



Hospitalization rates for different age categories are shown in Figure 3.17 below. Simply examining the graph, it appears that the most striking changes have been, between 2001-2013, the reductions on the number of alcohol-related hospitalizations for patients belonging to the age groups 30-39, 40-49, 50-64 years. Although for the remaining age groups hospitalization trends are diminishing over the years, they exhibit some ups and downs. For instance, as far as the class aged 25-29 is concerned, the number of hospitals discharges fluctuate of about 1.8% between 2008-2009, followed by a downward trend in the successive four years.[41]

Figure 3.17 Alcohol-related hospitalization rates, 2001-2013 (by age groups)



[41] See table II and IV in appendix for more details on the distribution of alcohol related hospitalizations.

Figure 3.18 draws trends on the number of alcohol-related hospitalizations by inpatients age and gender in the year previous the policy ban. As is evident, the highest number of alcohol related hospitalizations occur among those patients aged between 40 and 60 years.

Figure 3.18 Alcohol-related hospitalization rates, 2009 per specific age and gender

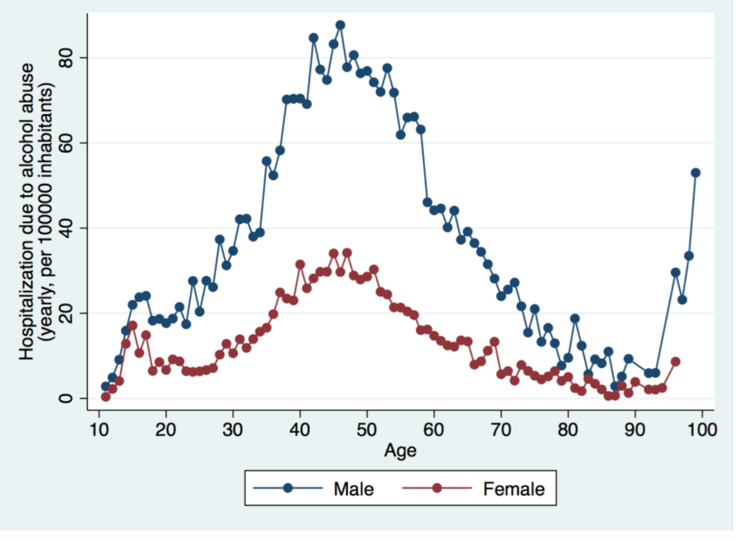


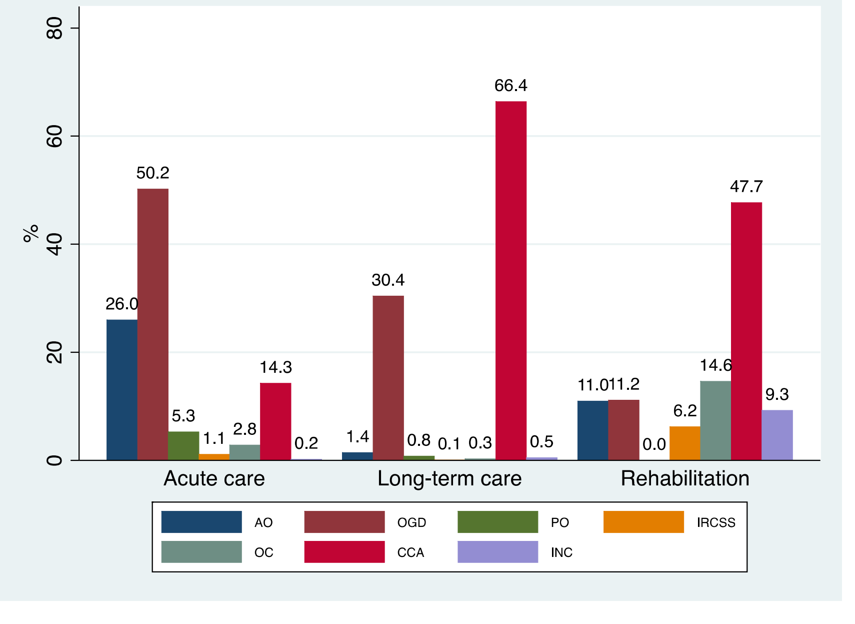
Table 3.5 shows the distribution of patients' admissions along the Italian regions and type of hospitalizations over the period 2001-2013. Of the 245,618 admissions due to alcohol assumption, 87.8% of them are classified as "acute", 9.3% as "rehabilitation" and only 2.8% as long-term care admissions. As regard to "acute" care admissions, 94.0% of these hospitalizations are "ordinary" while 6.0% of them are admissions in "day hospital" regime. Specifically, the highest percentage of alcohol related hospitalization in "ordinary" regime is shown in Lombardy (17.8%).

Fig 3.19 displays the distribution of alcohol related hospitalizations by type of discharges and hospital activities. As we can see from the figure, most hospitalizations in "acute" care are recorded both in Hospital managed by LHA (OGD) and Hospital Trust (AO). On the contrary, when looking at long-term care hospitalizations, most of alcohol-related admissions occur in private accredited hospitals (CCA).

Table 3.5 Number of hospitalizations due to alcohol related intoxication, by regions and discharge type

	Acute care				Rehabilitation				Long-term care			
	Ordinary admission		Day Hospital		Ordinary admission		Day Hospital		Ordinary admission		Day Hospital	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Piedmont	14341	7.01	573	5.38	4263	20.97	5	0.17	690	10.43	-	-
Aosta Valley	454	0.24	1030	7.59	-	-	-	-	-	-	-	-
Lombardy	36317	17.85	1630	12.00	4576	22.64	724	24.64	257	5.93	-	-
Prov. Bozen	9093	4.50	110	0.82	2	0.01	-	-	94	1.40	-	-
Prov. Trento	2110	1.04	4	0.03	1378	6.37	-	-	112	1.59	-	-
Veneto	19005	9.45	694	5.15	2284	10.66	2190	74.57	1405	19.44	9	13.24
Friuli V.G.	7203	3.57	184	1.36	2	0.01	-	-	124	1.86	-	-
Liguria	9445	4.68	658	5.06	66	0.46	2	0.07	9	0.21	-	-
E. Romagna	18266	9.19	670	5.04	1828	9.75	-	-	390	5.84	-	-
Tuscany	11910	5.91	743	5.53	4	0.02	-	-	12	0.21	-	-
Umbria	2445	1.23	40	0.30	1	0.00	-	-	5	0.08	-	-
Marche	8796	4.30	196	1.45	4913	26.09	-	-	1265	17.51	-	-
Lazio	12360	6.01	4315	34.65	18	0.08	3	0.10	71	1.06	-	-
Abruzzo	6599	3.21	445	3.31	538	2.45	4	0.14	23	0.33	-	-
Molise	1662	0.81	29	0.21	-	-	-	-	2	0.04	-	-
Campania	10643	5.23	469	3.51	16	0.07	1	0.03	2185	30.45	59	86.76
Apulia	10383	5.13	282	2.08	8	0.04	-	-	143	1.98	-	-
Basilicata	1464	0.70	58	0.44	-	-	-	-	8	0.11	-	-
Calabria	3717	1.80	253	1.88	62	0.28	1	0.03	44	0.62	-	-
Sicily	8471	4.14	509	3.73	20	0.10	5	0.24	8	0.11	-	-
Sardinia	8095	4.01	63	0.47	-	-	-	-	55	0.80	-	-

Figure 3.19 Distribution of alcohol-related hospitalizations, by type of discharge and hospital activity



3.7 Explorative analysis of the casual effect

For the purpose of our empirical analyses, we aggregate data on hospital discharges on a year basis, where our observational unit is represented respectively by the number of road accident hospitalizations (when we analyse the impact of the ban in road traffic discharges) and by the number of alcohol-related hospitalizations (when we focus on the effect of the night sale alcohol ban). Starting from that, since it was not possible to disentangle admissions recorded in the months pre and post ban adoption, data collected in 2010 have been excluded from the analysis.

Municipality level

In order to reduce variability across data, we aggregate patients' admissions (i.e. road traffic and alcohol related hospitalizations) at the hospital municipal level. This operation has required several steps.

Starting from the databases on hospital addresses, which provide information on all Italian public and private (accredited or not) hospitals over the years, hospital structures for the Lombardy region have been recorded by using their hospitals internal sub-codes.[42] Hence, since some hospitals exhibit an incorrect municipality name over the years, we check data for the presence of these inaccuracies renaming hospital municipality with the correct name.

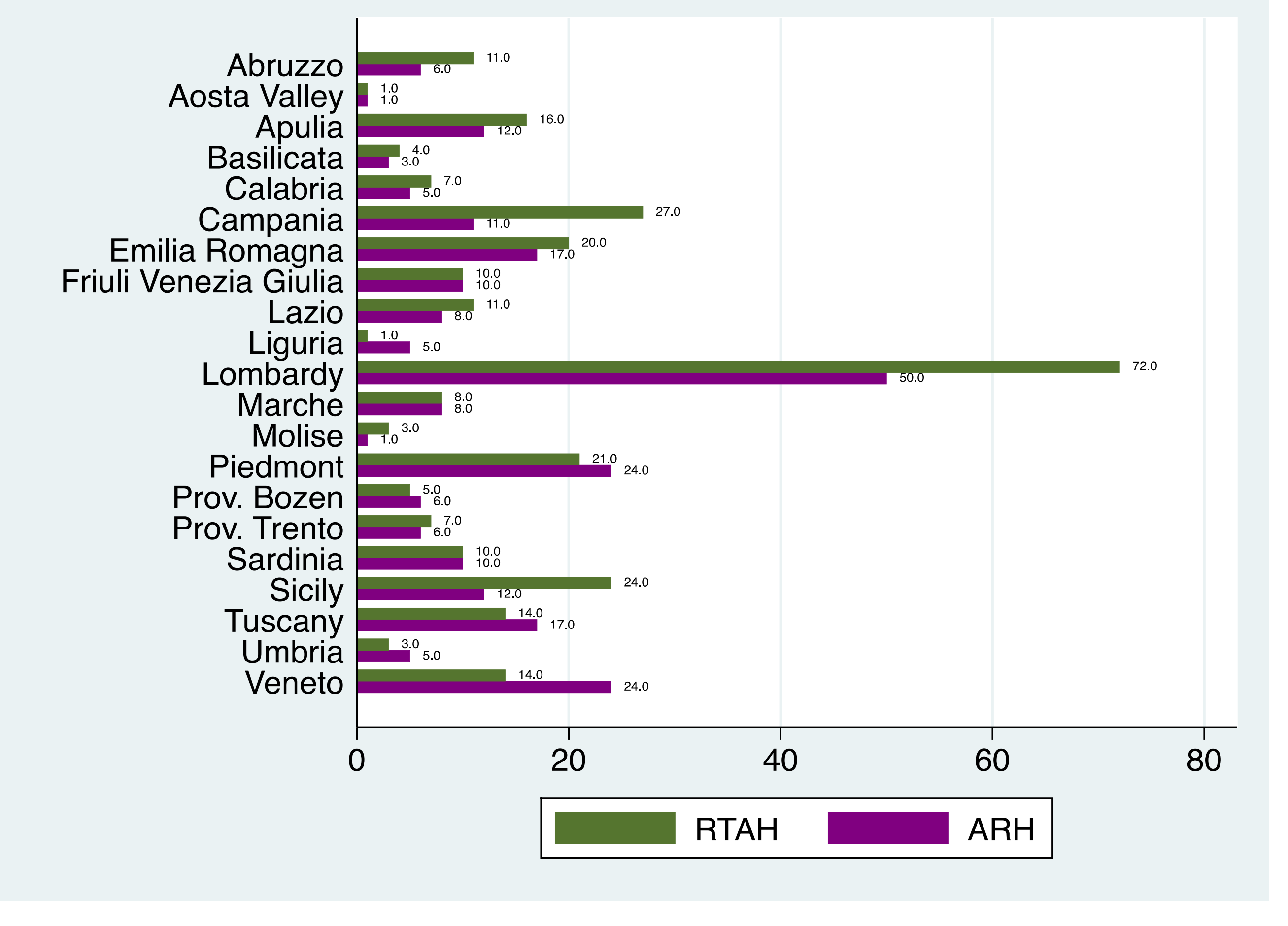
The next step consists of associating each municipality with its relative numeric code. However, in combining codes, we need to consider all territorial changes occurred over the period 2007-2013, which have led to changes in territorial constituencies, at the establishment of new territorial units and to municipalities name changes. For these reasons, all codes have been recoded using data on the municipality numeric codes with 110 provinces provided by the National Institute of Statistic (ISTAT).

[42] The database on hospital addresses are provided by the "*Nuovo sistema Informativo sanitario*" yearly (Ministry of Health. It contains information on each hospital denomination, address, municipality, LHA, type of structure (e.g. public or private) as well as day/month/year of the hospital first opening. Furthermore, it displays information on hospital internal sub-codes.

Finally, hospitalizations data have been merged with the dataset containing information at the municipality level. This procedure allows us to aggregate all hospitalizations data at municipality level over years. Table V in appendix II shows the distributions of municipalities both in the case of road traffic and alcohol related hospitalizations among the Italian regions.

For the purpose of the empirical analyses, we eliminate those municipalities which have not recorded any hospitalizations or less than 5 hospitalizations over the six years considered in the study. Figure 3.20 below shows the distribution of hospitals' municipalities along the Italian territory by type of discharge. Lombardy is the Italian region with the largest number of hospitals' municipalities for both alcohol-related problems and road traffic injuries (this result is determined, in part, by the presence in this region of the highest number of hospitals at Italian level).

Figure 3.20 Distribution of hospitals' municipality along the Italian territory, by type of discharge (RTAH and ARH)

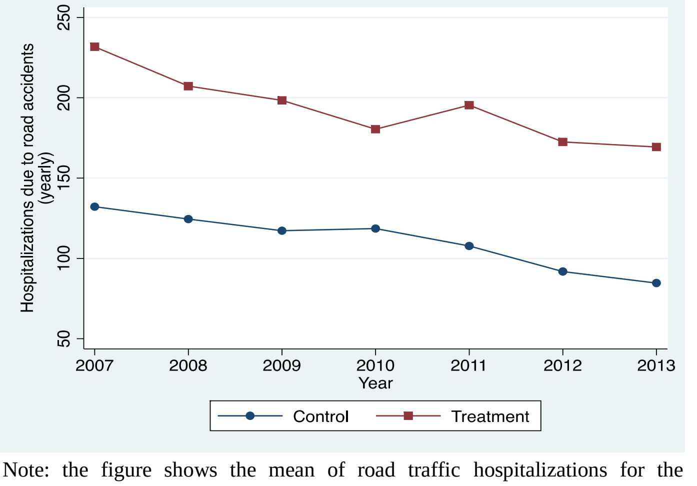


To sum up, we generate two very different panel of:

- 241 municipalities generating 1446 year observations about alcohol related hospitalizations
- a panel of 1734 municipality year observations in the case of road traffic accidents

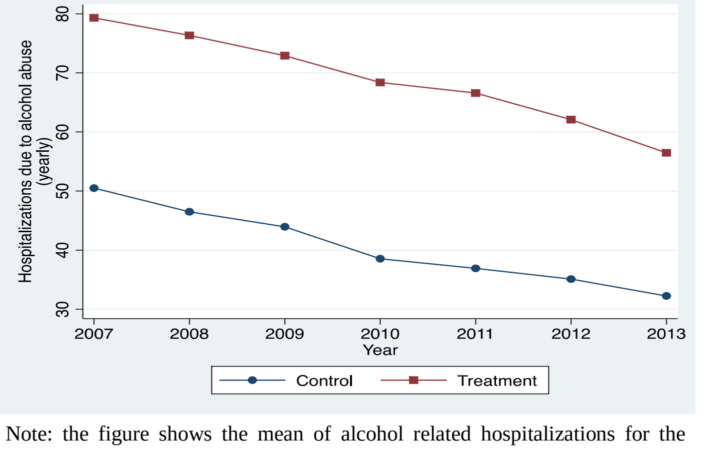
Figure 3.21 exhibits the number of accident-related hospitalizations for the treatment and control groups when defining a 10 km distance from the motorway access points. While for the control hospitals municipalities the reduction in the number of admissions due to accidents seems to decline over the years, in the treatment hospitals, admissions increased soon after the policy introduction to decline sharply one year after. Figure 3.22 refers to the number of alcohol related hospitalization grouped by treatment and control group. The mean of alcohol-related hospitalizations decrease constantly over the years in the control group, while some differences emerge for the treatments.

Figure 3.21 Road accident hospitalizations, 2007-2013



Note: the figure shows the mean of road traffic hospitalizations for the treatment and control group over the period 2007-2013 when considering aggregation at municipal level.

Figure 3.22 Alcohol related hospitalizations, 2007-2013



Note: the figure shows the mean of alcohol related hospitalizations for the treatment and control group over the period 2007-2013 when considering aggregation at municipal level.

Table 3.6 presents summary statistics of the dependent variable in the case of road traffic hospitalizations. In detail, this variable shows that in average there have been 156.24 discharges due to car and motor vehicle accidents, a value, this, much higher in the case of treated municipalities.

Table 3.6 Summary statistics of road traffic hospitalizations variable (municipal level)

Variable	Obs	Mean	St. Deviation	Min	Max
Number of road traffic accident (overall)	1734	156.24	272.89	5	4460
Number of road traffic accident in treated municipalities	954	194.30	346.77	5	4460
Number of road traffic accidents in control municipalities	780	109.69	120.89	5	923

Similarly, table 3.7 shows that the average number of alcohol-related hospitalizations is equal to 55.42, ranging from a minimum of 5 to a maximum of 715 hospitalizations.

Table 3.7 Summary statistics of alcohol-related hospitalizations variable (municipal leve)

Variable	Obs	Mean	St. Deviation	Min	Max
Number of road traffic accident (overall)	1446	55.42	85.17	5	715
Number of road traffic accident in treated municipalities	828	67.19	101.30	5	715
Number of road traffic accidents in control municipalities	618	39.65	52.89	5	316

Patients' Municipality level

To fully consider the effect of the policy prohibition, it is interesting to analyse hospital discharges using patient zip codes. The reason behind this choice originate on the idea that patients who are hospitalized in municipalities located within the bandwidth not necessarily live in those areas. Then, aggregating data by patient's zip codes, we can better distinguish among treated patients, and therefore more exposed to the policy ban, from controlled patients, less exposed to the policy.

Starting from the SDO data, which give information on patients' municipalities, a numerical code has been associated to each municipality. However, as it can be read in the previous subparagraph, since the Italian municipality boundaries have changed across time, changes in municipality names and codes have been evaluated. Considered that, all municipality codes have been checked and recoded using data on the numerical codes with 110 provinces provided by ISTAT.

In order to compare road traffic hospitalizations among the same patients' municipalities across years, we have excluded those patient's municipalities which are not repeatedly presents between 2007 and 2013. Thus, we consider only those patients' municipalities with at least 1 yearly discharge referable to road traffic or alcohol related hospitalizations. This procedure generates an additional reduction on analysis units (see Table VI in Appendix II to have an overview of the number of municipality before these data exclusions). Figure 3.23 provide us a descriptive statistics of the distribution of patients' municipalities along the Italian regions.

Overall, we have ended up with a panel of 3031 (1317) patients' municipalities over the period 2007-2013, generating 18186 (7902) patients' municipality observations about road traffic accident hospitalizations (alcohol-related hospitalizations).

Figure 3.23 Distribution of patients' municipalities along the Italian territory, by type of discharge (RTAH and ARH)

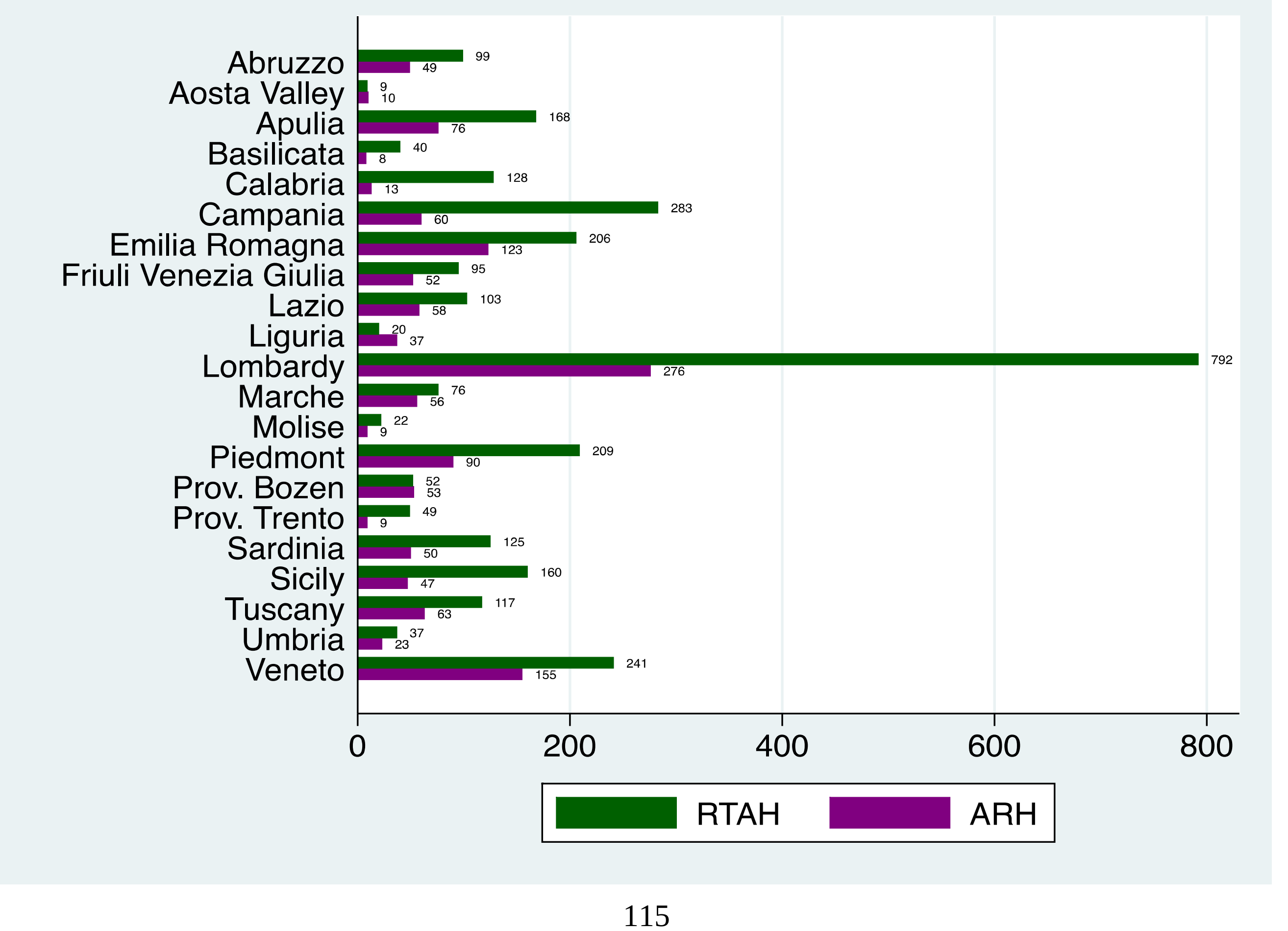
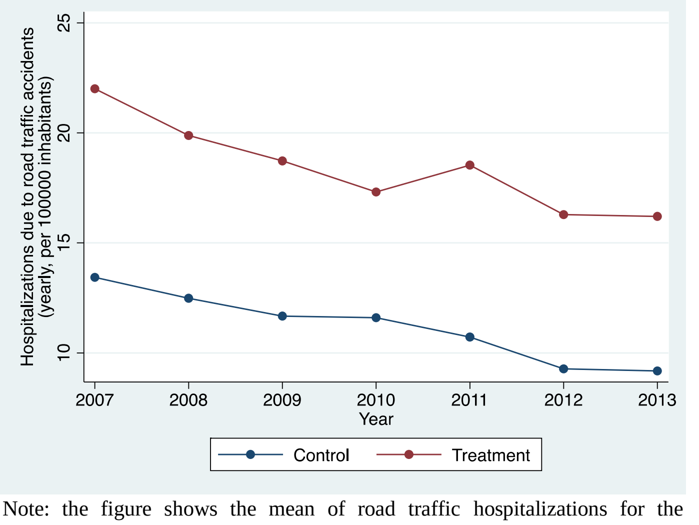


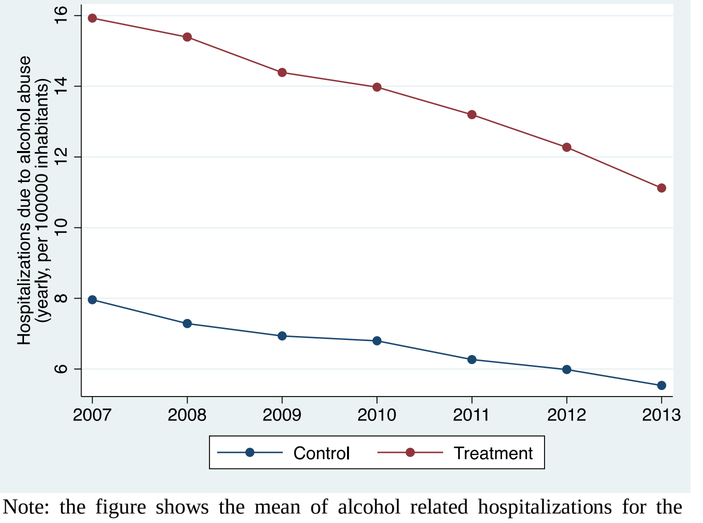
Figure 3.24 shows the mean of road traffic accident hospitalizations for the treatment and control group by patient' municipality of residence. It appears that hospitalizations decreased during the three years pre policy to significantly increase during the period 2010-2011. Looking at the mean of alcohol-related hospitalizations in Figure 3.25, we assist to similar trends to the ones showed in the hospitals' municipalities level analysis.

Figure 3.24 Road accident hospitalizations, 2007-2013



Note: the figure shows the mean of road traffic hospitalizations for the treatment and control group over the period 2007-2013 when considering aggregation at patient municipal level.

Figure 3.25 Alcohol related hospitalizations, 2007-2013



Note: the figure shows the mean of alcohol related hospitalizations for the treatment and control group over the period 2007-2013 when considering aggregation at patient municipal level.

More in detail, from the analysis of Table 3.8 (3.9) below, we can see that the average number of hospitalizations per patients' municipality is equal to 14.66 (9.91) in the case of road traffic hospitalizations (alcohol-related hospitalizations).

Table 3.8 Summary statistics of road traffic hospitalization variable (patient municipality level)

	Obs	Mean	St. Deviation	Min	Max
Number of road traffic accident (overall)	18186	14.66	63.10	1	2845
Number of road traffic accident in treated municipalities	8880	18.85	87.40	1	2845
Number of road traffic accidents in control municipalities	9306	10.66	21.44	1	460

Table 3.9 Summary statistics of alcohol related hospitalizations variable (patient municipality level)

	Obs	Mean	St. Deviation	Min	Max
Number of alcohol related hosp. (overall)	7902	9.91	34.90	1	796
Number of alcohol related hosp. in treated municipalities	4278	13.03	46.39	1	796
Number of alcohol related hosp. in control municipalities	3624	6.22	9.55	1	133

3.8 Control variables for confounding factors

In the next section a few control variables will be used, namely:

- the provincial (municipality) population density as proxy of traffic intensity (or, in the case of alcohol-related sample, as proxy of the internal demand of health care). We expect that the higher the population density is, the more people using the roads and abusing alcohol are.
- the provincial unemployment rate, as a proxy of the economic situation. The rational of using this variable is that variation in unemployment rate are associated with changes in road traffic and alcohol related outcomes.
- the provincial (municipality) income, as a proxy of population well being. We expect that the lower the value is, the less the money for health programs and the ability of people to drive new vehicle with wide safety features, and the higher the number of fatal road traffic and alcohol-related accident would be.

Table 3.10 Variables definition and data sources

Variable	Definition	Source
Unemployment rate	Ratio between the number of people unemployed/looking for a job and the labor force	ISTAT
Revenue	Total income	MoF
Population density	Ratio between the area population and the square km of the area	Own calculation using ISTAT data
Post	1 for the two post policy reform years (2011-2012-2013)	Own calculations
Treatment	1 if hospital h (municipality m) is within the 10 km bandwidth	Own calculations
Effect (post x treatment)	Interaction term between Post and Treatment	Own calculations
Closeness	Geographical distance from the centroid of each hospital (municipality) to the nearest motorway interchange/tollbooth	Own calculations GiS*
Exposure	Interaction term between Post and the variable Closeness	Own calculations

Notes: *ISTAT*= National Institute of Statistics , MEF= Ministry of Finance

3.9 Results

For each clinical condition (i.e. road traffic and alcohol-related hospitalizations), the model has been estimated twice considering different data aggregation *i*: hospitals and municipalities. At first, we estimate the following model:

$$Y_{it} = \alpha_0 + \alpha_1 \cdot post_t + \alpha_2 \cdot treat_i + \beta \cdot (Post_t \cdot treat_i) + \varepsilon_{it}, \tag{3.1}$$

where Y_{it} is the number of discharges (road traffic accidents, alcohol abuse) in hospital (patients) municipality *i* in year *t*. $Post_t$ is a binary variable equals 1 for the years after the implementation of the policy, 0 otherwise. $treat_i$ is a variable representing treated hospital (patients) municipality, equals to 1 if hospital (patient) municipality *i* is within the motorways bandwidth and 0 if it is located outside. $Policy_{it}$ represents our variable of interest calculated as the interaction between *post* and *treat*, which takes values 1 if the prohibition of the night-alcohol sales is force in hospital (patients) municipality *i* in year *t*. Hence, we will augment equation (1) by adding a set of control variables:

$$Y_{it} = \alpha_0 + \alpha_1 \cdot post_t + \alpha_2 \cdot treat_i + \beta \cdot (Post_t \cdot treat_i) + \gamma_i + \delta_t + X'_{it}\eta + \varepsilon_{it} \tag{3.2}$$

where γ_i are hospital (patient) municipality fixed effect, δ_t includes year fixed effects and X'_{it} is a vector of other covariates. Hence, since the DD model may have problem caused by the outcome's scale of measure (Lechner 2010, Marcus e Siedler 2015), we have also express our dependent variable in logarithm allowing us to interpret the estimated coefficients in terms of percentage change.

In an effort to study the effectiveness of the policy ban as the distance from the motorway service areas increases we estimate the following OLS equation:

$$Y_{it} = \alpha_0 + \beta \cdot (Post_t \cdot c_i)_{it} + X'_{it}\eta + \gamma_i + \delta_t + \varepsilon_{it} \tag{3.3}$$

where c_i is a continuous variable which takes a value equal to 1 if hospital (patients) municipality *i* is located within 1 km from the closer motorway access point and values gradually lower as the distance between hospitals (patients) municipality and motorway access points increases. We expect that the closer the hospital (patients) municipality to the motorways, the lower the number of hospitalizations due to road accidents and alcohol intoxication. Here the main variable of interest is the interaction term $Post_t \cdot c_i$, which detects the *exposure to the policy ban*, defined as the interaction between a dummy variable for the three post policy years and the distance between hospitals (patients) municipalities and motorways access points.

3.9.1 Municipality level analysis results

The baseline difference-in-difference model is used to analyse the effect of the policy ban on the number of admissions due to accidents in hospitals within the bandwidth with respect to the ones distant, before and after the policy introduction. The results of the basic difference-in-difference model are reported in Table 3.11 below. The first column shows the predicted number of hospitalizations for the pre-adoption period (years 2007-2008-2009), whereas column 2 reports specular hospitalizations for the post-treatment period (years 2011-2012-2013). As we can see from the table, when considering the municipality sample, it seems that the policy change has had an effect in reducing road accidents hospitalizations. Similarly, the values on the right hand side of the Table displays a reduction of -2.29 hospitalizations caused by alcohol related intoxication.

Table 3.11 Basic difference in difference results

	Before	After	Difference	Before	After	Difference
	Road traffic hosp.			Alcohol-related hosp.		
Overall						
Treatment	209.88	178.71	-31.18	74.26	60.12	-14.14
Control	124.64	94.75	-29.89	45.58	33.73	-11.74
Difference	85.24	83.96	-1.28	28.79	26.39	-2.29

In Table 3.12, we present the results from the DD estimator. Since the error term might not be independent within municipalities, leading, if the case, to wrong significance levels (Bertrand et al. 2004), we decided to cluster standard errors at the municipality level to control for serial correlation problems. Basically, the coefficient in the first column give us the same results of the basic model.

However, we can see that the effect coefficient is not statistically significant even though is showing the expected sign. The third column reports a fixed effect specification for the 289 hospitals and over 6 years. The estimate coefficient is similar to the basic DD regression, negative but statistically not significant. The last specification adds a set of additional time varying district variables such as unemployment, the municipality revenue and density. It is worth to notice, that after the inclusion of these control variables, as shown in column 5, the effect of the estimated coefficient became positive even though not statistically significant. As a robustness check, we re-estimate equation (3.2), defining *Y* as the natural logarithm of the number of road traffic accidents (see Table I in Appendix III). When considering the municipal-level analysis, the policy ban seems to affect positively the number of road traffic hospitalizations, however, the estimated coefficient is not statistically different from zero. Hence, the results appear to be not robust after the modification of the outcome variable. Looking at the estimates for alcohol-related hospitalizations, the results are always not significant. This seems to suggests that the night-sale alcohol ban do not have an effect in reducing alcohol related hospitalizations.

Table 3.12 The effect of the night alcohol sale prohibition in motorways service areas on road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013)

	Basic DID		+time/hospital dummies		+Controls	
	RTA	AR	RTA	AR	RTA	AR
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-29.8949 ***	-11.8479 ***				
	(7.6237)	(2.1457)				
Treat	85.2437 ***	28.6800 ***				
	(33.1068)	(10.7662)				
Effect ($post_t$ * $treat_i$)	-1.2812	-2.2874	-1.2812	-2.2874	4.1333	-0.8584
	(18.8967)	(3.1959)	(20.7195)	(3.5048)	(20.4675)	(3.4999)
Unemployment					-6.7711	0.4178
					(6.1194)	(0.7898)
Revenue					-0.0000 **	-0.0000
					(0.0000)	(0.0000)
Density					0.7951	0.3268 *
					(0.8731)	(0.1975)
Municipality FE	No	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes
Observations	1734	1,446	1734	1446	1662	1380
R^2	0.027	0.032	0.872	0.957	0.873	0.958
Number of municipalities	289	241	289	241	277	230

Note: Robust standard errors clustered at municipality level in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively.

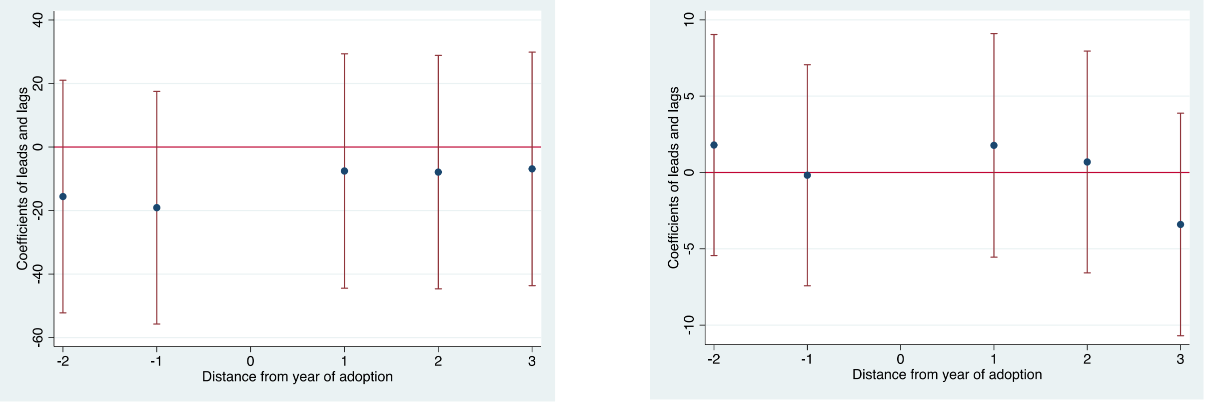
As we have mentioned before, the key assumption of the DD approach is that in the absence of the policy prohibition, both road accident and alcohol-related hospitalizations in areas closes to the motorway (the treatment group) would follow a similar trend as that in the control group. Considered that is not possible to directly test the common trend assumption, we firstly proceed to a visual inspection of the pre-policy trends for the treatment and control groups. As it has been shown above, Figure 3.21 reports the mean trends of road traffic accident hospitalizations for the treatment and control group. It appears that hospitalizations do not follow a similar trend in the pre-treatment period. In a similar manner, looking at the mean of alcohol-related hospitalizations in Figure 3.22, the lines display a similar trend, even though they are not close. Then, we estimate a different model where the condition of a hospital to be in a municipality within a bandwidth distance of 10 km is interacted with a year dummy as described in the following equation:

$$Y_{it} = \sum_{t=-m}^{-1} Treat_i * D_t + \sum_{t=0}^{q} Treat_i * D_t + \gamma_i + \delta_t + X'_{it} + \varepsilon_{it} \tag{3.4}$$

In this model, both leads and lags of the treatment are introduced as a way to verify whether before the introduction of the night sale alcohol ban there was a common trend of the outcome. We expect that, in presence of parallel trend, the coefficient of m leads would be not statistically significant. Figure 3.26 plots the results of equation 3.4 where each dot shows the estimated coefficient of a lead or lag with its relative 95% confidence interval respectively in the case of the road traffic and alcohol related sample. The results do not support the existence of a common trend, in fact, leads and lags coefficient are both not statistically different from zero.

Hence, a variety of alternative estimates have been performed. We begin analysing the differential effect of the ban for both men and women, to continue with DD estimates for different age classes. Transformation of the outcome variable by considering the number of road accidents(alcohol) hospitalization as share of the total hospitalizations have also been used. Results confirm the lack of clear evidence about the effectiveness of the policy restriction in reducing motor-vehicle accidents and alcohol-related hospitalizations.

Figure 3.26 Common Trend (Leads and Lags)



Road traffic accident sample **Alcohol related intoxication sample**

To better evaluate the effect of geographical distance on the number of road traffic accidents as well as alcohol related hospitalizations, we estimate equation (3.3). We expect that the closer the hospital to the motorway, the lower the number of hospitalizations caused by road traffic accidents and alcohol intoxication. Table 3.13 displays the results from the ordinary least square (OLS) estimation (see column 1 and 2). It appears that the proximity to the motorways, where the policy is in force, has lead to reductions in both road traffic and alcohol related hospitalizations. Despite the negative sign, the estimate coefficient is not statistically significant.

Table 3.13 Effect of exposure to the policy ban on the number of road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013)

	OLS				Poisson			
	RTA		AR		RTA		AR	
Exposition to policy ban	-72.8732		-4.6814		-0.0396		0.0011	
	(81.4465)		(5.8968)		(0.1896)		(0.0617)	
Unemployment	-6.0363	**	0.4629		-0.0178	*	0.0152	**
	(2.8430)		(0.4733)		(0.0099)		(0.0072)	
Revenue	-0.0000	**	-0.0000	**	-0.0000		-0.0000	
	(0.0000)		(0.0000)		(0.0000)		(0.0000)	
Density	0.7568		0.3224	***	0.0015		0.0029	**
	(0.4850)		(0.1188)		(0.0015)		(0.0012)	
Municipality FE	Yes		Yes		Yes		Yes	
Year FE	Yes		Yes		Yes		Yes	
Observations	1662		1380		1662		1380	
R^2	0.873		0.958		0.885		0.895	
Number of municipalities	277		230		277		230	

Note: Robust standard errors in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively.

3.9.2 Motor-vehicle accidents and alcohol–related hospitalizations by patients municipality of residence

The magnitude of the effect of the policy ban on the number of road accident and alcohol related hospitalizations varies when we estimate the basic DD model using aggregate data at patients' municipality of residence. As we have said before, the decision to consider the number of hospitalizations due to car accidents and alcohol-related hospitalizations by patient's zip codes as dependent variable originate on the idea that not necessarily patients who are hospitalized in municipalities located within the bandwidth live in those areas. Then, to fully consider the effect of the policy ban it could be more appropriate to consider hospital discharges using aggregations by patients' zip codes.

The values on the left hand side of Table 3.14 display an overall decrease of 0.68 hospitalizations caused by road traffic accidents. Similar reductions are reported when considering alcohol related hospitalizations (-1.50).

Table 3.14 Basic difference in differences results

	Before	After	Difference	Before	After	Difference
	Road traffic hosp.			Alcohol-related hosp.		
Overall						
Treatment	20.52	17.19	-3.33	14.47	11.59	-2.88
Control	11.98	9.33	-2.65	6.91	5.54	-1.37
Difference	8.54	7.86	-0.68	7.56	6.06	-1.50

Specifically, the values on the right hand side of the table show that in the municipalities closed to the motorways (*treated areas*), the average of patients' hospitalizations change over time, with about 14.47 alcohol-related hospitalizations before and roughly 11.59 after the implementation of the alcohol policy ban. Similarly, maybe due to the existence of other policies aimed at reducing alcohol misuse, as already mentioned above, the mean of alcohol-related hospitalizations decreases by 1.37 in those areas far from the motorways (*controlled areas*), causing an overall decrease of about 1.50 in alcohol-related hospitalizations due to the Italian night alcohol sale ban in motorways service areas.

In Table 3.15, we present the results from the estimate of the equation (3.2). Encouraging results can be noticed when referring to the alcohol related specifications. As we can see from the table, in fact, apparently the effect of the night sale alcohol ban is quite strong: the reduction of 1.5 corresponds to a general reduction in the number of admissions caused by alcohol-related hospitalizations of about 12.94% (=1.50/ (1.50 + 11.59)).[43] This result is confirmed after the introduction of fixed effect and control variables (column 4 and 6).

When looking at the estimates for road traffic accident hospitalizations, the estimates results are always not significant. This suggests that the night-sale alcohol ban seems to do not have an effect in reducing motor vehicle accident hospitalizations.

Table 3.15 The effect of the night alcohol sale prohibition in motorways service areas on road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013), by patient municipality of residence

	Basic DID		+time/hospital dummies		+Controls	
	RTA	AR	RTA	AR	RTA	AR
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-2.6516 ***	-1.3720 ***				
	(0.3481)	(0.1561)				
Treat	8.5357 ***	7.5596 ***				
	(2.6227)	(1.9434)				
Effect ($post_t$ * $treat_i$)	-0.6772	-1.5041 ***	-0.6772	-1.5041 ***	-0.5731	-0.9877 *
	(1.2235)	(0.4998)	(1.3404)	(0.5476)	(1.8612)	(0.4769)
Unemployment					-0.7686 **	0.0457
					(0.3832)	(0.1452)
Revenue					-0.0000	-0.0000 **
					(0.0000)	(0.0000)
Density					0.0603	0.0488
					(0.0372)	(0.0255)
Municipality FE	No	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes
Observations	18186	7,902	18186	7902	17376	7524
R^2	0.005	0.011	0.914	0.964	0.915	0.971
Number of municipaliti	3031	1317	3031	1317	2896	1254

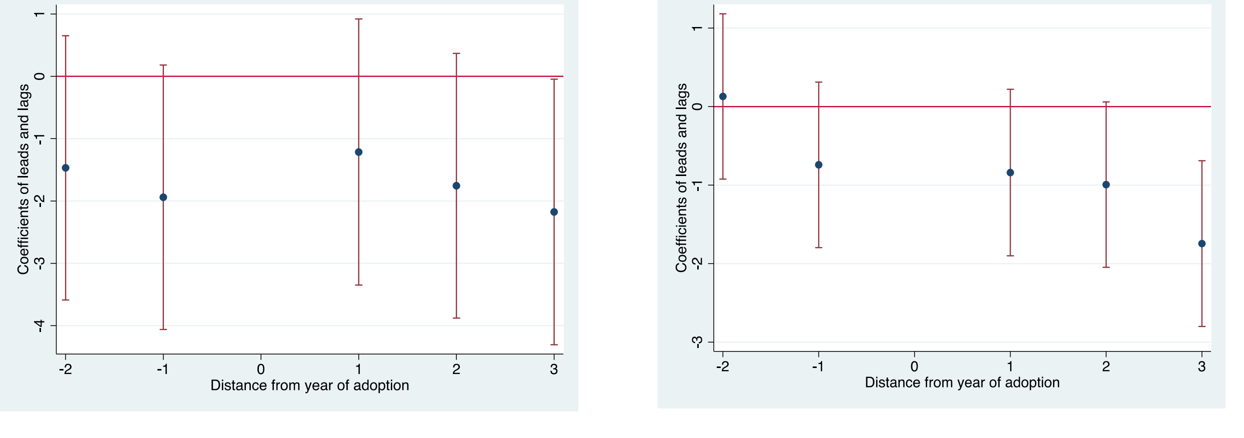
Note: Standard errors clustered at municipality level in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively

To verify the key assumption of DD method, equation (3.4) has been re-estimated in the case of aggregated data by patients' zip codes. Figure 3.27 shows the results of our estimates, where each dot displays the estimated coefficient for leads and lags. For the road traffic accident sample, the

[43] We refer to Marcus and Siedler (2015) for this calculation. Following their notation, in fact, in absence of the policy ban the number of alcohol-related hospitalizations would be equal to 13.09 (=1.50+11.59).

evidence does not support the existence of a common trend. On the contrary, for the alcohol-related sample, since leads coefficients are not significant while lags coefficients are statistically significant starting from the second year after the adoption of the policy ban, it seems that there is evidence in support of a common trend. However, to inspect whether the results of the estimates are sensitive to the choice of the control groups, several robustness checks have been performed. On the one hand, the sensitivity of our specification has been tested considering wider treatment areas. Suppose to consider the alcohol-related sample. Table 3.16, column 6, shows the results of DD estimate obtained using treated patients' municipalities within 14 km from motorway access points. This choice allows us to expand the treatment group, which increased from 679 municipalities, in the case of a 10 km bandwidth, to 801 municipalities considering a bandwidth of 14 km. Similarly, Table 3.17 reports the estimated coefficients using a bandwidth distance of 12km from the patients' zip codes to the motorway access points. As we can see from the Tables, these findings confirm the results of the main specification (Table 3.15). On the other hand, the robustness of our results have been checked considering reduced treatment groups. At first, only those municipalities within 8 km have been considered among the treated group (see estimates results on Table 3.18)[44]., while secondly, estimates have been performed simply using as treated municipalities those hospitalizations generated by patient living within 6 km from motorways access (Table 3.19)[45].

Figure 3.27 Common trends (Leads and Lags)



Road traffic accident sample **Alcohol related intoxication sample**

[44] In the case of a bandwidth of 8 km, the number of treated municipality is 601, while the control group is composed by 653 municipalities.
[45] The treated group is composed by 478 municipalities, while the control group is made by 776 municipalities.

Table 3.16 The effect of the night alcohol sale prohibition in motorways service areas on road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013), by patient municipality of residence (bandwidth at 14 km)

	Basic DID		+time/hospital dummies		+Controls	
	RTA	AR	RTA	AR	RTA	AR
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-2.9316***	-1.4447***				
	(0.4506)	(0.1836)				
Treat	5.2487**	5.9552***				
	(2.1820)	(1.6892)				
Effect ($post_t$ * $treat_i$)	-0.0826	-1.1613***	-0.0826	-1.1613**	-0.1387	-0.8345*
	(1.0374)	(0.4450)	(1.1366)	(0.4875)	(1.6476)	(0.4546)
Unemployment					-0.7718	0.0440
					(0.3895)	(0.1455)
Revenue					-0.0000	-0.0000***
					(0.0000)	(0.0000)
Density					0.0598	0.0488*
					(0.0371)	(0.0255)
Municipality FE	No	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes
Observations	18186	7,902	18186	7902	17376	7524
R^2	0.002	0.007	0.914	0.964	0.915	0.971
Number of municipalities	3031	1317	3031	1317	2896	1254

Note: Standard errors clustered at municipality level in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively

Table 3.17 The effect of the night alcohol sale prohibition in motorways service areas on road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013), by patient municipality of residence (bandwidth at 12 km)

	Basic DID		+time/hospital dummies		+Controls	
	RTA	AR	RTA	AR	RTA	AR
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-2.8320***	-1.4050***				
	(0.4023)	(0.1676)				
Treat	6.5112***	6.7834***				
	(2.3286)	(1.7920)				
Effect ($post_t$ * $treat_i$)	-0.2674	-1.3158***	-0.2674	-1.3158***	-0.2147	-0.9402**
	(1.0963)	(0.4660)	(1.2011)	(0.5106)	(1.6995)	(0.4660)
Unemployment					-0.7713**	0.0440
					(0.3870)	(0.1454)
Revenue					-0.0000	-0.0000***
					(0.0000)	(0.0000)
Density					0.0599	0.0489*
					(0.0371)	(0.0255)
Municipality FE	No	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes
Observations	18186	7902	18186	7902	17376	7524
R^2	0.003	0.009	0.914	0.964	0.915	0.971
Number of municipalities	3031	1317	3031	1317	2896	1254

Note: Standard errors clustered at municipality level in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively

Table 3.18 The effect of the night alcohol sale prohibition in motorways service areas on road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013), by patient municipality of residence (bandwidth at 8 km)

	Basic DID				+time/hospital dummies			+Controls			
	RTA		AR		RTA	AR		RTA		AR	
	(1)		(2)		(3)	(4)		(5)		(6)	
Post	-2.6710	***	-1.4625	***							
	(0.3115)		(0.1935)								
Treat	10.6332	***	8.0660	***							
	(3.0724)		(2.1797)								
Effect ($post_t$ * $treat_i$)	-0.7586		-1.5179	***	-0.7586	-1.5179	***	-0.5306		-0.9115	*
	(1.4256)		(0.5527)		(1.5618)	(0.6056)		(2.0919)		(0.5133)	
Unemployment								-0.7676		0.0445	
								(0.3777)		(0.1451)	
Revenue								-0.0000		-0.0000	***
								(0.0000)		(0.0000)	
Density								0.0602		0.0487	*
								(0.0373)		(0.0255)	
Municipality FE	No		No		Yes	Yes		Yes		Yes	
Year FE	No		No		Yes	Yes		Yes		Yes	
Observations	18186		7,902		18186	7902		17376		7524	
R^2	0.007		0.012		0.914	0.964		0.915		0.971	
Number of municipalities	3031		1317		3031	1317		2896		1254	

Note: Standard errors clustered at municipality level in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively

Table 3.19 The effect of the night alcohol sale prohibition in motorways service areas on road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013), by patient municipality of residence (bandwidth at 6 km)

	Basic DID				+time/hospital dummies			+Controls			
	RTA		AR		RTA	AR		RTA		AR	
	(1)		(2)		(3)	(4)		(5)		(6)	
Post	-2.4788	***	-1.7860	***							
	(0.3513)		(0.3252)								
Treat	12.4094	***	5.7821	***							
	(3.8779)		(2.3358)								
Effect ($post_t$ * treati)	-1.6464		-1.0693	*	-1.6464	-1.0693	*	-1.3280		-0.4920	
	(1.8253)		(0.5658)		(1.9997)	(0.6199)		(2.5089)		(0.6062)	
Unemployment								-0.7584	**	0.0450	
								(0.3704)		(0.1427)	
Revenue								-0.0000		-0.0000	***
								(0.0000)		(0.0000)	
Density								0.0607		0.0484	*
								(0.0370)		(0.0253)	
Municipality FE			No		Yes	Yes		Yes		Yes	
Year FE			No		Yes	Yes		Yes		Yes	
Observations	18186		7902		18186	7902		17376		7524	
R^2	0.0078		0.006		0.914	0.964		0.915		0.971	
Number of municipalities	3031		1317		3031	1317		2896		1254	

Note: Standard errors clustered at municipality level in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively

The results in table 3.18 are robust to the alternative treatment group. However, it is also evident that, considering a bandwidth of 6 km (Table 3.19), the effect coefficient is not statistically significant even though is showing the expected sign.

Another robustness check consists on applying a regional sample restriction to our data. Since the policy ban only applies to motorways petrol station and service areas, Sardinia have been excluded from the empirical study because it represents a particular case of region not crossed by motorways. Table 3.20 displays the results of this additional robustness check. Focusing on the alcohol-related hospitalizations, again, it can be noticed that the effect coefficient confirms the results of the main specification.

Table 3.20 The effect of the night alcohol sale prohibition in motorways service areas on road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013), by patient municipality of residence (excluding data from Sardinia)

	Basic DID		+time/hospital dummies		+Controls	
	RTA	AR	RTA	AR	RTA	AR
	(1)	(2)	(3)	(4)	(5)	(6)
Post	-3.1295	-1.3002 ***				
	(0.3577)	(0.1535)				
Treat	8.4607	7.7827 ***				
	(2.6305)	(1.9413)				
Effect (post * treat)	-0.1993	-1.5759 ***	-0.1993	-1.5759 ***	-0.2010	-1.0519 **
	(1.2262)	(0.4990)	(1.3434)	(0.5468)	(1.8662)	(0.4675)
Unemployment					-0.8405 ***	0.0264
					(0.4056)	(0.1564)
Revenue					-0.0000	-0.0000 ***
					(0.0000)	(0.0000)
Density					0.0646 *	0.0498 *
					(0.0386)	(0.0269)
Municipality FE	No	No	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes
Observations	17436	7,602	17436	7602	16884	7338
R^2	0.005	0.011	0.914	0.964	0.916	0.971
Number of municipalities	2906	1267	2906	1267	2814	1223

Note: Standard errors clustered at municipality level in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively

To evaluate the variation in the timing of the introduction of the night alcohol sale ban on road traffic and alcohol-related hospitalizations when it is difficult to clearly define a treatment area, a different approach has been used. Table 3.21 displays the results from the OLS *exposition* approach. It appears that an increase in the exposition to the police ban leads to significant reductions in the number of alcohol-related hospitalizations. In column 2, the coefficient of exposition to the policy ban has been reported along with some control variables for unemployment, municipality revenue and density. In addition, municipality and year fixed effect have been added to the estimate to control respectively for differences in municipalities organizations and to capture the influence of potential temporal shocks in the dependent variable. The OLS estimate seems to suggest that an increase in the degree of exposition leads to significant (5%) reductions in the number of alcohol related hospitalizations. On the contrary, in column 1, the result indicates the absence of a significant effect of the exposition to policy ban in the reduction of hospitalizations due to road accidents.

Table 3.21 Effect of exposure to the policy ban on the number of road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013), (by patient municipality of residence)

	OLS			
	RTA		AR	
	(1)		(2)	
Exposition to policy ban	-5.9342		-1.7235	**
	(5.8670)		(0.8941)	
Unemployment	-0.7488	***	0.0499	
	(0.1829)		(0.0849)	
Revenue	-0.0000		-0.0000	***
	(0.0000)		(0.0000)	
Density	0.0610	***	0.0483	***
	(0.0204)		(0.0132)	
Municipality FE	Yes		Yes	
Year FE	Yes		Yes	
Observations	17376		7524	
R^2	0.915		0.971	
Number of municipalities	2896		1254	

Note: Robust standard errors in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively.

It would have been interesting analyse the effect of the policy ban for different age classes and sex. However, because this type of administrative health data suffers of sensitivity problems, as mentioned in the first chapter of this thesis, it has not been possible to obtain data disaggregated per individual at municipal level.

Overall, no matter which specification has been used, it seems that the night alcohol sale ban in motorway services area have lead to an indirect reduction in the number of alcohol related hospitalizations.

3.10 Conclusions

This chapter investigated whether the article 53 (law n. 120), which has been introduced by the Italian Parliament in 2010 and restricts the hours and days of alcohol sale, has been effective in reducing the number of car-accidents and alcohol-related hospitalizations.

In order to gather evidence on this specific topic, administrative health data provided by the Italian Ministry of Health for the period 2007-2013 and information regarding hospitals locations and socio-economic characteristics of the surrounding areas have been combined and analysed using both DD models and an “exposure” approach.

Estimates at hospital municipality level seems to suggest the ineffectiveness of this policy ban in reducing both motor vehicle and alcohol-related hospitalization. Estimates from the “exposure” model, obtained after aggregating admissions in hospitals that are located within the same municipality, confirm the absence of an effect of the night sale ban in reducing alcohol-related hospitalizations.

Counts of hospitalizations by municipality of residence represent an interesting, and so far poorly explored, way of exploiting information contained in the hospital discharge data. Estimates obtained using those data reports a strong effect of the night sale alcohol ban in reducing alcohol-related hospitalizations. These results are confirmed after the introduction of several control variables and by adopting the alternative exposure approach.

Overall, there is evidence of the effectiveness of the policy ban in reducing alcohol-related hospitalization when aggregating data at patient municipality of residence.

Concluding remarks

This thesis has analysed the usefulness of administrative health data to perform economic research contributing to the empirical literature in a twofold manner. On the one hand, administrative health data have been used to evaluate whether patient's mobility toward distant hospitals represent a distinct phenomenon with its own intrinsic characteristics. On the other hand, Italian health data have been exploited to analyse the effectiveness of a policy ban, which prohibited the on-trade sale of hard liquors on motorway service areas, on road traffic accidents and alcohol-related hospitalizations.

The results obtained through this work of research can be summarize as follows.

In Chapter 2, health data coming from the Italian Ministry of Health have been used to check whether hospital characteristics are playing similar role independently of the hospital being close or distant from patient lieu of residence. Using conditional logit and mixed logit models, it has been found that hospitals with short distance seems to attract more patients. However, some differences can be found when imposing parameters within the two groups of *distant* vs *close* hospitals. On the one hand, looking at the sample of MDC admissions, the number of patient outflows seems to be restrained by both the number of beds and the location of hospitals in urban administrative centres. On the other hand, with regards to Cancer admission, an interesting effect is shown by distance who seems to not influence negatively distant hospital choices. Overall, attractiveness of distant hospital appears as a quite distinct phenomenon, for which many of the existing findings in the literature, dealing with short-run mobility, must be carefully re-considered. Sardinian insularity represents an excellent case study, however, future research could further explore this topic by including data on Southern Italian LHA to better understand the characteristics affecting patient choice of distant care and whether the presence of bordering mobility change the empirical evidence found in this work.

In the third chapter, using health data for the period 2007-2013 and information on hospital locations, DD methods and "*Exposure*" models have been estimated. The results obtained using count data at hospital municipality level seem to suggest the ineffectiveness of the policy in reducing the number of car accident and alcohol-related hospitalizations. However, estimates obtained using aggregated data at patient municipality of residents report strong effects of the night sale alcohol ban in reducing alcohol-related hospitalizations. Thus, although the policy ban was primary aimed at reducing road accidents caused by alcohol misuse, it has had an indirect effect in reducing alcohol-related hospitalization. These results are confirmed after the introduction of several control variables and by adopting the alternative exposure approach. These findings are consistent with those obtained by Marcus and Siedler (2015), which find that the prohibition of the night alcohol sale in the state of Baden-Wuttemberg, in Germany, have been effective in reducing alcohol-related hospitalizations.

It would be interesting to extend the analysis including data on emergency departments in order to better appraise the magnitude of the phenomenon and the real effectiveness of the policy. In fact, not all patients who visit accidents and emergency departments are then hospitalized, thus limiting the analysis only to a specific sample of severe patients.

Appendix I

Table I - Conditional logit model

	Model 1b			Model 2b		
	Coeff.		SE	Coeff.		SE
Distance	-0.0330	***	0.0002	-0.0176	***	0.0005
Case mix index	4.4112	***	0.0527	-0.0531		0.4175
Size (number of beds)	-0.0027	***	0.0000	-0.0042	***	0.0002
Teaching	1.3894	***	0.0199	-0.0065		0.1265
Private accredited	1.1998	***	0.0510			
Rac	0.3027	***	0.0251	1.0170	***	0.0838
Market share	-0.6092	***	0.0329	4.3272	***	0.2955
Ami Radj	0.0115	***	0.0026			
Smdcs Radj				-0.2257	***	0.0163
Dummy Regio	Yes					
Log Likelihood	-79284.27			-4512.85		
Pseudo R^2	0.484			0.433		
Degree of freedom	9			8		
N. of admissions	45138			2575		
N. hospitals	30			22		
Hausman-McFadden P-value	0.000			0.000		

Appendix II

Table I – Road traffic accident hospitalizations, by regions (values in %)

Region	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Piedmont	5.57	6.06	5.85	5.52	5.76	5.56	5.23	4.93	5.06	5.51	6.16	5.81	5.82
Aosta Valley	0.14	0.16	0.16	0.18	0.17	0.16	0.29	0.27	0.32	0.30	0.18	0.11	0.18
Lombardy	20.01	21.06	20.65	21.19	22.92	21.73	20.95	19.60	19.48	17.64	16.99	16.87	15.65
Prov. Bozen	0.88	1.14	0.94	0.77	1.44	1.29	1.32	1.39	1.06	1.10	1.25	1.42	1.34
Prov. Trento	1.07	0.70	0.65	1.11	1.20	0.98	1.12	0.92	1.11	1.08	1.13	1.26	1.34
Veneto	5.38	6.20	6.89	6.30	5.89	5.66	4.98	4.87	4.47	4.14	4.37	4.37	4.14
Friuli Venezia Giulia	1.95	2.31	2.36	2.62	2.89	2.65	2.90	2.78	2.55	2.62	2.39	2.49	2.51
Liguria	2.93	2.83	2.55	2.41	1.22	1.01	0.70	0.73	0.19	0.28	0.55	1.20	1.78
Emilia Romagna	9.81	9.36	8.34	8.03	8.12	8.03	7.10	6.69	7.03	6.95	6.75	8.63	7.22
Tuscany	5.42	5.27	5.35	5.41	4.70	4.61	4.55	3.46	3.87	4.43	4.76	4.59	5.02
Umbria	0.71	0.62	0.50	0.25	0.12	0.37	0.94	1.05	1.37	1.33	1.56	1.25	1.29
Marche	2.30	2.32	2.25	2.25	2.69	2.67	2.26	2.38	2.29	1.79	1.68	2.22	2.08
Lazio	11.64	5.60	4.13	3.96	3.54	3.28	8.52	9.07	9.24	9.21	8.94	8.95	8.07
Abruzzo	1.52	1.98	2.71	3.44	3.32	3.17	3.04	3.05	2.72	2.99	3.76	3.43	2.59
Molise	0.64	0.83	0.79	0.77	0.60	0.87	0.78	0.69	0.59	0.71	0.59	0.52	0.48
Campania	12.56	15.09	15.65	16.45	15.97	16.95	15.53	17.14	17.28	14.22	14.06	12.04	14.59
Apulia	7.62	8.20	6.97	6.25	6.25	6.89	6.76	7.04	7.14	9.00	7.29	5.97	8.15
Basilicata	0.64	0.29	0.16	0.16	0.24	0.54	0.56	0.81	0.93	1.04	0.97	0.91	0.89
Calabria	3.41	3.46	2.60	3.03	2.95	3.97	3.53	3.68	3.13	3.20	3.03	3.04	3.49
Sicily	5.81	6.50	7.35	7.02	7.10	6.77	6.29	6.50	6.57	6.58	8.51	10.01	8.63
Sardinia	-	-	3.15	2.88	2.91	2.83	2.66	2.95	3.60	5.89	5.09	4.90	4.75
Total (Freq.)	105005	87021	75447	65829	60589	59275	58868	53098	50854	48027	49023	43190	42769

Table II- Alcohol related hospitalization, by regions (values in %)

Region	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Piedmont	7.28	7.61	7.40	7.89	9.24	8.04	7.82	8.16	8.73	8.42	9.19	7.99	8.64
Aosta Valley	0.20	0.71	0.58	0.92	0.72	0.54	0.80	0.55	0.57	0.43	0.45	0.71	0.76
Lombardy	18.72	18.52	18.36	16.86	17.74	17.87	16.96	15.78	18.47	17.12	18.12	17.16	17.41
Prov. Bozen	3.42	3.91	3.87	4.12	4.07	3.91	4.49	4.52	3.26	3.05	3.06	3.33	3.64
Prov. Trento	1.68	1.20	1.78	1.42	1.57	1.65	1.48	1.95	2.14	0.90	0.86	0.80	1.05
Veneto	11.38	10.56	11.00	10.91	10.28	9.58	10.30	10.70	10.20	10.41	9.99	9.61	8.86
Friuli Venezia Giulia	2.73	2.65	2.52	2.80	2.71	2.83	3.30	3.04	3.49	3.78	3.70	4.04	3.66
Liguria	5.45	5.10	4.34	3.87	3.73	4.04	3.69	3.46	3.45	3.73	3.54	4.11	3.96
Emilia Romagna	6.82	6.93	6.62	7.53	8.00	7.46	8.74	9.89	9.66	10.59	11.44	12.31	12.36
Tuscany	5.88	6.20	5.93	5.57	4.53	4.10	4.74	5.01	4.82	4.61	4.23	4.89	5.13
Umbria	0.96	0.68	1.08	1.03	0.92	0.86	1.02	1.13	1.12	1.12	1.23	1.23	1.19
Marche	4.92	5.09	5.23	5.79	6.19	5.77	5.98	6.58	6.41	6.91	7.52	8.40	9.27
Lazio	6.79	5.98	6.84	6.67	6.77	8.03	6.46	6.93	6.99	8.42	6.69	6.40	5.99
Abruzzo	3.42	3.39	3.72	3.81	3.75	4.26	3.64	2.69	1.85	1.74	1.99	1.50	2.09
Molise	0.86	0.85	0.73	0.64	0.74	0.77	0.79	0.61	0.59	0.59	0.54	0.40	0.43
Campania	4.79	5.18	5.28	5.63	5.12	6.15	5.77	5.87	6.13	5.69	5.27	5.51	4.77
Apulia	5.12	5.07	4.57	4.12	4.38	4.39	4.30	4.24	4.19	4.17	3.68	3.85	4.02
Basilicata	0.88	0.80	0.76	0.74	0.69	0.54	0.41	0.47	0.55	0.62	0.61	0.27	0.19
Calabria	1.93	2.09	1.99	1.91	1.51	1.78	1.74	1.64	1.22	1.36	1.24	1.24	0.85
Sicily	4.04	4.06	4.20	4.16	3.57	3.64	3.68	3.40	3.17	3.17	3.67	2.88	2.69
Sardinia	2.73	3.45	3.20	3.59	3.75	3.79	3.90	3.37	2.98	3.18	2.97	3.39	3.04
Total (Freq.)	27444	25465	24170	23033	21527	19703	18201	16858	15948	14518	13940	12896	11915

Table III- Number of road accident hospitalizations, by age classes

Age interval in years	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
15-19	14947	12400	10763	8847	8257	8202	8314	7389	6776	6076	5511	4245	3844
20-24	14534	11940	9844	8058	6986	6652	6375	5513	5098	4727	4546	3657	3491
25-29	13661	10899	9172	7544	6464	5968	5634	4732	4252	3780	3714	3101	2920
30-39	17849	14889	13241	11687	10710	10235	10193	8619	8043	7496	7390	6191	5958
40-49	10844	8979	8221	7477	7127	7211	7440	7068	6989	6496	7104	6403	6363
50-64	12925	11010	9211	8263	7831	7807	7834	7342	7565	7511	8236	7682	8006
>64	13956	11765	10369	9208	8956	8910	9146	8832	8678	8746	9360	9118	9496

Table IV - Number of alcohol-related hospitalizations, by age classes

Age interval	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
10-14	159	150	104	180	165	184	165	179	147	153	121	107	121
15-19	1092	942	868	761	733	777	677	584	493	479	455	390	326
20-24	1192	954	920	820	722	663	513	480	441	376	368	310	257
25-29	1688	1471	1251	1294	1144	946	850	657	669	602	524	468	427
30-39	6647	6093	5840	5570	5061	4391	3701	3523	3160	2770	2629	2476	2118
40-49	7050	7001	6719	6558	6324	5967	5747	5362	5105	4682	4630	4202	4045
50-64	7213	6813	6483	5968	5655	5109	4930	4553	4397	4148	4013	3762	3508
>64	2403	2041	1985	1882	1723	1666	1618	1520	1536	1308	1200	1181	1113

Table V – Distribution of municipalities with hospitals

Region	RTAH	ARH
Piedmont	23	25
Aosta Valley	1	1
Lombardy	80	74
Prov. Bozen	7	7
Prov. Trento	7	7
Veneto	19	26
Friuli Venezia Giulia	10	10
Liguria	2	5
Emilia Romagna	20	23
Tuscany	17	27
Umbria	6	8
Marche	12	10
Lazio	13	19
Abruzzo	14	13
Molise	3	3
Campania	34	25
Apulia	22	25
Basilicata	5	4
Calabria	10	9
Sicily	29	25
Total	334	346

Note: The table shows the number of municipalities in which have been recorded hospitalizations for road traffic accidents (RTAH) and alcohol-related hospitalization (ARH).

Table VI - Distribution of patient's municipalities

Region	2007	2008	2009	2011	2012	2013
Piedmont	354	369	381	346	316	291
Aosta Valley	42	34	38	29	34	39
Lombardy	841	828	862	795	746	749
Prov. Bozen	96	97	86	94	87	84
Prov. Trento	93	82	89	72	56	61
Veneto	398	398	374	372	349	334
Friuli Venezia Giulia	135	133	145	128	135	125
Liguria	100	101	92	91	90	87
Emilia Romagna	248	256	241	243	243	231
Tuscany	165	179	159	147	159	154
Umbria	54	57	51	55	50	48
Marche	136	138	139	134	125	130
Lazio	174	189	176	145	168	135
Abruzzo	154	149	130	114	112	120
Molise	70	50	48	41	25	29
Campania	252	249	254	186	204	184
Apulia	176	173	168	158	141	150
Basilicata	57	55	47	50	32	26
Calabria	152	126	115	108	94	91
Sicily	177	174	158	155	131	114
Sardinia	192	178	185	155	159	137
Total	4066	4015	3938	3618	3456	3319

Appendix III

Table I – The effect of the night alcohol sale prohibition in motorways service areas on road traffic accidents (RTA) and alcohol related (AR) hospitalizations (2007-2013), (data at municipal-level)

	Log				Poisson			
	RTA		AR		RTA		AR	
Effect ($post_t$*$treat_i$)	0.1808	***	0.1393	***	0.1379	***	0.1024	
	(0.0485)		(0.0407)		(0.0525)		(0.0357)	
Unemployment	0.0064		0.0077		-0.0215	***	0.0126	*
	(0.0108)		(0.0094)		(0.0109)		(0.0071)	
Revenue	-0.0000	***	-0.0000	**	-0.0000		-0.0000	
	(0.0000)		(0.0000)		(0.0000)		(0.0000)	
Density	-0.0007		0.0005		0.0017		0.0029	**
	(0.0013)		(0.0017)		(0.0017)		(0.0012)	
Municipality FE	Yes		Yes		Yes		Yes	
Year FE	Yes		Yes		Yes		Yes	
Observations	1662		1380		1662		1380	
R^2	0.845		0.900		0.885		0.895	
Number of municipalities	277		230		277		230	

Note: Robust standard errors in parenthesis. *, **, *** indicate statistical significance at 10%, 5%, 1% respectively.

Bibliography

Aakvik, A. and Holmas, T.H. (2006). Access to primarly health care and health outcomes: the relationship between GP characteristics and mortality rates. *Journal of Health Economics*, 25: 1139-1153.

Abowd, J.M., Kramarz, F., Margolis, D.N. (1999). High wage workers and high wage firms. *Econometrica*, 67: 251–333

Acemoglu, D. and Finkelstein, A. (2008). Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector. *Journal of Political Economy*, 116 (5):837-880

Adams, E.K., Houchens, R., Wright, G.E., Robbins, J. (1991). Predicting hospital choice for rural Medicare beneficiaries: the role of severity of illness. *Health Service Research*, 26(5):583-612.

Agenas. (2012). I quaderni di monitor. Elementi di analisi e osservazione del Sistema salute.

American Statistical Assocation (1977). “Report of the Ad Hoc Committee on Privacy and Con dentiality,” *The American Statistician*, 31, pp. 59-78.

Anderson, P., Chisholm, D., Fuhr D.C. (2009). Effectiveness and cost-effectiveness of policies and programs to reduce the harm caused by alcohol. *Lancet*, 373: 2234-2246

Andrews, MR. (2015). Statewide Hospital Discharge Data: Collection, Use, Limitations, and Improvements. *Health Services Research Journal*, 50(1):1273-1299.

Angrist, JD. and Krueger, AB. (1999). Empirical strategies in labor economics. In: Ashenfelter, OC and Card, D. Handbook of Labor Economics, Elsevier, 3A: 1277-1366.

Ashenfelter, O. and Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, 67:648–660.

Arrow, K. (1962). The economic implications of Learning by doing. *Review of Economic Studies*, 29(3):155-173.

Arrow, KJ. (1963). Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, 53(5):941-973.

Autor, DH. (2003). Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. Journal of Labor Economics, 21(1):1-42.

Avdic, D., Lundborg, P.,Vikström, J. (2014). Learning-by-doing in a Highly skilled profession when stakes are high: evidence from advanced cancer surgery. IZA discussion Paper No 8099.

Ayanian, J.Z. and Weissman, J.S. (2002). Teaching hospitals and quality of care: a review of the literature. *The Milbank Quarterly*, 80(3):569-593.

Babor, T., Caetano, R., Casswell, S., Edwards, G. et al. (2010). ALCOHOL. No ordinary commodity, research and public policy. Second edition. Oxford University Press.

Bacharach, S.B., Bamberger, P., Biron, M. (2010). Alcohol consumption and workplace absenteeism: the moderating effect of social support. *Journal of Applied Psychology*, 95(2):334-348.

Balia, S., Brau, R., Marrocu, E. (2014). What Drives Patient Mobility Across Italian Regions? Evidence from Hospital Discharge Data. In Levaggi R. And Montefiori M. (eds.), *Health Care Provision and Patient Mobility, Developments in Health Economics and Public Policy* 23.

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

Ben-Akiva, M. and Bolduc, D. (1996). Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure. Working paper, Department of Civil Engineering.

Bertrand, M., E. Duflo, Mullainathan S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119:249-275.

Besley, T. and Ghatak, M. (2003). Incentives, Choice, and Accountability in the Provision of Public Services. *Oxford Review of Economic Policy*, 19(2), 235-249.

Beukers, P.D.C, Kemp, R.G.M., Varkevisser, M. (2014). Patient hospital choice for hip replacement: empirical evidence from Netherlands. *European Journal of Health Economics*. 15:927-936.

Biderman, C., De Mello, J.M.P., Schneider, A. (2009). Dry laws and homicides: evidence from the Sao Paulo metropolitan Area. *The Economic Journal*, 120:157-182.

Black, S., Devereux, P., Salvanes, K. (2007). From the cradle to the labour market? The effect of birth weight on adult outcomes. *The Quarterly Journal of Economics* 122, 409 -39.

Bockerman, P., Hyytinen, A., Maczulsku, T. (2015). Alcohol consumption and long-term labor market outcomes. *Health Economics.*

Botvin, G.J., Baker, E., Dusenbury L., Botvin, E.M., Diaz, T. (1995). Long-term follow-up results of a randomized drug abuse prevention trial in a white middle-class population. Journal of the American Medical Association, 273:1106-1112.

Boyce, T., Dixon, A., Fasolo, B., Reutskaja.(2010). Choosing a High Quality Hospital. The role of nudges, scorecard design and information. The Kings' Fund.

Bradley, C.J., Penberthy, L., Devers, K.J., Holden, D.J. (2010). Health services research and data linkages: issues, methods and directions for the future. *Health services research*, 45(5): 1468-1488.

Brekke, K.R., Gravelle, H., Siciliani, L., Straume, O.R. (2014). Patient choice, mobility and competition among health care providers. Levaggi R and Montefiori M (eds.), *Health Care Provision and Patient Mobility*, Developmentes in Health Economics and Public Policy 12. Springer-Verlag Italia.

Brett, C.E., Deary, I.J. (2014). Realising health data linkage from a researcher's perspective: following up the 6-Day sample of the Scottish mental survey 1947. *Longitudinal and Life Course Studies*, 5(3):283-298.

Buczko, W. (1992). What affects rural beneficiaries Use of urban and rural hospitals? *Health care Financing Review*. 14(2), 107-114.

Burge, P., Devlin, A., Appleby, J., Gallo, F., Rohr, C., Grant, J. (2004). Do patients always prefer quicker treatment? A discrete choice analysis of patients' stated preferences in the London patient choice project. *Applied Health Economics and Health Policy*, 3(4):183-194.

Burns, L.R. and Wholey, DR. (1991). The impact of physician characteristics in conditional choice models for hospital care. *Journal of Health economics*, 11:43-62.

Cadarette, S.Z., Wong, L. (2015). An introduction to health care administrative data. *The Canadian Journal of Hospital Pharmacy*, 68(3):232-237.

Card, D., Chetty, R., Feldstein, M., Saez, E. (2010). Expanding Access to Administrative Data for Research in the United States. NSF White Paper.

Cardell, S.N., Dobson, R., Dunbar, F. (1978). Consumer research implementations of random coefficient models. *Advances in Consumer Research*, 5. H.K. Hunt (ed.), Association for Consumer Research, Ann Arbor.

Carpenter, C. and Dobkin, C. (2011). The minimum legal drinking age and public health. *Journal of Economic Perspectives*, 25(2):133-156.

Carpenter C, Eiseberg D. (2009). Effects of Sunday sales restrictions on overall and day-specific alcohol consumption: evidence from Canada. Journal of Studies on Alcohol and Drugs, 70:126-133.

Carrell SE, Hoekstra M, West JE. (2011). Does drinking impair college performance? Evidence from a regression discontinuity approach. *Journal of Public Economics*, 95: 54–62.

Chaloupka FJ, Grossman M, Saffer H. (2002). The effect of Price on Alcohol Consumption and Alcohol-Related Problems. Alc Res Health, 26:22-34.

Chamberlayne R., Barer ML, Lawerence WJ. (1998). Creating a Population- based Linked Health Database: A resource for Health Services Research. *Canadian Journal of Public Health*, 89(4): 270-273.

Cherpitel, C.J., Ye, Y., Bond, J., Borges, G. (2003). The casual attribution of injury to alcohol consumption: a cross-national meta-analysis from the emergency room collaborative alcohol analysis project. *Alcohol Clinical and Experimental Research Journal*, 27:1805-1812.

Chetty, R., Friedman, J.N., Rockoff, J.E. (2014). Measuring the Impact of teachers I: evaluating bias in theacher value-added estimates. *American Economic Review*, 104(9):2593-2632

Christen, P. and Goiser, K. (2007). Quality and complexity measures for data linkage and deduplication. *Studies in Computational Intelligence***,** 43:127-151.

Christiadi, C. and Cushing, B. (2007). Conditional Logit, IIA, and Alternatives for Estimating Models of Interstate Migration. Research Paper 2007-4

Ciriani, V., De Capitani di Vimercati, S., Foresti, S., Samarati, P. (2007). K-Anonymity. In Secure Data Management in Decentralized Systems, T. Yu and S. Jajodia, Eds. Springer-Verlag, Berlin.

Conferenza Stato Regioni. Accordo tra Ministero della Salute e le Regioni e le Province autonome di Trento e Bolzano di approvazione delle Linee guida per la codifica delle informazioni cliniche presenti sulla scheda di dimissione ospedaliera (SDO). 6 Giugno 2002.

Connelly, R., Playford, C.J., Gayle, V., Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59:1-12.

Cook, P.J., Moore, M.J. (2002). The economics of alcohol abuse and alcohol-control policies. *Health affairs,* 21(2):120-133.

Cooper, Z., Gibbons, S., Jones, S., & McGuire, A. (2011). Does hospital competition save lives? Evidence from the English patient choice reforms. *Economic Journal*, 121(554), F228–F260. Patient Choice, Mobility and Competition Among Health Care Providers

Cuijpers, P. (2003). Three decades of drug prevention research. Drugs: Education, Prevention and Policy, 10:7_20.

Cutler D., Huckman R., Landrum M. (2004). The role of information in medical markets: an analysis of publicly reported outcomes in cardiac surgery. *American Economic Review*, 94:342-346.

Daley, J.I., Stahre, M.A., Chaloupka, F.J., Naimi, T.S. (2012). The impact of a 25-Cent_per_Drink Alcohol Tax Increase. *American Journal of Preventive Medicine*, 42(4):382-389.

Decreto Ministeriale 26 Luglio 1993 (in Gazzetta Ufficiale 3 Agosto 1993 n. 180). Disciplina del flusso informative sui dimessi dagli Istituti di ricovero pubblici e privati.

Decreto Ministeriale 28 Dicembre 1991. Istituzione della scheda di dimissione ospedaliera.

Delcher, C., Maldonado-Molina, M.M., Wagenaar A.C. (2012). Effects of alcohol taxes on alcohol-related disease mortality in New York State from 1969 to 2006. *Addictive Behaviors,* 37:783-789.

Dixon, A., Robertson, R., Appleby, K., Burge, P., Devlin, N., Magee, H. (2010). Patient choice. How patients choose and how providers respond. The King’s Fund

Dusheiko, M., Gravelle H., Jacobs, R. (2004). The effect of practice budgets on patient waiting times: allowing for selection bias. *Health Economics* 13, 941-58.

Einav, L. and Levin, J. (2014). Economics in the age of big data. Science, 346. DOI: 10.1126/science.1243089

Elias, P. (2014). Administrative data. In: Dus aA., Nelle D, Stock G, Wagner G. (Eds.), Facing the Future: European Research Infrastructures for the Humanities.

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

Farrell, S., Manning, W.G., Finch, M.D. (2003). Alcohol dependence and the price of alcoholic beverages. *Journal of Health Economics*, 22:117-147.

Flowers, J. and Ferguson, B. (2010). The future of health intelligence: challenges and opportunities. *Public Health*, 124(5):274-277.

Garrat, E., Barnes, H., Dibben, C. (2010). Health administrative data: exploring the potential for academic research, St Andrews: Administrative Data Liaison Service.

Gaynor, M., Moreno-Serra, R., Propper, C. (2013). Death by Market Power: Reform, Competition and Patient Outcomes in the National Health Service. *American Economic Journal: Economic Policy*, 5(4): 134–166.

Gaynor, M., Propper, C., Seiler, S. (2012). Free to Choose? Reform and Demand Response in the English National Health Service. Working Paper No. 18574, NBER.

Gaynor, M. and Town, R.J. (2012). Competition in Health Care Markets. In T. Mc Guire M. Pauly & P. P. Barros (Eds.), Handbook of Health Economics (Vol. 2, pp. 499–637). Amsterdam:Elsevier.

Gibbs, D.A., Sangl, J.A., Burrus, B. (1996). Consumer perspectives on information needs for health plan choice. *Health care Financing Review*. 18(1), 55-73.

Goldman, D. and Romley, JA. (2008). Hospitals as hotels: the role of patient amenities in hospital demand. NBER Working Paper 14619.

Goddard, M., Mannion, R., Smith, P. (2000). Enhancing performance in Health care: a theoretical perspective on agency and the role of information. *Health Economics,* 9:95-107.

Goodman, D.C,, Fisher, E., Stukel, T., Chang, C. (1997). The distance to community medical care and the likelihood of hospitalization: is closer always better? *American Journal of Public Health* 87(7), 1144–1150.

Green, C., Heywood, J.S., Navarro, M. (2014). Did liberalizing on bar hours' decrease traffic accidents? *Journal of Health Economics*, 35:189-198.

Green, C. and Navarro Paniagua, M. (2016). Play Hard, Shirk Hard? The effect of Bar Hours regulation on worker absence. *Oxford Bulletin of Economics and Statistics, 248-264.*

Gronqvist, H. and Niknami, S. (2014). Alcohol availability and crime: Lesson from liberalized weekend sales restrictions. *Journal of Urban Economics*, 81:77-84.

Hamermesh, D.S. (2013). Six decades of top economics publishing: who and how? *Journal of economics literature*, 51:162-172.

Hansen, D.J. (1994). Prevention of alcohol use and abuse. *Preventive Medicine*, 23: 683-687.

Hausman, J. and McFadden, D. (1984). Specification tests for the multinomial logit model. Econometrica, 52, 1219–1240.

Heaton, P. (2012). Sunday liquor laws and crime. *Journal of Public Economics*, 96:42-52.

Hensher, D.A., Rose, J.M., Green, W.H. (2005). Applied Choice Analysis: A primer. Cambridge: Cambridge University press.

Hibbard, J.H., Slovic, P., Jewett, J.J. (1997). Informing consumer decisions in health care: implications from decision-making research. *The Milbank Quarterly*, 75(3):395–406.

Hospital Episode Statistics. Health and Social care information centre. (http://content.digital.nhs.uk/hes).

Iezzoni, L.I. (1997). Assessing quality using administrative data. *Annual of internal medicine*, 127(8):666-674.

Iezzoni, L.I., Daley, J., Heeren, T., Foley, S.M., Fisher, E.S., Duncan, C., Hughes, J.S., Coffman, G. (1994). Identifying complications of Care using administrative data. *Medical care*, 32(7):700-715.

Istituto Superiore di Sanità (2014). L'alcol in Italia e nelle regioni: analisi e proposte per la prevenzione.http://www.epicentro.iss.it/alcol/apd2014/OK%20SINTESI%20FINALE%20DATI%20APD%20SCAFATO.pdf.

Johansson, Alho, Kiiskinen, Poikolainen (2007). The association of alcohol dependency with employment probability: evidence from the the population survey "Health 2000" in Finland. *Health Economics*, 16(7):739-754.

Jommi C., Cantù E., Anessi-Pessina E., (2001). New funding arrangements in the Italian National Health Service. Int J Health Plann Mgmt. 16, 347-368.

Jones, A. (2007). Panel data methods and applications to health economics. HEDG WP.

Kemm, J.R., Robinson, J., Verne, J. (2010). Social care data in England: what they tell us and what they do not tell us. *Public Health*, 124:265-268.

Kenkel, D. and Sindelar, J. (2011). Economics of Health Behaviors and Addictions: Contemporary Issues and Policy Implications. *The Oxford Handbook of Health Econ, 206-231.*

Kunn, S. (2015). The challenges of linking survey and administrative data. IZA World of Labor, 214. doi: 10.15185/izawol.214

Lansley, A. (2010). 'My ambition for patient-centred care'. Speech, 8 June 2010. Available at: www.dh.gov. uk/en/MediaCentre/Speeches/DH_116643.

Lechner, M. (2010). The estimation of causal effects by difference-in-difference methods. University of St. Gallen Department of Economics Working Paper Series. 28.

Lester, R.A. (1946). Shortcomings of marginal analysis for the wage-employment problems. *American Economic Review*, 36:63-82.

Levaggi, R. and Menoncin, F. (2008). Fiscal Federalism, patient mobility and soft budget constraint in Italy. Politica Economica. *Italian journal of Economic Policy*. 3: 367-388.

Levitt, S.D. and Porter, J. (2001). How dangerous are drinking drivers? *Journal of Political Economy*, 109(6):1198-1237.

Lichtman, J.H., Leifheit-Limson, E.c, Goldstein, L.B. (2015). Centers for medicare and medicaid services medicare data and stroke research: goldmine or landmine? *Stroke*, 46(2):598-604.

Lindo, J.M.., Swensen, I.D., Waddell, (2013**).** Alcohol and student performance: Estimating the effect of legal access. *Journal of Health Economics*, 32(1): 22–32.

Lynge E, Lynge Sandegaard J, Reboly M. (2011) The Danish National Patient Register. *Scandinavian Journal of Public Health*, 39(7):30-33.

Lovenheim, M.F. and Steefel, D.P. (2011). Do blue laws save lives? The effect of Sunday alcohol sales bans on fatal vehicle accidents. *Journal of Policy Analysis and Management*, 30(4):798-820.

Ludvigsson, J.F, Otterblad-Olausson, P., Pettersson, B.U., Ekbom, A. (2009). The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *European Journal of Epidemiology*, 24(11):659-667.

Luft, H.S., Garnick, D., Mark, D., Peltzman, D., Phibbs, C., Lichtenberg, E., McPhee S. (1990). Does quality influence choice of hospital? *The Journal of the American Medical Association*. 263, 2899-2906.

Lunde, A.S. (1975). The birth number concept and record linkage. *American Journal of Public Health*, 65(11):1165-1169.

MacDonald, Shields. (2004). Does problem drinking affect employment? Evidence from England. *Health Economics*, 13(2):139-155.

Manning, W.G., Blumberg, L., Moulton, L.H. (1995). The demand for alcohol: The differential response to price. *Journal of Health Economics*, 14(2):123-148.

Marcus, J. and Siedler, T. (2015). Reducing binge drinking? The effect of a ban on late-night off-premise alcohol sales on alcohol-related hospital stays in Germany. *Journal of Public Economics*, 123:55-77.

Mason, J.K. and Laurie, GT. (2010). Mason and McCall-Smith's Law and Medical Ethics (7th ed) Oxford: Oxford Press.

Mazzali, C. and Duca, P. (2015). Use of administrative data in healthcare research. *Internal Emergency Medicine*, 10:517-524.

McFadden, D.L. (1974). Conditional logit analysis of qualitative choice behaviour. in P. Zarembka (ed.), FRONTIERS IN ECONOMETRICS, 105-142, Academic Press: New York.

McFadden, D. (1984). Econometric analysis of qualitative response models. In: Griliches Z and Intriligator MD (eds) Handbook of Econometrics, II. Amsterdam: Elsevier, pp. 1395–1457.

McFadden, D.L., Train, K.E. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15:447-470.

Messina, G., Vigiani, N., Lispi, L., Nante, N. (2003). Patient migration among the Italian regions in 2003. *Italian Journal of Public Health*. 5(1), 45-52.

Meyer, B.D. (1995). Natural and Quasi-Experiments in Economics. *American Statistical Association*, 13(2):151-161.

Middleton, J.C., Hahn, R.A., Kuzara, J.L., Elder, R., Brewer, R.D., Chattopadhyay, S., Fielding, J., Naimi, T.S., Toomey, T.L., Lawrence, B. (2010). Effectiveness of policies maintaining or restricting days of alcohol sales on excessive alcohol consumption and related harms. *American Journal of Preventive Medicine*, 39(6):575-589.

Moise, P. (2001). Using hospital administrative databases for a disease-based approach to studying health care systems. OECD Ageing related disease study. https://www.oecd.org/denmark/1889879.pdf.

Moscone, F., Tosetti, E., Vittadini, G. (2012). Social interaction in patients' hospital choice: evidence from Italy. *Journal of the Royal Statistical Society*, 175:453-472.

Mullahy, J. and Sindelar, J. (1996). Employment, unemployment, and problem drinking. *Journal of Health Economics*, 15:409-434.

National Institute for Health and Clinical Excellence (NICE). (2007). A review of the effectiveness and cost-effectiveness of interventions delivered in primary and secondary schools to prevent and/or reduce alcohol use by young people under 18 years old. Alcohol and schools: review of effectiveness and cost-effectiveness. NICE: main report (PHIAC 14.3a). http://www.nice.org.uk/nicemedia/pdf/ AlcoholSchoolsConsReview.pdf (accessed March 2, 2008).

National Institute on Ageing, National Institutes of Health. (2016). Expert Meeting on the potential value of Centers for Medicare and Medicaid services data as a resource for National institutes of aging studies.

Nelson, J.P. (2013). Meta-analysis of alcohol price and income elasticities – with corrections for publication bias. *Health Economics Review*, 3-17.

Neely, S.K. and McInturff, W.D. (1988). What Americans Say about the Nation's Medical Schools and Teaching Hospitals. Report on Public Opinion Research, Part I.

Newhouse, J,P. (1994). Frontier estimation: How useful a tool for health economics? *Journal of Health Economics*, 13:317:322.

Newton, A., Sarker, S.J., Pahal, G.S., van den Bergh, E., Young, C. (2007). Impact of the new UK licensing law on emergency hospital attendances: a cohort study. *Emerg Med J*, 24(8):532-534.

Nguyen, L.L. and Barshes, N.R. (2010) Analysis of large databases in vascular surgery. *Journal of Vascular Surgery*, 52(3):768:774.

Obenauer, M. and von der Nienburg, B. (1915). Effect of minimum- wage determinations in Oregon. *Bulletin of the U.S. Bureau of Labor Statistics*, 176, Washington, D.C.: U.S. Government Printing Office.

OECD. (2000) Terminology on Statistical Metadata. Economic Commission for Europe of the United Nations (UNECE), Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva.

Peters, E., Dieckmann, N., Dixon, A., Hibbard, J.H., Mertz, C.K. (2007). Less is More in Presenting Quality Information to Consumer. Medical Care Research and Review, 64(2)169-190.

Piketty, T. and Saez, E. (2014). Inequality in the long run. *Science*, 344: 838–843.

Pischke, J.S. (2005). Empirical methods in applied economics. Lecture notes.

Pope, D.G. (2009). Reacting to rankings: evidence from America’s Best Hospitals. *Journal of Health Economics*, 28(6):1154–1165.

Popova, S., Giesbrecht, N., Bekmuradov, D., Patra, J. (2009). Hours and days of sale and density of alcohol outlets: impacts on alcohol consumption and damage: a systematic review. *Alcohol and Alcoholism*, 44(5):500-516.

Porell, F.W. and Adams, EK. (1995). ‘‘Hospital Choice Model: A Review and Assessment of Their Utility for Policy Impact Analysis.’’ *Medical Care Research and Review,* 52 (2): 158–95.

Quantin, C., Allaert, F.A., Avillach, P., Fassa, M., Riandey, B., Trouessin, G., Cohen, O. (2008). Building application-related patient identifiers: what solution for a European country? Internal Journal of Telemedicine and Applications. http://dx.doi.org/10.1155/2008/678302.

Rademakers, J., Delnoij, D., Boer, D.de. (2011). Structure, process or outcome: which contributes most to patients’ overall assessment of healthcare quality? *BMJ Quality and safety*, 20(4), 326-331.

Regidor, E. (2004). The use of personal data from medical records and biological materials: ethical perspectives and the basis for legal restrictions in health research. *Social Science & Medicine,* 59:1974-1984.

Reyes, J.W. (2007). Environmental Policy as Social Policy? The Impact of Childhood Lead Exposure on Crime. *B.E. Journal of Economic Analysis and Policy,* 7(1).

Rhum, C. (1996). Alcohol policies and highway vehicle fatalities. *Journal of Health Economics,* 15: 435-454.

Roche, A.M. et al. (2008). Workers’ drinking patterns: the impact on absenteeism in the Australian workplace. Addiction, 103:738–748.

Roh, C.Y., Lee, KH., Fottler, M.D., (2008). Determinants of hospital choice of rural hospital patients: the impact of networks, service scopes, and market competition. J Med Syst. 32, 343-353.

Roh, C.Y. and Moon, M. (2005). Nearby, but not wanted? The bypassing of rural hospitals and policy implications for rural health care systems. *Policy Studies Journal*, 33:477-494.

Romano, P.S. and Mutter, R. (2004). The evolving science of quality measurement for hospitals: implications for studies of competition and consolidation. Int. J. of Hlth Care Finan. Econ. 28: 1154-1165.

Romley, J.A. and Goldman, D.P. (2011). How costly is hospital quality? A revealed-preference approach. The Journal of Industrial Economics, 578-608.

Ryan, R.M., Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being’. American Psychologist, 55(1):68–78.

Santos, R., Gravelle, H., Propper, C. (2016). Does quality affect patients' choice of doctor? Evidence from UK. *The economic Journal*, DOI:10.1111/ecoj.12282.

Sassi, F. (2015). Tackling Harmful Alcohol Use: Economics and Public Health Policy. OECD.

Schneider, E.C. and Epstein, A.M. (1996). Influence of cardiac-surgery performance reports on referral practices and access to care. *New England Journal of Medicine*, 335:251-256.

Schmidt, M., Johannesdottir Schmidt, S.A., Sandegaard, J.L., Ehrenstein, V., Perdse, L., Sorensen, H.T. (2015). The Danish National Patient Registry: a review of content, data quality, and research potential. Clinical Epidemiology, 5(7):449-490.

Schoenman, J.A., Sutton, J.P., Kintala, S., Love, D., Maw, R. (2005). The Value of Hospital Discharge Databases. Rockville, MD: Agency for Healthcare Research and Quality; and Final report submitted to the Agency for Healthcare Research and Quality under contract number 282-98-0024. Available at http://www.hcup-us.ahrq.gov/reports/final_report.pdf.

Scottish Government. (2010). Linking social care, housing & health data. Data linkage literature review 2010. Edinburgh: Scottish Government.

Simon, J.L. (1966). The price elasticity of liquor in the U.S. and a simple method of determination. *Econometrica*, 34:193-205.

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

Smith C, Stefan A, Valdmanis V, Verheyen P. (1997). Principal-agent problems in health care systems: an international perspective. *Health policy*, 41:37–60.

Snow, J. (1855). On the Mode of Communication of Cholera. 2nd edition. London: John Churchill.

Stockwell T, Chikritzhs TN. (2009) .Do relaxed trading hours for bars and clubs means more relaxed drinking? A review of the international research on the impacts of changes to permitted hours of drinking. *Crime Prevention and Community Safety*, 11:153-170.

Syverson C. (2011). What determines productivity? *Journal of Economic Literature*, 49: 326–365.

Schwartz, B. (2004). The Paradox of Choice: Why more is less. New York: Harper Perennial.

Schneider, E.C. and Epstein, A.M. (1998). Use of public performance reports: a survey of patient undergoing cardiac surgery. *JAMA*, 279(20):1638-1642.

Sweeney, L. (2002). k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5): 557-570.

Tai, W.C., Porell, F.W., Adams, K. (2004). Hospital Choice and Rural Medicare Beneficiaries: Patients, Hospital Attributes, and the Patient-Physician Relationship. *Health Services Research*, 39(6), 1903-1922.

Tay, A. (2003). Assessing Competition in Hospital Care Markets: The Importance of Accounting for Quality Differentiation. *The RAND Journal of Economics*. 34(4), 786-814.

Taylor, B., Irving, H.M., Kanteres, F. et al. (2010). The more you drink the harder you fall: a systematic review and meta-analysis of how acute alcohol consumption and injury or collision risk increase together. *Drug Alcohol Depend*, 110(1-2):108-116.

Taubman, S.L., Allen, H.L., Wrighy, B.J., Baicker, K., Finkelstein, A.N. (2014). Medicaid increases emergency-department use: evidence from Oregon's health insurance experiment. *Science*, 343: 263-268.

Terza, J.V. (2002). Alcohol abuse and employment: a second look. *Journal of Applied Economics*, 17:393-404.

Train, K. (2003). Discrete choice Methods with Simulation. (http://eml.berkeley.edu/choice2/ch6.pdf).

Train K. (2009). Discrete choice models with simulation. 2nd edn. Cambride university press.

Van Gootheest, G., de Groot, M.C.H., van der Laan, D.J., Smit, J.H., Bakker, B.F.M. (2015). Record linkage for health studies: three demonstration projects. *Statistics Netherlands*. http://www.biolink-nl.eu/public/2015_recordlinkageforhealthstudies.pdf.

Varkevisser, M., van der Geest, S.A., Shut, F.T. (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.

Victor, A., Delnoij, D.M.J., DFriele, E., Rademakers, J.J.D.J.M. (2012). Determinants of patient choice of healthcare providers: a scoping review. *BMC Health Services Research*, 12:272.

Vingilis E, McLeod AI, Seeley J, Mann RE, Beirness D, Compton CP. (2005). Road safety impact of extended drinking hours in Ontario. *Accident Analysis and Prevention*, 37:549-556.

von Wachter T, Song J, Manchester J. (2009). Long-Term Earnings Losses due to Mass Layoffs During the 1982 Recession: An Analysis Using U.S. Administrative Data from 1974 to 2004. Mimeo, Columbia University.

Wahlstedt E and Ekman B. (2016) Patient choice, Internet based information sources, and perceptions of health care: Evidence from Sweden using survey data from 2010 and 2013. *BMC Health Services Research*, 16:325.

Webb G, shakeshaft A, Sanson-Fisher R, Havard A. (2009). A systematic review of work-place intervention for alcohol-related problems. *Addiction*, 104:365-377.

Wicki, M. and Gmel, G. (2011). Hospital admission rates for alcoholic intoxication after policy changes in the canton of Geneva, Switzerland. *Drug and Alcohol Dependence*, 118:209-215.

Winglee M, Valliant R, Scheuren F. (2005). A case study in record linkage. *Survey Methodology*, 31(1):3-11.

Woollard M. (2014). Administrative data: problems and benefits. A perspective from the United Kingdom. In: Dus aA., Nelle D, Stock G, Wagner G. (Eds.), Facing the Future: European Research Infrastructures for the Humanities and Social Sciences. SCIVERO, Berlin.

Wooldridge, J.M. (2011). Econometric analysis of cross section and panel data. Second Edition. The MIT Press.

World Health Organization (2014). Global Status report on Alcohol and Health 2014. World Health Organization, Geneva.

Zuckerman M, Porac J, Lathin D, Smith R, Deci EL. (1978). On the importance of self-determination for intrinsically motivated behavior. Personality and Social Psychology Bulletin, 4(3): 443–46.