Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN METODOLOGIA STATISTICA PER LA RICERCA SCIENTIFICA

Ciclo XXVI

Settore Concorsuale di afferenza: 13/D2
Settore Scientifico disciplinare: SECS-S/03

MATERNAL SMOKING IMPACT ON THE DELIVERY COST:

a population-based study in the Emilia-Romagna region

Presentata da: Violeta Balinskaitė

Coordinatore Dottorato:
Prof. Angela Montanari

Relatore: Prof. Furio Camillo
Co-Relatore: Prof. Michel Mouchart

Esame finale anno 2014

Contents

List of Figures

List of Tables

Introduction

This doctoral thesis is devoted to the study of the causal effects of the maternal smoking on the delivery cost.
Causality is still a challenging topic and problematic as a concept leading to different opinion of its foundations.
A look back over the centuries and even in more recent times, we see philosophers and scholars debating causal relations. Aristotle distinguished four "causes" of a thing in his *Physics*: the *maternal* cause (that *out of which* the thing is made), the *formal* cause (that *into which* the thing is made), the *efficient* cause (that *which makes* the thing), and the *final* cause (that *for which* the thing is made). He emphasized the *causes* of a thing rather than the effects of causes. Hume (1740,1748) underlined that causation is a relation between experiences rather than one between facts. He argued that it is not empirically verifiable that the cause produces the effect, but only that the experienced event called the cause is invariably followed by the experienced event called the effect. John Stuart Mill's (1843) thinking, driven by an experimental model, was close to the idea that the effect of a cause is always relative to another cause, unlike Hume. The philosophical debate on causality took a significant turn fifty years ago, when *probabilistic* accounts were advanced by I.J.Good and P. Suppes. They both attempt to carry out the construction of causal relations on the basis of probabilistic relations among event, without resorting to the physical processes connecting the cause to the effect, employed statistical relevance on the basic concept and assumed temporal precedence of causes (Russo (2009)).
Rather than trying to formulate views on some underlying philosophical issues, statisticians are often faced with the concrete problem of finding empirical support in favor, against, or trying to prove or disprove, a causal claim, made in some substantive scientific or non-scientific context. The dominating paradigm was that 'statistics can only tell us about association and not causation', and for most of the 20th century, causality remained an ill-defined concept, and empirical researchers who wanted to draw causal conclusions from data had to resort to informal reasoning and justification. During the 1960s an early development of a theory of causal inference in observational studies, by Bradford Hill (1965), emerged. He proposed a

set of guidelines to strengthen the case for a causal interpretation of the results of a given observational study. In the late 1970s, a formal theory of causal inferences was founded by Rubin, based on *potential outcomes*.
During the 1990s, researchers gathered from such disciplines as statistics, philosophy, economics, social sciences, machine learning, and artificial intelligence, proposed a more aggressive approach to causality. They transformed cause-effect relationships into objects that can be manipulated mathematically (Pearl (2000)).
In the USA, the economic consequences of smoking in pregnancy have been studied fairly extensively while very little is known about cost in the European Union context. Most of the works available shows, the objects of inference are "smoking attributable fractions", "conduct attributable fractions", "relative expenditure risks", etc., all of which are based on a comparison of specific health-care expenditures (or disease rates) for a particular type of smoker with a non smoker.
An earlier studies estimated expenditures on neonatal care based on relationships between maternal smoking and low birth weight (Manning et al. (1989); Oster et al. (1988); Lightwood et al. (1999); Office of Technology Assessment (1988)). Oster et al in 1988, estimated that the mean cost of neonatal care was $288 higher for infants born to smokers than non-smokers. In the same year, the US Office of Technology Assessment estimated the long-term effects on childhood health and educational costs. Results estimated smoking-attributable neonatal expenditure of $366 million in the United States in 1996, or $704 per maternal smoker (at 1996 prices),indicating wide variations in expenditures attributable to smoking amongst US states. Manning and colleagues used parameter estimates from published sources to estimate the impact of maternal smoking during pregnancy on the risk for low birth weight infants. The results showed an estimated $652 million in additional annual costs were incurred for neonatal care of infant born low birth weight because of maternal smoking. Miller et al. (2001) estimated the costs attributable to smoking during pregnancy for mothers and infants in a US-based study. The model estimated smoking-attributable costs for eleven different infant and maternal outcomes. The authors concluded that maternal smoking during pregnancy resulted in higher health care costs both for the treatment of maternal and infant smoking related disease. Also in the United States, Adams et al. (2002) used Pregnancy Risk Assessment Monitoring System (PRAMS) data on smoking behaviour, birth outcomes and resource utilization to estimate neonatal costs attributable to maternal smoking during pregnancy. The results showed that amongst mothers who smoke, smoking added is excess of $700 in neonatal costs. A greater burden was estimated by Aligne and Stodddard (1997) using relative risk estimates to calculate direct medical expenditures and costs for loss of life. The authors concluded that the overall cost of

medical expenditures was $4.6 billion and loss of life costs of $8.2 billion.
The economic costs of maternal smoking have received little attention in the EU, as we are aware of only two studies conducted in the UK. Petrou et al. (2002) looked at the longer term economic impacts using linked birth and death data. The study population was compromised of all infants born to women who both lived and delivered in Oxforshire or West Berkshire during the period January 1980-December 1989. The cost of each hospital admission, including the initial birth admission, was estimated by multiplying the length of stay by the unit cost of the respective specialty. The findings showed that infants born to women who reported smoking during pregnancy were hospitalized for a significantly greater number of days than infants born to women who had either never smoked or had smoked in the past. Over the first 5 years of life, the adjusted mean cost difference was estimated at £462 when infants born to women who smoked at least 20 cigarettes per day were compared to infants of non-smoking mothers, and £307 when infants born to women who smoked 10-19 cigarettes per day were compared to infants of non-smoking mothers. The second study, made by Godfrey et al. (2010), focused on attribution of cases to smoking in which the authors calculated attributable risks and estimated the economic costs of smoking in pregnancy for maternal (increased risk of spontaneous abortion, ectopic pregnancy and etc.) and infant outcomes (increased risk of preterm delivery, low birth weight and etc.) during pregnancy and in the year following birth. The total annual cost of smoking during pregnancy was estimated to be approximately £8.1 million for maternal and £23.5 million for infant outcomes.

The aim of the study is to identify the causal relation between different maternal smoking status and the delivery cost in the Emilia-Romagna region.
The main questions to be answered in this thesis are:

Research question [1]: *What are the overall causal effects on delivery expenditures of the maternal smoking?*

Research question [2]: *What are the differences of health care expenditures between women that stopped smoking before pregnancy, and those that stopped smoking at the beginning of pregnancy or continued to smoke during pregnancy?*

Discussion: *Can data driven approach uncover causal relationship or should we model causal mechanism to model a given phenomenon?*

Outline of the study

The thesis is composed of four chapters.

Chapter 1 gives an overview of the Italian health care system and presents the data set used in analysis. In this chapter we describe the national and regional, with a deeper look into Emilia-Romagna region, health care system. The last section of the first chapter describes the data which arises from a complex work of data manipulation of different type of data: administrative data based on survey (CedAP) and purely administrative data (SDO).

Chapter 2 and *Chapter 3* are dedicated to answer the **Question [1]** and **Question [2]**. More particulary, the second chapter is devoted to the application of the geometric multidimensional method, which is not based on underlying theories or assumptions about the selection process, but instead uses the existing variability within the data and lets the data speak. While the third chapter is devoted to an estimation of structural model. It describes a step by step construction of a conceptual model of the cost and then the conceptual model is estimated with the use of the administrative data set described in chapter 1.

Chapter 4 presents the discussion if purely statistical approach (also called associational or descriptive models, explanatory data analysis or data mining) can undercover causal relations or we need to model structures in order to analyze complex networks of causal relations.

Acknowledgments

The work of this thesis is not a result of individual work, many people have contributed to its realization. I wish to thank them in the next few lines.
Firstly, I would like to thank to my supervisor professor Furio Camillo for his constant attention, care and belief. I appreciate all his contributions of time, ideas and funding to make my Ph.D. experience productive and stimulating.
Then I would like to thank to my co-supervisor professor Michel Mouchart for his support during my stay in the Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA) at the Université catholique de Louvain (Louvain-la-Neuve, Belgium) and for encouraging my research. Your advices have been invaluable.
I would also like to thank to professor Daniela Cocchi for her attention and support, specially during the first year of my studies .
Further more, special thanks to my committee members, professor Giampiero Gallo, professor Eduardo Rossi and professor Margherita Carlucci for letting my defence be an enjoyable moment, and for your brilliant comments and suggestions.
Many thanks to all my colleagues, with whom I have shared my Ph.D. experience: Arianna, Lucia, Irene, Federico, Oltion, BaBak. Special thanks to Lucia and Irene for their care and support during my stay in Italy.
I am also grateful to all my friends from all around the world. I am grateful for their friendship, support and love.
And finally, a special thanks to my family. Words can not express how grateful I am to my mother for all of the sacrifices that you have made on my behalf.

Chapter 1

The Italian Health Care System and Data Description

1.1 The National Italian Health Care System

The National Health Service (*Servizio Sanitario Nazionale*, SSN) in Italy was established in 1978 to replace the earlier system of social health insurance and was based on the principles of universalism, equity and solidarity with two fundamental aims:

- to provide quality assistance with appropriate, timely and adequate services to guarantee health protection , care and recovery, while respecting citizens' needs;
- to promote health by contrasting environmental, social and work conditions that hinder it, and by encouraging the diffusion of health-respecting behaviours and lifestyles.

The reform defined an integrated, centralized system in which a few specific administrative responsibilities were allocated to the regional and local levels. The central and regional governments had clashed since 1978 about financing and jurisdiction. Following a process of informal expansion of regional power in the early Nineties the process of health care decentralization started and regional governments have been granted broad discretion in planning and organizing health care services in their own territory. The individual regions have thus been able to choose among various organization models, differing from each other in a variety of aspects: the size of the local health care authorities, the level of integration between local authorities and autonomous hospital facilities, the involvement of private providers (Jommi et al. (2001); Fiorentini et al. (2008)).
In 2000 the so-called 'fiscal federalism' was represented by Legislative Decree No.

56 of February 18^{th} (2000) which established that financing of the regional health care systems would no longer depend, as in the past, exclusively on transfers from the central government. the regions can now rely on a blend of their own resources and central government transfers. The regions' revenues consists of a regional tax on productive activities and a regional surtax on the national personal income tax. The process of regionalization was further strengthened by Law No. 42/2009 delivered by the Italian parliament in 2009, which provided regions with significant autonomy in organizing health care services, allocating financial resources to their local health authorities, and in monitoring and in assessing performance (Formez (2007); Antonini and Pin (2009)). The central government retains overall responsibility for ensuring that services, care, and assistance are equitable distributed to citizens across the country.
Now the Italian system is organized into three levels: national, regional and local. The state has exclusive power to set the 'essential levels of care' (*livelli essenzali di assistenza (LEAs)*), introduced in November 2001, which provides and guarantees to all citizens, free-of-charge or with only shared cost through resources collected by the general system of taxation. LEAs were revised in April 2008 and they now amount to more than 5,700 rehabilitation, treatment and health care procedures. LEAs are structured in three main areas:

- public health - community prevention in work and life environments, food safety, injury prevention, etc. are also listed;
- primary healthcare - general practitioner, pharmaceutical care, specialist medicine to out-patient diagnostics, prostheses supply to disable people, domiciliary services provision to elderly and seriously ill citizens, territorial counselling services, semi-residential and residential facilities;
- hospital care - given through emergency care, ordinary hospitalization, day hospital, day surgery, long-term care and rehabilitation facilities, etc.

The main central institution is the Ministry of Health and it is responsible for five different functions:

- health care planning
- health care financing
- framework regulation
- monitoring
- general governance of the National Institutes for Scientific Research (IRCCS).

Figure 1.1: Overview of the Italian health care system

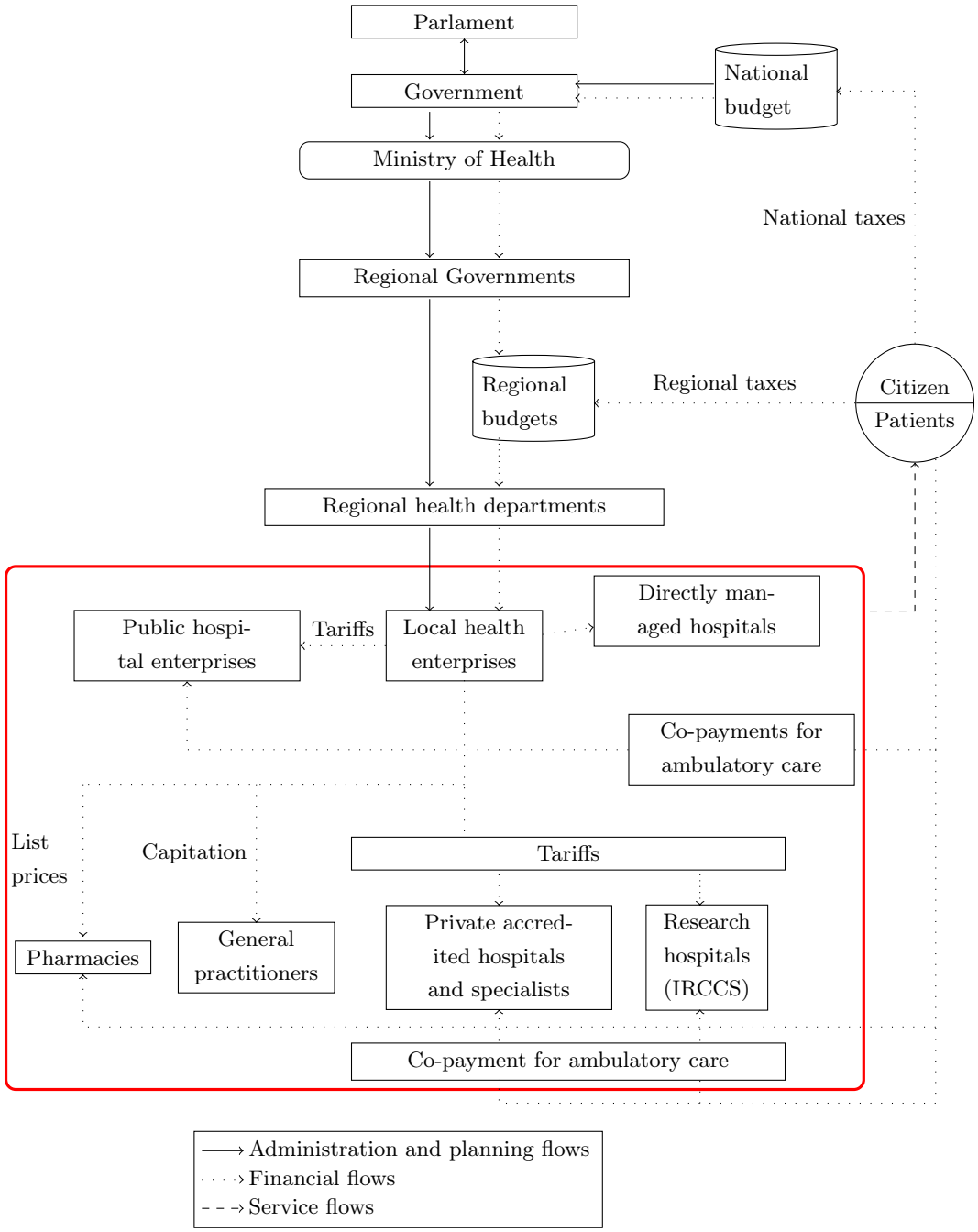


Figure 1.1 summarizes the main organizational actors, as well as the relationships between them.

1.2 The Regional Health Care System

From the early '90s the regional level has legislative and executive functions, technical support, as well as evaluation functions. The legislative functions are shared between the elected Regional Council and the regional government. According to Legislative Decree No. 229/1999, regional legislation should define:

- the principles for organizing health care providers and for providing health care services;
- the criteria for financing all health care organizations (public and private) providing services financed by the regional health departments;
- the technical and management guidelines for providing services in the regional health departments, including assessing the need for building new hospitals, accreditation schemes and accounting systems.

Regional governments, mainly through their departments of health, outline a three-year regional health plan. Regional governments use this plan, based both on the National Health Plan and on assessed regional health care needs, to establish strategic objectives and initiatives, together with financial and organizational criteria for managing health care organizations. Regional health departments are also responsible for:

- allocating resources to various local health enterprise (ASL, *Azienda sanitaria locale*) and public hospitals (AOs, *Azienda ospedaliera*);
- applying national framework rules to define the criteria for authorizing and accrediting public and private health care settings in the region;
- technically coordinating health care activities through a Standing Conference for Regional Health and Social Care Planning;
- monitoring the efficiency, effectiveness and appropriateness of the services provided by accredited public and private organizations;
- defining the geographical boundaries of health districts inside each ASL;
- appointing the general managers of ASLs and AOs;

- defining a regulatory framework governing how the general directors of hospitals and ASLs exercise autonomy in the strategic planning process.

The regional health departments in some regions provide technical support directly to the ASLs and to public and private hospitals. Other regions have formed a regional agency for health responsible for assessing the quality of local health care and providing technical and scientific support to the regional health departments and to the ASLs. The regional agencies also provide technical support to the regional health departments during the planning process to assess population needs,to define the range of services to be supplied to address these needs and to assess the quality of services provided by providers in region.
Public and private health care providers (whether they provide in-patient and/or out-patient care) are remunerated through a fee-for-service system based on two Formulary Lists, both based on the ICD9-CM WHO Classification of Deceases and Procedures:

- the NTPA (*Nomenclature Tariffario delle Prestazioni Ambulatotiali*), i.e. an out-patient formulary list of medical acts and procedures which can be delivered in out-patient facilities under SSN' finding (Ministerio della Sanita (1992)), and

- the NTPO (*Nomenclature Tariffario delle Prestazioni Ospidaliere*) which is a DRG (diagnosis-related group) based in-patient formulary list (Ministerio della Sanita (1997)), covering all hospital activity from acute or day-hospital admissions to long-term and nursing home assistance.

By this system, providers are funded through pre-determined tariffs which are established by the regions or, in the absence of specific regional legislation, by tariffs defined at the central level by the Ministry of Health.
The Emilia-Romagna region issued a law (no. 29/2004,"General norms on the organization and activities of the Regional Health Service") in which the role of the region in the Regional Health Service (SSR) is reinforced, the collaboration with communities and municipalities is strengthened; a larger contribution of health professionals in clinical governance is promoted; education, research and organization and technological innovation are considered important functions of the health system.
The Emilia-Romagnia Government rules the Regional Health Service with the Department for Health Policies which is the seat for planning, qualification and address of resources and activities. Since 2005, it also deals with planning and managing policies for non self-sufficiency and with coordinating social-and-health services. It

is supported by some regional Commissions and Committees on planning, coordination and control of specific activities, and by the Regional Agency for Health and Social Care, which has a function of technical-scientific support for the health system.
Health Trusts represent the local articulation of the SSR, that can count on 11 Local Health Trusts (AUSL), 4 University Hospital Trusts (AOU), 1 Hospital Trust (AO) and 1 Research Hospital (IRCCS - Istituti Ortopedici Rizzoli in Bologna). AUSL are geographically organized in Health Districts to guarantee access to first level health and social services and assistance, and are structured in Departments that offer services; at territorial level, there are the Primary Care Department, the Department of Mental Care and the Department of Public Health. the small and medium size hospitals belong to the Local Health Trust.
The Regional Health Service can also count on accredited (authorized by the public service) for profit or no profit private hospital, residential and outpatient structures, where citizens can refer for free on the basis of specific agreement with the RHS (MIGHRER (2012)).

1.3 Data

The collection of proprietary information of the Health Information System (SIS) of the Italian Ministry of Health began in 1984. The proprietary information represents a unique source of information, which takes into account the reorganization processes of the National Health Service that have taken place over the years. In 2001 the Permanent Conference for the Relations between the State, the Regions and the Self-governing Provinces of Trento and Balzano redefined the structural characteristics and the objectives of the New Health Information System - NSIS, which is based on the cooperation and integration of the various information systems managed independently by the individual regional or local authorities, which represent the various governing levels of National Health Service, and on the sharing of information.
The final dataset arises from a complex work of data manipulation of different types of data: administrative data based on survey (CedAP) and purely administrative data (SDO).
CedAP-Birth Assistance Certificate ("Il Certificato di Assistenza al Parto") provides health, epidemiological and socio-demographic information through the survey on births. The questionnaire is filled, not later than 10 days after birth, by a midwife or a doctor who attended the birth or a physician responsible for operating unit where the birth took place. In the case of stillbirth and/or the presence of fetal

malformations, the specific information in the questionnaire is filled by responsible medical doctor.

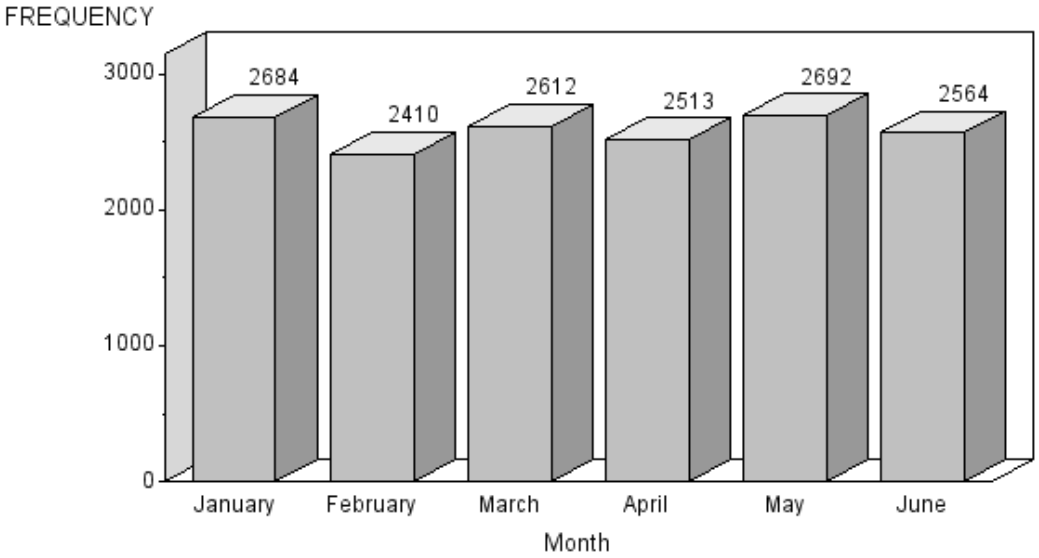
At the national level, the general criteria is dictated by Ministry of Health (decree n.349, July 16, 2001 and act n.15, 19/12/2001). All documentation can be found here : www.saluter.it/siseps/sanita/cedap/documentazione.

SDO - Hospital Admissions ("La Scheda di Dimissione Ospedaliera") consist information on hospital admissions recorded by hospitals and nursing homes through discharge papers (decree of Ministry of Health, December 28, 1991). The information describes clinical (relevant symptoms and diagnosis, surgery, diagnostic and therapeutic procedures, implants, methods of discharge) and organizational (for example: an unit admission and discharge, internal transfers) aspects of hospitals, as well payments based on diagnosis-related groups. All documentation concerning SDO can be found here:

www.saluter.it/siseps/sanita/sdo/documentazione.

Data - from each data set SDO and CedAP- are linkable via anonymized unique patient identifiers. A population of interest is newborns born between January and June in 2010 and the final data set includes 15,475 observations (Figure 1.2). Table

Figure 1.2: Number of Newborns in Emilia-Romagna, January-June (2010)



1 in Appendix A gives the distribution of each indicator (frequencies).

In the last 10 years the number of immigrants in Emilia-Romagna increased significantly, from 3.3% in 2000 to 11.3% of the resident population in 2010. Growth was much greater that at a national level, where new resident citizens accounted for 7% of the population (*latest figure available as of* 31^{st} *December 2009*). In our data set 28% of women declared non Italian nationality (23.71% of male, Table 1). Since

2004, the average age of women giving birth in Europe has risen, but the proportions of mothers of 35 years and older vary between countries from 10.9% in Romania to 34.7% in Italy (Figure 1). According to ISTAT the mean age at the childbearing over the last 15 years increased by 1.2 in Emilia-Romagna (1.5 in Italy). Figure 2 presents a distribution of mothers by nationality and age group: only 25% of Italians are younger than 30 years (EU citizen 47.76%, citizen of LMLIC 55.05% and other-63.29%).

In the context of maternal and perinatal health education level has many advantages. According to Eurostat (Figure 5) every Italian region (for which data are available) reported 20% or less of their resident population aged 25-64 had attained a tertiary level education and are at the bottom end of the ranking. Analysis of the Emilia-Romagna data set shows that over 20% of women and 16% of men (for 9% of men information is not available) have a tertiary level education. The Italian women have the highest university degree rate, at 29.67%, followed by other EU citizen with 18.27%, LMLIC citizen with 11.39%, and Other with 9.19%, as Figure 3 shows.

Over the past 2 decades, smoking among pregnant women has declined by about 60-75% in developed countries (Cnattingius (2006)), but still in many European countries, more than 10% of women smoke during their pregnancy. The annual report on smoking "Smoke in Italy 2009" (survey DOXA) indicated that about 25% of Italian population were smokers (28.9% men, 22.3% women), which corresponds, on average, that every fourth person smokes.

The smoking status of mother was collected in two stages (using questionnaire, CedAP):

Did you smoke (tobacco) in the last 5 years prior to pregnancy:

- *YES*
- *NO*

if YES, specify whether:

- *stopped smoking before pregnancy;*
- *stopped smoking at the beginning of pregnancy;*
- *continued to smoke during pregnancy.*

Data shows that 6.64% of women smoked during the pregnancy, 4.85% stopped smoking at the beginning of pregnancy, and 3.93% of women stopped smoking prior to pregnancy (Table 1). Majority of women who indicated one of latter smoking category have high school (47.45%) or secondary (30.73%) education level.(Figure

4).
Another interesting point is the high overall ceasarean and pregnancy ultrasound rates. In 2010, the highest ceasarean rate in EU was in Cyprus (52.2%), followed by Italy (38.0%, Figure 6). Almost 29% of deliveries was made using ceasarean section in Emilia-Romagna (Table 1). Pregnancy ultrasound are a regular part of prenatal medical care and most healthy women need very few scans. Ultrasound can be used in late pregnancy to assess the baby's condition when there are complications, but carrying out scans on all women is controversial. For more than 60% of women the number of pregnancy ultrasound reaches four or more scans per pregnancy.
The variable which indicates the cost, consists of the cost of delivery and the cost of hospitalization during the first six months after the birth. The cost is calculated by the administration office of Emilia-Romagna according to the regional and national normative documents (act no. GPG/2011/119).

Smoking status	**Average cost**	**Std Dev**
Stopped smoking before pregnancy	2435.83	2996.72
Stopped smoking at the beginning of pregnancy	2208.29	2326.01
Continued to smoke during pregnancy	2647.68	3560.30
Did not smoke	2610.59	3774.58

Table 1.1: Average Cost by smoking status

Table 1.1 represents average (naive) cost and standard deviation by smoking status. It shows that the average delivery cost of mother who smoked during pregnancy is slightly higher than the one of non-smoker.

1.4 Summary

This chapter has presented a structure of Italian National and Regional Health Care System. The Italian National Health System (NHS), established in 1978, follows a model similar to the Beveridge model developed by the British NHS. Like the British NHS, healthcare coverage for the Italian population is provided and financed by the government through taxes. Since the early 1990s cause of strong decentralization, power shifted from the state to the Italian regions. Consequently, the state now retains limited supervisory control and continues to have overall responsibility for the NHS in order to ensure uniform and essential levels of health services across the country.
In the last section of this chapter we presented the data used later in the thesis. The

collection of proprietary information of the Health Information System (SIS) of the Italian Ministry of Health began in 1984. The proprietary information represents a unique source of information, which takes into account the reorganization processes of the National Health Service that have taken place over the years. In 2001 the Permanent Conference for the Relations between the State, the Regions and the Self-governing Provinces of Trento and Balzano redefined the structural characteristics and the objectives of the New Health Information System - NSIS, which is based on the cooperation and integration of the various information systems managed independently by the individual regional or local authorities, which represent the various governing levels of National Health Service, and on the sharing of information.
Our final dataset arises from a complex work of data manipulation of different types of data: administrative data based on survey (CedAP) and purely administrative data (SDO).

Chapter 2

Counterfactual Approach

2.1 Introduction

In the past decades, a counterfactual model of causality, which is also known as the potential outcome model, has been developed and widely used in statistics and economics, and with increasing frequency it is used in sociology, psychology, and political science. Its origins trace back to early works on experimental design of Neyman (1935) and Fisher (1935). The causal analysis of observational data was introduced and formalized in a series of papers by Rubin (1973a, 1974, 1977).

The main idea of the counterfactual approach for observational data analysis is simple. Suppose that each unit in a population of interest can be exposed to two or more alternative states of a cause and each state is characterized by a distinct set of conditions, exposure to which potentially affect an outcome of interest. In the counterfactual framework, each unit has a potential outcome under each treatment state, even though each unit can be observed in only one treatment state at any point in time.

Lets denote potential outcome Y and two-stage treatment state T. The observable outcome variable Y_{obs} can be therefore defined as:

$$Y_{obs} = Y(1) \quad if \quad T = 1,$$

$$Y_{obs} = Y(0) \quad if \quad T = 0.$$

As shown in table below, the fundamental problem of causal inference is one of missing data.

Consider that one can never observe the potential outcome under both treatment and control states, it implies impossibility to calculate individual-level causal effects. Thereby, as a consequence, researchers focus their attention on the estimation of

Group	Y(1)	Y(0)
Treatment group	Observable	Counterfactual
Control group	Counterfactual	Observable

Table 2.1: The fundamental problem of causal inference

aggregated causal effects, usually average causal effects:

$$E[\delta_i] = E[Y_i(1)] - E[Y_i(0)], \tag{2.1}$$

where $E[.]$ denotes the expectation operator. This changing of interest from individual level to average level was introduced by Holland (1986).

Outline of the chapter

This chapter is composed of three main sections. In the next section, a brief overview of conventional methods is given with profounder attentions on *The Potential Outcome Approach* (section 2.2). The section 2.3 is dedicated to the description of geometric multidimensional approach. In the last section (section 2.4), application of multivariate method to data set is presented.

2.2 Conventional Methods

2.2.1 Introduction

There is three basic strategies for estimating causal effects: first,conditioning on variables that block all back-door paths from the causal variable to the outcome variable; second, using exogenous variation in an appropriate instrumental variable to isolate covariation in the causal and outcome variable, and third, establishing an isolated and exhaustive mechanism that relates the causal variable to the outcome variable and then calculate the causal effect as it propagates through the mechanism. According to Rosenbaum (2002), *an observational study is an empiric investigation of treatments, policies, or exposures and the effects they cause, but it differs from an experiment in that the investigator cannot control the assignment of treatments to subjects*, and from here arises the problem of selection bias.
In this chapter the attention is focused on two basic conventional methods for estimating causal effects, when randomized experiments are impossible. In the next section we focus on the potential outcome approach as it represents a starting point of method introduced in section 2.3.

2.2.2 The Potential Outcome Approach

The intuitive and transparent definition of causal effects via potential outcomes is known as Rubin's Causal Model (Holland (1986)). The purpose of this model is to construct a model that is complex enough and that allows to formalize basic intuitions concerning cause and effect.

In the simplest case, the logical elements of Rubin's model form a quadruple (U, T, D, Y) where U is a population of units, T is a set of causes or treatments to which each one of the units in U may be exposed, $D(u) = d$ if d is the cause in T to which unit u is actually exposed, and $Y(u, d)$ is the value of the response that would be observed if unit $u \in U$ were exposed to cause $d \in T$.

For simplicity, let assume that there are just two causes or levels of treatment, denoted by t_1 (the treatment) and t_2 (the control). The role of time is important because of the fact that when a unit is exposed to a cause this must occur at some specific time or within specific time period. Variables now divided into two classes: pre-treatment - whose values are determined prior to exposure to the cause; post-treatment - whose values are determined after exposure to the cause. The response variable, Y, must fall into the post-treatment class. Cause the values of post-treatment variables are potentially affected by the particular cause (t_1 or t_2), to represent the notion of causation, we need not a *single*, but *two* potential responses (Y_{t_1} and Y_{t_2}). The interpretation of $Y_{t_1}(u)$ and $Y_{t_2}(u)$ for a given unit u, is that $Y_{t_1}(u)$ is the value of the response that would be observed if the unit were exposed to t_1 and $Y_{t_2}(u)$ is the value that would be observed on *the same unit* if it were exposed to t_2.

In Rubin's model, causes are taken as undefined elements of the theory, and effects are defined in terms of the elements of the model. The effect of the cause t_1 on u measured by Y and relative to cause t_2 is defined as:

$$Y_{t_1}(u) - Y_{t_2}(u).$$

As the impossibility of observing both $Y_{t_1}(u)$ and $Y_{t_2}(u)$ exits, the statistical solution is used - to calculate the *average causal effect* (ATE). The ATE of t_1 (relative to t_2) over U is the expected value of the difference $Y_{t_1}(u) - Y_{t_2}(u)$ over the u's in U; that is,

$$ATE(Y) = E(Y_{t_1}(u) - Y_{t_2}(u)) = E(Y_{t_1}(u)) - E(Y_{t_2}(u)).$$

The right side of the last equation reveals that information on *different* units that *can be observed* can be used to gain knowledge about ATE. The important point here is that the statistical solution replaces the impossible-to-observe causal effects of t_1 on a specific unit with the possible-to-estimate average causal effect of t_1 over a population of units.

In observational studies the active experimenter is replaced by a passive observer who cannot arrange the values of $D(u)$ to achieve independence. The Rubin's model still can be applied, but now D is not automatically independent of Y. In such studies usually we have a covariate, X, and in this case we may check the distribution of X in each exposure group by comparing the values of

$$P(X = x|D = d)$$

across the values of $d \in T$. If there is evidence that $P(X = x|D = d)$ depends on d, then we may not believe that the independence assumption holds in an observational study. However, we might use a weaker conditional independence assumption, called *strong ignorability* (Rosenbaum and Rubin (1983a)). It is the basis for all covariate-adjusted causal effects in observational studies and is based on the conditional expectations or regression functions:

$$\begin{aligned} & E\{E(Y_{t_1}|S = t_1, X) - E(Y_{t_2}|S = t_2, X)\} \\ =& E\{E(Y_{t_1}|X) - E(Y_{t_2}|X)\} \\ =& E(Y_{t_1}) - E(Y_{t_2}) = ATE(Y) \end{aligned}$$

Rubin's model was really developed to address the problem of causal inference in observational studies and these types of studies can be found in Rubin (1974, 1977), Holland and Rubin (1983), Rosenbaum and Rubin (1983a,b, 1984b, 1985a,b), and Rosenbaum (1984a,b,c, 1987).

2.2.3 Propensity Score Methodology

In an observational study comparing two treatments, the use of propensity score methodology is very common (Rubin (2001); Rubin and R.P.Waterman (2006), Normand et al. (2001); Lee (2006); Morgan and Harding (2006); Peck (2007); Austin (2008); Wyse et al. (2008); Ye and Kaskutas (2009)).
Propensity score methods were proposed by Rosenbaum and Rubin (1983a) as central tools to help assess the causal effects of interventions. With no missing data in the covariates X_i, the propensity score e_i is defined as the probability that the ith unit is treated given that its vector of covariates is X_i,

$$e_i \equiv e(X_i) \equiv Pr(T_i = 1|X_i).$$

In a randomized experiment, the propensity scores are known, whereas in an observational study, they must be estimated from the data on T_i and X_i. No outcome data are required or desired; even if available in the data set.
The mapping from X_i to e_i is generally a many-one function. The central result of

Rosenbaum and Rubin (1983a) is that if a group of treated units and control units have the same value of the propensity score, e_i, then they have the same distribution of multivariate X_i, no matter what the dimension of X_i.
Figure 2.1 presents a flow diagram for implementing the basic propensity score methodology. In the step three, probabilities may be estimated by a variety of methods, the most common probably being logistic regression. For the step six there are many available algorithms Gu and Rosenbaum (1993); Rosenbaum (1989); Rubin (1979, 1980); Rubin and Thomas (1992a,b, 1996, 2000).

Despite the broad utility of propensity score methods, it still have few limitations. It is important to keep in mind that propensity score methods can only adjust for observed confounding covariates and not for unobserved ones. Another limitation is that they work better in lager samples. With more than two treatment conditions, the propensity score usually differs for each pair of treatment groups being compared (that is, with three treatment groups labelled A, B, and C, there are three propensity scores: A compared with B, A compared with C, and B compared with C). And a final possible limitation of propensity score methods is that a covariates related to treatment assignment but not to outcome is handled the same as a covariate with the same relation to treatment assignment but strongly related to outcome.

2.2.4 The Economic Approach

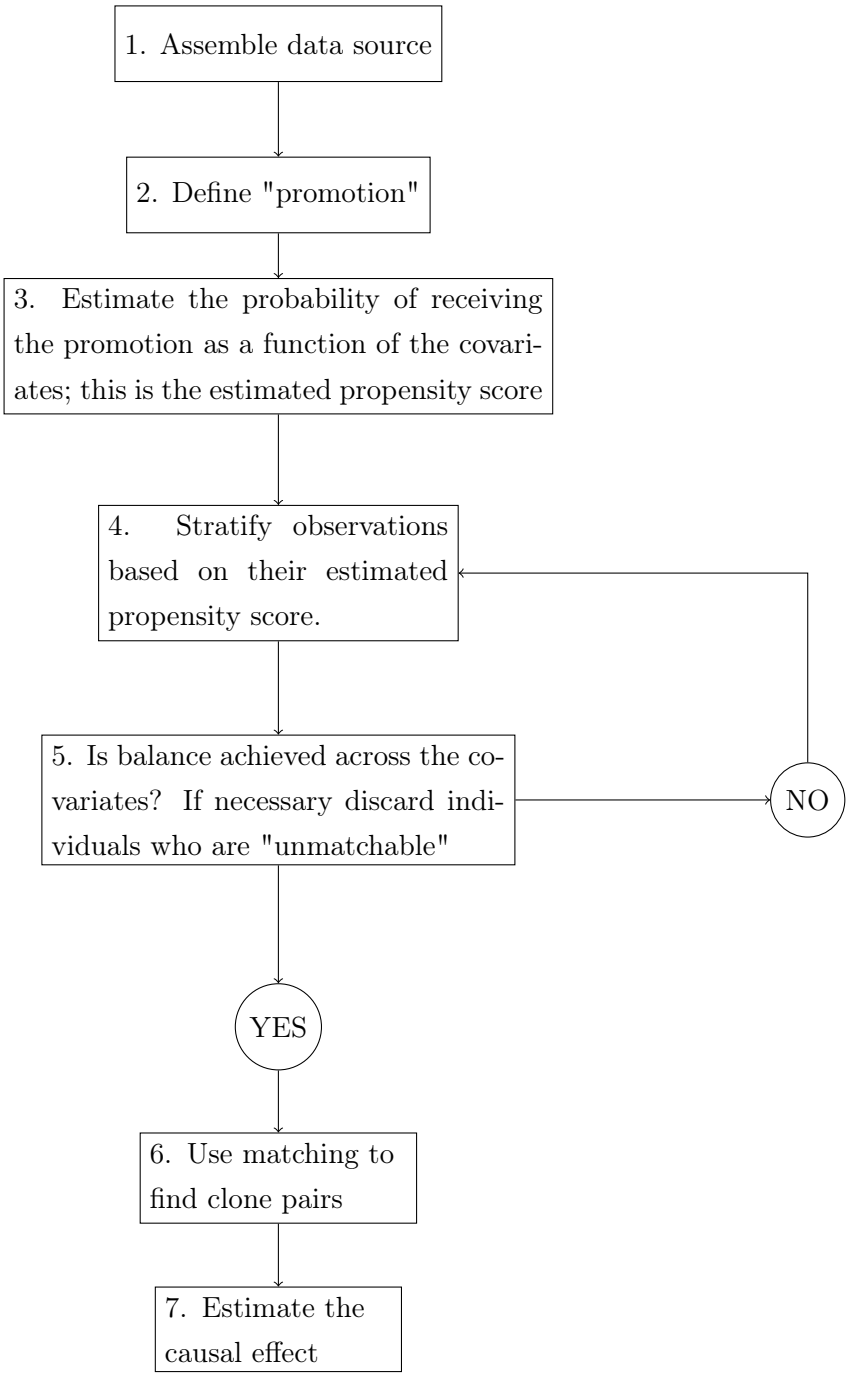
The economic approach focus its interest on the framework that motivates estimators. In particular, economists are interested on covariates involved in both outcome and participation equations. More precisely, they suggest specific functional forms of estimating equations motivated by a *priori* theory.
One of the most used model is the Heckman's selection model (Heckman (1979); Heckman and Robb (1985); Heckman and Hotz (1989)). Even if it deals with sample selection, the same approach can be used in dealing with non-random assignment to treatment as well. The selection model takes into account two equations: a selection equation and an outcome equation.
The outcome equation is represented as linear equation:

$$Y_{it} = X_{it}\beta + d_i\alpha_t + U_{it}, \;\; t > k$$

$$= X_{it}\beta + U_{it}, \;\; t \leq k$$

with U as an error term ($E[U_{it}|X_{it}] = 0$); Y_{it} as an observed outcome for unit i in period t; α_t as the impact of the treatment under evaluation, and a dummy variable d_i, which equals one if the ith unit participates in the treatment and is zero otherwise, and the convention is that treatment occurs in period k.

Figure 2.1: The roadmap



When assigment to treatment is nonrandom, selection bias in the estimation of α_t can arise because of dependence between d_i and U_{it}. The bias, in a model without regressors, is present if $E(U_{it}|d_t) \neq 0$; and in a model with regressors if $E(U_{it}|d_i, X_i) \neq 0$. In this case, an ordinary least squares regression of Y_{it} on X_{it} and d_i does not yield consistent estimates of α_t (or β) and this can arise for a variety of reasons.
The Heckman's selection model assumes that the participation decision can be described in terms of an indext funtion, NI_i, of observed (Z_i) and unobserved (V_i) variables, as well known as the selection equation:

$$IN_i = Z_i\gamma + V_i.$$

Then, the ith unit's treatment status is:

$$d_i = 1 \; iff \; IN_i > 0,$$

$$= 0 \; otherwise,$$

and it is assumed that V_i is independently and identically distributed across units. Dependence between U_{it} and D_i can arise for one of two reasons: dependence between Z_i and U_{it} (selection on the observables) or dependence between V_i and U_{it} (selection on the unobservables). Selection on observable occurs when the dependence between U_{it} and d_i is due to a set of observed variables, Z_i, which influence selection into treatment; while selection on unobservable may occur when the dependence between the treatment indicator variable and U_{it} is not eliminated even after controlling for Z_i. Then, selection is said to depend on unobservable. Such selection bias estimators are formed by invoking assumptions about the distribution of V_i, Z_i and U_{it}.
Two important features of economic models are the following: on one hand, alternative non-experimental estimation procedures should produce approximately the same program estimates, but this requirement is not always matched. On the other hand, there is no objective way to choose among alternative non-experimental estimates.
The economic models differ from propensity score methods in the sense that latter one do not require any model for outcome. But both propensity score method and economic selection models are model dependent: economists use a model for both the selection process and outcomes; whereas, propensity score methods use a model for the assignment mechanism.

2.3 Geometric Multidimensional Method

2.3.1 Introduction

In the previous section we presented some statistical techniques for solving the *counterfactual problem* in order to estimate causal effects. This chapter aims to define a geometric multidimensional method.

The strategy mainly based on the multivariate approach introduced in Camillo and D'Attoma (2010); D'Attoma (2009). In the absence of random assignment , the dependence between the available covariates involved in the selection process and the assignment-to-treatment indicator variable exits. As a solution they proposed to use a conditional analysis (Escofier (1988)), which allows to isolate the part of the variability of the X-space due to the assignment-to-treatment mechanism. Specifically, the conditioning was applied to the problem of measuring imbalance with categorical covariates where all or some of them may be linked to an external categorical variable T. Authors adopted the concept of *inertia* as a measure of association among categorical covariates, where the term *inertia* is used by analogy with the definition in applied mathematics of *moment of inertia* which stands for the integral of mass times squared distance to the centroid (Greenacre (1984)).

In brief, the method involves computing and testing the global imbalance, classifying cases in order to generate well-matched comparison groups, and then computing the treatment effect.

2.3.2 Global Imbalance Measure and Test

The Global Imbalance Measure formula is defined as:

$$GI = \frac{1}{Q} \sum_{t} \sum_{j} \frac{b_{jt}^2}{k_{.t} k_{.j}} - 1, \tag{2.2}$$

where Q denotes the number of pre-treatment covariates, t is the number of treatment levels, j presents the total number of the Q pre-treatment covariates categories, b_{tj} is the number of units with category j in the treatment group t, $k_{.t}$ is the size of group t, and $k_{.j}$ is the number of units with category j. (for more detailed notation,see D'Attoma (2009)).

In order to determine if the detected imbalance is significant with respect to the hypothetical case of a random structure on the data, the null hypothesis of no dependence between covariates X and treatment T is specified as:

$$H_0 : Inertia_{\text{within}} = Inertia_{\text{total}}.$$

If the null hypothesis has not been rejected then the observed covariates are not related to the assignment-to-treatment indicator variable, meaning that data are

balanced.

On the basis of the asymptotic distribution function of $Inertia_{\text{between}}$ expressed as in:

$$Inertia_{\text{between}} \sim \frac{\chi^2_{(T-1)(J-1),\alpha}}{nQ},$$

the interval of plausible values for GI is defined as:

$$GI \in \left(0, \frac{\chi^2_{(T-1)(J-1),\alpha}}{nQ}\right). \tag{2.3}$$

If the measured GI is outside the interval, then the null hypothesis of no dependence among X and T is rejected and data are unbalanced.

The main advantage of the GI measure is its simplicity of interpretation. The proposed measure varies in $[0, I_t]$. Perfect balance occurs when $Inertia_{\text{between}} = 0$; whereas, perfect imbalance occurs when $Inertia_{\text{within}} = 0$ and $Inertia_{\text{between}} = Inertia_{\text{total}}$ which indicates that the observed total variability of the X-space is completely due to the influence of conditioning (T).

2.3.3 Multiple Correspondence Analysis and Cluster Analysis

When data are not balanced, the estimation of an unbiased treatment effect is represented by the transition from the global space to local space. In this step they adopt a *tandem approach* (Arabie and Hubert (1994)), which involves executing a cluster analysis on the basis of the low-dimensional multiple correspondence analysis (MCA) coordinates in order to identify homogeneous groups. Using MCA a data matrix can be decomposed into eigenvalues and eigenvectors, where the eigenvalues (λ_k) are the principal inertias of a Burt table. The use of MCA coordinates before clustering gives the advantage to work with continuous variables rather than categorical covariates.

The cluster analysis (CA) is used to group objects or individuals described by a number of variables or characteristics and this form of data analysis is not new on evaluation field (Henry and McMillan (1993); Peck (2005)).

This approach uses an agglomerative hierarchical clustering (AHC) which produces sequences of nested partitions of increasing heterogeneity, between partition into n clusters where each object is isolated and partition into one cluster which includes all the objects. The AHC algorithm using a well-defined proximity measure search for the closest clusters at each step and merge them. Here cluster analysis is carry out employing Ward's method which uses an analysis of variance approach to evaluate the distances between clusters. In short, it attempts to minimize the sum of squares of any two hypothetical clusters that can be formed at each step.

The sequence of partitions can be presented by a tree diagram (dendrogram) and

the appropriate number of clusters can be identified by examining the cut points in groups imagined on the dendrogram.
Once the specified cluster solution set is selected, the balance within each cluster is tested using GI measure.

2.3.4 Average Treatment Effect

In the final step of the procedure, first a local average treatment effect within balanced groups (and excluding observations in unbalanced clusters) is computed and then the Average Treatment Effect on the Treated (ATT) and Average Causal Effects (ATE) are calculated according to the following formulas:

$$ATT = \sum_{c=1}^{C} l_c \frac{\sum_{i \in I(c) T_i}}{\sum_{\forall i} T_i}, \tag{2.4}$$

and

$$ATE = \sum_{q=1}^{Q} l_q \frac{n_q}{N}, \tag{2.5}$$

where n_q is the number of units in cluster q, N is the number of units in the sample considered,T_i is the number of treated units, and l_q is the local ATE in cluster q.

2.4 Application. The impact analysis

2.4.1 Introduction

This section is dedicated to the application of the geometric multidimensional approach to our data set.
We perform four separate analysis: two analysis for question **Research Question [1]**

- consider two different potential groups (Case 1): one control (did not smoke in the last 5 years prior to pregnancy) and one treatment (smoked in the last 5 years prior to pregnancy) group,
- consider four different potential groups (Case 2): control (did not smoke in the last 5 years prior to pregnancy) and three treatment (stopped smoking before pregnancy, stopped smoking at the beginning of pregnancy,continued to smoke during pregnancy) groups,

and two for **Research Question [2]**

- consider three different potential groups (Case 3): one control (stopped smoking before pregnancy) and two treatment (stopped smoking at the beginning of pregnancy,continued to smoke during pregnancy) groups,

- consider two differnt potential groups (Case 4):one control (stopped smoking at the beginning of pregnancy) and one treatment (continued to smoke during pregnancy) groups.

Analyzing the impact of smoking on the cost, first we selected the available pre-treatment covariates and then we assumed that there is no confounding, and that all variables are causally prior to treatment assigment. As well, we assumed that bias arises only due to difference in observed covariates. We considered the following 18 pre-treatment covariates:

- maternal age
- marital status
- mother's nationality
- birth place (mother)
- educational level (mother)
- current occupation status (mother)
- profession (mother)
- sector of employment (mother)
- parental age
- father's nationatily
- educational level (father)
- current occupation status (father)
- profession (father)
- sector of employment (father)
- number of previous live births
- number of spontaneous abortions
- number of induce abortions
- number of Caesarean

The aim is to find groups of non-smokers and smokers as similar as possible on which estimate the causal effects of interest.
First we will check if the data is balanced, meaning that the empirical distribution of the covariates in the groups are more similar. Then, if balance is not detected, we will try to balance data by controlling for X by performing a cluster procedure to find local groups of balanced and comparable units. And finally, we will estimate the causal effects of interest.
Analysis in this chapter have been performed using the statistical softwares SPAD and SAS 9.3.

2.4.2 Overall balance

First we measure the level of selection bias in each case that arises from the non-random selection mechanism, and we do this by computing GI measure. Results (Table 2.2) show the presence of imbalace in data in all four defined cases. The Global Imbalance measure falls in the critical region, thereby demanding adjustment in order to estimate a treatment effect that is not biased by selection.

2.4.3 MCA and cluster analysis

The multiple correspondence analysis was carried out using all 18 pre-treatment covariates. The results of the MCA are a set of factorial coordinates that are continuous and orthogonal to one another. On the basis of these new coordinates, we perform a cluster analysis to find groups of comparable units on which estimate local causal effects. We used a hierarchical clustering method and the Ward's method as group proximity measure. The approriate number of clusters is chosen by examining the cut of dendrogram. The basic idea is that going deeper in the cut of the tree diagram, is more likely that groups are balanced in terms of pre-treatment covariates. The units belonging to non balanced groups are discarded.
The MCA and cluster analysis was carried out in SPAD and the GI measure and the multivariate test of imbalance was performed using $\%GI$ SAS macro (Camillo and D'Attoma (2012)).
Tables 2.3 and 2.4 present the examined and selected cluster solutions. For Case 1 and Case 2 we selected the 45-clusters solution set because it discards the smallest amount of units with respect to other solutions. Going more deeper than 45-clusters in the cut of the tree diagram (51-,60- or 70- clusters solution), higher number of clusters are not balanced or common support is not satisfied, which leeds to higher number of discarded units.
It is clear that, on one hand, if there are too many clusters, more observations may be discarded due to the lack of common support. On the other hand, if the chosen

Table 2.2: Balance in the Overall Data Set

Case	Treatment1	Treatment2	Treatment3	Treatment4	Control	GI	Interval	Balance
1	-	-	-	**2381**	**13094**	**0.0105**	**(0, 0.0004)**	**no**
				15.39%	84.61%			
2	**607**	**747**	**1027**	-	**13094**	**0.0137**	**(0,0.0012)**	**no**
	3.92%	4.83%	6.64%		84.61%			
3	-	**747**	**1027**	-	**607**	**0.0219**	**(0, 0.0054)**	**no**
		31.37%	43.13%		25.5%			
4	-	-	**1027**	-	**747**	**0.0139**	**(0, 0.0038)**	**no**
		57.89%			42.11%			

number of clusters is too small, more observations may be discarded due to lack of balance.

Case	Examined n-clusters solutions	Selected n-clusters solution
1	19-,24-,36-,45-,51-,60-,72-	45-clusters
2	19-,24-,36-,45-,51-,60-,72-	45-clusters
3	4-,6-,11-,12-,16-	16-clusters
4	2-,3-,4-,6-,8-	8-clusters

Table 2.3: Analyzed n-clusters solutions

Case	Discarded units	Discarded units %
1	860	5.56
2	1791	11.57
3	0	0
4	0	0

Table 2.4: Discarded units

For Case 3 and Case 4 we selected the 16- and 8-clusters solutions, respectively. In both cases, all clusters are balanced in terms of pre-treatment covariates, and no observations were discarded.

Tables 2 - 5 in Appendix B show the results of selected clusters solution, including the number of observations in treatment and control groups, GI measure and interval of plausible values, as well the balance. In Case 1, five of the clusters result in having unbalanced characteristics by GI measure. While in Case 2, in three of the clusters no common support was didected and five result in having unbalanced characteristics. All these clusters will not be included in the futher analysis.

2.4.4 ATT and ATE

Once we chose the clusters solutions we then calculated average treatment effects (local) separately within each balanced clusters (mean difference of treatment and control groups).

Tables 6 - 9 in Appendix B include local treatment effects and information if difference is statistically significant (p-value). Then we computed ATT and ATE according to the formulas 2.4 and 2.5 using only clusters where difference was significant.

	ATT	**ATE**
Case 1		
did not smoke in the last 5 years prior to pregnancy **vs** smoked in the last 5 years prior to pregnancy	-141.75	-1.1
Case 2		
did not smoke in the last 5 years prior to pregnancy **vs** stopped smoking before pregnancy	-26.3	645.32
did not smoke in the last 5 years prior to pregnancy **vs** stopped smoking at the beginning of pregnancy	-308.39	-361.47
did not smoke in the last 5 years prior to pregnancy **vs** continued to smoke during pregnancy	30.12	14.03
Case 3		
stopped smoking before pregnancy **vs** stopped smoking at the beginning of pregnancy	-59.25	-93.23
stopped smoking before pregnancy **vs** continued to smoke during pregnancy	88.38	121.01
Case 4		
stopped smoking at the beginning of pregnancy **vs** continued to smoke during pregnancy	184.63	130.45

Table 2.5: ATT and ATE

Table 2.5 represents the average treatment effects for treated and the average treatment effects for all cases. At the beginning of this section we identified four different analysis to perform to answer to **Question [1]** (Case 1, Case2) and **Question [2]** (Case 3, Case 4).

The results in the first case show no maternal smoking impact on the cost where we compare mothers who smoked in the last five years prior to pregnancy with non-smokers. However, the analysis of case 2, with more specified maternal smoking status, gives us distinct results. It shows that there is a causal effect on delivery expenditure of the mother who stopped smoking before pregnancy (ATE=645.32), and mother who continued to smoke during pregnancy (ATE=14.03). The first one shows that, on average, delivery cost for women who stopped smoking before pregnancy was 645.32 euro more expencive than for non-smoker women. Particulary, in this case, we want to focus more on the cluster 40. As shown in Table 10, cluster 40 is composed of women age 25-35 (86.81%), citezen of LMLIC (91.68%), married (93.85%) and not occupied (unemployed (3.8%), student (1.81), housewife (94.03%), n/a (0.36)), and 70% of women already have children. Furthermore, the analysis of

this case indicates that, on average, health care expenditure for women who stopped smoking at the beginning of pregnancy was 361.47€less than for non-smoker.
The results of case 2 look questionable. This can be caused by the selected pre-treatment covariates. Recall that in this case we have fourr smoking status and all of them started at different time. Moreover, if woman indicated that she stopped smoking in the last 5 years prior to pregnancy, we do not know the exact moment. For tha analysis we considered 18 pre-treatment covatiates, one of them is education. This covariate indicates the highest education level, but we do not have information when it was obtained.
The second part of Table 2.5 represents the estimated treatment effects between women who stopped smoking before pregnancy (control groups) and women who stopped smoking at the beginning of pregnancy or continued to smoke during pregnancy (treatment groups); and women who stopped smoking at the beginning of pregnancy (control group) and continued to smoke during pregnancy (treatment group). In both cases, Case 3 and Case 4, we found an evidence of effect of women who was smoking during pregnancy on the cost. Tables 11 and 12 notes the main features of covariates in the clusters where difference was statistically significant.
Table 11 shows that only in three clusters the difference was significant. Women within clusters 6 and 16 are mostly Italian, born in north-east Italy and 30+ years old. More than 70% of women in cluster 16 have tertiary education level, but there is no at all information about their partner. In contradistinction to clusters 6 and 16, cluster 12 is composed of other EU citizen, younger than 30 years old and working women. If we look at Table 12, in both clusters there is no or very little information about father. The main common maternal characteristics are nationality (Italian), birth place (north-east Italy) and occupation status (occupied). In addition, cluster 7 is composed of younger and more educated mothers.

2.4.5 Summary

We began this chapter with a profound description of the Potential Outcome Approach (known as Rubin's Causal Model), as it was a starting-point for geometric multidimentional approach. As well, we gave a brief overview of conventional methods (propensity score, Heckman's selection model).
The objective of this part of thesis was to estimate the maternal smoking impact on the delivery cost using geometric multidimentional method. This approach involves first identification whether bias due selection mechanism exist, then execution of cluster analysis on MCA coordinates, and finally the comparison of treatment and control cases within balanced clusters to estimate treatment effect.
First of all, we observed that there is no maternal smoking impact on the delivery

cost where we compare mothers who smoked in the last five years prior to pregnancy with non-smokers. However, the analysis with more specified smoking status, gave us different results. It showed that there is a causal effect on the delivery expenditure of the mother who continued to smoke during pregnancy.

Chapter 3

Structural Modelling

3.1 Introduction

A model is an abstract object which may contain statements, figures and mathematical expressions designed to obtain an increase knowledge of some aspects of reality. A structural model is a model which uncovers a structure underlying the data generating process. Structural models incorporate not only observable, or manifest, variable but also, in many instances, unobservable, or latent, variables. Thus such models may capture an underlying structure of the world. Structural models are also called 'causal models'.
The delivery cost can be seen as the outcome of a complex process that involves socio-economic, environmental and biological variables - all these variables take a part at particular stages of the process. The causal mechanism of the cost is presented within a conceptual framework. Once it is determined, estimation of the mechanisms included in the model is the next natural step which is based on the decomposition of the joint distribution of the set of variables through conditional and marginal distributions. This decomposition is based on theoretical knowledge of the delivery cost process.

Outline of the chapter

The presentation of this chapter is organized around these four steps. Next section represents the theoretical concept of structural modelling. Sections 3.3 and 3.4 present and highlight the importance of the theoretical process leading to the delivery cost. The last section presents the available variables, the operational framework and the main results.

3.2 Theoretical Framework

In this section, we will describe the theoretical concept of structural modelling introduced by Mouchart et al. (2009, 2010). In Mouchart et al. (2010), they stated that '*Structural modelling, instead, does not denote a particular (statistical) model (e.g. structural equation models, covariance models, multilevel models, etc.) but refers to a general methodological account of model-building and model-testing. In this sense, we take structural modelling to be a general methodological framework for causal analysis*'.
The main idea of this approach is to decompose the global mechanism into submechanisms through a recursive decomposition of a multivariate distribution.
The formal framework of structural modelling is presented by the *hypothetico-deductive* methodology. Hypothetico-deductivism is a view according to which the scientist first formulates a hypothesis and then tests it by seeing whether the consequences derived from the hypothesis obtain or not. Model building and model testing is performed through two stages:

1. formulating the causal hypothesis;
2. iteratively:
    - building the statistical model;
    - drawing consequences to conclude to the empirical validity or invalidity of the causal hypothesis.

In the first stage, we formulate causal hypothesis from background theories, from knowledge concerning the phenomenon at issue, and from a preliminary analysis of data. The hypothesis, which is also called the 'conceptual hypothesis', however, is not analysable a priori.
Once the conceptual hypothesis is formalized, estimation of the mechanisms included in the model is the next stage. The statistical model is based on the decomposition of the joint distribution of the set of variables through conditional and marginal distributions. The conditioning variables of each conditional component of the decomposition are exogenous variables for the corresponding sub-process.
A model to consider as structural, at least two conditions have to be fulfilled. It has to be coherent from a theoretical point of view and the parameters of the model should be stable. The stability, or invariance, condition is actually a complex issue. This is a condition of stability not of the causal variables, but of the causal relation itself. The idea is that each variable is determined by a set of other variables through a relationship that remains invariant when those other variables are subject

to external influence. This condition allows us to predict the effects of changes and interventions. Stability of distributions is also assumed to ensure that the (conditional) independencies between variables will not be jeopardized by variations in the parameters.
The whole recursive decomposition can be interpreted as characterizing a global mechanism, whereas each conditional distribution within the recursive decomposition can be interpreted as characterizing sub-mechanism within the global one. If we can identify sub-mechanisms within a global one, this means that we are able to decompose the global mechanism and thus disentangle the action of each component.

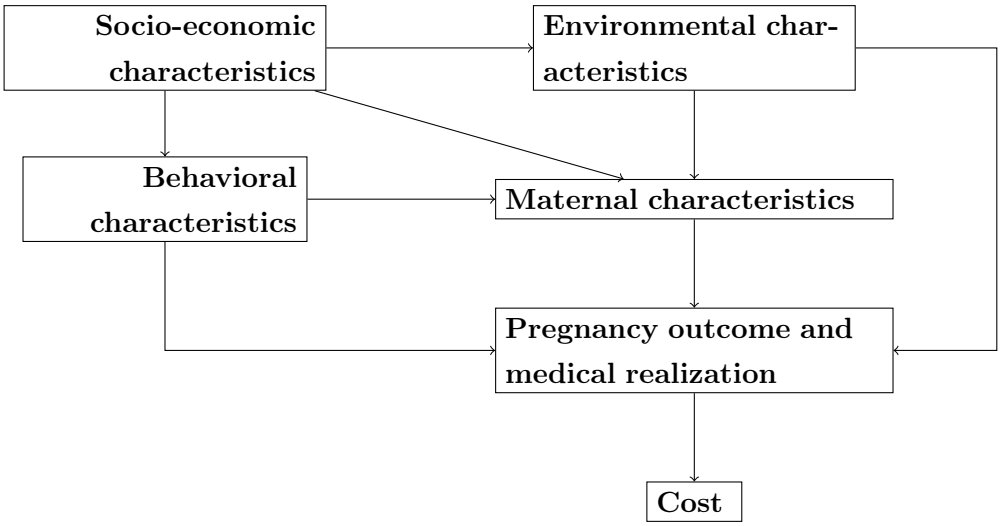
3.3 Conceptual Framework

The general idea behind the construction of the conceptual framework is to highlight the causal mechanisms leading to a particular event, delivery cost in the present case. Causal framework on the determinants of the delivery cost was not described in previous studies. Nevertheless, the vast numbers of researches have been undertaken in order to detect the influence of social and biological factors on pregnancy outcome. The socioeconomic, environmental factors, the wide range of maternal health and behaviour factors were associated with adverse pregnancy outcome (Anandalakshmy et al. (1993); W.J. Graham and S.F. Murray (1997); Hajo and Wildschut (1995); McCarthy and Maine (1992); Magadi (1999); Magadi et al. (2001); Mohamed et al. (1998); Mosley and Chen (1984); Magadi et al. (2004)).
We consider that only pregnancy outcome and medical realization have influence on the delivery cost and the set of causal pathways leading to it is represented by demographic, epidemiological and economical factors.
The *directed acyclic graph* (DAG) corresponding to the conceptual framework represented in Figure 2.1 describes all causal relations between the distinct determinants of the cost. The model includes six groups of variables: socio-economic, environmental and behavioral characteristics, maternal characteristics, pregnancy outcome and medical realization, and cost. Each directed arrow represent the causal relation between variables based on a review of the literature.

The concepts of the model are defined in the following way:

- Socio-economic characteristics **[SEC]** includes parents socioeconomic circumstances, education, partnership and parenthood histories.
- Environmental characteristics **[EC]** encompass environmental tobacco smoke, air pollutants from motor vehicles and industrial facilities, water, food and a myriad of consumer products and other substances that individuals come into direct contact with each day.

Figure 3.1: The Conceptual Framework



- Behavioral characteristics [**BC**] combine preventive health behavior such as prenatal care, nutrition, and use of drugs, tobacco, and alcohol.
- Maternal characteristics [**MC**] include the general physical and psychological health state of mother at impregnation and during pregnancy, and the reproductive history of the mother.
- Pregnancy outcome and medical realization [**POMR**] include all variables linked to the pregnancy outcome and medical assistance. Pregnancy outcome is related to information about delivery and newborn (fetal position before birth, weight at birth and etc.). During delivery there can be some extra medical assistance as pediatrician, anesthetist or other. As well there exists differences across the hospitals themselves. Some of them are more equipped due to their financial support or to their competencies.
- Cost [**COST**] refers to the cost of delivery and the cost of hospitalization during the first six months after the birth.

3.4 Description of the causal mechanisms

3.4.1 Socio-economic characteristics

Socio-economic characteristics on behavioral characteristics

Socio-economic characteristics such as SEC in childhood, education, adult SEC, and

partnership history have an effect on maternal behavioral characteristics. According to Graham and Der (1999), women's smoking status in adulthood is sensitive to both past and current socio-economic circumstances. They found that smoking status is significantly associated with father's social class, school leaving age and highest school qualification. As well adult SEC, as housing tenure, car ownership, cohabitation status and being independent on means-tested benefits, have a significant influence on smoking status. Later Graham et al. (2006), implicated women's domestic trajectories and circumstances (childbearing histories and cohabitation status) into analysis of the socio-economic effect on women's current and former smoking status in early adulthood. Geographical area of residence and SEC has also been the focus of investigation of alcohol consumption, although results are not conclusive. For example, some studies indicate that women with lower SEC are more likely to consume alcohol in pregnancy (Leonardson and Loudenburg (2003); Leonardson et al. (2007); Raatikainen et al. (2006); Stansberg-Larsen et al. (2006)), while others have found the opposite (Center fo Desease Control and Prevention [CDC] (1995); Chang et al. (2006); Pevalin et al. (2001)). The illicit drug as cannabis, cocaine, and other stimulants were used more by younger with lower level of education and lower household income women (Gelder et al. (2010)).
Martin et al. (2007) examined whether women whose partners are involved in their pregnancy are more likely to receive early prenatal care and reduce cigarette consumption over the course of the pregnancy. The study showed a positive influence of father involvement into prenatal care. Among women who smoked at conception, those whose partners were involved in their pregnancy reduced their cigarette consumption 36% more than women whose partners were not involved. As well, study suggested that fathers with less than a high school diploma are less likely to be involved in their partner's pregnancy.

Socio-economic characteristics on environmental characteristics

The association between environmental conditions and socioeconomic status has been an important subject for the scientific community during the last decade. Several studies have found that disadvantaged groups experience the worst environmental conditions , while on the other hand, some other studies have identified greater exposure in areas of higher socioeconomic status.
Environmental tobacco smoke (ETS) is steeply and inversely associated with socioeconomic variables as education, occupation, and median neighborhood income. ETS exposure is highest among individuals in the lowest occupational level, and lowest among university-educated (Whitlock et al. (1998)). Cesaroni et al. (2010) , char-

acterized the residents of Rome according to different measure of traffic exposure and socioeconomic position. In all areas of the city there is evidence of association between older age and living in proximity of an high traffic roads (HTR), however, the association between traffic exposure and socioeconomic position had a different sigh in the two areas: the central part of Rome and the rest of the city. In the city center where the traffic is higher and 25% of the residents live close to HTR, less affluent and less educated people tend to live closer to HTR than more affluent and highly educated people, the opposite is seen in the rest of the city.
Branis and Linhartova (2012), analyzed differentials in exposure to sulfur dioxide (SO_2),atmospheric particulate matter (PM_{10}) and nitrogen dioxide (NO_2) among Czech population categorized according to education level, unemployment rate, population size and average annual salary. They concluded that inhabitants with low education level and high unemployment rate mainly reside in smaller cities with higher concentration levels of combustion-related air pollution, while residents with higher socioeconomic status reside in large cities with expose to higher levels of traffic-related air pollution. Using the Population and Housing Census 2001, Fernández-Somoano et al. (2013) , investigated relationship between area-level socioeconomic characteristics and outdoor NO_2 concentrations in rural and urban areas of northern Spain. They concluded that outdoor concentrations of NO_2 are higher for higher level of education and with higher socioeconomic index based on occupation in census tracts with over 50% urban area. While in more urban area, they found association of higher NO_2 with a lower socioeconomic index. Vrijheid and Martinez (2012) examined whether socioeconomic inequalities exist in exposure to multiple common environmental contaminants in air (traffic-related air pollution), water (total trihalomethane concentration level in tap water) and food (total mercury level, levels of organochlorine pollutants, p,p'-DDE and PCB congeners). Their study demonstrated that association between socioeconomic status and levels of exposure to environmental pollutants is weak and inconsistent, with some exposures more prevalent in higher and others in lower social and educational classes. They concluded that the assumption that more disadvantaged groups have higher exposure levels does not always hold and will depend on the type of exposure, and on the location of both early-life and current residence of the person.

Socio-economic characteristics on maternal characteristics

Study of the relationships between socioeconomic status and health has had a long scientific history and still continues to be a top public health priority (U.S. Department of Health and Human Services (2010)). A lot of disciplines like epidemiology,

demography, sociology, and economics attended in establishing the strong and consistent relations between SES and health. Individuals who are less educated, have lower-status job, and earn less or no income are at greater risk for poor health than their higher-SES counterparts. The associations extend from relatively minor illnesses to serious and life-threatening disease. Cundiff et al. (2013) examined interactions between different indicators of SES and ambulatory blood pressure (ABP) during daily life. Their findings show that education and income are independently and inversely associated with ABP in women, and that education may offset the risk for high blood pressure associated with low income. Hughes and Waite (2009), Zhang and Hayward (2006) showed that marital history over the life time course shapes a range of health outcomes, including cardiovascular disease, chronic conditions, and depressive symptoms. Using nationally representative sample of U.S. adults, Zajacova et al. (2009) , identified socioeconomic and racial/ethnicity disparities in the burden of multiple persistent infections. Adults with higher education and income have a lower burden of persistent infections, and non-Hispanic white have lower burden than minority adults. The burden of persistent infection may be one pathway through which lower socioeconomic position "gets under the skin" and leads to an earlier onset of disease and mortality.

3.4.2 Environmental characteristics

Environmental characteristics on maternal characteristics

The relationship between women's health and the environment is complex, including association between home, workplace, public environment and women's health. Women still carry the greatest burden of household work and are exposed to cleaning substances and other toxins in the home environment. Almost half the world's population still relies for its everyday household energy needs on inefficient and highly polluting solid fuel , mostly biomass and coal. Biomass and coal smoke emit many health-damaging pollutants, including particular matter, carbon monoxide, sulfur oxides, nitrogen oxides, aldehydes, benzene, and polyaromatic compounds (Smith (1987)). Women who spend many hours cooking and are exposure to biomass smoke, develop chronic obstructive lung disease (COPD), asthma, respiratory track infection, including tuberculosis and lung cancer. As well it had been argued that exposure to biomass fuel smoke is a bigger risk factor for COPD than tobacco smoking (Kodgule and Salvi (2012); Smith et al. (2004); Jamison et al. (2006)). Unfortunately, the research in this area is not proceeding very rapidly and there is no research made concerning the impact of violence in the home or other factors on women's physical and mental health.

Occupation health research is increasingly focused on the impact of the workplace on women's health and in addition occupational exposures to toxin and in particular their impact on women's reproductive health (Surdu et al. (2013); Foss et al. (2011)).
Recently there is increasing interest in exploring the links between work environment and the increasing incidence of breast cancer (Labreche et al. (2010); Snedeker (2006); Labreche and Goldberg (1997); Cantor et al. (1995)). Miyake et al. (2012) studied the association between employment and the prevalence of depressive symptoms during pregnancy. They concluded that employment, whether full-time or part-time, and holding a professional or technical job or a clerical or related occupation may be inversely associated with the prevalence of depressive symptoms during pregnancy. As well, it is showed that pregnant women who worked with patients or children or food product had an excess risk of sick leave during pregnancy (Morales-Suarez-Varela et al. (2010)).
The effects of outdoor air pollutants on health and mortality is well known (Atkinson et al. (2013); Carey et al. (2013); Cohen (2000)). Orru et al. (2012) investigated the impact of particulate air pollution on health in Vilnius and Kaunas. The study results indicated that besides premature mortality, 272 cases of hospitalization due to cardiovascular disease and 564 cases due to cardiovascular complains can be predicted annually. More recent epidemiological studies have raised concerns about the potential impact of air pollution on central nervous system (CNS) outcomes including chronic brain inflammation, microglia activation, and white matter abnormalities leading to increased risk of neurodegenerative diseases, multiple sclerosis (Block and Calderon-Garciduenas (2009)). Some human studies have shown that living in conditions with elevated air pollution is linked to decreased cognitive function (Ranft et al. (2009); Power et al. (2011); Weuve et al. (2012)), AD- and PD-like neuropathology (Calderon-Garciduenas et al. (2004, 2010, 2012)), increased stroke incidence (Villeneuve et al. (2006); Mateen and Brook (2011)), and elevated autism risk (Volk et al. (2011)).

Environmental characteristics on pregnancy outcome and medical realization

It is well known that environmental exposures plays an important role in the causation of disease. Various studies have suggested associations between environmental characteristics such as air pollution, environmental tobacco smoke, water contaminants, metals, radiation and pregnancy outcomes such as pregnancy loss, stillbirth, fetal growth, preterm birth and congenital anomalies. The main findings of system-

atic review and meta-analysis of association between maternal ETS exposure and pregnancy outcome, is significantly associated with greater increase in the risk of low birth weight (<2500 g), an increase risk of congenital anomalies, but did not find significant increase in small-for-gestational age (Leonardi-Bee et al. (2008); Salmasi et al. (2010)). Lakshimi et al. (2013) investigated the association between indoor air pollution (liquid petroleum gas/electricity, kerosene, biomass) and risk of stillbirth. They proved that women who cook with biomass fuels were found to be at approximately 20-40% higher risk of delivering a stillbirth. Pope et al. (2010) carried out systematic review on risk of low birth weight and stillbirth associated with indoor air pollution in developing countries. They identified a significantly increased risk of LBW of 38% from exposure to IAP, an associated reduction in birth weight of 96.6 g, and a 51% increase in risk of stillbirth associated with IAP. The International Collaboration on Air Pollution and Pregnancy Outcomes (ICAPPO) was formed to better understand relationships between air pollution and adverse birth outcomes in different countries. Fourteen research groups from nine countries developed a protocol to estimate odds ratios (ORs) for association between PM_{10} and LBW among term births, adjusted for socioeconomic status and second for additional location-specific variables. Primary studies showed that ORs ranged from 0.63 [95% confidence interval (CI), 0.30-1.35] for Netherlands to 1.15 [95% CI, 0.661-2.18] for Vancouver (2011). Using targeted maximum likelihood estimation analysis, Padula et al. (2012) , provided a semiparametric estimate of the causal association between traffic exposure during pregnancy and term LBW. Even if the results did not show a clear exposure-response relation; however, they found a significant difference in the predicted probability of LBW, showing that higher traffic density is associated with increased probability of LBW. Increased level of ozone during the first trimester increased the risk of pre-eclampsia and preterm birth (Olsson et al. (2013)).
Morales-Suarez-Varela et al. (2010) investigated of women working in occupation where exposure to infections agents is common have a higher risk of adverse pregnancy outcome. they concluded that prevalence of congenital anomalies was slightly higher in children of women who worked with patients, and prevalence of small for gestational age infants was higher among women who worked with food products.

3.4.3 Behavioral characteristics

Behavioral characteristics on maternal characteristics

Maternal physical and psychological health status is influenced by the smoking, alcohol consumption, and illicit drug use. It is well known that smoking can cause a lot of health problems as cancer, heart disease, stroke, and others. Tobacco smoke

is the most important known cause of cancer and has been associated with an extensive list of specific cancer (International Agency for Research on Cancer (2004, 2012); U.S. Department of Health and Human Services (2004)). There are number of studies on the association between tobacco smoke and breast cancer risk (Palmer and Rosenberg (1993); Terry and Rohan (2002); Reynolds et al. (2004); Xue et al. (2011)). Bjerkaas et al. (2013) examined association between smoking initiation before the first childbirth and breast cancer . Using Cox proportional hazard models, they concluded that ever smokers compared with never smokers, had a 15% (hazard ratio (HR)=1.15, 95% CI 1.10-1.21) increased risk of breast cancer and ever smokers who started to smoke more than 1 year after the first childbirth had nor increased risk (HR=0.93, 95% CI 0.86-1.02), while those who initiated smoking more than 10 years before their first childbirth had a 60% (HR=1.60, 95% CI 1.42-1.80) increased risk of breast cancer, compared with never smokers. Some studies showed that a consumption of one cigarette was followed by a decrease in the distensibility of the aorta and medium-sized arteries (Faila et al. (1997); Stefanadis et al. (1997)), and that chronic smoking, independent of the duration, is associated with a decrease in aortic distensibility (Sassalos et al. (2006)). Research has also identified that smoking has a negative impact on skin health and effects wound healing at a cellular level (McRobert (2013)). Women smokers had significantly higher rates of posttraumatic stress disorder, past year depression and anxiety, suicidality and co-occurring disorders (Holma et al. (2013); Jessup et al. (2012)).

Kroll et al. (2012) examined association of subtypes of haematological malignancy with alcohol drinking and tobacco smoking in the prospective Million Women Study. Based on Cox regression estimates, they concluded that *among predominantly moderate drinkers, greater alcohol intake was associated with reduce risk of lymphoid malignancies, and cigarette smoking was associated with increased risk of Hodgkin lymphoma, mature T-cell malignancies and myeloproliferative/myelodysplastic disease.* Poli et al. (2013) in their consensus paper reviewed the available evidence on the association between moderate alcohol use, health and disease. Based on the review of epidemiological studies, they concluded that there is a statistically significant association between moderate alcohol consumption (one or two drinks/d or 12-24 g of ethanil/d for women) and risk reduction of atherosclerosis, myocardial infarction or ischaemic stroke. Moderate alcohol use on the other hand is associated with modest but significant increased risk of cancer of the upper respiratory and gastrointestinal tracts, as well as of breast cancer. The authors mentioned that the association between ethanol intake and all-cause mortality follows the well-known 'J'-shaped curve: mortality among moderate drinkers is significantly lower than that of teetotallers and that of heavy drinkers. The National Epidemiological Study of

Alcoholism and Related Conditions (NESARC, USA) reported that persons with drug dependence and alcohol dependence has approximately 9- and 4-fold increased risk of major depressive disorder, respectively . Based on the 2008 Thai National Mental Health survey, Suttajit et al. (2012) tried to determine whether alcohol use disorders were associated with major depressive and anxiety disorders, and whether the use of illicit drugs increased these associations. Using logistic regression they came to conclusion that individuals with alcohol use disorders (compared with the general population) alone had significantly increased risk of major depressive (OR 4.09, 95% CI 2.31-7.26 in women) and anxiety disorders (OR 4.34, 95% CI 2.35-8.03 in women). The risk became higher among individuals with both alcohol use disorders and illicit drug use (OR 11.53, 95% CI 1.32-100.65 in women for major depressive disorder, and OR 13.10, 95% CI 1.48-115.60 in women for anxiety disorders).
Individuals with a history of heroin dependence have poorer health and functioning than their counterparts in the general population. At a younger age, women reported poorer overall health status and more chronic health and mental health problem than men (Grella and Lovinger (2012)).

Behavioral characteristics on pregnancy outcome and medical realization

Maternal behavior before and during pregnancy is associated with complications during pregnancy, delivery, health status of newborn and long-term deficits in physical and mental development of the offspring. The effects of tobacco use during pregnancy have been extensively studied over many years. The first compelling scientific evidence that smoking could harm fetus appeared in 1957. Simpson (1957) in his primary report demonstrated that the *prematurity rate for smokers is significantly higher than for nonsmokers, and that there is a relationship between incidence of prematurity and the number of cigarettes smoked per day.* The US Department of Health and Human Services listed over 12 adverse pregnancy outcomes associated with smoking in the 1980 Surgeon General's Report. They reported that *babies born to women who smoke during pregnancy are, on the average, 200 grams lighter than babies born to comparable nonsmoking women*; *the risk of spontaneous abortion, fetal death, and neonatal death increases directly with increasing levels of maternal smoking during pregnancy*; as well maternal smoking results a highly significant increase in *the risk of abruptio placentae, placenta previa, bleeding early or late in pregnancy, premature and prolonged rupture of membranes, and preterm delivery.* The biological mechanisms of how tobacco smoke effects fetal development have been examined in extensive human and laboratory studies, which show that many of the 7000 chem-

icals can cross the placental barrier and have a direct harmful effect on the unborn baby (British Medical Association, 2004; A.E. Quinton et al., 2008; P. Talbot, 2008; J.M. Rogers, 2009). Arias and Viner-Brown (2012) examined whether there is relationship between maternal smoking and birth defects in Rhode Island. Results from their study showed a significant association between maternal smoking and all birth defects (adjusted odds ratio(aOR)=1.27), specifically a strong significant associations between maternal smoking and clubfoot (aOR=2.24) and pulmonic stenosis (aOR=4.75). In the early neonatal period, parental smoking has been shown to increase the rates of respiratory tract infections and sudden infant death syndrome (Tong et al. (2009); Kafouri et al. (2009)). Tobacco use during pregnancy has an effect on the development of preterm premature rupture of membranes (PPROM). England et al. (2013) analyzed the effect of cigarette smoking on PPROM categorized by gestational age. The unconditional logistic regression analysis and Wald test were used to estimate the adjusted risk of PPROM according to smoking status. They founded that smoking more than 10 cigarettes per day is associated with an increased risk of PPROM at all gestational age categories (<28 week: OR 3.19, 95% CI 2.20-12.7; <32 weeks: OR 2.36, 95% CI 1.09-5.11; < 37 weeks: OR 1.97, 95% CI 1.32-2.94; and >37 weeks: OR 3.19, 95% CI 0.92-11.0), but there was no association with significant risk of PPROM and smoking 1 to 10 cigarettes per day.

Cigarettes smoking appears to be the most important maternal behavior characteristic associated with adverse pregnancy outcomes; alcohol and drugs use are also have an impact, but with less importance. Alcohol is the risk factor of premature deliveries, abortions, and placenta-associated pathologies. Disorders of children with prenatal exposure to alcohol are described as fetal alcohol syndrome, alcohol related neurodevelopmental disorders and alcohol related birth defects (Jagielska et al. (2012)). O'Leary et al. (2010) investigated the association between dose, pattern, and timing of prenatal alcohol exposure (PAE) and birth defects. They used data from a randomly selected, population-based cohort of nonindigenous women who gave birth to a live infant in Western Australia between 1995 and 1997. they concluded a significant fourfold increase in birth defects followed heavy PAE during the first trimester, predominantly attributable to ventricular and atrial septal defects. The medical literature on the effects of prenatal drug use is less conclusive than that for smoking or alcohol consumption. This we can explain by the fact that reliable data on drug use is rare. Noonan et al. (2007) estimated the effect of maternal illicit drug use (cocaine, heroin, marijuana and other unspecified drugs) on low birth weight and abnormal infant health conditions using economic framework. The results of this study showed that prenatal illicit drug use increases the likelihood of low birth weight by 4-6 percentage points and increases the likelihood of abnormal infant

conditions by 7-12 percentage points. Jansson and Velez (2011) reviewed the current and relevant scientific literature regarding the effects of maternal substance use on the developing child. Cocaine/opioids-exposed infants are at risk for preterm birth and LBW. A dose-response relationships between cocaine and negative association with motor and state regulation capabilities has been reported. The most notable opioid effect on the neonate is neonatal abstinence syndrome. Regular marijuana use can result prolonged fetal exposure, minor physical anomalies as ocular hypertelorism and epicanthus,while heavy marijuana use can result shorter gestational periods. A research of Black et al. (2013) showed that women using illegal drugs are significantly more likely to have an antepartum hemorrhage, preterm delivery at any gestation, and a low birth weight infant than women who smoke cigarettes. Nutritional status of woman before and during pregnancy is a fundamental determinant of foetal growth, birth weight and infant morbidity (Merchant and Kurz, 1993; WHO 1995a).

Bukowski et al. (2009); Catov et al. (2011); Liu et al. (2011); Phithakwatchara and Titapant (2007) examined the effects of pre-pregnancy and/or periconceptional nutrition on the risk of the developing pre-eclampsia later in pregnancy. The latter authors reported that the risk of pre-eclampsia was significantly increased in overweight Thai women compared with normal weight women. The association between maternal obesity and fetal macrosomia has been established by many studies (Birdsall et al. (2009)). Maternal weight and insulin resistance before pregnancy affect fetal growth, as is reflected in the birth weight (Catalano et al. (1995)) . In some studies, an up to five-fold increase in intrauterine death as well as increased infant mortality rate has been recorded in obese women (Nohr et al. (2005); Salihu et al. (2007)). Czeizel and Dudas (1992) compared the risk of neural tube defect (NTD) births among women receiving vitamin supplement and those receiving trace-element supplements daily from at least 1 month before conception and until the date of the second missed menstrual period or later in a randomized controlled trial and showed significant reductions in congenital malformations and the first time occurrence of NTD.

Several observational studies have examined the association with maternal nutrition status based on weight and height and/or vitamin supplement use during the periconceptional period and birth size (Ronnenberg et al. (2003); Liu et al. (2011)). The most recent prospective study from Vietnam (Ota et al. (2011)) reported a significantly higher risk of delivering a SGA infant among women who were underweight before conception compared with those with BMI (body mass index) between 18.5 and 23.0 kg/m^2.

3.4.4 Maternal characteristics

Maternal characteristics on pregnancy outcome and medical realization

All the deaths and diseases of infants during perinatal and neonatal period are strongly associated with maternal biological characteristics and with problems during pregnancy. Lao et al. (2013) conducted retrospective cohort study to examine the relationship between maternal hepatitis B surface antigen status (HBV) with pregnancy-induced hypertension and pre-eclampsia. Using multiple logistic regression they concluded the increased adverse pregnancy outcome associated with maternal HBV infection and a reduced incidence of pre-eclampsia. Morton et al. (2013) using matched cohort study established relationship between women with physical disabilities and pregnancy outcomes. They pointed out that women with physical disabilities have *a statistically significantly higher rate of pregnancy complications, including infections during pregnancy, 2 IUFDs (instances of intrauterine fetal demise), and low birth weight infants compared with non-disabled controls.* In the early 1990s, Offenbacher's group using a bacteremia model and a "chamber" model to mimic a focal infection on pregnant hamsters demonstrated that periodontal bacteria an inflammatory mediators have the ability to disseminate systematically to the foetal-placenta unit, via the blood circulation and induce pregnancy complications (Collins et al. (1995a,b)). Since these first landmark, many investigators have tried to explain whether this causal relation of an oral pathogen inducing adverse pregnancy outcomes in an animal model has analogy with periodontal disease and adverse pregnancy outcomes in humans. Cruz et al. (2012) and Guimaraes et al. (2012) showed that maternal periodontitis was associated with low birth weight. The study of Kumar et al. (2012) showed a significant association between periodontitis and an increased risk of pre-eclampsia, intrauterine growth restriction, preterm delivery and low birth-weight infants. Mulder et al. (2002) conducted a review on how perinatal maternal stress effects pregnancy and (unborn) child. They concluded that maternal psychological factors as high stress and anxiety levels, can increase the risk for spontaneous abortion and preterm labour and for having a malformed or growth-retarded baby. Several studies examined association between reproduction history and adverse pregnancy outcome. A previous induced abortion is associated with a significantly increased risk of LBW (Shah and Zao (2009); Klementti et al. (2012); Hardy et al. (2013)) and preterm birth (Shah and Zao (2009); Klementti et al. (2012)). The association is positive, meaning that the risk increases as the number of induced abortion increase. While repeated cesarean sections increases the risk of uterine rupture and intraoperative complications (Gasim et al. (2013)).

3.4.5 Pregnancy outcome and medical realization

Pregnancy outcome and medical realization on the cost

The pregnancy outcome and medical realization influence directly the cost. According to act no. GPG/2011/119, the cost can vary depending on the hospital type or the difficulty of intervention, for example, cesarean section is much more expensive than vaginal method of delivery.

3.5 Decomposition of the multivariate distribution

The Figure 3.1 represents cost as a recursive system in which all the variables are linked through a direct acyclic chain, that is, each variable depends on its predecessors and there are no feedback relations. The absence of an arrow between two determinants of cost indicates that these two indicators are conditionally independent. this set of conditional independencies is grouped by the following relations:

1. $COST \perp\!\!\!\perp BC, MC, EC, SEC \,|\, POMR$
2. $POMR \perp\!\!\!\perp SEC \,|\, MC, BC, EC$
3. $BC \perp\!\!\!\perp EC \,|\, SEC$

where the symbol $\perp\!\!\!\perp$ is read as 'independent of' and the symbol | as 'conditionally on'. For example, [2] means that pregnancy outcome and medical realization is independent of socio-economic characteristics conditionally on maternal, behavioral and environmental characteristics. Graphically, this condition is represented by the absence of directed arrow between pregnancy outcome and medical realization and socio-economic characteristics.
Then the process leading to the cost can be decomposed into the product of marginal and conditional probabilities:

$$\begin{aligned} p(COST, POMR, MC, EC, BC, SEC|\theta) \quad &= p(COST|POMR, \theta_C) \\ \times p(POMR|MC, EC, BC, \theta_{POMR}) \\ \times p(MC|EC, BC, SEC\theta_{MC}) \\ \times p(BC, EC|SEC, \theta_{BC,EC}) \\ \times p(SEC|\theta_{SEC}) \end{aligned} \tag{3.1}$$

where $\theta_{COST} = \theta_{(COST|POMR)}$,
$\theta_{POMR} = \theta_{(POMR|MC,EC,BC)}$,
$\theta_{MC} = \theta_{(MC|EC,BC,SEC)}$,

$\theta_{BC,EC} = \theta_{(BC,EC|SEC)}$,
and $\theta = (\theta_{COST}, \theta_{POMR}, \theta_{MC}, \theta_{BC,EC}, \theta_{SEC})'$ is the vector of the parameters of the multivariate distribution. This decomposition represented by (3.1) corresponds exactly to the DAG presented by the conceptual framework in Figure 3.1.
In the latter formula each of 5 conditional densities represent a structural process and the parameters of these conditional densities are independent of each other, this means that the conditional densities could consequently be individually estimated without loss of information: the parameter of conditional density is not influenced by parameters of the other conditional densities. In the model described by (3.1) some of the variables are exogenous in a conditional density while endogenous in another. In our complete process,only the socio-economic variables are exogenous with respect to all the set of parameters. Consequently we are interested in a process characterized by a multivariate distribution composed of 5 endogenous variables (COST, POMR, MC, BC, EC) and 1 exogenous variable (SEC).

3.6 Application

3.6.1 Introduction

The structure of the cost can be viewed as a complex mechanism where socio-economic, environmental and biological variables enter into account, those variables having a role at a particular moment of the mechanism. Once a conceptual framework has been determined, it is replaced by the operational framework using the available database.

3.6.2 Data

When the concepts of the conceptual framework are substituted by available data, we get the so-called *operational framework*. Before going into statistical analysis, the set of selected indicators and their description are presented.

Table 3.1: Description of the selected data

Label	Code	Categories
Socio-economics Characteristics		
Mother:		
Maternal age	MA	<25,25-29,30-34,35-39,41+

Continued on next page

Table 3.1 – *Continued from previous page*

Label	**Code**	**Categories**
Marital status	MS	single,married,separated,divorced, widowed,not stated
Nationality	MN	Italian,other EU citizen,citizen of LMLIC*,other
Place of birth	MPB	north-east Italy, north-west Italy, central Italy, south Italy, islands(Italy), abroad, n/a
Education level	ME	primary or no education,secondary,highschool,tertiary
Current occupation status	MCOS	occupied,unemployed,in search of first job,student,housewife,other,n/a
Father:		
Paternal age	FA	<25,25-29,30-34,35-39,41+,n/a
Nationality	FN	Italian,other EU citizen,citizen of LMLIC*,other
Education level	FEL	primary or no education,secondary,highschool,tertiary
Current occupation status	FCOS	occupied,unemployed,in search of first job,student,housewife,other,n/a
Environmental Characteristics		
Mother:		
Profession	MP	self-employed, manager or director, employee, worker, other dependent employer , n/a
Sector of employment	MSE	agriculture,hunting and fishing, industry, services, public administration, other private sector, n/a
Father:		
Profession	FP	self-employed, manager or director, employee, worker, other dependent employer,n/a
Sector of employment	FSE	agriculture,hunting and fishing, industry, services, public administration, other private sector,n/a

Continued on next page

Table 3.1 – *Continued from previous page*

Label	Code	Categories
Behavioral Characteristics		
Smoking	SM	stopped smoking in the last 5 years before pregnancy, stooped smoking at the beginning of pregnancy, was smoking before and during pregnancy, non-smoker
Amniocentesis	AC	yes,no,n/a
Chorionic villus sampling	CVS	yes,no,n/a
Fetoscopy/Funiculocentesis	FT	yes,no,n/a
Number of control visits	NV	0-5,6,7,8,9+
Number of pregnancy ultrasound	NE	0,1-3,4+
Maternal Characteristics		
Number of previous:		
live births	NB	0,1,2,3+
spontaneous abortion	NSA	0,1,2,3+
induce abortion	NIA	0,1,2,3+
cesarean	NC	0,1,2,3+
Hospitalization during pregnancy	HP	yes,no,n/a
Pregnancy Outcome,Medical Realization		
Type of hospital	TH	Public type A, Public type B, Private
Method of labour	ML	Spontaneous, Induce, Without
Reason of induction	RI	Prolonged pregnancy,Oligodramnios, Premature rupture of membranes (PROM), maternal pathology, fetal pathology,n/a
Type of induction	TI	prostaglandin, oxytocin, other drug, amnioressi, other mechanical method
Anti-pain method used in labour	AntiM	no method used, epidural analgesia, other type of pharmacological analgesia, non-pharmacological method,n/a

Continued on next page

Table 3.1 – *Continued from previous page*

Label	**Code**	**Categories**
Method of delivery	MD	vaginal (non-assisted), vaginal with use of forceps, vaginal with use of ventouse, cesarean
Presence during delivery:		
Pediatrician/monatologist	PP	yes,no,n/a
Anesthetist	AP	yes,no,n/a
Neonatal Nurse	NP	yes,no,n/a
Fetal position before birth	FPBB	facing down 1, facing upward 1, facing upward 2, facing down 2, feet first, lying sideways
Weight at birth	WB	<1500,1500-2499,2500-3499,3500+
Presence of malformation	PM	yes,no
Apgar score	AS	10-7,6-4,3-0
Type of resuscitation	TR	manual ventilation, intubation, not required
Cost		
Cost	C	*continuous*

The socio-economic characteristics are represented by maternal age (MA), marital status (MS), nationality (MN) and place of birth (MPB) of the mother, maternal education (ME), mother's occupation status (MCOS), paternal age (FA) and education (FEL), father's nationality (FN) and occupation status (FCOS). The indicators of the behavioural characteristics are maternal smoking status (SM) and prenatal care: amniocentesis (AC), chorionic villus sampling (CVS), fetoscopy (FT), number of control visits (NV) and number of pregnancy ultrasound (NE). For maternal characteristics, the indicators are the numbers of previous live births (NB), spontaneous (NSA) and induce abortion (NIA), ceasarean (NC), and hospitalization during pregnancy (HP). The environmental characteristics are represented by maternal profession (MP) and sector of employment (MSE), as well by paternal profession (FP) and sector of employment (FSE). And finally, for the pregnancy outcome and medical realization, the indicators are type of hospital (TH), method of labour (ML), reason (RI) and type (TI) of induction, anti-pain method used in labour (AntiM), method of delivery (MD), presence of pediatrician (PP), anesthetist (AP) and neonatal nurse (NP), fetal position before birth (FPBB), weight at birth

(WB), apgar score (AS), type of resuscitation (TR) and presence of malformation (PM).
The indicators available from the CedAP and SDO data sets are far from being perfect. For the behavioural characteristics, the database includes only the smoking status and prenatal care characteristics. Alcohol and drug habits as well as nutrition are not available. For the maternal variables, the past reproductive history is well represented, however, we do not have much information on the health status of the mother or any information about exposure to stress during pregnancy. Information on the environmental characteristics is also relatively substantial: parent's profession and sector of employment.
By passing from the conceptual framework to the operational framework, some relations do not have sense any more. For example, the weight at birth (indicator of the pregnancy outcome and medical realization) do not influence the cost. The operational framework has to be revised accordingly.

3.6.3 Operational Framework

The operational framework corresponds to the conceptual framework where the theoretical determinants are replaces by their indicators. The causal mechanisms are represented in the operational framework (Figure 10) on the basis of the observed variables. The vertices or nodes in the Figure 10 represent variables while the directed edges or links between nodes represent assumed causal relations.
As our research interest is the maternal smoking impact on the delivery cost, we 'reduced' operational framework which graphical representation is given in Figure 3.2.
Once the operational framework has been determined, the estimation of statistical model is based on the decomposition of the joint distribution of the set of variables through conditional and marginal distributions:

$$
\begin{aligned}
&p(COST, TI, MD, AP, PP, NP, AntiM, TH, TR, PM, ML, RI, FPBB, WB, AS, SM, HP, \theta) \\
&= p(COST|TI, MD, AP, NP, PP, TH, TR, PM, Ml, AntiM, \theta_{COST}) \\
&\times p(RI|SM, \theta_{RI}) \times p(HP|SM, \theta_{HP}) \times p(ML|RI, \theta_{ML}) \\
&\times p(MD|FPBB, \theta_{MD}) \times p(PM|SM, HP, \theta_{PM}) \times p(WB|SM, HP, \theta_{WB}) \\
&\times p(AS|SM, HP, WB, \theta_{AS}) \times p(TR|MD, TI, \theta_{TR}) \times p(NP|AS, WB, \theta_{NP}) \\
&\times p(PP|WB, \theta_{PP}) \times p(AP|AntiM, TI, MD, TR, \theta_{AP}) \\
&\times p(HP, \theta_{HP}) \times p(AntiM, \theta_{AntiM}) \times p(FPBB, \theta_{FPBB}) \\
&\times p(SM, \theta_{SM}) \times p(TI, \theta_{TI}). \qquad (3.2)
\end{aligned}
$$

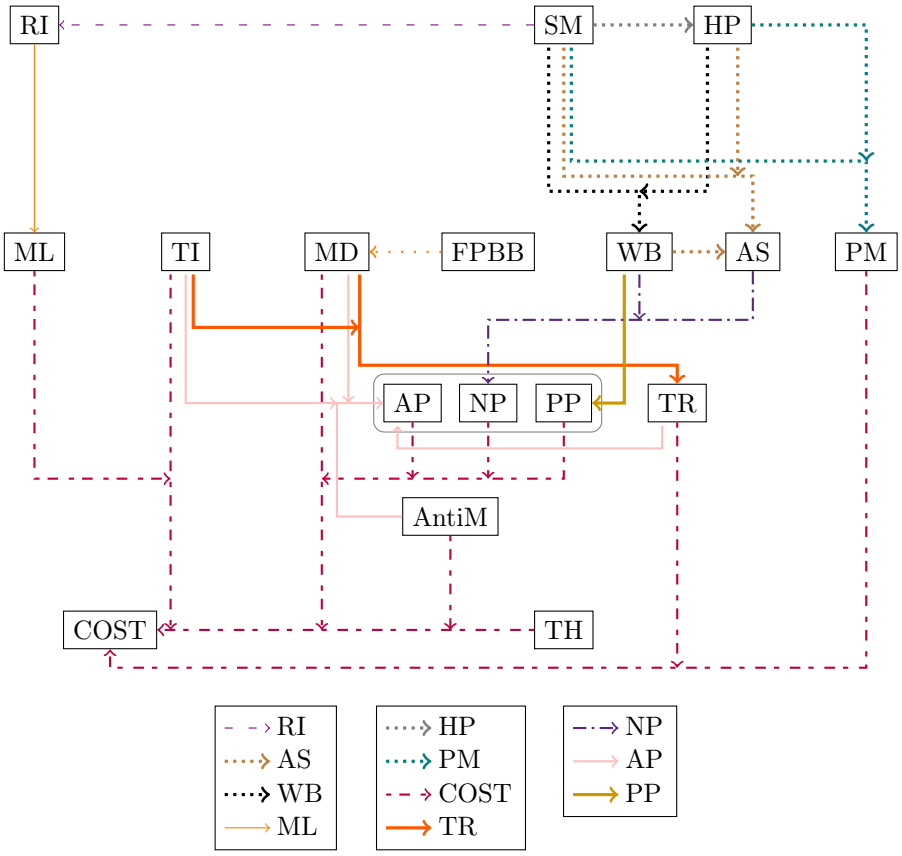


Figure 3.2: 'Reduced' operational framework

3.6.4 Estimation method

The method presented in Section 3.2 denote a general theoretical way of model-building, but do not employ any specific empirical method for estimation.
Here is presented chosen estimation method (other analyzed methods briefly presented in Appendix C).
The data can be presented by a matrix Z of dimension $N \times M$ where N is the number of observation and M is the number of variables. We can say that Z is composed of a set of endogenous variables Y and a set of exogenous variables, X. In other words, the process of interest is the distribution of Y with respect to X. Each of the endogenous variables is estimated through ordered logit model. Such a model estimates the conditional probability that a particular event is below a given level k.

$$Logit(RI) = \alpha_{RI} + \beta_{RI,SM} * SM \tag{3.3}$$

$$Logit(HP) = \alpha_{HP} + \beta_{Hp,SM} * SM \tag{3.4}$$

$$Logit(ML) = \alpha_{ML} + \beta_{ML,RI} * RI \tag{3.5}$$

$$Logit(MD) = \alpha_{MD} + \beta_{MD,FPBB} * FPBB \tag{3.6}$$

$$Logit(PM) = \alpha_{PM} + \beta_{PM,SM} * SM + \beta_{PM,HP} * HP \tag{3.7}$$

$$Logit(WB) = \alpha_{WB} + \beta_{WB,SM} * SM + \beta_{WB,HP} * HP \tag{3.8}$$

$$Logit(AS) = \alpha_{AS} + \beta_{AS,SM} * SM + \beta_{AS,HP} * HP + \beta_{AS,WB} * WB \tag{3.9}$$

$$Logit(TR) = \alpha_{TR} + \beta_{TR,MD} * MD + \beta_{TR,TI} * TI \tag{3.10}$$

$$Logit(PP) = \alpha_{PP} + \beta_{PP,WB} * WB \tag{3.11}$$

$$Logit(NP) = \alpha_{NP} + \beta_{NP,AS} * AS + \beta_{NP,WB} * WB \tag{3.12}$$

$$\begin{aligned} Logit(AP) = \alpha_{AP} &+ \beta_{AP,AntiM} * AntiM + \beta_{AP,TI} * TI \\ &+ \beta_{AP,MD} * MD + \beta_{AP,TR} * TR \end{aligned} \tag{3.13}$$

$$\begin{aligned} Logit(COST) = \alpha_{COST} &+ \beta_{COST,TH} * TH + \beta_{COST,TR} * TH + \beta_{COST,PM} * PM \\ &+ \beta_{COST,AntiM} * AntiM + \beta_{COST,ML} * ML \\ &+ \beta_{COST,TI} * TI + \beta_{COST,MD} * MD + \beta_{COST,AP} * AP \\ &+ \beta_{COST,PP} * PP + \beta_{COST,NP} * NP \end{aligned} \tag{3.14}$$

In the equation (3.14), the variable COST have been discretized, and replaced with five categories variable (using 20,40,60,80 percentiles).

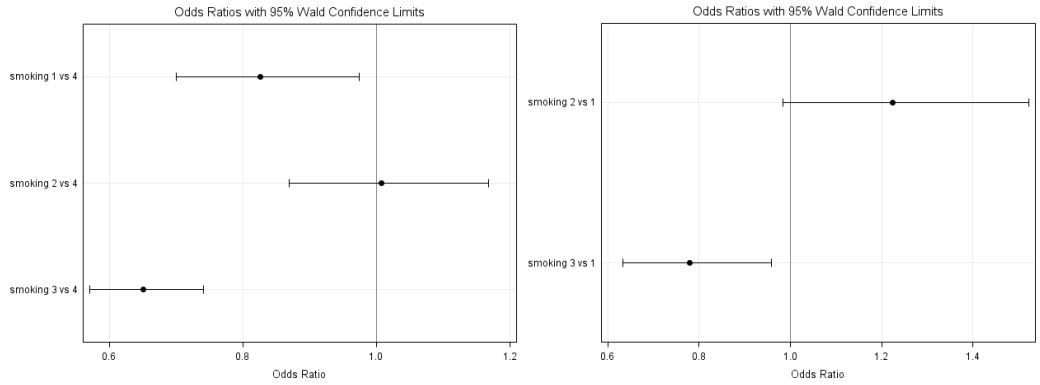
3.6.5 Presentation of the results

Tables 13- 16 present estimation of equation (3.3) to equation (3.14). These results (the odds ratios and coefficients) represent the effect of each selected explanatory

variable (first column in the table) on each dependent variable (first line in the table). The value of the parameters are presented in Tables 14 and 16 (in Appendix C) where significant results are represented by the symbols *(p-value <0.001) or **(p-value <0.05). For analysis we consider two populations of interest: first, newborns born between January and June in 2010 (N=15,475); and second, only newborns whose mothers smoked in the last 5 years prior to pregnancy (N=2,381). The data analysis for this section was generated using SAS 9.3.

As Figure 3.2 shows, maternal smoking status has an impact on hospitalization during pregnancy, reason of induction weight at birth, presents of malformation and Apgar score, but only impact on the weight at birth is significant (tables 14 and 16). Equation for the weight at birth is estimated by ordered logistic regression (eq. 3.8). Parameters with a positive/negative sign mean that the category of the explanatory variable increases/decreases the probability of being in a higher category of the dependent variable (weight at birth in this case). For short, in eq. (3.8), a positive/negative coefficient (β) increases/decreases, with respect to the category of reference, the probability of a higher weight at birth. The odds ratio [smoked in the last 5 years prior to pregnancy/did not smoke in the last 5 years prior to pregnancy] is 0.79, with small confidence interval. This shows that women, who smoked in the last 5 years prior to pregnancy, have a higher risk to deliver a baby with lower birth weight. Figure 3.3 presents the graphical representation of the odds ratio for

Figure 3.3: Weight at birth: odds ratio for maternal smoking status

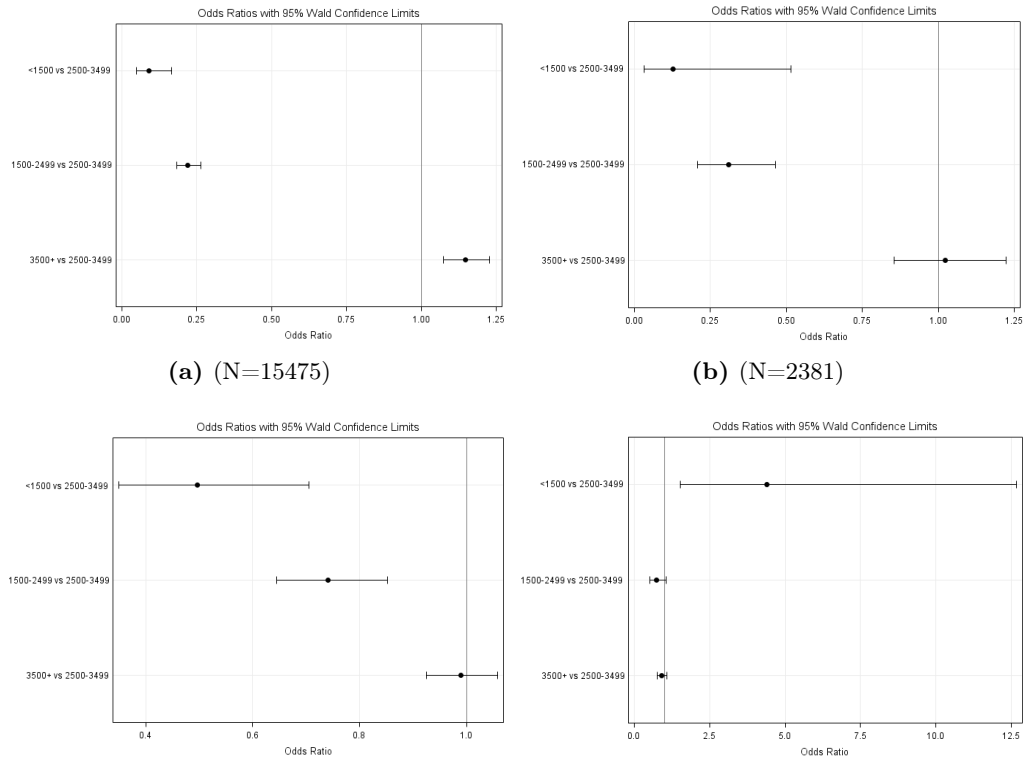


(a) Four smoking categories **(b)** Three smoking categories

smoking variable. The four-categories maternal smoking status is characterized by a non linear relation (Figure 3.3 (a)) while the relation between three-categories maternal smoking status and weight at birth appears as linear (Figure 3.3 (b)). A detrimental effect is observed for women who continued to smoke during pregnancy (OR=0.65 with respect to non-smoker, and OR=0.78 with respect to women who

stopped smoking before pregnancy), and for women who stopped smoking before pregnancy (OR=0.83 with respect to non-smoker).
Graphical representation (Figure 3.4) of the odds ratios for weight at birth indicates that it influences significantly the presence of pediatrician and neonatal nurse (except case (d)). In the figure 3.4 (a) and (b), the relation looks like exponential, although there are not enough points to affirm the real existence of an exponential relationship. The odds ratio goes up from 0.09 (0.13) for a category [<1500] to 1.15 (1.02) for category [3500+] (category of reference is [2500-3499]), respectively. Equation for the cost is estimated by an ordered logit model. Medical realization

Figure 3.4: Pediatrician and/or Neonatal nurse presence during delivery: odds ratio for weight at birth



(a) (N=15475) **(b)** (N=2381)

(c) (N=15475) **(d)** (N=2381)

indicators such as type of hospital, method of delivery and type of resuscitation have a marked impact on the cost.
For the type of resuscitation, the odds ratio goes down from 5.59 (10.68) for manual ventilation to 3.87 (6.58) for an intubation (category of reference is 'not required'); though the confidence intervals for second case are rather large. The effect of type of hospital and method of delivery are easily explained, as the cost is defined according to the regional and national normative documents (act no. GPG/2011/119). For

example, the cost of the ceasarean section in hospital type A starts from 2031.48 €(1895.05 €in hospital type B), while the cost of vaginal delivery in hospital A and B starts from 220 €.
In both cases, the odds ratio for presence of pediatrician at birth with respect to reference category 'not presence' is 1.26, 95% IC: 1.15-1.37 (1.39, 95% IC: 1.11-1.74), which indicates that presence of pediatrician during delivery is associated with higher odds of the cost.
Finally, the presence of malformation appears as one of the determinants of the cost. The odds ratio [presence of malformation/no presence of malformation] is 3.51 (3.75), showing that newborns with malformation are more likely to increase the health care expenditure.
However, we have to point out that this approach is based on the background knowledge and that the cost is estimated by administrative office of Emilia-Romagna which can cause that not all causal relations were detected.

3.6.6 Summary

The objective of this chapter was to evaluate the impact of maternal smoking on delivery cost through structural modelling.
A first step for detecting causal effects is to evaluate a conceptual framework of the determinants of the delivery cost (Section 3.3). The construction of a conceptual framework is based on a review of the literature, is independent of the available data and represents all theoretically possible causal effects of the determinants of the cost.
The operational framework corresponds to the conceptual framework where the theoretical determinants are replaces by their indicators and the estimation of statistical model is based on the decomposition of the joint distribution of the set of variables through conditional and marginal distributions. The estimation of parameters was calculated through ordered logit model.

Chapter 4

Conclusion

As mentioned in the Introduction, the main objective of this study was the analysis of the maternal smoking impact of the delivery cost and to analyze this relation, two research questions have been arise. The conclusion of this thesis is composed of three main parts. In the first step, we summarize the main findings presented in the 2nd and 3rd chapters. In the second step, we will recall **Discussion** presented in the Introduction. In the third and final step, we briefly introduce possible future works and some proposals to administration office of Emilia-Romagna region.

Main findings

Causality is fundamental to our understanding of the nature world. Causal statements are part of everyday speech, as well as legal, scientific and philosophical vocabulary. Human being reach an intuitive consensus on the meaning of many causal utterances and there have been numerous attempts to formalize causality in a way that it is faithful to this consensus.
In the Introduction, two research questions have been put forward. We will answer these questions according to the results obtained throughout this research. The goal of the first question was to evaluate the causal effects on delivery expenditures of the maternal smoking. Furthermore, the second query whether there are the differences of health care cost according to distinct smoking history.
Based on the set of results obtained using geometric multidimensional approach (Chapter 2), we can claim that there is no causal effect on the delivery expenditure of the maternal smoking when the smoking status is not precisely defined. However, in the case when smoking status is defined explicitly, we have found an evidence of maternal smoking effect on the cost. The estimated treatment effect on the delivery cost of the woman who continued to smoke during pregnancy was 14.03 € (compare

with the one that did not smoke in the last 5 years prior to pregnancy), 121.01 € (compare with the one that stopped smoking before pregnancy), and 130.45 € (compare with the one that stopped smoking at the beginning of pregnancy). Furthermore, the analysis showed that, on average, health care expenditure for women who stopped smoking at the beginning of pregnancy was 361.47 € less than for non-smoker, and 93.23 € less than for women whose stopped smoking before pregnancy. The latter results look doubtful, but it can be caused by the selected pre-treatment covariates.

Chapter 3 is dedicated to structural modelling. The constructed conceptual framework is based on a review of the literature, is independent of the available data and represents all theoretically possible causal effects of the determinants of the cost. When we replaced determinants in conceptual framework by available data, the estimation method was based on decomposition of joint distribution and the estimation of parameters through ordered logit model.

The maternal smoking status has an impact on hospitalization during pregnancy, reason of induction weight at birth, presents of malformation and Apgar score, but only impact on the weight at birth is significant. The woman, who smoked in the last 5 years prior to pregnancy, have a higher risk to deliver a baby with lower birth weight respect to the one that did not smoke (OR=0.79, 95% CI: 0.73-0.87). Therefore, if woman continued to smoke during pregnancy the risk to deliver a baby with lower birth weight was even higher OR=0.65 (OR=0.78 with respect to women who stopped smoking before pregnancy). The weight at birth influences the presence of pediatrician during pregnancy in a significant way for newborns weighting less than 2500g (OR=0.09 (0.13) for a category [<1500] and 0.22 (0.31) for category [1500-2499] (category of reference is [2500-3499])).

The presence of malformation, type of resuscitation,presence of pediatrician and type of hospital appear as one of the important determinants of the cost. For the type of resuscitation, the odds ratio goes down from 5.59 (10.68) for manual ventilation to 3.87 (6.58) for an intubation (category of reference is 'not required'); though the confidence intervals for second case are rather large. The odds ratio for presence of pediatrician at birth with respect to reference category 'not presence' is 1.26, 95% IC: 1.15-1.37 (1.39, 95% IC: 1.11-1.74), which indicates that presence of pediatrician during delivery is associated with higher odds of the cost. The odds ratio [presence of malformation/no presence of malformation] is 3.51 (3.75), showing that newborns with malformation are more likely to increase the health care expenditure.

However, we have to point out that this approach is based on the background knowledge and that the cost is estimated by administrative office of Emilia-Romagna which can cause that not all causal relations were detected.

Discussion

In the Introduction we raised a question:

> *Can data driven approach uncover causal relationship or should we model causal mechanism to model a given phenomenon?*

and this step is devoted for a discussion of *pro et contra* of each approach.
A main advantage of geometric multidimensional approach to assess a causal link between the maternal smoking and the delivery cost is that causal inference from observational data can be made without model dependence. In particular, it is not needed to specify *a priori* any model and just let data speak.
In addition, our data mainly consist of categorical variables which need to ne treated with unusual metrics. Using geometric multidimensional method this problem is easily solved. In this case the result of the multiple correspondence analysis is a set of new variables that are continuous and orthogonal to one another.
However, it has several disadvantages. In this case, an important disadvantage of this approach is to determine proper pre-treatment covariates. Recall, that the smoking status describe four treatment levels which started at different time point. And in this case to identify the covariate which is measured before the actual assignment of treatments is rather challenging. Variables measured after treatment assignment can be ambiguous and including them into analysis can give unexpected results (Balinskaite (2013)).
Although the use of cluster analysis can be an issue. That is, the cluster analysis can produce subgroups where either treated or controls are absent (no common support) and a huge number of observations may be discarded.
On the other hand, the structural modelling is model based approach which does not denote a particular statistical model, but refers to a general methodological framework for causal analysis. One of the main advantages of this approach is that it can deal both with the effects of causes and with the causes of effects. And the quality of the results depends upon the quality of the process of model building and model testing.
Furthermore, it is based on a conceptual model (or diagram) constructed independently of the available data which allows to be aware of the effects of the non observed variables.
Nonetheless, it also has its limitations. First of all, it requires reliable prior information which is base on actual knowledge of the theoretical pathway and it does not

detect unknown or unexpected causal relations, and each significant causal relation could still partly be explained by unknown (from a theoretical point of view) common causes.
A second issue is related to the known confounders which can be incorporated into the model only upon the condition that indicators of these confounders are available in the data set. Another issue is concerning the stability of the model which must be fulfilled. To do that, different databases with similar sample sizes is needed, and sometimes it may be difficult to repeat the research (or to have data) on comparable population, in order to check if the results remain stable under changes of context.

Future works

Future works in this area might concern these steps:

- In Section 3.5.2 we mentioned that our database are far from being perfect and particularly information on the environmental characteristics is relatively substantial. This problem could be solved using data from Regional Agency for Environmental Protection in the Emilia-Romagna region (ARPA).

- For this research we used data from Emilia-Romagna region and the time of interest was January-June 2010. It may be of interest to evaluate and compare maternal smoking impact on the delivery cost in different time and region. The Italian regions have always been different in terms of size, geographical character, economic development, civic culture, and institutional performance, with a sharp cleavage between the North and the South of the country (Putnam (1993); Cotta and Verzichelli (2007)). From 2000 (the legalistative decree no. 56) the regional healthcare system is not depending exclusively on transfers from the central government, but as well from their own resources. The regions' revenues consist of a regional tax on productive activities and a regional surtax on the national personal income tax. The gap between the health care systems of the northern and southern regions are increasing in the last decades. Furthermore, according to DOXA (ASSFAD - Survey DOXA-ISS) 19.3% of females were smokers in 2006 (17.3% south Italy and islands; 28.2% center Italy; 17.0% north Italy), 22.3% in 2009 (23.1%; 26.1%;20.0%) and 17.2% in 2012 (20.9%; 14.7%, 15.5%).

- Chapter 3 was dedicated to structural modelling and here we described the causal mechanism of the delivery cost which involves socio-economic, environmental and biological indicators. Due to research and time limit we 'reduced' the causal mechanism of the cost. In future, could be created a research group

composed of medical doctor, social scientist, statistician and economist to estimate causal relations presented in operational framework (Figure 10).

With this research we attend to estimate the maternal smoking impact on the delivery cost, but as well to give some proposals to administration office of Emilia-Romagna region. In this context, following recommendations may be of interest:

- To include additional questions in the questionnaire corresponding CedAP data:
  - if women indicated that she continued to smoke during the pregnancy, add a question to evaluate the intensity (for example, number of cigarettes smoked per day);
  - add a question to assess information about alcohol consumption before and/or during pregnancy.
- The first cessation services (SCS) begun to operate at the end of the 90s and in 2012 there were 372 SCS in Italy. More than half of these centers are located in the northern Italy (56%, 44 SCS centers in Emilia Romagna). But still cessation probabilities among Italian smokers, particularly for those aged 30-59 years, have been very low and stalled (Pucchio et al. (2009); Carreras et al. (2012); Istituto Superiore Di Sanita (2012)). According to DOXA survey, the average age at which female starts to smoke decreased from 19 in 2003 to 18.2 in 2011. The policy makers may consider the development of appropriate cessation activities related to maternal smoking; the opportunities for health and medical professionals to learn more effective means of assisting pregnant women to stop smoking; and the implementation of educational programs against tobacco use in schools.

Bibliography

E.K. Adams, V.P. Miller, C. Ernst, B.K. Nishimura, C. Melvin, and R. Merritt. Neonatal health care costs related to smoking during pregnancy. *Health Economics*, 11:193–206, 2002.

C.A. Aligne and J.J. Stodddard. Tobacco and children. an economic evaluation of the medical effects of parental smoking. *Archives of Pediatrics and Adolescent Medicine*, 151(7):648–653, 1997.

P.N. Anandalakshmy, P.P. Talwar, K. Buckshee, and V. Hingorani. Demograpic,socio-economic and medical factors affecting maternal mortality - an indian experience. *Journal of Family Welfare*, 39(3):1–4, 1993.

L. Antonini and A. Pin. The italian road to fiscal federalism. *Italian Journal of Public Law*, 1:1–16, 2009.

P. Arabie and L. Hubert. Cluster analysis in marketing research. In R.P. Bagozzi, editor, *Advanced methods of marketing research*. Oxford,UK:Blackwell, 1994.

W. Arias and S. Viner-Brown. Maternal smoking and birth defects in rhode island. *Medicine and Health/Rhode Island*, 95(8):262–263, 2012.

R.W. Atkinson, I.M. Carey, A.J. Kent, T. van Staa, H.R. Anderson, and D.G. Cook. Long-term exposure to outdoor air pollution and incidence of cardiovascular diseases. *Epidemiology*, 24(1):44–53, 2013.

P.C. Austin. The performance of different propensity score methods for estimating relative risks. *Journal of Clinical Epidemiology*, 61:537–545, 2008.

V. Balinskaite. Short-term costs of smoking during pregnancy: geometric multidimensional approach. In *SAS Global Forum 2013 Proceedings*, 2013.

K.M. Birdsall, S. Vyas, N. Khazaezadeh, and E. Oteng-Ntim. Maternal obesity:a review of interventions. *International Journal of Clinical Practice*, 63:494–507, 2009.

E. Bjerkaas, R. Parajuli, E. Weiderpass, A. Engeland, G. Maskarinec, R. Selmer, and I.T. Gram. Smoking duration before first childbirth: an emerging risk factor for breast cancer? results from 302,865 norwegian women. *Cancer Causes and Control*, 2013.

M. Black, S. Bhattacharya, T. Fairley, D. M. Campbel, and A. Shetty. Outcomes of pregnancy in women using illegal drugs and in women who smoke cigarettes. *Acta Obstetricia et Gynecologica*, 92:47–52, 2013.

M.L. Block and L. Calderon-Garciduenas. Air pollution: mechanism of neuroinflammation and cns disease. *Trends in Neurosciences*, 32:506–516, 2009.

M. Branis and M. Linhartova. Association between unemployment, income, education level, population size and air pollution in czech cities: evidence for enviromental inequitity? a pilot national scale analysis. *Health and Place*, 18(5): 1110–1114, 2012.

R. Bukowski, F.D. Malone, F.T. Porter, D.A. Nyberg, C.H. Comstock, and G.D.V. Hankins. Preconceptional folate supplementation and the risk of spontanuous preterm birth: a cohort study. *PLOS Medicine*, 6(5):1–11, 2009.

L. Calderon-Garciduenas, W. Reed, R.R. Maronpot, C. Henriquez-Roldan, R. Delgado-Chavez, and A. Calderon-Garciduenas. Brain inflammation and alzheimer's-like pathology in individuals exposed to severe air pollution. *Toxicologic Pathology*, 32:650–658, 2004.

L. Calderon-Garciduenas, M. Franco-Lira, C. Henriquez-Roland, N. Osnaya, A. Gonzalez-Maciel, and R. Reynoso-Robles. Urban air pollution: influences on olfactory function and pathology in exposed children and young adults. *Experimental and Toxicologic Pathology*, 62:91–102, 2010.

L. Calderon-Garciduenas, M. Kavanaugh, M. Block, A. D'Angiulli, R. Delgado-Chavez, and R. Torres-Jardon. Neuroinflammation, hyperphosphorylated tau, diffuse amyloid plaques, and down-regulation of the cellular prion protein in air pollution exposed children and young adults, 2012.

F. Camillo and I. D'Attoma. A new data mining approach to estimate causal effects of policy interventions. *Expert Systems with Applications*, 37(1):171–181, 2010.

F. Camillo and I. D'Attoma. %gi sas marco: a sas macro for measuring and testing global imbalance of covariates within subgroups. *Journal of Statistical Software*, 51:1–21, 2012.

K.P. Cantor, P.A. Stewart, L.A. Brinton, and M. Dosemeci. Occupational exposures and female breast cancer mortality in the united states. *Journal of Occupational and Environmental Medicine*, 37(3):336–348, 1995.

I.M. Carey, R.W. Atkinson, A.J. Kent, T. van Staa, D.G. Cook, and H.R. Anderson. Mortality associations with long-term exposure to outdoor air pollution in a national english cohort. *American Journal of Respiratory and Critical Care*, 187 (11):1226–1233, 2013.

G. Carreras, S. Gallus, L. Lannucci, and G. Gorini. Estimating the probabilities of making a smoking quit attempt in italy: stall in smoking cessation levels, 1986-2009. *BMC Public Health*, 12:1–6, 2012.

P. Catalano, N. Drago, and S. Amini. Maternal carbohydrate metabolism and its relationship to fetal growth and body composition. *American Journal of Obstetrics and Gynecology*, 172:1464–1470, 1995.

J.M. Catov, L.M. Bodnar, J. Olson, S. Olsen, and E.A. Nohr. Periconceptional multivitamin use and risk of preterm or small-to-gestational-age births in the danish national birth cohort. *American Journal of Clinical Nutrition*, 94:906–912, 2011.

Center fo Desease Control and Prevention [CDC]. Sociodemographic and behavioral characteristics associated with alcohol consumption during pregnancy. *U.S. Morbidity and Mortality Weekly Report*, 14:261–264, 1995.

G. Cesaroni, C. Badaloni, V. Romano, E. Donato, C.A. Perucci, and F. Forastiere. Socioeconomic position and health status of people who live near busy roads: the rome longitudinal study (rols). *Environmental Health*, 9(41):1–12, 2010.

G. Chang, T.K. McNamara, E.J. Orav, and L. Wilkins-Haug. Alcohol use by pregnant women: partners, knowledge, and other predictors. *Jounal of Studies on Alcohol*, 67:245–251, 2006.

S. Cnattingius. The epidemiolody of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine and Tobacco Research*, 6(2):125–140, 2006.

A.J. Cohen. Outdoor air pollution and lung cancer. *Environmental Health Perspectives*, 108(4):743–750, 2000.

J.G. Collins, M.A. Smith, R.R. Anorld, and S. Offenbacher. Effects of escherichia coli and porphyromonas gingivalis lipopolysac-charide on pregnancy outcome in the golden hamster. *Infection and Immunity*, 62:4652–4655, 1995a.

J.G. Collins, H.W. III Windley, R.R. Anorld, and S. Offenbacher. Effects of porphyromonas gingivalis infection on inflammatory mediator response and pregnancy outcome in hamsters. *Infection and Immunity*, 62:4356–4361, 1995b.

M. Cotta and L. Verzichelli. *Political Institutions in Italy.* Oxford: Oxford University Press, 2007.

R.G. Cowell, P. David, S.L. Lauritzen, and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems.* Springer, 1999.

I. Santa Cruz, D. Herrera, C. Martin, A. Herrero, and M. Sanz. Association between periodontal status and pre-term and/or low-birth weight in spain: clinical and microbiological parameters. *Journal of Periodontal Research*, 40(4):339–345, 2012.

J.M. Cundiff, B.N. Uchino, T.W. Smith, and W. Birmingham. Socioeconomic status and health: education and income are independent and joint predictors of ambulatory blood pressure. *Journal of Behavioral Medicine*, pages 1–8, 2013.

A.E. Czeizel and I Dudas. Prevention of the first occurence of neural-tube defects by periconceptional vitamin supplementation. *New England Journal of Medicine*, 327:1832–1835, 1992.

I. D'Attoma. *A partial dependence factorial analysis to deal with selection bias in observational studies.* PhD thesis, University of Bologna, 2009.

M.C. England, A. Benjamin, and H.A. Abenhaim. Increased risk of preterm premature rupture of membranes at early gestational ages among maternal cigarette smokers. *American Journal of Perinatology*, Jan 17, 2013.

B. Escofier. Analyse des correspondances multiples conditionelle. In E. Diday, editor, *Data Analysis and Informatics.* North Holland, Amsterdam: Elsevier Science, 1988.

M. Faila, A. Grappiolo, S. Carugo, I. Calchera, C. Giannattasio, and G. Mancia. Effects of cigarette smoking on carotid and radial artery distensibility. *Journal of Hypertension*, 15:1659–1664, 1997.

A. Fernández-Somoano, G. Hoek, and A. Tardon. Relationship between area-level socioeconomic characteristics and outdoor no2 concentrations in rural and urban areas of northern spain. *BMC Public Health*, 13(71):1–19, 2013.

G. Fiorentini, M. Lippi Bruni, and C. Ugolini. Health system and health reforms in europe: the case of italy. *Intereconomics-Review of European Economic Policy*, 43(4):205–212, 2008.

A.R. Fisher. *The design of experiments*. Edinburgh: Oliver and Boyde, 1935.

Formez. *I Sistemi Di Governance Dei Servizi Sanitari Regionali*. Rome: Formez, 2007.

L. Foss, H.M. Gravseth, P. Kristensen, B. Claussen, I.S. Mehlum, S. Knardahl, and K. Skyberg. The impact of workplace risk factors on long-term musculoskeletal sickness absence : a registry-based 5-year follow-up from the oslo health study. *Journal of Occupational and Environmental Medicine*, 53(12):1478–1482, 2011.

T Gasim, F.E. Al Jama, M.S. Rahman, and J. Rahman. Multiple repeat cesarean sections: operative difficulties, maternal complications and outcome. *The Journal of Reproductive Medicine*, 58:312–318, 2013.

M.M.H.J. Van Gelder, J. Reefhuis, A.R. Caton, M.M. Werler, C.M. Druschel, and N. Roeleveld. Characteristics of pregnant illicit drug users and associations between cannabis use and perinatal outcome in a population-based study. *Drug and Alcohol Dependence*, 109:243–247, 2010.

C. Godfrey, K.E. Pickett, S. Parrott, N.D. Mdege, and D. Eapen. Estimating the costs to the nhs of smoking in pregnancy for pregnant women and infants. Technical report, Department of Health Science, The University of York, 2010.

H. Graham and G. Der. Influence on women's smoking status: the contribution of socioeconomic status in adolescence and adulthood. *European Journal of Public Health*, 9(2):137–141, 1999.

H. Graham, B. Francis, H. M. Inskip, and J. Harman. Socioeconomic lifecourse influences on women's smoking status in early adulthood. *Journal of Epidemiology and Community Health*, 60:228–233, 2006.

M.J. Greenacre. *Theory and Applications of Correspondence Analysis*. Harcourt Brace Jovanovich, 1984.

C.E. Grella and K. Lovinger. Gender differences in physical and mental health outcomes among an aging cohort of individuals with history of heroin dependence. *Addictive Behaviors*, 37:306–312, 2012.

X. Gu and P. R. Rosenbaum. Comparison of multivariate matching methods: structures, distances, and algoriths. *Journal of Computational and Graphcal Statistics*, 2:405–420, 1993.

A.N. Guimaraes, A. Silva-Mato, F.M. Siqueira, R.M. Cyrino, L.O.M. Cota, and F.O. Costa. Very low and low birth weight associated with maternal periodontitis. *Journal of Clinical Periodontology*, 39:1024–1031, 2012.

I. Hajo and J. Wildschut. socio-economic and demographic factors: age, parity, social class and ethnicity. In D.K. James, P.J. Steer, C.P. Weiner, and B. Gonik, editors, *High Risk Pregnancy*. W.B. Saunders, London, 1995.

G. Hardy, A. Benjamin, and H.A. Abenhaim. Effect of induced abortions on early preterm births and adverse perinatal outcomes. *Obstetrics*, 35(2):138–143, 2013.

J. J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1): 153–161, 1979.

J.J. Heckman and V.J. Hotz. Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training. *Journal of American Statistical Association*, 84(408):862–874, 1989.

J.J. Heckman and R. Robb. Alternative methods for evaluating the impact of interventions. *Journal of Econometrics*, 30:239–267, 1985.

G.T. Henry and J.H. McMillan. Performance data: Three comparison methods. *Evaluation Review*, 17:643–652, 1993.

P.W. Holland. Statistics and causal inference. *Journal of American Ctatistical Association*, 81:945–970, 1986.

P.W. Holland and D.B. Rubin. On lord's paradox. In H. Wainer and S. Messick, editors, *Principals of Modern Psychological Measurement*. Hillsdale, NJ: Lawrence Erlbaum, 1983.

I.A. Holma, K.M. Holma, T.K. Melartin, M. Ketokivi, and E.T. Isometsa. Depression adn smoking: a 5-year prospective study of patients with major depressive disorder. *Depression and Anxiety*, 2013.

M.E. Hughes and L.J. Waite. Marital biography and health at mid-life. *Journal of Health and Social Behavior*, 50:344–358, 2009.

International Agency for Research on Cancer. Vol. 83: tobacco smoke and involuntary smoking. Technical report, Lyon: IARC, 2004.

International Agency for Research on Cancer. Vol. 100e: a review of human carcinogens: personal habits and indoor combustions. Technical report, Lyon: IARC, 2012.

Istituto Superiore Di Sanita. Guida ai servizi territoriali per la cessazione dal fumo di tabacco. Technical report, Istituto Superiore Di Sanita, 2012.

I. Jagielska, A. Kazdepka-Ziemińska, M. Stankiewicz, and J. Kaźmierczak. Alcohol-woman,pregnancy and a newborn child. *Przegl Lek*, 69(10):1108–1110, 2012.

D.T. Jamison, J.G. Breman, and A.R. Measham. *Disease Control Priorities in Developing Countries. 2nd edition*. World Bank, Washington, 2006.

L.M. Jansson and M.L. Velez. Infants of drug-dependent mothers. *Pediatrics in Review*, 32(1):5–13, 2011.

M.A. Jessup, S.L. Dibble, and B.A. Cooper. Smoking and behavioral health of women. *Journal of Women's Health*, 21(7):783–791, 2012.

C. Jommi, E. Cantu, and E Anessi-Pessina. New funding arrangements in the italian national health service. *International Journal of Health Planning and Management*, 16(4):347–368, 2001.

S. Kafouri, G. Leonard, and M. Perron. Maternal cigarette smoking during pregnancy and cognitive performance in adolescence. *international Journal of Epidemiology*, 38:158–172, 2009.

R. Klementti, M. Gissler, M. Niinimaki, and E. Hemminki. Birth outcomes after induced abortion: a national wide register-based study of first births in finland. *Human Reproduction*, 27(11):3315–3320, 2012.

R. Kodgule and S. Salvi. Exposure to biomass smoke as a cause for airway disease in women and children. *Current Opinion in Alergy and Clinical Immunology*, 12 (1):82–90, 2012.

M.E. Kroll, F. Murphy, K. Pirie, G.K. Reeves, J. Green, and V. Beral. Alcohol drinking, tobacco smoking and subtypes of haematological malignancy in the uk million women study. *British Journal of Cancer*, 107:879–887, 2012.

A. Kumar, M. Basra, N. Begum, V. Rani, S. Prasad, A.K. Lamba, M.Verma, S. Agarwal, and S. Sharma. Association of maternal periodontal health with adverse pregnancy outcome. *The Journal of Obstretics and Gynaecology Research*, 39(1):40–45, 2012.

F. Labreche and M.S. Goldberg. Exposure to organic solvents and breast cancer in women: a hypothesis. *American Journal of Industrial Medicine*, 32:1–14, 1997.

F. Labreche, M.S. Goldberg, M.F. Valois, and L. Nadon. Postmenopausal breast cancer and occupational exposures. *Occupational and Environmental Medicine*, 67:263–269, 2010.

P.V.M. Lakshimi, N.K. Virdi, A. Sharma, J.P. Tripathy, K.R. Smith, M.N. Bates, and R. Kumar. Household air pollution and stillbirth in india: analysis of the dlhs-ii national survey. *Environmental Research*, 121:17–22, 2013.

T.T. Lao, D.S. Sahota, Y.K.Y. Cheng, L.W. Law, and T.Y. Leung. Meternal hepatitis b surface antigen status and incidence of pre-eclampsia. *Journal of Viral Hepatitis*, 20:343–349, 2013.

S. Lee. Propensity score adjusments as a weighting schime for volunteer panel web surveys. *Journal of Official Statistics*, 22(2):329–349, 2006.

J. Leonardi-Bee, A. Smyth, J. Britton, and T. Coleman. Environmental tobbaco smoke and fetal health: systematic review and meta-analysis. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, 93:351–361, 2008.

G. R. Leonardson, R. Loudenburg, and J. Struck. Factors predictive of alcohol use during pregnancy in three rural states. *Behavioral and Brain Functions*, 3(8):1–6, 2007.

G.R. Leonardson and R. Loudenburg. Risk factors for alcohol use during pregnancy in a multistate area. *Neurotoxicology and Teratology*, 25:651–658, 2003.

J.M. Lightwood, C.S. Phibbs, and S.A. Glantz. Short-term health and economic benefits of smoking cessation: low birth weight. *Pediatrics*, 104(6):1312–1320, 1999.

X. Liu, J. Du, G. Wang, Z. Chen, W. Wang, and Q. Xi. Effect of pre-pregnancy body mass index on adverse pregnancy outcome in the north of china. *Archives of Gynecology and Obstetrics*, 283:65–70, 2011.

M. Magadi, N. Madise, and I Diamond. Factors associated with unfavourable birth outcomes in kenya. *Journal of Biosocial Science*, 33(2):199–225, 2001.

M. Magadi, I. Diamond, N. Madise, and P. Smith. Pathways of the determinants of unfavourable birth outcomes in kenya. *Journal of Biosocial Sciences*, 36:153–176, 2004.

M.A. Magadi. *The determinants of poor maternal health care and adverse pregnancy outcomes in Kenya.* PhD thesis, University of Southampton, 1999.

W.G. Manning, E.B. Keeler, J.P. Newhouse, E.M. Sloss, and J. Wasserman. The taxes of sin. do smokers and drinkers pay their way? *JAMA*, 261(11):1604–1609, 1989.

L.T. Martin, M.J. McNamara, A.S. Milot, T. Halle, and E.C. Hair. The effects of father involment during pregnancy on receipt of prenatal care and maternal smoking. *Maternal and Child Health Journal*, 11:595–602, 2007.

F.J. Mateen and R.D. Brook. Air pollution as an emerging global risk factor for stroke. *JAMA*, 305:1240–1241, 2011.

J. McCarthy and D. Maine. A framework for analyzing the determinants of maternal mortality. *Studies in Family Planning*, 23(1):23–33, 1992.

J. McRobert. Smoking and its effects on the healing process of chronic wounds. *British Journal of Community Nursing*, 18(3):18–23, 2013.

MIGHRER. Migrants and health care: Responses by european regions (mighrer). Technical report, World Health Organization, Regional Office for Europe, Regions for Health Network, 2012.

D.P. Miller, K.F. Villa, S.L. Hogue, and D. Sivapathasundaram. Birth and first-year costs for mothers and infants attributable to maternal smoking. *Nicotine and Tobacco Research*, 3(1):25–35, 2001.

Ministerio della Sanita. Nomenclatore tariffario delle prestazioni ambulatoriali dm 7.11.91. Technical report, Ministerio della Sanita, 1992.

Ministerio della Sanita. Aggiornamento delle tariffe delle prestazioni di assistenza ospedaliera di cui al dm 14.12.1994; dm 30.697. Technical report, Ministerio della Sanita, 1997.

Y. Miyake, K. Tanaka, and M. Arakawa. Employment, income, and education and prevalence of depressive symptoms during pregnancy: the kyushu okinawa maternal and child health study. *BMC Psychiatry*, 12:117–122, 2012.

W.N. Mohamed, I. Diamond, and P.W.F. Smith. The determinants of infant mortality in malaysia: A graphical chain modelling approach. *Journal of Royal Statistical Society*, 161(3):349–366, 1998.

M. Morales-Suarez-Varela, L. Kaerlev, J.K. Zhu, A. Llopis-Gonzalez, N. Gimeno-Clemente, E.A. Nohr, J.P. Bonde, and J. Olsen. Risk of infection and adverse outcomes among pregnant working women in selected occupational groups: a study in the danish national birth cohort. *Environmental Health*, 9:70–80, 2010.

S.L. Morgan and D.J. Harding. Matching estimators of causal effects: prospects and pitfalls in theory and practice. *Sociological Methods and Research*, 35(1): 3–60, 2006.

C. Morton, J.T. Le, L. Shahbandar, C. Hammond, E.A. Murphy, and K.L.Kirschner. Pregnancy outcomes of women with physical disabilities: a matched cohort study. *PM&R*, 5:90–98, 2013.

W.H. Mosley and L.C. Chen. Analytical frameworks for the study of child survival in developing countries. *Population and Development Review*, 10:25–48, 1984.

M. Mouchart, F. Russo, and G. Wunsch. Structural modelling, exogeneity and causality. In H. Engelhardt, H.P. Kohler, and A. Prskawetz, editors, *Causal Analysis in Population Studies:concepts, methods, applications*. Dordrecht: Springer, 2009.

M. Mouchart, F. Russo, and G. Wunsch. Inferring causal relations by modelling structures. *Statistica*, 4, 2010.

E.J.H. Mulder, P.G. Robles de Medina, A.C. Huizink, B.R.H. Van den Bergh, J.K. Buitelaar, and G.H.A. Visser. Prenatal maternal stress: effects on pregnancy and (unborn) child. *Early Human Development*, 70:3–14, 2002.

J.S. Neyman. Statistical problems in agricultural experimentation with discussion. *Journal of the Royal Statistics Society, Series B*, 2:107–180, 1935.

E.A. Nohr, B.H. Bech, M.J. Davies, M. Frydenberg, T.B. Henriksen, and J. Olsen. Pregnancy obesity and fetal death. *Obstetrics&Gynecology*, 106(2):250–259, 2005.

K. Noonan, N.E. Reichman, H. Corman, and D. Dave. Prenatal drug use and production of infant health. *Health Economics*, 16(4):361–384, 2007.

S.L.T. Normand, M.B. Landrum, E. Guadagnoli, J.Z. Ayanian, T.J. Ryan, P.D. Cleary, and B.J. McNeil. Validating recommendations for coronary angiography following an acute myocardial infarction in the elderly: a macthed analysis using propensity scores. *Journal of Clinical Epidemiology*, 54:387–398, 2001.

Office of Technology Assessment. Healthy children: investing in the future. Technical report, U.S. Government Printing Office, US Congress, Washington, DC, 1988.

C.M. O'Leary, N. Nassar, J.J. Kurinczuk, N. de Klerk, E. Geelhoed, E.J. Elliott, and C. Bower. Prenatal alcohol exposure and risk of birth defects. *Pediatrics*, 126 (4):843–850, 2010.

D. Olsson, I. Mogren, and B. Forsberg. Air pollution exposure in early pregnancy and adverse pregnancy outcomes: a register-based cohort study. *BMJ*, 3:1–8, 2013.

H. Orru, A. Laukaitiene, and I. Zurlyte. Particulate air pollution and its impact on health in vilnius and kaunas. *Medicina*, 48(9):472–477, 2012.

G. Oster, T.E. Delea, and G.A. Colditz. Maternal smoking during pregnancy and expenditures on neonatal health care. *American Journal of Preventive Medicine*, 4(4):216–219, 1988.

E. Ota, M. Haruna, M. Suzuki, D.D. Anh, L.H. Tho, and N.T.T. Tam. Maternal body mass index and gestational weight gain and their association with perinatal outcomes in vietnam. *Bulletin of the World Health Organization 2011*, 89:127–136, 2011.

A.M. Padula, K. Mortimer, A. Hubbard, F. Lurmann, M. Jerrett, and I.B. Tager. Exposure to traffic-related air pollution during pregnancy and term low birth weight: estimation of causal association in a semiparametric model. *American Journal of Epidemiology*, 176(9):815–824, 2012.

J.R. Palmer and L. Rosenberg. Cigarette smoking and the risk of breast cancer. *Epidemiologic Review*, 15:145–156, 1993.

J. Pearl. *Causality: models, reasoning, and inference*. Cambridge:Cambridge University Press, 2000.

L.R. Peck. Using cluster analysis in program evaluation. *Evaluation Review*, 29(2): 178–196, 2005.

L.R. Peck. What are the effects of welfare sanction policies? *American Journal of Evaluation*, 28(3):256–274, 2007.

S. Petrou, P. Cooper, L. Murray, and L.L. Davidson. Economic costs of post-natal depression in a high-risk british cohort. *The British Journal of Psychiatry*, 181: 505–512, 2002.

D. J. Pevalin, T.J. Wade, A. Brannigan, and R. Sauve. Beyond biology: the social context of prenatal behaviour and birth outcomes. *Soz.-Präventivmed*, 46:233–239, 2001.

N. Phithakwatchara and V. Titapant. The effect of pre-pregnancy weight on delivery outcome and birth weight in potential diabetic patients with normal screening for

gestational diabetes mellitus in siriraj hospital. *Journal of the Medical Association of Thailand*, 90:229–236, 2007.

A. Poli, F. Marangoni, A. Avogaro, G. Barba, S. Bellentani, M. Bucci, R. Cmabieri, A.L. Catapano, and S. Costanzo. Moderate alcohol use and health: a consensus paper. *Nutrition, Metabolism and Cardiovascular Diseases*, 13:1–18, 2013.

D.P. Pope, V. Mishra, L. Thomson, A.R. Siddiqui, E.A. Rehfuess, M. Weber, and N.G. Bruce. Risk of low birth weight and stillbirth associated with indoor air pollution from solid fuel use in developing countries. *Epidemiologic Reviews*, 32: 70–81, 2010.

M.C. Power, M.G. Weisskopf, S.E. Alexeeff, B.A. Coull, A. 3rd Spiro, and J. Schwartz. Traffic-related air pollution and cognitive function in a cohort of older men. *Environmental Health Perspectives*, 119:682–687, 2011.

A. Di Pucchio, E. Pizzi, G. Carosi, M. Mazzolo, D. Mattioli, R. Pacifici, and S. Pichini. National survey of the smoking cessation services in italy. *International Journal of Environmental Research and Public Health*, 6:915–926, 2009.

R.D. Putnam. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press, 1993.

K. Raatikainen, N. Heiskanen, and S. Heinonen. Does unemployment in family affect pregnancy outcome in conditions of high quality maternal care? *BMC Public Health*, 6:46–53, 2006.

U. Ranft, T. Schikowski, D. Sugiri, J. Krutmann, and U. Kramer. Long-term exposure to traffic-related particulate matter impairs cognitive function in the elderly. *Environmental Research*, 109:1004–1011, 2009.

P. Reynolds, S. Hurley, D.E. Goldberg, H. Anton-Culver, and L. Berstein. Active smoking, household passive smoking, and breast cancer: evidence from the california teachers study. *Journal of the National Cancer Institute*, 96:29–37, 2004.

A.G. Ronnenberg, X. Wang, H. Xing, C. Chen, and D. Chen. Low preconception body mass index is associated with birth outcome in a prospective cohort of chinese women. *Journal of Nutrition*, 133:3449–3455, 2003.

P.R. Rosenbaum. From association to causation in observational studies: the role of tests of strongly ignorable treatment assignment. *Journal of American Statistical Association*, 79:41–48, 1984a.

P.R. Rosenbaum. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, ser. A*, 147:656–666, 1984b.

P.R. Rosenbaum. Conditional permutation tests and the propensity score in observational studies. *Journal of the American Statistical Association*, 79:565–574, 1984c.

P.R. Rosenbaum. The role of a second control group in an observational study. *Statistical Science*, 2:292–316, 1987.

P.R. Rosenbaum. Optimal matching for observational studies. *Journal of the American Statistical Association*, 84:1024–1032, 1989.

P.R. Rosenbaum. *Observational studies*. New York: Springer, 2002.

P.R. Rosenbaum and D.B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55, 1983a.

P.R. Rosenbaum and D.B. Rubin. Assesing sensitivity to an unobservable binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, ser. B*, 45:212–218, 1983b.

P.R. Rosenbaum and D.B. Rubin. Reducing bias in observational studies using subclasification on the propensity score. *Journal of the American Statistical Association*, 79:516–524, 1984b.

P.R. Rosenbaum and D.B. Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, 39:33–38, 1985a.

P.R. Rosenbaum and D.B. Rubin. The bias due to incomplete matching. *Biometrics*, 41:103–116, 1985b.

D. B. Rubin. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74:318–328, 1979.

D. B. Rubin. Bias reduction using mahalanobis' metric matching. *Biometrics*, 36: 1980, 1980.

D. B. Rubin and N. Thomas. Affinely invariant matching methods with ellipsoidal distributions. *Annals of Statistics*, 20:1079–1093, 1992a.

D. B. Rubin and N. Thomas. Characterizing the effect of matching using linear propensity score methods with nomal covariates. *Biometrika*, 79:797–809, 1992b.

D. B. Rubin and N. Thomas. Matching using estimated propensity scores: relating theory to practice. *Biometrics*, 52:249–264, 1996.

D. B. Rubin and N. Thomas. Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, 95:573–585, 2000.

D.B. Rubin. Matching to remove bias in observational studies. *Biometrics*, 29: 159–183, 1973a.

D.B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational psychology*, 66:688–701, 1974.

D.B. Rubin. Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2:1–26, 1977.

D.B. Rubin. Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2:169–188, 2001.

D.B. Rubin and R.P.Waterman. Estimating the causal effects of marketing interventions using propensity score methodology. *Statistical Science*, 21(2):206–222, 2006.

F. Russo. *Causality and Causal Modelling in the Social Sciences: Measuring Variations*. Springer, 2009.

H.M. Salihu, A. Dunlop, M. Hedayatzadeh, A.P. Alio, R.S. Kirby, and G.R. Alexander. Extreme obesity and risk of stillbirth among black and white gravidas. *Obstetrics&Gynecology*, 110(3):552–557, 2007.

G. Salmasi, R. Grady, J. Jones, and S.D. McDonald. Enviromental tobacco smoke exposure and perinatal outcomes: a systematic review and meta-analyses. *Acta Obstetricia et Gynecologica*, 89:423–441, 2010.

K. Sassalos, C. Vlachopoulos, N. Alexopoulos, T. Gialernios, K. Aznaouridis, and C. Stefanadis. The acute and chronic effect of cigarette smoking on the elastic properties of the ascending aorta in healthy male subject. *The Hellenic Journal of Cardiology*, 47:263–268, 2006.

P.S. Shah and J. Zao. Induced termination of pregnancy and low birth-weight and preterm birth: a systematic review and meta-analyses. *An International Journal of Obstetrics and Gynaecology*, 116(11):1425–1442, 2009.

W.J. Simpson. A preliminary report on cigarette smoking and the incidence of prematurity. *Obsterical and Gynecological Survey*, 12(6):868–869, 1957.

K.R. Smith. *Biofuels, Air Pollution, and Health: A Global Review*. Plenum Press, New York, 1987.

K.R. Smith, S. Mehta, and M. Feuz. Indoor air pollution from household use of solid fuels. In M. Ezzati, A.D. Lopez, and A. Rodgers, editors, *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Geneva, Switzerland: World Health Organization, 2004.

S.M. Snedeker. Chemical exposures in the workplace: effect on breast cancer risk among women. *AAOHN*, 54(6):280–281, 2006.

K. Stansberg-Larsen, N. Rod Nielsen, A.M. Nybo Andersen, J. Olsen, and M. Gronbaek. Characteristics of women who binge drink before and after they become aware of their pregnancy. *European Journal of Epidemiolody*, 23:565–572, 2006.

C. Stefanadis, E. Tsiamis, and C. Vlachopoulos. Unfavorable effect of smoking on the elastic properties of the human aorta. *Circulation*, 97:31–38, 1997.

S. Surdu, E.F. Fitzgerald, M.S. Bloom, F.P. Boscoe, D.O. Carpenter, R.F. Haase, and E Gurzau. Occupational exposure to arsenic and risk of non-melonoma skin cancer in a multinational european study. *International Journal of Cancer*, 2013.

S. Suttajit, P. Kittirattanapaiboon, B. Junsirimongkol, S. Likhitsathian, and M. Srisurapanont. Risks of major depressive disorder and anxiety disorders among thais with alcohol use disorders and illicit drug use: findings from the 2008 thai national mental health survey. *Addictive Behaviors*, 37:1395–1399, 2012.

P.D. Terry and T. Rohan. Cigarette smoking and the risk of breast cancer in women:a review of the literature. *Cancer Epidemiology, Biomarkers and Prevention*, 11: 953–971, 2002.

V.T. Tong, J.R. Jones, P.M. Dietz, D. D'Angelo, and J.M. Bombard. Trends in smoking before, during, and after pregnancy - pregnancy risk assessment monitoring system (prams), united states, 31 sites, 2000-2005. *MMWR Surveill Summ*, 58:1–29, 2009.

U.S. Department of Health and Human Services. The health consequences of smoking: a report of the surgeon general. Technical report, Washington DC: U.S. Department of Health and Human Services, US Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office of Smoking and Health, 2004.

U.S. Department of Health and Human Services. Healthy people 2010. Technical report, U.S. Department of Health and Human Services, 2010.

P.J. Villeneuve, L. Chen, D. Stieb, and B.H. Rowe. Associations between outdoor air pollution and emergency department visits for stroke in edmonton, canada. *European Journal of Epidemiology*, 21:689–700, 2006.

H.E. Volk, I. Hertz-Picciotto, L. Delwiche, F. Lermann, and R. McConnell. Residential proximity for freeways and autism in the charge study. *Environmental Health Prospectives*, 119:873–877, 2011.

M. Vrijheid and D. Martinez. Socioeconomic status and exposure to multiple enviromental pollutants during pregnancy: evidence for environmental inequility? *Journal of Epidemiology and Community Health*, 66:106–113, 2012.

J. Weuve, R.C. Puett, J. Schwartz, J.D. Yanosky, F. Laden, and F. Grodstein. Exposure to particulate air pollution and cognitive decline in older women. *Archives of Internal Medicine*, 172:219–227, 2012.

G. Whitlock, S. MacMahon, S.V. Hoorn, P. Davis, R. Jackson, and R. Norton. Association of environmental tobbaco smoke exposure with socioeconomic status in a population of 7725 new zealanders. *Tobbaco Control*, 7:276–280, 1998.

W.J. Graham and S.F. Murray. A question of survivla? review of safe motherhood in kenya. Technical report, Ministry of Health [Kenya], Division of Primary Health Care, Kenya, 1997.

A.E. Wyse, V. Keesler, and B. Schneider. Assessing the effects of small school size on mathematics achievements: a propensity score-matching approach. *Teachers College Record*, 110:1879–1900, 2008.

F. Xue, W.C. Willett, B.A. Rosner, S.E. Hankinson, and K.B. Michels. Cigarette smoking and the incidence of breast cancer. *Archives of Internal Medicine*, 171: 125–133, 2011.

Y. Ye and L.A. Kaskutas. Using propensity scores to adjust for selection bias when assessing the effectiveness of alcoholics anonymous in observational studies. *Drug and Alcohol Dependence*, 104:56–64, 2009.

A. Zajacova, J.B. Dowd, and A.E. Aiello. Socioeconomic and race/ethnic patterns in persistent infection burden among u.s. adults. *J Gerontol A Biol Sci Med Sci*, 64(2):272–279, 2009.

Z. Zhang and M.D. Hayward. Gender, the marital life course, and cardiovascular disease in late midlife. *Journal of Marriage and Family*, 68(3):639–657, 2006.

!TEX encoding = UTF-8

Appendix A

Table 1: Descriptive statistics (frequencies) of the selected indicators

Variables	Freq (%)	Variables	Freq (%)
Mother:		**Number of control visits**	
Maternal age		0-5	22.06
<25	11.20	6	22
25-29	22.36	7	21.55
30-34	34.67	8	18.32
35-39	25.56	9+	16.07
40+	6.21	**Number of pregnancy ultrasound**	
Marital status		0	1.22
single	27.59	1-3	34.68
married	65.07	4+	64.10
separated	1.28	**Amniocentesis**	
divorced	0.94	yes	17.62
widowed	0.14	no	81.84
not stated	4.97	n/a	0.54
Nationality		**Chorionic villus sampling**	
Italian	72.01	yes	7.48
other EU citizen	4.09	no	91.9
citizen of LMLIC[1]	14.28	n/a	0.62
other	9.63	**Fetoscopy/Funiculocentesis**	
Place of birth		yes	0.47
north-west Italy	3.57	no	98.67
north-east Italy	48.44	n/a	0.87
central Italy	2.18	**Type of hospital**	
south Italy	10.22	public type A	67.79
islands (Italy)	3.07	public type B	30.93
abroad	32.03	private	1.28

Continued on next page

[1]Low middle and low income countries according to UN

Table 1 – *Continued from previous page*

Variables	Freq (%)	Variables	Freq (%)
n/a	0.5	**Method of labour**	
Education level		spontenous	59.82
primary or no education	7.22	induce	21.05
secondary	25.56	without	19.13
high school	42.6	**Reason of induction**	
terniary	24.62	prolonged pregnancy	25.87
Current occupation status		oligodramnios	15.16
occupied	64.84	PROM	26.06
unemployed	4.31	maternal pathology	15.38
in search of first job	0.11	fetal pathology	7.57
student	1.28	n/a	9.96
housewife	23.81	**Type of induction**	
other	0.07	prostaglandin	75.65
n/a	5.58	oxytocin	21.98
Profession		other drug	0.31
self-employed	12.49	amnioressi	2.05
manager or director	2.42	other mechanical method	-
employee	53.06	**Anti-pain method used in labour**	
worker	25.53	no method used	38.82
other dependent employer	6.06	epidural analgesia	7.77
n/a	0.44	other type of pharmacological analgesia	1.26
Sector of employment		non-pharmacological method	34.47
agriculture,hunting and fishing	1.94	n/a	17.67
industry	17.71	**Method of delivery**	
services	31.29	vaginal (non-assisted)	67.35
public administration	11.62	vaginal with use of forceps	0.12
other private sector	35.49	vaginal with use of ventouse	3.68
n/a	1.95	cesarean	28.84
Father:		**Presents during delivery:**	
Paternal age		**Pediatrician/monatologist**	
<25	2.82	yes	51.04
25-29	11.52	no	46.28
30-34	25.95	n/a	2.68
35-39	29.30	**Anesthetist**	
40+	21.05	yes	60.86
n/a	9.35	no	39.14
Nationality		n/a	0.01

Continued on next page

Table 1 – *Continued from previous page*

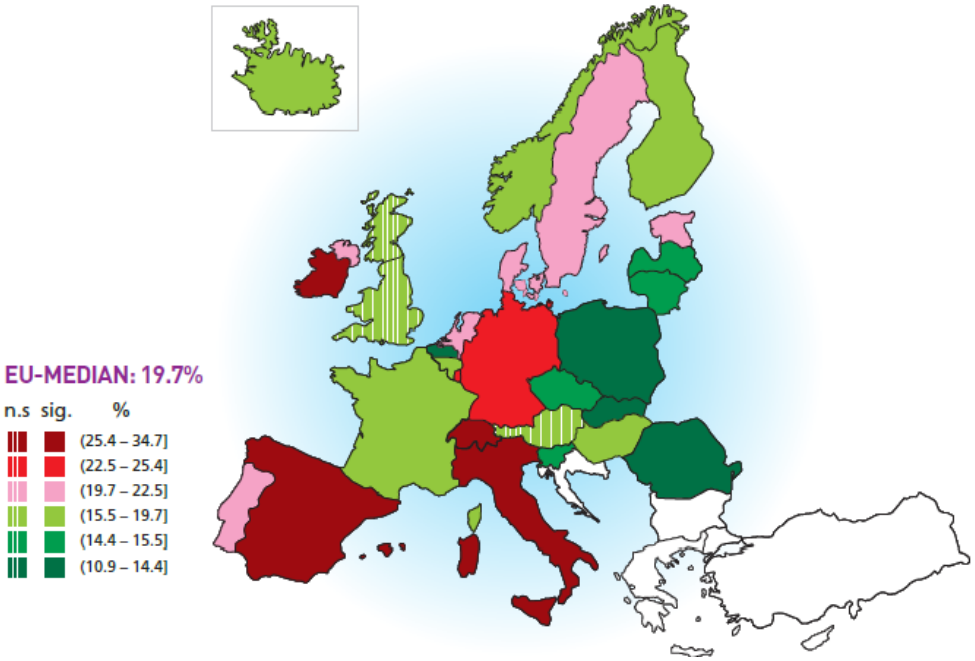
Variables	Freq (%)	Variables	Freq (%)
Italian	63.31	**Neonatal Nurse**	
other EU citizen	2.66	yes	43.35
citizen of LMLIC	12.91	no	53.74
other	8.14	n/a	2.91
n/a	12.98	**Smoking**	
Education level		stopped smoking in the last 5 years before pregnancy	3.93
primary or no education	3.12	stooped smoking at the beginning of pregnancy	4.85
secondary	32.06	was smoking before and during pregnancy	6.64
high school	39.22	non-smoker	84.58
tertiary	16.40	**Fetal position before birth**	
n/a	9.2	facing down 1	94.93
Current occupation status		facing down 2	4.26
occupied	88.08	facing upward 1	0.03
unemployed	3.14	facing upward 2	0.33
in search of first job	0.03	feet first	0.12
student	0.17	lying sideways	0.33
housewife	0.01	**Weight at birth**	
other	0.14	<1500	0.81
n/a	8.43	1500-2499	4.73
Profession		2500-3499	59.96
self-employed	24.42	3500+	34.5
manager or director	3.17	**Presents of malformation**	
employee	24.61	yes	1.07
worker	44.07	no	98.93
other dependent employer	3.25	**Apgar score**	
n/a	0.48	10-7	99.52
Sector of employment		6-4	0.4
agriculture, hunting and fishing	3.66	3-0	0.08
industry	34.59	**Type of resuscitation**	
services	19.85	manual ventilation	1.54
public administration	5.35	intubation	0.88
other private sector	34.57	not required	97.58
n/a	1.97		
Number of previous:			
live births		spontaneous abortion	
0	51.98	0	81.39

Continued on next page

Table 1 – *Continued from previous page*

Variables	Freq (%)	Variables	Freq (%)
1	36.97	1	14.57
2	8.3	2	3.04
3+	2.75	3+	1
still birth		induce abortion	
1+	0.57	0	92.16
0	99.43	1	6.07
cesarean		2	1.28
0	88.5	3+	0.49
1	9.77	**Hospitalization during pregnancy**	
2	1.49	yes	6.95
3+	0.24	no	92.30
		n/a	0.75

Figure 1: Mothers aged 35 years and above as a percentage of all pregnancies with known maternal age in 2010[a]



[a]Source:European Perinatal Health Report 2010

Figure 2: Percent of women, by age group and nationality

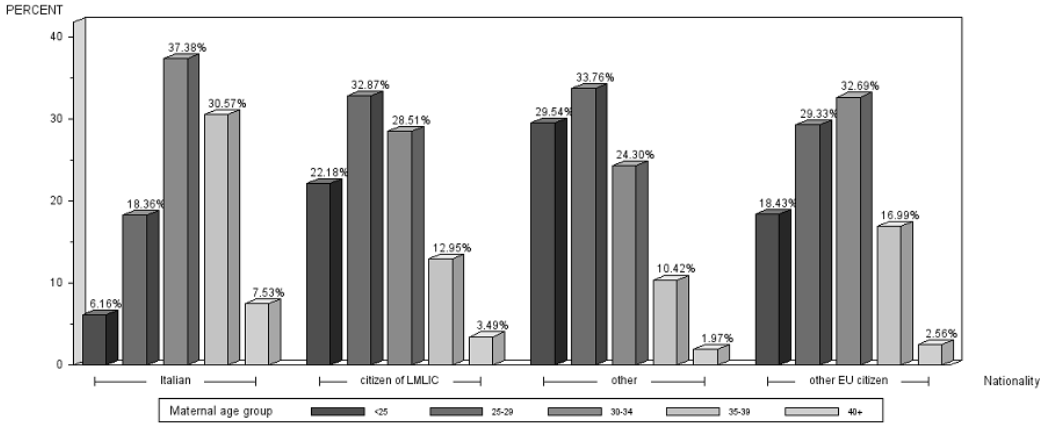


Figure 3: Percent of women, by education and nationality

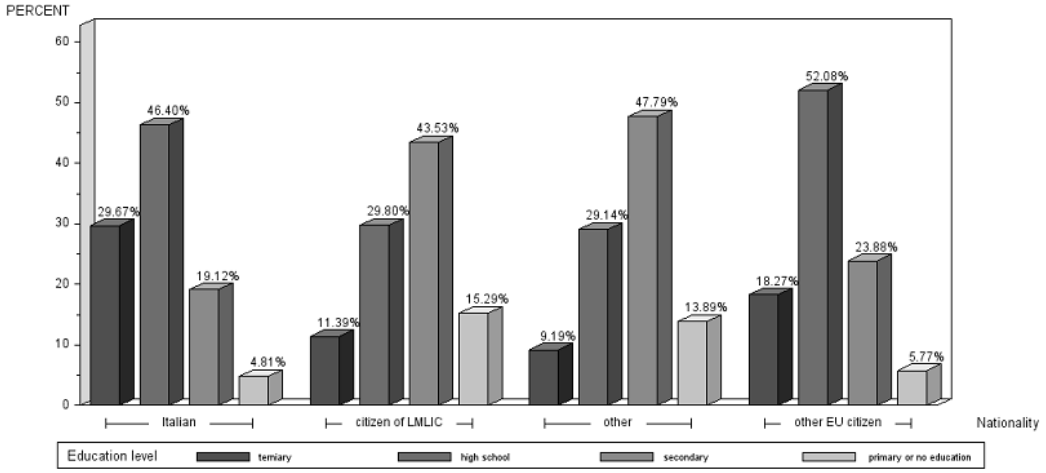


Figure 4: Percent of women, by education and smoking status

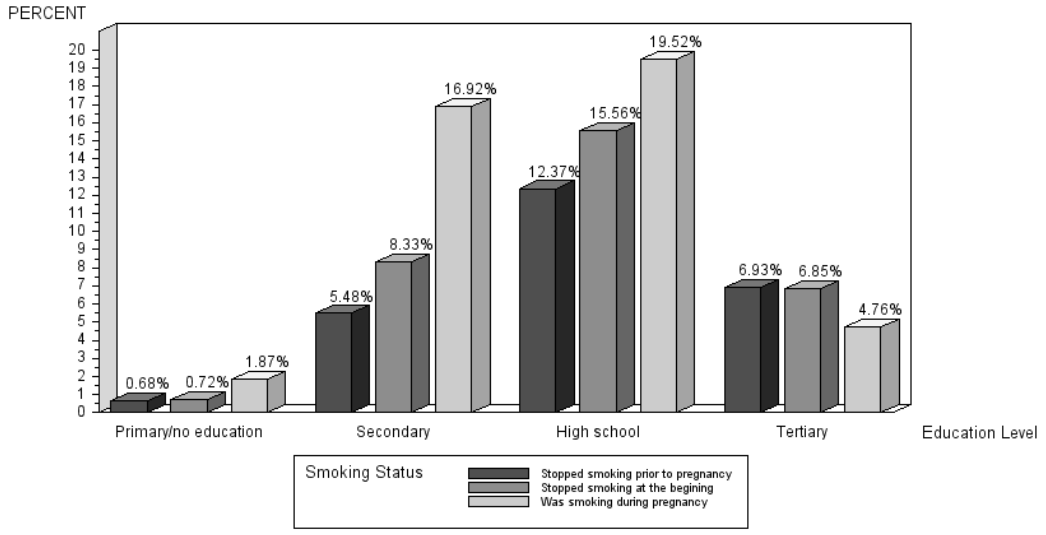
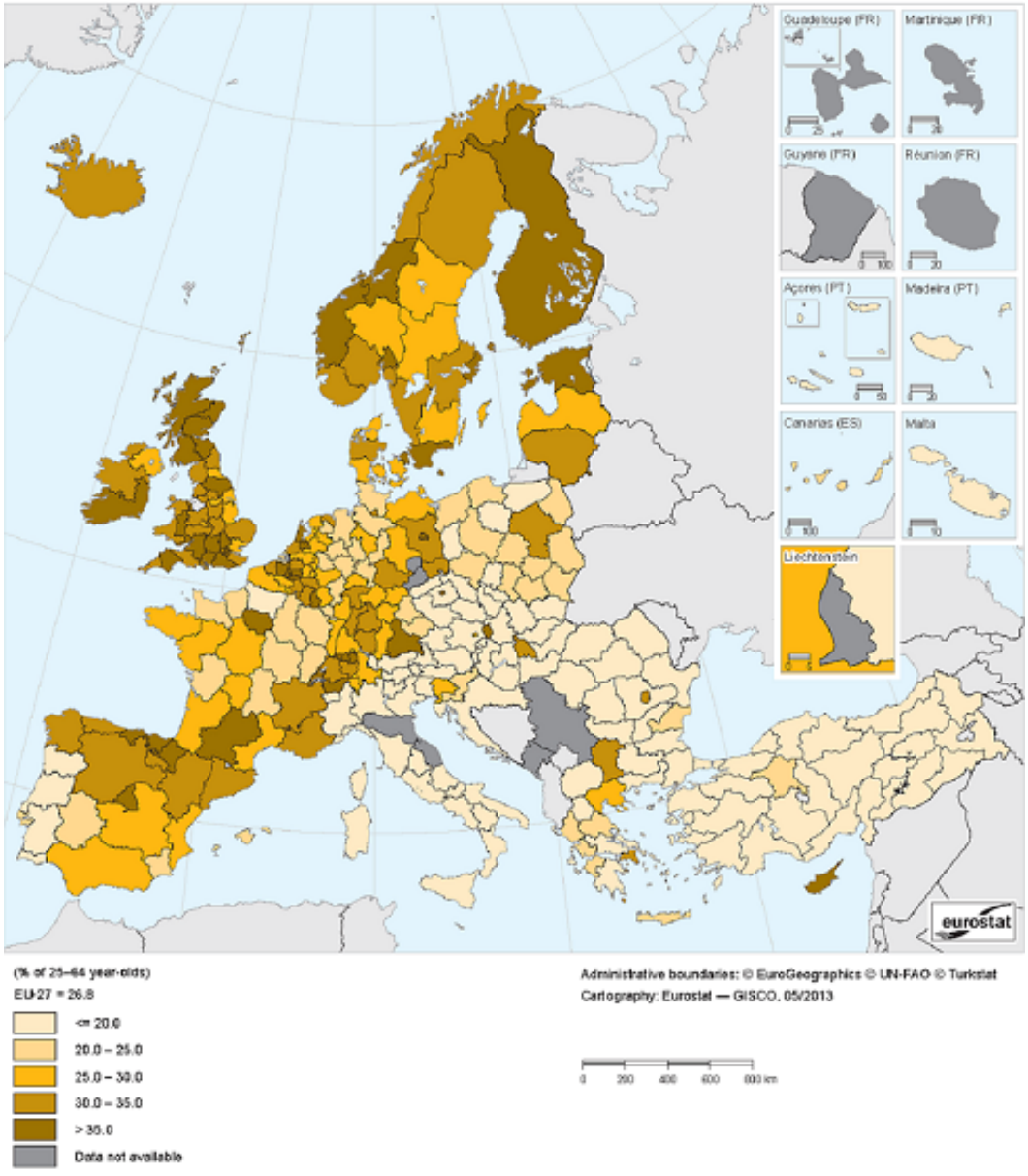
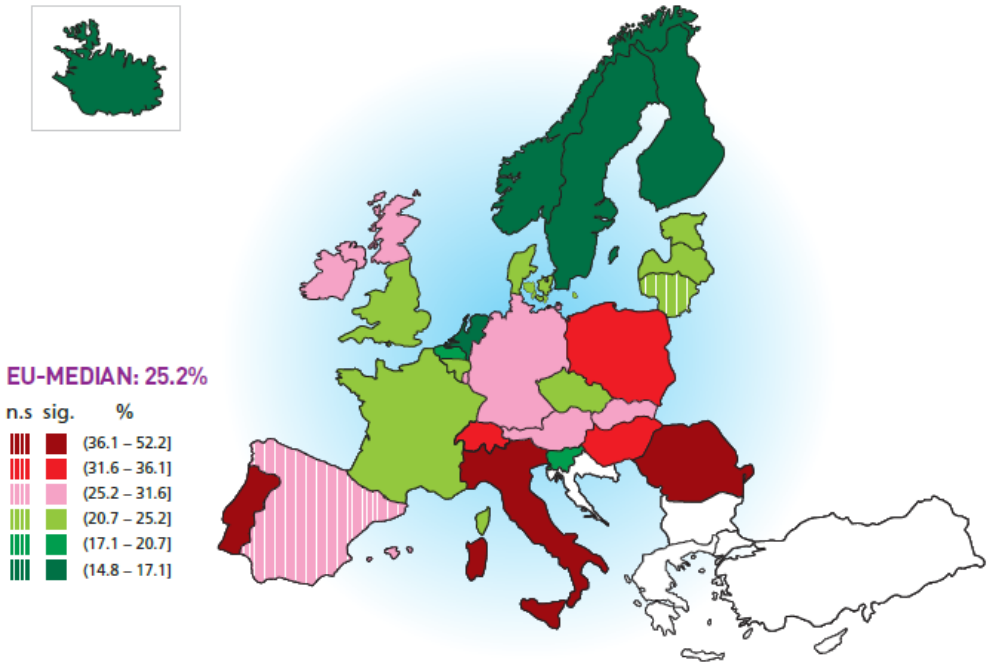


Figure 5: Persons aged 25-64 with tertiary education attainment (% of 25-64 year-old)[a]



[a]Source: Eurostat (online data code *edat_lfse_11*)

Figure 6: Caesareans as a percentage of all births in 2010[a]



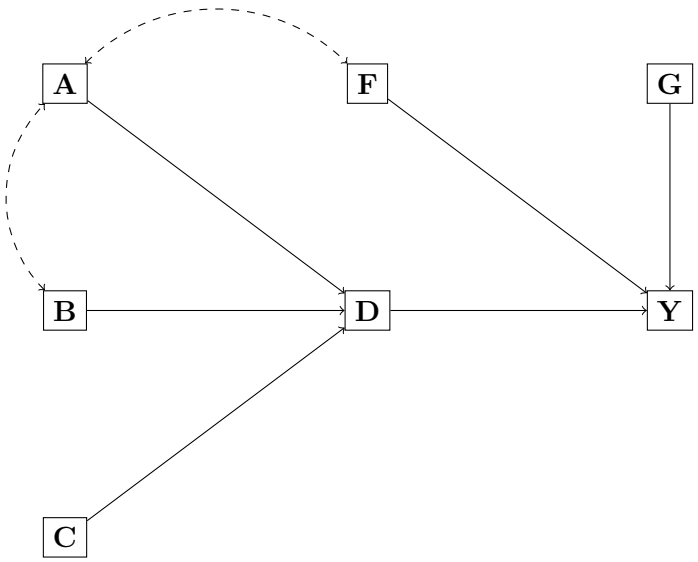
[a]Source:European Perinatal Health Report 2010

Appendix B

The three strategies to estimate causal effects

Consider the causal relationships presented by Figure 7 and suppose that these relationships are derived from a set of theoretical propositions. In Figure 7, each node represents an observable random variable and each directed edge from one node to another signifies that the variable at the origin of the directed edge causes the variable at the terminus of the directed edge. Each curved and dashed bidirected edge signifies the existence of common unobserved nodes that cause both terminal nodes. Suppose that the causal variable of primary interest is D and that the causal effect that we wish to estimate is the effect of D on Y. Suppose we want

Figure 7: Causal diagram **A**



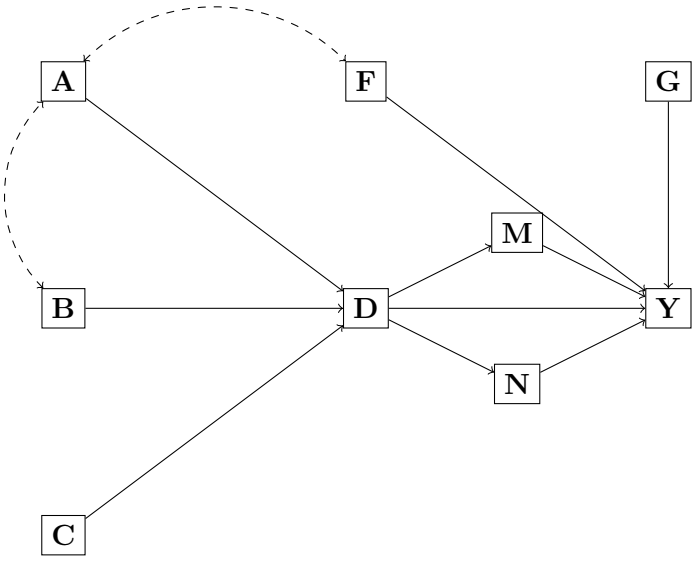
to estimate the causal effect of D on Y with conditioning on variables that block

all back-door paths[2] from the causal variable to the outcome variable. There are two two back-door paths from D to Y in Figure 7 that generate a supplemental noncausal association between D and Y: (1) D to A to F to Y and (2) D to B to A to F to Y. Both of these back-door paths can be blocked in order to eliminate the supplemental noncausal association between D and Y by observing and then conditioning on A and B or by observing and then conditioning on F.

Now consider we want to estimate the effect of D on Y using exogenous variation in an appropriate instrumental variable to isolate covariation in the causal and outcome variables. The goal is to use a localized exogenous shock to both the causal variable and the outcome variable in order to estimate indirectly the relationship between the two. In Figure 7, the variable C is a valid instrument for D because it causes D but does not have effect on Y except though its effect on D. As a result, one can estimate consistently the causal effect of D on Y by taking the ratio of the relationships between C and Y and between C and D. If the goal is to obtain the causal effect of D on Y, then the variables A, B, F, and G do not need to be observed.

Consider Figure 8 and we still want to estimate the causal effect of D on Y using

Figure 8: Causal diagram **B**



mechanistic estimation strategy. The causal effect of D on Y can be calculated by estimation of the causal effect of D on M and N and then subsequently the causal

[2] A path is any sequence of edges pointing in any direction that connects one variable to another.A back-door path is then defined as a path between any causally ordered sequence of two variables that includes a directed edge -> that points to the first variable

effects of M and N on Y. This can be done, because the mediating variables M and N completely account for the causal effect of D on Y, and because M and N are not determined by anything other than D. If the goal is to obtain the causal effect of D on Y, then the variables A, B, C, F, and G can be ignored.

Strong Ignorability Assumption

The property of *strong ignorability* of T given X holds if, for potential outcomes Y_1 and Y_0, the distribution of these potential outcomes is conditionally independent of T given X, and for any value of the covariates, there is a possibility of a unit receiving the treatment or not receiving the treatment. That is,

$$(Y_1, Y_0) \perp T|X$$

and

$$0 < Pr(T = 1|X = x) < 1 \ \forall x.$$

MCA and cluster analysis

Table 2: Case 1, 45-Cluster solution set (smoked in the last 5 years prior to pregnancy (T=1), did not smoke in the last 5 years prior to pregnancy (T=0))

Cluster	n	n(T=1)	n(T=0)	GI	Interval	Balance
1	791	172	622	0.005	(0,0.007)	Yes
2	112	11	101	0.063	(0,0.05)	No
3	636	188	448	0.006	(0,0.009)	Yes
4	529	76	453	0.007	(0,0.01)	Yes
5	490	106	384	0.006	(0,0.011)	Yes
6	503	71	432	0.007	(0,0.011)	Yes
7	497	118	379	0.013	(0,0.014)	Yes
8	610	89	521	0.007	(0,0.009)	Yes
9	530	92	435	0.007	(0,0.009)	Yes
10	411	63	348	0.013	(0,0.014)	Yes
11	314	103	211	0.013	(0,0.017)	Yes
12	577	166	411	0.011	(0,0.012)	Yes
13	572	60	512	0.005	(0,0.009)	Yes
14	312	38	274	0.008	(0,0.017)	Yes
15	308	27	281	0.01	(0,0.019)	Yes
16	658	81	577	0.004	(0,0.008)	Yes
17	360	39	321	0.017	(0,0.018)	Yes
18	63	6	57	0.063	(0,0.081)	Yes

Continued on next page

Table 2 – *Continued from previous page*

Cluster	n	n(T=1)	n(T=0)	GI	Interval	Balance
19	183	28	155	0.036	(0,0.028)	No
20	74	4	70	0.027	(0,0.066)	Yes
21	35	10	25	0.098	(0,0.127)	Yes
22	90	36	54	0.043	(0,0.056)	Yes
23	147	4	143	0.118	(0,0.03)	No
24	138	42	96	0.03	(0,0.037)	Yes
25	124	24	100	0.02	(0,0.034)	Yes
26	203	51	152	0.022	(0,0.028)	Yes
27	160	32	128	0.02	(0,0.03)	Yes
28	194	53	141	0.011	(0,0.024)	Yes
29	520	129	391	0.009	(0,0.01)	Yes
30	269	50	219	0.021	(0,0.022)	Yes
31	541	123	418	0.01	(0,0.011)	Yes
32	182	25	157	0.021	(0,0.03)	Yes
33	268	19	249	0.029	(0,0.021)	No
34	280	13	267	0.007	(0,0.018)	Yes
35	194	22	172	0.018	(0,0.026)	Yes
36	434	12	422	0.007	(0,0.011)	Yes
37	213	10	203	0.015	(0,0.025)	Yes
38	264	18	246	0.02	(0,0.021)	Yes
39	434	2	432	0.005	(0,0.011)	Yes
40	558	7	551	0.008	(0,0.009)	Yes
41	403	8	395	0.012	(0,0.013)	Yes
42	136	33	103	0.021	(0,0.029)	Yes
43	323	59	264	0.01	(0,0.015)	Yes
44	150	12	138	0.045	(0,0.025)	No
45	682	46	636	0.004	(0,0.005)	Yes

Table 3: Case 2, 45-Cluster solution set (stopped smoking before pregnancy (T=1), stopped smoking at the beginning of pregnancy (T=2), continued to smoke during pregnancy (T=3), did not smoke in the last 5 years prior to pregnancy (T=0))

Cluster	n	n(T=1)	n(T=2)	n(T=3)	n(T=0)	GI	Interval	Balance
1	794	47	75	50	622	0.013	(0,0.018)	Yes
2	112	4	2	5	101	0.17	(0,0.136)	No
3	636	52	63	73	448	0.017	(0,0.023)	Yes
4	529	27	33	16	453	0.017	(0,0.028)	Yes
5	490	33	35	38	0.018	(0,0.029)	Yes	
6	503	19	18	34	432	0.019	(0,0.029)	Yes
7	497	27	28	63	379	0.028	(0,0.031)	Yes
8	610	34	26	29	521	0.017	(0,0.024)	Yes
9	530	26	36	33	435	0.015	(0,0.025)	Yes
10	411	23	20	20	348	0.029	(0,0.035)	Yes
11	314	10	41	52	211	0.031	(0,0.045)	Yes

Continued on next page

Table 3 – *Continued from previous page*

Cluster	n	n(T=1)	n(T=2)	n(T=3)	n(T=0)	GI	Interval	Balance
12	577	35	51	80	411	0.024	(0,0.026)	Yes
13	572	24	15	21	512	0.016	(0,0.025)	Yes
14	312	18	10	10	274	0.023	(0,0.046)	Yes
15	308	10	7	10	281	0.038	(0,0.052)	Yes
16	658	27	34	20	577	0.015	(0,0.021)	Yes
17	360	13	13	13	321	0.048	(0,0. 043)	No
18	63	2	.	4	57	-	-	no common support
19	183	5	7	16	155	0.092	(0,0.075)	No
20	74	.	1	3	70	-	-	no common support
21	35	1	4	5	25	0.19	(0,0.342)	Yes
22	90	4	11	21	54	0.123	(0,0.151)	Yes
23	147	.	1	3	143	-	-	no common support
24	138	10	13	19	96	0.081	(0,0.101)	Yes
25	124	4	6	14	100	0.056	(0,0.091)	Yes
26	203	9	15	27	152	0.052	(0,0.076)	Yes
27	160	6	13	13	128	0.059	(0,0.08)	Yes
28	194	6	18	29	141	0.037	(0,0.065)	Yes
29	520	32	36	61	391	0.024	(0,0.027)	Yes
30	269	21	18	11	219	0.051	(0,0.052)	Yes
31	541	31	30	62	418	0.022	(0,0.026)	Yes
32	182	3	9	13	157	0.057	(0,0.081)	Yes
33	268	3	4	12	249	0.063	(0,0.057)	No
34	280	1	4	8	267	0.02	(0,0.049)	Yes
35	194	3	9	10	172	0.063	(0,0.071)	Yes
36	434	1	4	7	422	0.024	(0,0.029)	Yes
37	213	3	2	5	203	0.046	(0,0.068)	Yes
38	264	5	5	8	246	0.039	(0,0.057)	Yes
39	434	2	.	.	432	0.005	(0,0.011)	Yes
40	558	2	2	3	551	0.021	(0,0.023)	Yes
41	403	2	1	5	395	0.024	(0,0.032)	Yes
42	136	1	8	24	103	0.042	(0,0.077)	Yes
43	323	9	9	41	264	0.024	(0,0.041)	Yes
44	150	1	4	7	138	0.093	(0,0.065)	No
45	682	11	6	29	636	0.011	(0,0.014)	Yes

Table 4: Case 3, 16-Cluster solution set (stopped smoking at the beginning of pregnancy (T=1), continued to smoke during pregnancy (T=2), stopped smoking before pregnancy (T=0))

Cluster	n	n(T=1)	n(T=2)	n(T=0)	GI	Interval	Balance
1	300	121	132	47	0.026	(0,0.03)	Yes
2	402	146	131	125	0.014	(0,0.022)	Yes
3	10	2	5	3	0.385	(0,0.666)	Yes
4	234	67	111	56	0.027	(0,0.041)	Yes
5	311	91	108	112	0.024	(0,0.03)	Yes
6	229	75	72	82	0.03	(0,0.044)	Yes
7	152	35	74	43	0.042	(0,0.058)	Yes
8	192	63	85	44	0.038	(0,0.044)	Yes
9	65	18	32	15	0.086	(0,0.142)	Yes
10	59	18	32	9	0.092	(0,0.146)	Yes
11	53	18	26	9	0.136	(0,0.156)	Yes
12	77	27	35	15	0.063	(0,0.115)	Yes

Continued on next page

Table 4 – *Continued from previous page*

Cluster	n	n(T=1)	n(T=2)	n(T=0)	GI	Interval	Balance
13	61	16	32	13	0.093	(0,0.133)	Yes
14	86	23	51	12	0.116	(0,0.117)	Yes
15	93	17	66	10	0.065	(0,0.088)	Yes
16	57	10	35	12	0.073	(0,0.108)	Yes

Table 5: Case 4, 8-Cluster solution set (continued to smoke during pregnancy (T=1), stopped smoking at the beginning of pregnancy (T=0))

Cluster	n	n(T=1)	n(T=0)	GI	Interval	Balance
1	656	378	278	0.007	(0,0.008)	Yes
2	412	185	227	0.009	(0,0.013)	Yes
3	267	164	103	0.019	(0,0.02)	Yes
4	88	54	34	0.029	(0,0.059)	Yes
5	145	92	53	0.019	(0,0.037)	Yes
6	76	52	24	0.053	(0,0.073)	Yes
7	83	66	17	0.042	(0,0.052)	Yes
8	47	36	11	0.055	(0,0.084)	Yes

ATT and ATE

Table 6: Case 1, Local effects and p-value

Cluster	Local effects	p-value	Cluster	Local effects	p-value
1	-698.5	<0.01	24	736.69	0.33
2	-	-	25	-1502.86	0.07
3	-57.35	0.82	26	-429.87	0.45
4	-2.91	0.99	27	41.23	0.94
5	204.41	0.54	28	-363.16	0.42
6	-372.2	0.22	29	-16.41	0.97
7	-529.25	0.05	30	-286.07	0.36
8	-111.12	0.74	31	-41.17	0.9
9	353.95	0.31	32	3021.84	0.05

Continued on next page

Table 6 – *Continued from previous page*

Cluster	Local effects	p-value	Cluster	Local effects	p-value
10	18.21	0.93	33	-	-
11	-1035.83	0.04	34	-419.78	0.35
12	-329.28	0.36	35	-252.5	0.36
13	-198.79	0.71	36	-441.3	0.3
14	-367.93	0.13	37	363.15	0.77
15	-653.85	0.07	38	-1048.31	<0.01
16	-465.68	0.11	39	-1884.92	<0.01
17	-344.84	0.61	40	4492.16	<0.03
18	2303.7	0.27	41	-741.48	0.03
19	-	-	42	-1665.13	0.01
20	707.66	0.43	43	-819.71	0.01
21	-1283.91	0.02	44	-	-
22	569	0.32	45	1283.28	0.13
23	-	-			

Table 7: Case 2,Local effects and p-value

Cluster	Local effects$_{10}$	p-value	Local effects$_{20}$	p-value	Local effects$_{30}$	p-value
1	-479.23	0.24	-882.53	<0.01	-628.58	0.04
2	-	-	-	-	-	-
3	234.14	0.63	-374.18	0.22	8.44	0.98
4	520.72	0.47	-278.87	0.31	-317.36	0.73
5	304.85	0.58	233.99	0.66	89.93	0.87
6	-910.78	<0.01	-225.54	0.6	-148.87	0.76
7	-587.21	0.06	-466	0.28	-532.52	0.11
8	-329.4	0.24	208.16	0.81	-141.47	0.62
9	51.48	0.88	-90.8	0.83	1077.44	0.23
10	-484.9	0.01	523.3	0.29	91.69	0.79
11	-1803.67	<0.01	-858.99	0.14	-1027.61	0.05
12	32.04	0.97	-704.05	0.03	-248.43	0.64
13	416.41	0.62	-880.38	0.01	-415.02	0.34
14	-280.26	0.43	-667.35	<0.01	-226.33	0.74
15	-1049.99	<0.01	-147.4	0.86	-612.24	0.24
16	-560.37	0.13	-741.73	<0.01	131.41	0.85
17	-	-	-	-	-	-
18	-	-	-	-	-	-

Continued on next page

Table 7 – *Continued from previous page*

Cluster	Local effects$_{10}$	p-value	Local effects$_{20}$	p-value	Local effects$_{30}$	p-value
19	-	-	-	-	-	-
20	-	-	-	-	-	-
21	-1454.6	0.58	-1047.27	0.07	-1439.1	0.03
22	-613.24	0.05	-368.26	0.24	1285.13	0.05
23	-	-	-	-	-	-
24	1533.39	0.03	-289.17	0.57	1019.27	0.46
25	-1669.96	0.07	-1542.18	0.07	-1438.26	0.09
26	35.07	0.96	-983.22	0.03	-277.44	0.76
27	2001.76	0.34	-666.88	0.06	-155.52	0.74
28	-1299.32	<0.01	-775.73	0.05	86.6	0.88
29	-581.97	0.17	103.34	0.92	209.6	0.76
30	424.04	0.37	92.88	0.84	-642.8	0.15
31	-550.7	0.06	57.99	0.93	165.61	0.75
32	-560.07	0.71	573.77	0.52	5543.26	0.05
33	-	-	-	-	-	-
34	-1705.23	0.58	-104.35	0.95	-416.82	0.5
35	119.18	0.92	-540.46	0.13	-104.85	0.81
36	-1196.2	0.66	-554.83	0.68	-268.67	0.79
37	3377.15	0.01	-992.31	0.52	-903.06	0.04
38	-941.81	<0.01	-903.29	0.03	-1205.5	<0.01
39	-	-	-	-	-	-
40	16838.19	<0.01	735.08	0.73	-1233.82	<0.01
41	-1182.24	<0.01	827.52	0.81	-878.98	<0.01
42	-2746.48	0.63	-2780.57	<0.01	-1248.26	0.09
43	-993.94	0.07	-1301.63	<0.01	-675.67	0.04
44	-	-	-	-	-	-
45	-622.38	0.08	-1019.72	<0.01	2482.6	<0.01

Table 8: Case 3, Local effects and p-value

Cluster	Local effects$_{10}$	p-value	Local effects$_{20}$	p-value
1	-353.68	0.51	46.87	0.93
2	-333.89	0.27	-109.05	0.75
3	-2933.63	0.19	1635.18	0.78
4	83.14	0.7	541.02	0.16
5	-500.52	0.16	-319.53	0.44
6	633.55	0.08	717.09	0.05

Continued on next page

Table 8 – *Continued from previous page*

Cluster	Local effects$_{10}$	p-value	Local effects$_{20}$	p-value
7	635.29	0.2	1106.27	0.08
8	379.42	0.57	328.97	0.47
9	-2690.53	0.3	-2162.34	0.42
10	417.44	0.74	1602.2	0.33
11	192.54	0.62	181.62	0.57
12	-1639.32	0.02	-1511.75	0.01
13	-1310.81	0.33	-647.13	0.67
14	-1524.6	0.3	-1331.51	0.41
15	-415.37	0.34	564.08	0.32
16	-320.99	0.27	2630.11	0.02

Table 9: Case 4, Local effects and p-value

Cluster	Local effects	p-value
1	306.17	0.12
2	223.28	0.38
3	263.42	0.61
4	997.15	0.2
5	380.39	0.36
6	216.73	0.77
7	979.44	<0.01
8	3471.62	<0.01

Table 10: Maternal characteristics, Cluster 40

Variables	**Freq (%)**	**Variables**	**Freq (%)**
Maternal age		**Education level**	
<25	9.76	primary or no education	20.61
25-29	51.18	secondary	33.45
30-34	35.62	high school	32.73
35-39	3.44	terniary	13.2
40+	-	**Current occupation status**	

Continued on next page

Table 10 – *Continued from previous page*

Variables	Freq (%)	Variables	Freq (%)
Marital status		occupied	-
single	2.35	unemployed	3.8
married	93.85	in search of first job	-
separated	0.18	student	1.81
divorced	0.18	housewife	94.03
widowed	-	other	-
not stated	3.44	n/a	0.36
Nationality		**Profession**	
Italian	8.32	self-employed	12.49
other EU citizen	-	manager or director	2.42
citizen of LMLIC	91.68	employee	53.06
other	-	worker	25.53
Place of birth		other dependent employer	6.06
north-west Italy	-	other dependent employer	6.06
north-east Italy	0.36	n/a	0.44
central Italy	-	**Sector of employment**	
south Italy	0.72	agriculture,hunting and fishing	1.94
islands (Italy)	-	industry	17.71
abroad	98.01	services	31.29
n/a	0.9	public administration	11.62
		other private sector	35.49
Number of previous:		n/a	1.95
live births		spontaneous abortion	
0	30.74	0	89.33
1	55.15	1	8.14
2	13.38	2	1.63
3+	0.72	3+	0.9
cesarean		induce abortion	
0	85.9	0	93.67
1	12.66	1	4.34
2	1.27	2	-
3+	0.18	3+	1.99

Table 11: Case 3, Clusters description

Covariates	Cluster 6	Cluster 12	Cluster 16
Maternal age	35-39 (45.41)	25-29 (36.36)	30-34 (43.86)

Continued on next page

Table 11 – *Continued from previous page*

Covariates	Cluster 6	Cluster 12	Cluster 16
Marital status	married (65.5)	married (51.95)	not stated (40.35)
Mother's nationality	Italian (99.13)	Other EU citizen (94.81)	Italian (78.95)
Birth place (mother)	north-east Italy (59.83)	abroad (97.4)	north-east Italy (50.88)
Education level (mother)	primary/no education (62.45)	secondary (58.44)	tertiary (73.68)
Current occupation status (mother)	occupied (97.82)	occupied (100)	occupied (100)
Profession (mother)	employee (83.93)	worker (53.25)	n/a (100)
Sector of employment (mother)	public administration (42.86)	services (37.66)	n/a (100)
Paternal age	35-39 (49.34)	30-34 (36.36)	n/a (100)
Father's nationality	Italian (96.51)	Italian (50.65)	n/a (100)
Education level (father)	primary/no education (48.03)	secondary (55.84)	n/a (100)
Current occupation status (father)	occupied (100)	occupied (100)	n/a (100)
Profession (father)	employee (65.94)	employee (54.55)	n/a (100)
Sector of employment (father)	services (29.69)	other private sector (41.56)	n/a (100)
Number of previous:			
live births	no (63.76)	no (61.04)	no (59.65)
spontaneous abortion	no (84.28)	no (74.03)	no (70.18)
induce abortion	no (95.2)	no (84.42)	no (96.49)
caesarean	no (94.32)	no (97.4)	no (94.74)
Local ATT	50.27	-51.52	89.63
Local ATE	100.5	-71.24	91.75

Table 12: Case 4, Clusters description

Covariates	Cluster 7	Cluster 8
Maternal age	25-29 (30.12)	30-34(36.17)
Marital status	single (87.95)	not stated(40.43)
Mother's nationality	Italian (86.75)	Italian (76.6)
Birth place (mother)	north-east Italy (66.27)	north-east Italy (55.32)
Education level (mother)	high school (45.78)	primary/no education (72.34)

Continued on next page

Table 12 – *Continued from previous page*

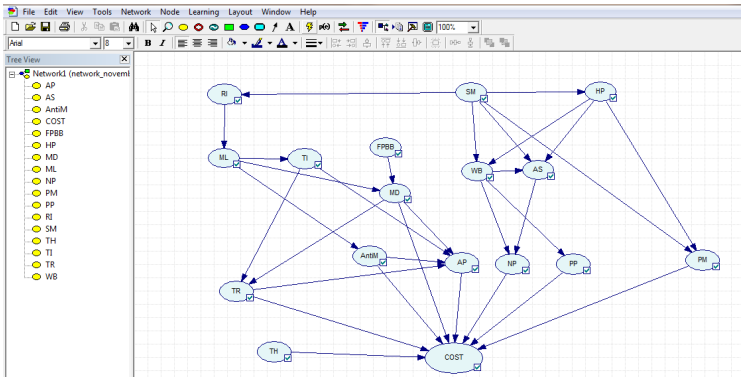
Covariates	**Cluster 7**	**Cluster 8**
Current occupation status (mother)	occupied (61.45)	occupied (100)
Profession (mother)	employee (41.18)	n/a (100)
Sector of employment (mother)	services (43.14)	n/a (100)
Paternal age	n/a (100)	n/a (94.74)
Father's nationality	n/a (97.58)	n/a (95.74)
Education level (father)	n/a (100)	n/a (97.87)
Current occupation status (father)	n/a (100)	n/a (95.74)
Profession (father)	n/a (100)	n/a (95.74)
Sector of employment (father)	n/a (100)	n/a (95.74)
Number of previous:		
live births	no (73.49)	no (57.45)
spontaneous abortion	no (77.11)	no (63.93)
induce abortion	no (80.72)	no (95.74)
caesarean	no (96.39)	no (93.62)
Local ATT	62.94	121.69
Local ATE	43.38	87.07

Appendix C

Other estimation methods

- Probabilistic Networks. It is a graphical models of (causal) interactions among a set of variables, where the variables are represented as nodes of a graph and the interactions as directed links between nodes. Any pair of unconnected/non-adjacent nodes of such a graph indicates (conditional) independence between the variables represented by these nodes under particular circumstances that can easily be read from the graph. Hence, probabilistic networks capture a set of (conditional) dependence and independence properties associated with the variables represented in the network (Cowell et al. (1999)). The estimation was made using GeNIe 2.0 software.

Figure 9: GeNIe 2.0



- The cost was estimated using multiple regression (formula presented below):

$$\begin{aligned} COST = \alpha_{COST} &+ \beta_{COST,TH} * TH + \beta_{COST,TR} * TH + \beta_{COST,PM} * PM \\ &+ \beta_{COST,AntiM} * AntiM + \beta_{COST,ML} * ML \\ &+ \beta_{COST,TI} * TI + \beta_{COST,MD} * MD + \beta_{COST,AP} * AP \\ &+ \beta_{COST,PP} * PP + \beta_{COST,NP} * NP + \varepsilon. \end{aligned}$$

  Calculation was made using *proc gml* in SAS 9.3, and obtained results were difficult to interpret.

Table 13: Determinants of delivery cost in Emilia-Romagna January-June, 2010: Results (odds ratios) of an ordered logit model (*smoker1:stopped smoking before pregnancy, smoker2: stopped smoking at the beginning of pregnancy; smoker3:continued to smoke during pregnancy*)

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
Two smoking categories												
SM												
smoker/non-smoker	0.87	1.01	0.75	0.79	1.29							
	(0.77-1.05)	(0.9-1.12)	(0.51-1.11)	(0.73-0.87)	(0.72-2.35)							
HP												
yes/no			0.99	0.28	0.02							
			(0.55-1.79)	(0.24-0.32)	(0.53-1.95)							
n/a/no			1.24	1.04	3.79							
			(0.17-8.92)	(0.24-0.32)	(0.98-14.67)							
WB												
<1500/2500-3499					84.75	0.09	0.49					
					(42.09-170.64)	(0.05-0.17)	(0.35-0.71)					
1500-2499/2500-3499					6.54	0.22	0.74					
					(3.25-13.16)	(0.18-0.26)	(0.64-0.85)					
3500+/2500-3499					1.36	1.15	0.99					
					(0.72-2.57)	(1.07-1.23)	(0.93-1.06)					
AS												
6-4/10-7							0.42					
							(0.24-0.72)					
3-0/10-7							2.11					
							(0.65-6.83)					
Four smoking categories												
SM												
smoker1/non-smoker	0.93	0.91	0.77	0.83	1.53							
	(0.69-1.24)	(0.74-1.11)	(0.38-1.58)	(0.7-0.97)	(0.53-4.42)							
smoker2/non-smoker	0.94	1.17	0.63	1.01	1.36							
	(0.72-1.23)	(0.98-1.39)	(0.35-1.14)	(0.87-1.17)	(0.48-3.85)							
smoker3/non-smoker	0.85	0.95	0.87	0.65	1.16							
	(0.68-1.06)	(0.81-1.11)	(0.48-1.57)	(0.57-0.74)	(0.49-2.69)							
HP												

Continued on next page

Table 13 – *Continued from previous page*

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
yes/no			0.99 (0.55-1.79)	0.28 (0.24-0.32)	1.02 (0.53-1.95)							
n/a/no			1.23 (0.17-8.85)	1.05 (0.73-1.5)	3.92 (1.01-15.17)							
WB												
<1500/2500-3499					85.18 (42.29-171.52)							
1500-2499/2500-3499					6.58 (3.27-13.25)							
3500+/2500-3499					1.36 (0.72-2.56)							
RI												
prolonged pregnancy/no								4.21 (3.69-4.8)				
oligodramnios/no								4.21 (3.56-4.98)				
PROM/no								4.21 (3.69-4.79)				
maternal pathology/no								4.21 (3.56-4.97)				
fetal pathology/no								4.21 (3.34-5.31)				
n/a/no								4.21 (3.43-5.17)				
FPBB												
facing upward 1/facing down 1									117.3 (72.46-189.9)			
facing upward 2/facing down 1									2.59 (0.38-17.59)			
facing down 2/facing down 1									3.85 (2.23-6.66)			
feet first/facing down 1									7.05 (2.56-19.42)			
lying sideways/facing down 1									15.28			

Continued on next page

Table 13 – *Continued from previous page*

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
									(7.68-30.42)			
TI												
prostaglandin/no										0.81	0.99	0.95
										(0.62-1.06)	(0.89-1.09)	(0.61-1.49)
oxytocin/no										1.31	0.72	1.03
										(0.73-2.35)	(0.6-0.86)	(0.65-1.63)
other drug/no										999.9	0.65	1.38
										(0.001-999.9)	(0.14-3.04)	(0.77-1.32)
amnioressi/no										999.9	2.02	
										(0.001-999.9)	(1.12-3.63)	
MD												
vaginal with use of forceps/vaginal										0.11	1.24	1.01
										(0.02-0.48)	(0.45-3.41)	(0.45-2.26)
vaginal with use of ventouse/vaginal										0.21	1.24	1.06
										(0.14-0.31)	(0.51-0.73)	(0.9-1.25)
cesarean/vaginal										0.24	0.03	18.29
										(0.19-0.3)	(0.02-0.04)	(16.1-20.77)
TR												
manual ventilation/not required											0.68	5.59
											(0.47-0.98)	(4.26-7.33)
intubation/not required											1.2	3.87
											(0.72-2.0)	(2.7-5.53)
AntiM												
epidural analgesia/no method used											0.25	1.03
											(0.22-0.29)	(0.91-1.17)
other type of pharmacological analgesia/no method used											2.95	1.01
											(2.05-4.25)	(0.77-1.32)
non-pharmacological/no method used											2.3	0.94
											(2.11-2.51)	(0.88-1.01)
n.a/no method used											0.31	2.19
											(0.22-0.43)	(1.69-2.83)
ML												

Continued on next page

Table 13 – *Continued from previous page*

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
induce/spontaneous												1.29
												(0.83-2.0)
without/spontaneous												0.6
												(0.46-0.79)
AP												
yes/no												0.73
												(0.66-0.81)
n.a/no												1.26
												(0.69-2.25)
NP												
yes/no												0.97
												(0.9-1.03)
n.a/no												0.72
												(0.48-1.09)
PP												
yes/no												1.26
												(1.15-1.37)
n.a/no												0.97
												(0.57-1.63)
TH												
public type A/public type B												6.35
												(5.92-6.82)
private/public type B												1.76
												(1.35-2.29)
PM												
yes/no												3.51
												(2.6-4.74)

Table 14: Determinants of delivery cost in Emilia-Romagna January-June, 2010: Coefficients of the ordered logit model (**: significant at 0.1%, **: significant at 5%; smoker1:stopped smoking before pregnancy, smoker2: stopped smoking at the beginning of pregnancy; smoker3:continued to smoke during pregnancy*)

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
Two smoking categories												
SM												
smoker/non-smoker	-0.11	0.005	-0.28	-0.23*	0.26							
HP												
yes/no			-0.003	-1.28*	0.02							
n/a/no			0.21	0.04	1.33**							
WB												
<1500/2500-3499					4.44*							
1500-2499/2500-3499					1.88*							
3500+/2500-3499					0.31							
Four smoking categories												
SM												
smoker1/non-smoker	-0.08	-0.09	-0.26	-0.19**	0.43							
smoker2/non-smoker	-0.06	0.15	-0.46	0.01	0.31							
smoker3/non-smoker	-0.16	-0.05	-0.14	-0.43*	0.14							
HP												
yes/no			-0.01	-1.28*	0.02							
n/a/no			0.21	0.05	1.36**							
WB												
<1500/2500-3499					4.44**	-2.39*	-0.7*					
1500-2499/2500-3499					1.88**	-1.51*	-0.29*					
3500+/2500-3499					0.31	0.14*	-0.01					
AS												
6-4/10-7							-0.87**					
3-0/10-7							0.75					
RI												
prolonged pregnancy/no								1.44*				
oligodramnios/no								1.44*				
PROM/no								1.44*				
maternal pathology/no								1.44*				
fetal pathology/no								1.44*				
n/a/no								1.44*				
FPBB												
facing upward 1/facing down 1									4.76*			
facing upward 2/facing down 1									0.95			
facing down 2/facing down 1									1.35*			
feet first/facing down 1									1.95*			

Continued on next page

Table 14 – *Continued from previous page*

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
lying sideways/facing down 1									2.73*			
TI												
prostaglandin/no										-0.21	-0.04	-0.05
oxytocin/no										0.27	-0.33**	0.03
other drug/no										11.71	-0.43	0.32
amnioressi/no										11.39	0.7**	
MD												
vaginal with use of forceps/vaginal										-2.22**	0.21	0.01
vaginal with use of ventouse/vaginal										-1.56*	-0.5*	0.06
cesarean/vaginal										-1.42*	-3.54*	2.91*
TR												
manual ventilation/not required											-0.39**	1.72*
intubation/not required											0.18	1.35*
AntiM												
epidural analgesia/no method used											-1.37*	0.03
other type of pharmacological analgesia/no method used											1.08*	0.01
non-pharmacological/no method used											0.83*	-0.06
n.a/no method used											-1.18*	0.79*
ML												
induce/spontaneous												0.26
without/spontaneous												-0.5**
AP												
yes/no												-0.32*
n.a/no												0.23
NP												
yes/no												-0.03
n.a/no												-0.33
PP												
yes/no												0.23*
n.a/no												-0.03
TH												
public type A/public type B												1.85*
private/public type B												0.57*
PM												
yes/no												1.26*

Table 15: Determinants of delivery cost in Emilia-Romagna January-June, 2010: Results (odds ratios) of an ordered logit model (*smoker1:stopped smoking before pregnancy, smoker2: stopped smoking at the beginning of pregnancy; smoker3:continued to smoke during pregnancy*)

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
SM												
smoker2/smoker1	1.02	1.29	0.82	1.22	0.81							
	(0.68-1.51)	(0.99-1.67)	(0.33-2.02)	(0.98-1.52)	(0.19-3.44)							
smoker3/smoker1	0.91	1.04	1.13	0.78	0.72							
	(0.63-1.32)	(0.81-1.34)	(0.46-2.77)	(0.63-0.96)	(0.19-2.75)							
HP												
yes/no			1.24	0.25	1.82							
			(0.29-5.22)	(0.18-0.34)	(0.47-6.98)							
n/a/no			999.9	0.35	4.05							
			(0.001-999.9)	(0.09-1.27)	(0.27-61.63)							
WB												
<1500/2500-3499					999.9	0.13	4.93					
					(0.001-999.9)	(0.03-0.52)	(1.52-12.67)					
1500-2499/2500-3499					999.9	0.31	0.75					
					(0.001-999.9)	(0.21-0.46)	(0.53-1.05)					
3500+/2500-3499					999.9	1.02	0.9					
					(0.001-999.9)	(0.86-1.22)	(0.76-1.08)					
AS												
6-4/10-7							0.37					
							(0.12-1.17)					
RI												
prolonged pregnancy/no								4.15				
								(3.0-5.73)				
oligodramnios/no								4.15				
								(2.68-6.43)				
PROM/no								4.15				
								(2.99-5.75)				
maternal pathology/no								4.15				
								(2.73-6.29)				
fetal pathology/no								4.15				
								(2.45-7.02)				
n/a/no								4.15				
								(2.03-8.48)				
FPBB												

Continued on next page

Table 15 – *Continued from previous page*

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
facing upward 1/facing down 1									54.66 (22.25-134.3)			
facing upward 2/facing down 1									0.001 (0.001-999.99)			
facing down 2/facing down 1									2.96 (0.82-10.71)			
feet first/facing down 1									999.9 (0.001-999.9)			
lying sideways/-facing down 1									999.9 (0.001-999.9)			
TI												
prostaglandin/no										1.12 (0.55-2.29)	1.12 (0.85-1.48)	0.23 (0.07-0.75)
oxytocin/no										1.53 (0.37-6.39)	0.86 (0.56-1.33)	0.29 (0.09-0.97)
amnioressi/no										999.9 (0.001-999.9)	2.92 (0.59-14.46)	
MD												
vaginal with use of ventouse/vaginal										0.26 (0.09-0.71)	0.7 (0.44-1.12)	1.39 (0.93-2.09)
cesarean/vaginal										0.31 (0.19-0.52)	0.03 (0.02-0.04)	16.95 (12.35-23.27)
TR												
manual ventilation/not required											0.57 (0.22-1.45)	10.68 (5.17-22.06)
intubation/not required											2.62 (0.85-8.08)	6.58 (2.68-16.18)
AntiM												
epidural analgesia/no method used											0.29 (0.21-0.42)	1.35 (0.9-1.82)

Continued on next page

Table 15 – *Continued from previous page*

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
other type of pharmacological analgesia/no method used											2.59	1.19
											(1.11-6.08)	(0.64-2.25)
non-pharmacological/no method used											2.51	1.09
											(2.01-3.14)	(0.9-1.3)
n.a/no method used											0.12	2.01
											(0.03-0.52)	(1.09-3.73)
ML												
induce/spontaneous												5.33
												(1.64-17.3)
without/spontaneous												0.77
												(0.4-1.6)
AP												
yes/no												0.85
												(0.66-1.09)
n.a/no												0.88
												(0.21-3.62)
NP												
yes/no												1.1
												(0.91-1.33)
n.a/no												1.06
												(0.32-3.48)
PP												
yes/no												1.39
												(1.11-1.74)
n.a/no												1.24
												(0.38-4.06)
TH												
public type A/public type B												5.16
												(4.32-6.15)
private/public type B												1.58
												(0.92-2.71)
PM												
yes/no												3.75

Continued on next page

Table 15 – *Continued from previous page*

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
												(1.92-67.32)

Table 16: Determinants of delivery cost in Emilia-Romagna January-June, 2010: Coefficients of the ordered logit model (**: significant at 0.1%, **: significant at 5%; smoker1:stopped smoking before pregnancy, smoker2: stopped smoking at the beginning of pregnancy; smoker3:continued to smoke during pregnancy*)

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
SM												
smoker2/smoker1	0.02	0.25	-0.19	0.2	-0.21							
smoker3/smoker1	0.09	0.04	0.12	-0.25**	-0.32							
HP												
yes/no			0.21	-1.4*	0.59							
n/a/no			11.96	-1.04	1.39							
WB												
<1500/2500-3499					16.48	-2.06**	-1.48**					
1500-2499/2500-3499					14.35	-1.17*	-0.29					
3500+/2500-3499					13.3	0.02	-0.1					
AS												
6-4/10-7							-0.99					
RI												
prolonged pregnancy/no								1.42*				
oligodramnios/no								1.42*				
PROM/no								1.42*				
maternal pathology/no								1.42*				
fetal pathology/no								1.42*				
n/a/no								1.42*				
FPBB												
facing upward 1/facing down 1									4.0*			
facing upward 2/facing down 1									-12.52			
facing down 2/facing down 1									1.08			
feet first/facing down 1									14.36			
lying sideways/facing down 1									14.36			
TI												
prostaglandin/no										0.12	0.12	-1.47**
oxytocin/no										0.43	-0.15	-1.25**
amnioressi/no										11.74	1.07	
MD												
vaginal with use of ventouse/vaginal										-1.33**	-0.36	0.33
cesarean/vaginal										-1.17*	-3.35*	2.83*
TR												
manual ventilation/not required											-0.56	2.37*
intubation/not required											0.96	1.88*
AntiM												
epidural analgesia/no method used											-1.21*	0.29**

Continued on next page

Table 16 – *Continued from previous page*

Variables	HP	RI	PM	WB	AS	PP	NP	ML	MD	TR	AP	COST
other type of pharmacological analgesia/no method used											0.95**	0.18
non-pharmacological/no method used											0.92*	0.08
n.a/no method used											-2.14**	0.69**
ML												
induce/spontaneous												1.67**
without/spontaneous												-0.27
AP												
yes/no												-0.67
n.a/no												-0.13
NP												
yes/no												0.09
n.a/no												0.06
PP												
yes/no												0.33**
n.a/no												0.22
TH												
public type A/public type B												1.64*
private/public type B												0.46
PM												
yes/no												1.32*

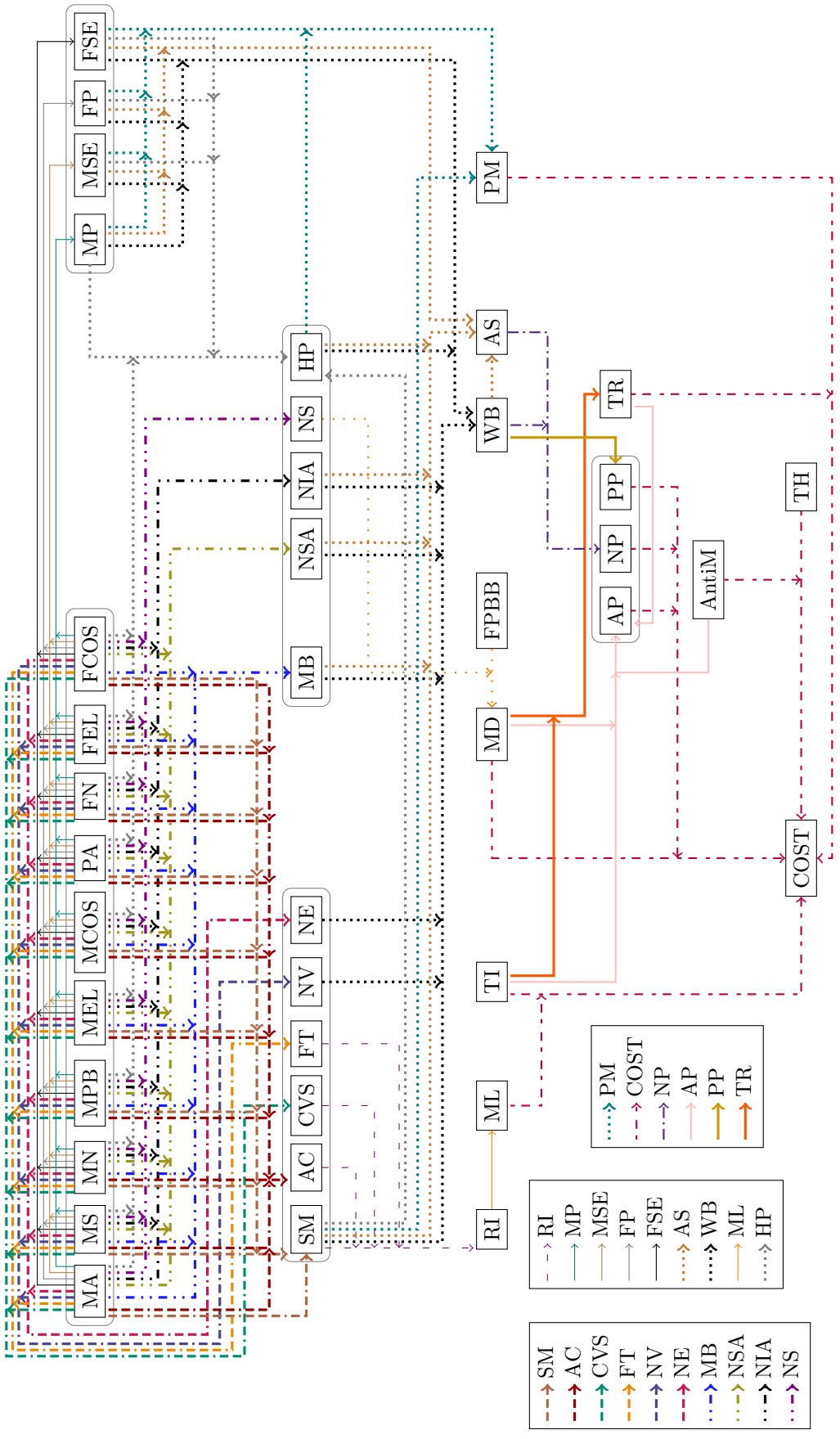


Figure 10: The Operational Framework