UNIVERSITÀ DEL PIEMONTE ORIENTALE

"AMEDEO AVOGADRO"

Department of Translation Medicine

**PhD in Medical Sciences and Biotechnology**

# Study of occupational and environmental exposure to asbestos through linkage methods and spatial analysis.

**Supervisor**:
Professor Corrado Magnani
Professor Daniela Ferrante

**Candidate**:
Chiara Airoldi
(20032068)

**XXXIV course**

**Academic years 2019-2022**

~~A te, che ancora non sei nato ma che sei già per noi un’esplosione di gioia.~~

A te, che sei luce. Così piccolo ma così forte.

# TABLE OF CONTENTS

ABSTRACT

INTRODUCTION

The identification and monitoring of cancers caused by occupational or environmental exposure to asbestos is an important aspect of public health. In Italy, to evaluate the impact of exposure to carcinogens (included asbestos) different surveillance systems are implemented: a National Mesothelioma Registry (ReNaM), a registration system for occupational cancers with ‘lower attributable fraction’ (ReNaLOCCAM) and a list of Italian Polluted Sites (SIN). The ReNaM includes Malignant Mesothelioma (MM) incident cases and assesses asbestos exposure at individual level while the ReNaLOCCAM is based on a research activity (OCM- Occupational Cancer Monitoring) that, throught population-based case-control studies, identifies economic activities with major cancer risk. Despite OCM analyses different cancer types, data on cases’ occupational histories are limited to name and industrial sector of firms where patients worked without direct information on asbestos exposure.

The city of Casale Monferrato (northwest Italy), where a high incidence of MM was observed due to the presence of the largest Italian asbestos cement plant (Eternit), is one of the sites of national interest for environmental remediation. Several studies have been conducted in this area regarding the workers, their wives, and the residents.

AIMS

The first aim of the study was to identify subjects, included in the OCM as cases, who worked in firms (identified using the ReNaM databases) with at least possible occupational asbestos exposure [BRIC 59].

The second aim was to assess the spatial variation of MM risk in the area of Casale Monferrato and to detect clustering associated to the secondary sources of asbestos [LATENT].

METHODS

For the first project [BRIC 59], we considered the ReNaM registries from seven Italian Regions, while the OCM cases were available only for two of them. As ReNaM and OCM could not be merged directly due the absence of a linkage key, a third database (Italian

Social Security – archive) was used, and different matching methods were implemented and evaluated. The outcome was a list (‘dictionary’) of firms with likely exposure to asbestos according to the ReNaM evaluation, identified by official firm names and codes, and asbestos exposure rating. Then, the dictionary was applied to the occupational histories of the cancer cases from the OCM project for the identification of subjects who worked in firms with at least possible asbestos exposure.

For the second project [LATENT], we used data from a population-based case–control study conducted in Casale Monferrato between 2001 and 2006. Demographic and occupational data along with residential information were recorded. Bivariate Kernel density estimation was used to map spatial variation in disease risk while adjusted logistic models were applied to estimate the impact of residential distance from the Eternit plant. Kulldorf test and Cuzick Edward test were then performed.

RESULTS

In BRIC 59, we included 4134 cases of MM recorded in ReNaM, corresponding to 12801 occupational records. The linkage procedure allowed to select 5327 (41.61%) records that were used to prepare the dictionary. After deleting the duplicates, 2606 firms were identified with certain (n=1826, 70.07%), probable (n=222, 8.52%) and possible (n=558, 21.41%) asbestos exposure. Of 35010 cases included in the OCM system, 6546 (18.7%) worked in firms selected for asbestos exposure.

One hundred ninety-six cases and 322 controls were included in the analyses of LATENT project. The contour plot of the cases to controls ratio showed a well-defined peak of MM incidence near the asbestos cement factory, and the risk decreased monotonically in all directions when large bandwidths were used. However, considering narrower smoothing parameters, several peaks of increased risk were reported. A constant trend of decreasing Odds Ratio with increasing distance was observed, with estimates of 10.9 (95% CI 5.32-22.38) and 10.48 (95%CI 4.54-24.2) for 0-5 km and 5-10 km, respectively (reference >15 km). Finally, a significant ($p<0.0001$) excess of cases near the pollution source was identified and cases are spatially clustered relative to the controls until 13 nearest neighbours.

## CONCLUSION

The project has allowed to deepen the study of the effect of two types of asbestos exposure: occupational [BRIC 59] and environmental [LATENT]. Firstly, through the use of available databases (ReNaM and OCM) we improved the knowledge about the firms with at least possible asbestos exposure and the identification of cancer cases related to possible occupational exposure to this carcinogen. Secondly, we analysed the Italian National Priority Contaminated site of Casale Monferrato in terms of spatial variation of MM risk to better clarify the public health impact of the Eternit and asbestos related products.

## PUBLICATIONS

Three papers were published in international journals [Airoldi et al., 2020a; Airoldi et al., 2020b; Airoldi et al., 2021] and the work was presented at different National congresses.

# RIASSUNTO


INTRODUZIONE

Un aspetto importante di sanità pubblica è l'identificazione e il monitoraggio dei tumori causati dall'esposizione professionale o ambientale ad amianto. In Italia, per valutare l'impatto dell'esposizione ad agenti cancerogeni (incluso l'amianto) sono stati implementati diversi sistemi di sorveglianza: un Registro Nazionale dei Mesoteliomi (ReNaM), un sistema di registrazione dei tumori professionali con bassa frazione eziologica (ReNaLOCCAM) e un elenco di Siti Nazionali Inquinati (SIN). Il ReNaM include i casi incidenti di Mesotelioma Maligno (MM) e valuta l'esposizione ad amianto a livello individuale mentre il ReNaLOCCAM si basa su un'attività di ricerca (OCM-monitoraggio dei cancri occupazionali) che, attraverso studi caso-controllo di popolazione, identifica le attività economiche a maggior rischio di tumore. Nonostante OCM analizzi diversi tipi di cancro, i dati sulla storia professionale dei casi sono limitati al nome e al settore industriale delle aziende in cui i pazienti hanno lavorato senza informazioni dirette sull'esposizione ad amianto. Infine, la città di Casale Monferrato (Italia nord-occidentale), dove è stata osservata un'elevata incidenza di MM per la presenza del più grande impianto italiano di cemento amianto (Eternit), è uno dei siti di interesse nazionale per le bonifiche ambientali a causa del suo inquinamento. Diversi studi sono stati condotti nell'area e hanno interessato i lavoratori, le loro mogli e i residenti.

OBIETTIVI

Il primo obiettivo dello studio era identificare i soggetti, inclusi nell'OCM come casi, che hanno lavorato in aziende (identificate utilizzando le banche dati ReNaM) con possibile esposizione professionale ad amianto [BRIC 59].

Il secondo obiettivo era valutare la variazione spaziale del rischio di MM nell'area di Casale Monferrato e rilevare gruppi di casi associati alle fonti secondarie di amianto [LATENT].

## METODI

Per il primo progetto [BRIC 59], abbiamo considerato i registri ReNaM di sette regioni italiane mentre i casi OCM erano disponibili solo per due di esse. Poiché ReNaM e OCM non potevano essere uniti direttamente a causa dell'assenza di una chiave di collegamento, è stato utilizzato un terzo database (INPS - Istituto nazionale della previdenza sociale) e sono stati implementati e valutati differenti metodi di linkage. Il risultato è stato una lista (dizionario) di aziende con probabile esposizione ad amianto secondo la valutazione ReNaM, costituito da nomi e codici ufficiali delle aziende e il grado di esposizione ad amianto. Il dizionario è stato poi applicato alle storie occupazionali dei casi di cancro del progetto OCM per l'identificazione dei soggetti che hanno lavorato in aziende con possibile esposizione ad amianto.

Per il secondo progetto [LATENT] abbiamo utilizzato i dati di uno studio caso-controllo di popolazione condotto a Casale Monferrato tra il 2001 e il 2006. Sono stati registrati dati demografici e occupazionali insieme a informazioni residenziali. Per mappare la variazione spaziale del rischio di malattia è stata utilizzata la stima della densità con un Kernel bivariato mentre per stimare l'impatto della distanza residenziale dall'impianto Eternit sono stati applicati modelli logistici aggiustati. Sono stati quindi eseguiti il test di Kulldorf e il test di Cuzick Edward.

## RISULTATI

Nel progetto BRIC 59, abbiamo incluso 4134 casi di MM registrati in ReNaM che corrispondono a 12801 storie occupazionali. La procedura di linkage ha consentito di ottenere 5327 (41,61%) storie lavorative che sono state utilizzate per la creazione del dizionario. Dopo aver cancellato i duplicati, sono state identificate 2606 imprese con esposizione all'amianto certa (n=1826, 70,07%), probabile (n=222, 8,52%) e possibile (n=558, 21,41%). Su 35010 casi inclusi nel sistema OCM, 6546 (18,7%) hanno lavorato in aziende selezionate per l'esposizione ad amianto.

Centonovantasei casi e 322 controlli sono stati inclusi nelle analisi del progetto LATENT. Il grafico del rapporto tra casi e controlli, utilizzando bande ampie, mostrava un picco ben definito di incidenza di MM vicino alla fabbrica di cemento amianto e il rischio diminuiva in modo monotono in tutte le direzioni. Tuttavia, considerando parametri di lisciamento più ristretti, sono stati segnalati diversi picchi con rischio aumentato. È stata osservata una tendenza costante alla diminuzione dell'Odds Ratio all'aumentare della distanza, con stime rispettivamente di 10.9 (IC 95% 5.32-22.38) e 10.48 (IC 95% 4.54-24.2) per 0-5 km e 5-10 km (riferimento > 15 km). Infine, è stato identificato un eccesso significativo ($p<0,0001$) di casi vicino alla fonte di inquinamento e i casi sono maggiormente raggruppati rispetto ai controlli fino a 13 vicini più vicini.

CONCLUSIONE

Il progetto ha permesso di approfondire lo studio dell'effetto di due tipologie di esposizione all'amianto: quella professionale [BRIC 59] e quella ambientale [LATENT]. In primo luogo, attraverso l'utilizzo delle banche dati disponibili (ReNaM e OCM) abbiamo migliorato la conoscenza delle aziende con possibile esposizione ad amianto e l'identificazione di casi di cancro associati alla possibile esposizione professionale a questo cancerogeno. In secondo luogo, abbiamo analizzato il sito italiano di Casale Monferrato in termini di variazione spaziale del rischio di MM e abbiamo chiarito meglio l'impatto dell'Eternit e dei prodotti correlati ad amianto sulla salute pubblica.

PUBBLICAZIONI

Tre articoli sono stati pubblicati in giornali internazionali [Airoldi et al., 2020a; Airoldi et al., 2020b; Airoldi et al., 2021] e il lavoro è stato presentato a differenti congressi nazionali.

# INTRODUCTION

## Asbestos

Asbestos (from the Greek word meaning “inextinguishable”) is the generic commercial designation for a group of naturally occurring mineral silicate fibers of the serpentine and amphibole series. These respectively include the serpentine mineral chrysotile (also known as ‘white asbestos’), and the five amphibole minerals – actinolite, amosite (also known as ‘brown asbestos’), anthophyllite, crocidolite (also known as ‘blue asbestos’), and tremolite. Chrysotile has long relatively flexible fibers, whereas amphiboles are characterized by shorter, rigid fibers. Natural fiber types sometimes occur in combination.

This group of fibrous minerals have physical characteristics such as large aspect ratio (length to diameter ratio), flexibility, separability and wearability, and chemical and physical durability, that are associated to the definition of asbestos and their industrial use. Asbestos is made of thin but very dense fibers, which confer high mechanical resistance and flexibility. It is resistant to chemical and biological agents, abrasion, and wear. Finally, it has a satisfactory behavior in both thermal and mechanical wear, even at very high temperatures. All asbestos types have been widely used as fireproofing and insulation materials.

Asbestos use was reported since ancient times, mostly for its fire resistance properties. Industrial use begun in the late XIX century, thanks to its unique chemical and physical properties, as well as to its low price and easy accessibility. Its consumption markedly increased from the 1920s until the 1980s [Virta, 2016]. Asbestos was employed in the manufacture of a large list of products, most notably including textiles, building materials, gaskets, insulation, and brake linings. At a global scale, the highest level of asbestos consumption occurred in 1977, when approximately 4.7 million tons were reached worldwide. Then, asbestos health risks triggered country-wide bans and stringent regulations, which resulted in a worldwide asbestos consumption decline until the late 1990s, when it leveled at two million tons, a consumption level that has been maintained since then with some minor fluctuations [Valenzuela et al., 2016].

The occasions of asbestos contact can be different and generally are classified in occupational, environmental, and domestic, according to the different exposure sources.

Occupational exposure occurs in the mining and milling of asbestos, the manufacturing or use of products containing asbestos, e.g., in construction, automotive industry, the asbestos-abatement industry and in other industries using asbestos as a raw material or in an industrial setting. The general population is exposed to asbestos fibres from outdoor air (environmental exposure) or via domestic contact (familiar/household exposure). Environmental exposures include among others industrial emissions, natural outcroppings, or erosion of asbestos-containing building materials. The term household exposure includes all exposure opportunities in the domestic setting: the installation, degradation, removal, or repair of asbestos-containing products in the domestic environment both with the para-occupational exposures, i.e., those exposures brought home by asbestos workers or due to the use of asbestos-containing tools and products in the home [IARC, 2012].

## Asbestos-related diseases

The summary of human evidence stated explicitly that all forms of asbestos are carcinogenic for humans causing mesothelioma, lung, laryngeal, and ovarian cancer, as well as non-cancer diseases such as asbestosis, and pleural fibrosis/plaques. The evidence of a causal relationship was judged to be limited for cancers of the pharynx, stomach and colorectum [IARC, 2012]. According to global estimates, about 125 million people in the world are exposed to asbestos at the workplace and at least 255000 people die each year from asbestos-related lung cancer, mesothelioma and asbestosis resulting from occupational exposures [WHO, 2014; Furuya et a., 2018].

Mesothelioma is commonly recognized as the primary asbestos-related cancer type. Malignant mesothelioma is an aggressive, diffuse, fatal, asbestos-associated neoplasm originating from the lining cells (mesothelium) of the pleural, peritoneal or, rarely, pericardial cavities around the lung tissue, abdominal cavity and heart. Despite the introduction of modern therapeutic interventions, it has a very poor prognosis with a median survival from the time of presentation of approximately 9-12 months [Robinson et al., 2005].

In Italy, in December 2018, 31572 cases of mesothelioma referring to the period of incidence between 1993 and 2015 were recorded in ReNaM: 93.2% of them affected the pleura while 6.3% involved the peritoneum. Moreover, in 2017, the pleural malignant mesothelioma incidence rate was 3.41 and 0.99 cases per 100,000 person/years in men and women respectively [Marinaccio et al., 2021]. Even though it is not a common disease, it is of extreme relevance due to its known relationship with a well-defined cause (asbestos). A case of malignant mesothelioma (MM) may be considered a "sentinel event", a warning signal that in a population there is a presence of asbestos. A large body of literature exists on the risk of MM after occupational exposures to asbestos, as summarized in several reviews [IARC, 2012; Hodgson and Darnton, 2000; Magnani et al., 2013]. The consequences of para-occupational exposures, i.e. those experienced by family members of asbestos workers, have also been widely studied [Goswami et al., 2013; Xu et al., 2018]. Neighborhood exposures had already been associated with MM in the seminal study by Wagner et al. [Wagner et al., 1960]. Later, several epidemiological studies detected an increased MM risk in residents living near asbestos mines and asbestos products factories [Berry, 1997; Hansen et al., 1998; Howel et al., 1997; Kielkowski et al., 2000; Kurumatani and Kumagai, 2008; Magnani et al., 1995; Magnani et al., 2000; Maule et al., 2007; Musti et al., 2009; Newhouse and Berry, 1976; Rees et al., 1999].

## Italian surveillance system

Italy, due to long-term asbestos use and consequent excess of MM incidence and mortality, is one of the countries more sensitive to the issues of monitoring and preventing asbestos-related health effects.

The Italian law on the protection of Workers (D.Leg. 81/2008) includes three systems for the registration of occupational cancers based on: the National Mesothelioma Registry (ReNaM) [Marinaccio et al., 2021], the National Sinonasal Cancer Registry (ReNaTuNS) [Binazzi et al., 2020] and on a registration system for occupational cancers with 'lower attributable fraction' (ReNaLOCCAM) [Massari et a., 2021]. ReNaM and ReNaTuNS are active surveillance systems and refer to occupational cancers with 'high attributable fraction', respectively mesothelioma and sinonasal cancer. Otherwise, the active research of other occupational cancers (lung, bladder, laryngeal, leukemia and lymphomas, in addition to skin cancers and melanomas) presents some implementational difficulties due to lower occupational etiological component and very large overall number of cases per year, so a different approach has been implemented.

Moreover, in Italy, the Health Ministry (D.Leg. 52/1992) proposed the National strategic program "Environment and Health" that included the SENTIERI project (Epidemiological Study of Residents in National Priority Contaminated Sites). The aim of the project was the evaluation of epidemiological evidence on the association between 63 causes of death and selected environmental exposures (including asbestos) to find sites of national interest for environmental remediation, defined as Italian Polluted Sites (SIN). One of these Italian National Priority contaminated sites is the area including the city of Casale Monferrato (northwest Italy) and 48 surrounding villages, where a high incidence of MM was observed.

## Italian Law on the protection of Workers

### ***ReNaM***

The Italian National Register of Malignant Mesotheliomas (ReNaM) is a permanent surveillance system of malignant mesothelioma occurrence, with Regional Operating Centers (CORs) active in each Italian region, that identifies mesothelioma incident cases and assesses asbestos exposure at the individual level.

CORs actively search for MM incident cases in health care institutions that diagnose and treat cases (especially pathology and histology units, pulmonology, and chest surgery wards), analyze the pathological diagnosis and classify cases in four categories: certain, probable, possible and no MM. Subjects included in the first three groups are registered in ReNaM. A standardized questionnaire is administered by a trained interviewer to the subjects (direct interview) or the next of kin (indirect interview). The questionnaire includes sections on demographic characteristics, lifelong occupational and residential histories, selected leisure time activities and characteristics of the home environment possibly relevant for asbestos exposure. It also includes a brief clinical history with focus on occupational chest diseases. Occupational exposure to asbestos is classified as definite, probable, or possible. Definite occupational exposure is assigned to the subjects whose work involved the use of asbestos or asbestos containing materials. Probable occupational exposure is attributed to the subjects who worked in a company where asbestos was certainly used, but whose exposure cannot be documented, and possible occupational exposure to the subjects who have worked in a company referring to an economic sector where asbestos has been used but without more direct information [ISPESP, 2003; Marinaccio et al., 2020].

## ***ReNaLOCCAM***

The ReNaLOCCAM is a registration system of occupational cancers with ‘lower attributable fraction’. It was started as a research activity (OCM – Occupational Cancer Monitoring) for identifying occupational cancer risks and cancer cases in Italy that are likely to be of occupational origin using information available in the Italian Social Security (INPS) archives both with administrative health data and, subsequently, contributed to the design of the registry.

In details, OCM is a set of population-based case-control studies, where cases are obtained from cancer registries, regional mortality archives and regional hospital discharge records while controls are sampled from the electronic archives for National Health System. Cancer cases are drawn from routinely available sources and are of both genders, aged 35-69 with a social security occupational history. Subjects aged more than 69 are excluded because too much time has elapsed since retirement, whereas cases younger than 35 are excluded as their occupational origin is unlikely considering the short latency. Controls are identified by random sampling of the case base, i.e. the National Health System archives of the same areas and calendar year as the cases. Only subjects with occupational information are included in the study, considering only people who had worked in the private sector, because of the corresponding limit of the INPS database. Then, employment histories, consisting of company names, industrial sector codes, and periods of employment, are obtained by automatic linkage to the Social Security (INPS) files. An individual is considered as “exposed” to a given industrial sector if he/she worked for a company in that sector for at least one year. People employed in banks, shops, hotels and restaurants, insurance, education, and social services are considered “unexposed” [Crosignani et al., 2006]. The OCM project estimates the cancer risks in different primary sites by economic activity of private sectors.

Then, the results obtained are sent to the operative regional centers in terms of economic sector with increased risk of cancer. Ideally, industrial hygienists should verify the occupational exposure of patients and implement public health policy. Currently, due to the large numbers of cases identified and organizational difficulties, no cases interview were done so the ReNaLOCCAM is not completely active.

### *Gap in knowledge*

The ReNaM provides a rich databases of mesothelioma cases with an accurate description of their occupational exposure to asbestos. Particularly, it is possible to obtain a list of industries with evidence of asbestos exposure, based on precise and individual information from detailed interviews. On the contrary, OCM analyzes different cancer types and data on cases' occupational histories are limited to name and industrial sector of firms where patients worked. No information on asbestos exposure is directly available in the OCM data.

If we were able to link the two databases we could identify cancers, other than MM, related to a possible asbestos exposure of occupational origin in an automatic way. Unfortunately, ReNaM and OCM could not be merged directly due to lack of information standardized reporting and the absence of a linkage key. Therefore, one of the objectives of the current project was to devise a linkage procedure to get as much information as possible from the ReNaM registry for the estimation of asbestos exposure of OCM subjects.

## Italian National Priority Contaminated sites – Casale Monferrato

The SENTIERI Project studies mortality of residents at sites of national interest for environmental remediation (SIN). SINs correspond to municipalities located in the vicinity of industrial areas, either active or dismissed, near incinerators or dumping sites of industrial or hazardous waste. [Pirastu et al., 2010]. Asbestos is specifically mentioned in the regulatory acts of recognition for 10 NPCSs and it is the only agent that has determined environmental contamination in 3 of them (Casale Monferrato, Broni and Bari) [Zona et al., 2016].

The area of Casale Monferrato (northwest Italy) is of particular interest due to the presence of the largest Italian asbestos cement (AC) plant, owned by Eternit, that was active from 1907 to 1986. The size of the work force varied over time and was up to 1500 workers. In 1981 the company reported the use of 15000 tons of asbestos (10% crocidolite) [Magnani et al., 1996]. In the period 2000-2011, 558 cases of MM were recorded in subjects who lived in the area and the median age a diagnosis was 67 and 74 years for males and females, respectively. The standardized incidence ratio was 911 [90%CI 831-996] for males and 1.34 [90%CI 1.20-1.49] for females. Two hundred thirty-five cases with occupational exposure were mainly attributable to the industrial sectors of asbestos-cement production, construction and engineering industry. Moreover, 109 cases were identified as residential exposed as they lived near the Eternit plant and 87 patients had a familial exposure.

In this area where the incidence was 10 times higher than the corresponding Italian rates, [CPO] as well as the results reported by SENTIERI, several studies have been conducted regarding the AC workers [Magnani et al., 2008], their wives [Ferrante et al., 2007] and the residents [Ferrante et al., 2016]. Moreover, Magnani et al. [Magnani et al., 2001] and Maule et al. [Maule et al., 2007] in a case control study including cases diagnosed in 1987 - 1993 provided strong evidence of an increased MM risk related to the distance of dwellings from the Eternit plant, suggesting a causal role of the asbestos environmental pollution. Residents near the plant had an odds ratio (OR) for mesothelioma of 10.5 (95% CI 3.8-50.1), adjusted for occupational and domestic exposure. Furthermore, the results of spatial clustering tests gave some support for the hypothesis of exposure (and risk) associated with secondary sources of asbestos. Such secondary sources included some separate, accessory departments of the factory (i.e. the finished products warehouse and rail yard, both located on the eastern side of downtown Casale Monferrato while the factory was on its western border), the transport systems across the town of raw asbestos and AC products and a landfill for asbestos wastes discharge. They also included private buildings, yards and lanes where AC tailings from the work process had been incorrectly disposed of: for instance, to cover small and sometimes large land areas after having been finely broken, or as fine dust (locally known as "polverino") as attic insulation [Arpa].

### *Gap in knowledge*

Despite the area was already studied, further assessments were needed. We wished to investigate clustering consequent to secondary sources exposure, by taking advantage of the more recent and larger dataset of our 2001-2006 case-control study [Ferrante et al., 2016]. We also wanted to overcome some limitations of the analyses of our first spatial case-control study [Maule et al.,2007], namely the fact that the estimates of environmental exposures had been adjusted for occupational and domestic exposure by classifying such exposures only in a dichotomous way and that allowance had been made only for a fixed lag time period of 20 years.

# AIMS

The first aim of the study was to devise a procedure for the identification of firms with asbestos exposure using the information available in the National Mesothelioma Registry – ReNaM. The purpose was to identify subjects, included in the OCM as cases, with at least possible asbestos exposure as they worked in this firms with possible occupational asbestos exposure. This activity was a task of the **BRIC 59** project funded by the Italian Ministry of Health and coordinated by the Italian National Institute of Health.

The second aim, corresponding to the **LATENT** (environmenta**L** Asbes**T**os **E**xposure malig**N**ant meso**T**helioma) project, was to assess the spatial variation of MM risk in the asbestos exposed area of Casale Monferrato and to detect clustering associated with secondary sources of asbestos. We used individual data generated by a case-control sampling design [Ferrante et al., 2016] and we applied different statistical approaches to study the environmental asbestos exposure.

# MATERIALS AND METHODS

## BRIC 59 project

The first part of the project (BRIC 59) consisted in the analysis of ReNaM registries from seven Italian Regions (Piedmont, Lombardy, Emilia-Romagna, Tuscany, Lazio, Apulia and Sicily) and their linkage with OCM, available only for Sicily and Lazio. The other Italian regions, despite being contacted in several ways, did not participate at the project due to lack of time and/or resources.

The project was conducted following four steps:

1. MM cases included in the ReNaM database were linked to the INPS database, in order to capture the firm names where they worked.

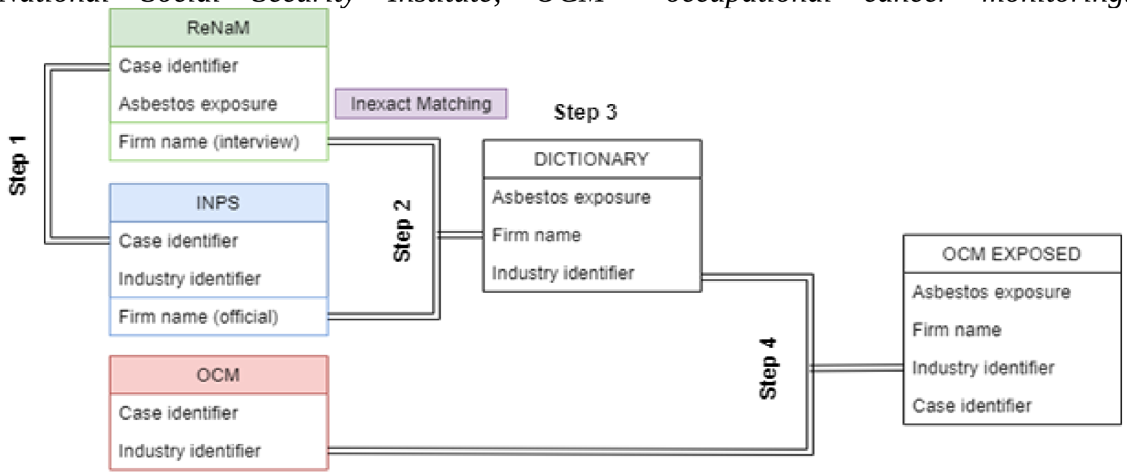
2. The firm names (reported with possible imprecisions in the ReNaM's interview) were associated with the official firm names recorded in the INPS database. The procedure was devised to capture the similarities in firm names using different linkage methods.

3. Firms with likely exposure to asbestos according to the ReNaM evaluation were listed, identified by INPS firm names and codes, and asbestos exposure rating.

4. The list (thus called 'dictionary' in the following) was applied to the occupational histories of the cancer cases from the OCM project for the identification of subjects who worked in firms with possible asbestos exposure.

Figure 1 summarizes the flow of different record linkage activities.

*Figure 1 Graphical description of the different record linkage activities. INPS— Italian National Social Security Institute; OCM— occupational cancer monitoring.*



## Step 1. INPS records

To obtain the "official" working histories of subjects with MM, the required databases were the registry of the MM cases (ReNaM) and the records of pension contributions (INPS). The ReNaM database, for this step, provided the names and dates of birth of cases recorded in the period 1993-2012 for the regions included in our project. We limited the analysis to the MM cases with complete demographic data and evidence of work periods with occupational asbestos exposure according to ReNaM evaluation. The INPS database included all pension contributions from private firms from 1974. Quality of INPS data was considered high as it is used to check the eligibility criteria for pension benefits. Contributions from the public sector, agriculture, military forces, and self-employed workers were not included, as they were administered by other institutions. For each subject the INPS database recorded all the firms (in term of official registration name and code) in which he/she had worked.

The linkage was conducted nominally. The MM cases' names and dates of birth from ReNaM were linked to the INPS database. The output was the list of firms where each MM case had worked. It included all firms where a MM case had worked and paid pension contributions, with dates of entry and dismissal, the official name of the firms, the address, the Value Added Tax (VAT) code, the economic activity code (coded using both ATECO 91 and ATECO 81), and the job duties of the index subject in broad categories (blue or white-collar).

## Step 2. ReNaM and INPS

The data entered in step 2 were the records of the job history of MM cases from ReNaM and the list of firms where the MM cases had worked according to INPS (output from step 1). For the purpose of the project, we had to transfer the information on asbestos exposure rated by ReNaM to the official firm names as in the INPS database. The activity was complex because the firm names in the ReNaM records were obtained during interviews and were not standardized.

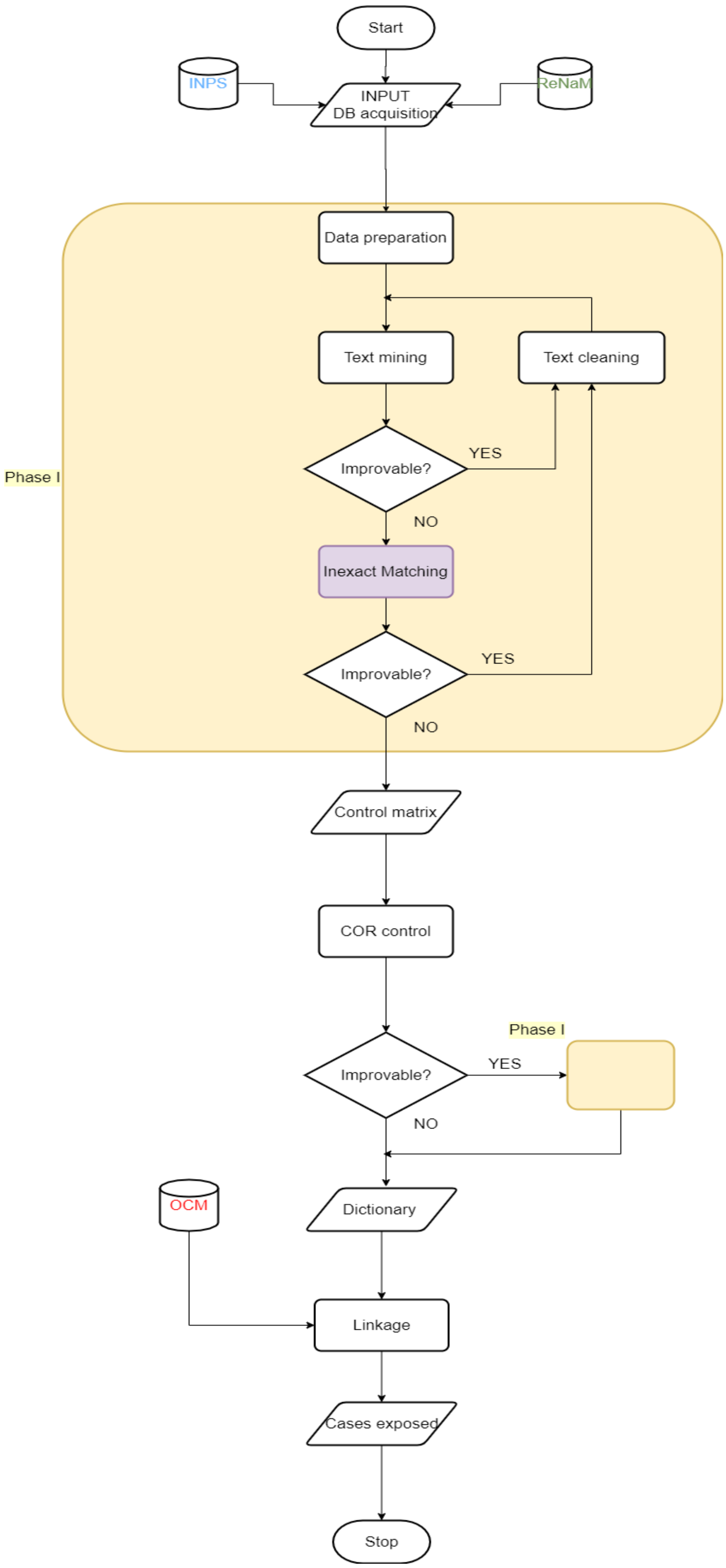
Particularly, the ReNaM database reported the names of the firms in a text string composed by open words. As always happens in interviews, imprecisions in the wording of firm names were very common and subjects rarely reported exactly the official firm names used by INPS. ReNaM occupational hygienists' rated the level of asbestos exposure for each firm from the job descriptions collected at interview, but kept the firm names as they were reported, with no standardization. The methodology of ReNaM for asbestos exposure assessment was described in detail elsewhere [Marinaccio et al., 2015; ISPESL, 2003]. Instead, INPS recorded the official firm's name with a unique identifier.

The process implemented consisted in the link of the ReNaM database to the INPS obtained from step 1, by firm names. The linkage cannot be deterministic, due to the different collection of firm identifiers in the ReNaM (interview of MM cases) and in the INPS (official administrative registration) databases, as explained before. Moreover, a ReNaM record can have zero, one or more linkages with INPS.

A methodology had to be devised and applied to maximize the number of records matched. To perform this, we first evaluated and amended the data wording, implementing a recursive and iterative pre-processing procedure (text mining and text cleaning [Silge and Robinson, 2017]). Second, we linked the databases using different approaches and, third, we sent the results to the different ReNaM regional centers for a check by local experts. This last step was used to evaluate and supervise the performance of our algorithm, examining the linked and non-linked data. We also checked a random sample of records to estimate the percentage of correct matches. The procedure was repeated after receiving the amendments by the regions. Eventually, we prepared the final list of asbestos - using firms with their official name and INPS registration code.

The procedure was implemented separately for each region, and the partial results obtained for each unit were used to increase the capability and competence of the algorithm. The flowchart in Figure 2 shows the procedure, and further details are provided in the next paragraphs.

*Figure 2. Flowchart of the record linkage process.*



## ***Step 2.1: Text Mining***

Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection by using a suite of analysis tools. In a manner analogous to data mining, text mining seeks to extract useful information from data sources through the identification and exploration of interesting patterns. In our case it helped us to understand and discover patterns and anomalies of the firms' names.

Prior to more sophisticated processing, we split the firms' names into separated words (tokens/grams), in a process called "tokenization". Then, we decided to analyze our words evaluating word similarity, term frequency, word difference using the log odds ratio (OR) and the bi-grams.

### *Word similarity*

The term "word similarity" indicates how similar are two words (tokens) in a semantical way, that is, how close they are. Generally, the tools that evaluate the word similarity examine inflections, synonyms, and stemming. The term stemming indicates the modifications of a word to express different grammatical categories such as tense, case, voice, aspect, person, number, gender, and mood. In our study, particularly for ReNaM records, we were interested in the identification of spelling mistakes or typographical errors that were reported during the interviews of cases and to the identification of the proper nouns. Moreover, we wanted to find the singular/plural forms or the different verbs conjugations that needed standardization.

To investigate the words that shared a common part and similitudes between terms, the Sound Like operators were used both with other indexes [Roesch, 2012]. Particularly we implemented the soundex algorithm, the spelling distance, the generalized edit distance and then proportion common letters algorithm.

The soundex algorithm converts each word to a code according to a rule that was originally developed by Odell and Russel [Odell, 1956]. Particularly, it returns a copy of the argument that is encoded having same key for similar sounding words in English language. The algorithm keeps the first letter of the string and discards any of subsequent letters of (A E H I O U W Y) in the string. Remaining letters in the string are encoded by numbers 1 - 6 grouped as follows: (1 for B F P V; 2 for C G J K Q S X Z; 3 for D T; 4

for L; 5 for M N; 6 for R) while discarding two or more adjacent letters in the same group. Then, different words with equal codes could be compared to evaluate if they were equal in semantic term.

The spelling distance determines the likelihood of two words matching, express as the asymmetric spelling distance between the two words. It is calculated as the normalized "cost" for converting the first word to the second one by using a sequence of operations. Particularly, the operations include character insertion, deletion and replacement and different "costs" are associated at each conversion. Then these "costs" are summed and divided by the length of the query. All possible ways to translate the first word into the second were calculated but only the smallest was considered. Lower values indicate more similar terms [Roesch, 2012].

The generalized edit distance is the measure of dissimilarity between two strings and can be considered a generalization of Levenshtein distance [Levenshtein, 1965]. It calculates the sum of "costs" of converting the first string into the second string. The rationale for determining the generalized edit distance is based on the number and type of typographical errors that can occur.  Both with the "cost", we can only calculate and compare the number insertions, deletions, or replacements. In a similar way to spelling distance where "costs" were different, low numbers suggest semantic similarities.

The proportion common letters (PCL) algorithm calculates the sum of the number of letters that match in the two character strings from "right to left" and "left to right". Then, this value is divided by the total number of letters in the two strings. Values near 1 indicate that two words compared are similar [Rai, 2012]

These procedure were able to identify some errors as "FIATT" instead of "FIAT", or singular/plural terms as OFFICINA/OFFICINE, and so on.

### *Term frequency*

The term frequency (TF) measures the number of times that a word appears in a text divided by the number of terms in that text; this frequency is the base of other math formulas and indices. In our project, the TF was useful to understand which words were very common in ReNaM and INPS and could be removed as they were too much generic.

The approach was also able to identify the firms more frequent, in this case no modifications were done.

A brief analysis of term frequency was performed separately for each region and the terms with higher frequencies were plotted using Pareto charts and word clouds. Finally, we prepared a list of stop words, terms which are not specific or discriminatory to perform a linkage as “IMPRESA”, “NEGOZIO”, and a list of industries where many people worked as “PIRELLI”, “OLIVETTI”, and so on.

### *Word difference using the log odds ratio (OR)*

To compare the word usage among two documents and identify which words are more or less likely to come from each document, a common approach is the calculation of how many times each database used each word. In our analysis, this step allowed to find specific words for each data source and to identify terms with different syntax but similar semantic meaning.

Particularly, the word difference was evaluated implementing the log(OR) as $\ln\left(\frac{\left[\frac{n+1}{total+1}\right]_{INPS}}{\left[\frac{n+1}{total+1}\right]_{ReNaM}}\right)$, where n is the number of times the word in question is used by each dataset, and the total indicates the total words for each dataset. Values higher than 0 indicates that the word is more frequent in INPS while values lower than 0 show that it is more used in ReNaM.

We analyzed the 10 most distinct words for each dataset to reduce the number of records and we plotted them to better understand which words were more common in one registry. The procedure was able to detect some different way to indicate the same firms, such as: “FIAT” in ReNaM and “FCA” in INPS.

### *Bi-grams*

Previously described approaches use the single word as unit, but terms that tend to co-occur within a database are also of particular interest. Particularly, tokenizing by bi-grams is a useful way to explore pairs of adjacent words and examine which words tend to immediately follow others. There are advantages and disadvantages to examining the bigrams rather than individual words. Pairs of consecutive words might capture structures that are not captured when counting single words and may provide context than makes tokens more understandable. However, the level of complexity grows up and strings of text with only one term cannot be analyzed with this procedure.

In our analysis we searched which terms were never reported alone through the comparison of bi-grams and single words. The methodology applied was very useful to identify terms as for example “CASALE MONFERRATO” that could be considered inseparable.

## ***Step 2.2: Text Cleaning***

Text cleaning is the process of preparing raw text for further analysis, in our case for linkage procedure. Using the information provided by the previous text mining, we preprocessed the text data using different approaches as describe below.

### *Number and delimiters’ removal*

All character variables were converted to uppercase to avoid case-sensitive problems. Numbers were removed from the strings (without replacing with blanks). The stressed vowels were replaced with the corresponding vowels without accents.

Several choices have been made about the delimiters, based on the previous data mining results. We decided to replace some of these (“” & -) with a blank while others (. , / * ‘) were removed without spacing.

### *Record removal*

The records not useful for the “dictionary building” were removed. In INPS, records without the Value Added Tax code and without company name were deleted. Generally, these units refer to agricultural sector and the company name was reported as “SA AGRICOLO”, where SA corresponds to anonymous corporation and it included the direct growers. Moreover, records that reported no specific company name as “UNSPECIFIED”, “UNKNOWN" or “IGNOTO” and so on were not included in subsequent analyses.

### *Stop words removal*

In the databases analyzed (ReNaM and INPS) we found many common words that occurred many times and were not specific or discriminatory for the linkage steps.

Although there are some automatic and software-implemented functions that remove the common words that usually do not contribute to the semantics of the documents and have no real added value, we decided to prepare an ‘ad hoc’ list of stop words, based on the results of previous text mining procedures. First of all, we deleted the majority of Italian prepositions and articles such as “DI”, “DEL, “E”, “IL” and so forth. Then, we removed words that were irrelevant to the categorization task and can be dropped with no harm to the classifier performance, although in a normal context can be interesting (as SOCIETA’, DITTA, AZIENDA). Excluding these words may even result in improvement of linkage, owing to noise reduction. Moreover, when a job position was composed only by not specific or useless terms, the record was first manually controlled and eventually deleted.

### *Short and repeated words removal*

The strings with repeated words were modified deleting the double term. For example, the familiar society “MARIO ROSSI e ANTONIO ROSSI” was replaced with “MARIO e ANTONIO ROSSI”

We also removed the words with length shorter than two letters because we did not consider them very useful for the linkage steps. One unique exception was made for the acronyms of the company written using comma or dot because in these situations the procedure splits the name into more than one column. A specific command was implemented to recognize this type of strings and avoid the mistakes. For example, “I S A C” was merged into “ISAC”.

### *Term correction*

Manual text normalization was then implemented using all the results obtained in the text mining phases. Both with the correction of mistakes and singular/plural terms, other changes were made using all the knowledge available. For example, “BREDA”, “ANSALDO BREDA” or “ANSALDO TERMOSUD SPA” refer to the same company and they were harmonized.

### *Tokens join*

After the analysis of the bi-grams and the study of their relationships with the string composed by one term we decided to join the words never used alone. Particularly, when two terms were present always together (they always co-occur) were joined into a single token to improve the performance of the linkage procedure (as explained after). The most common example was “Casale Monferrato” that was replaced in “CASALE_MONFERRATO”.

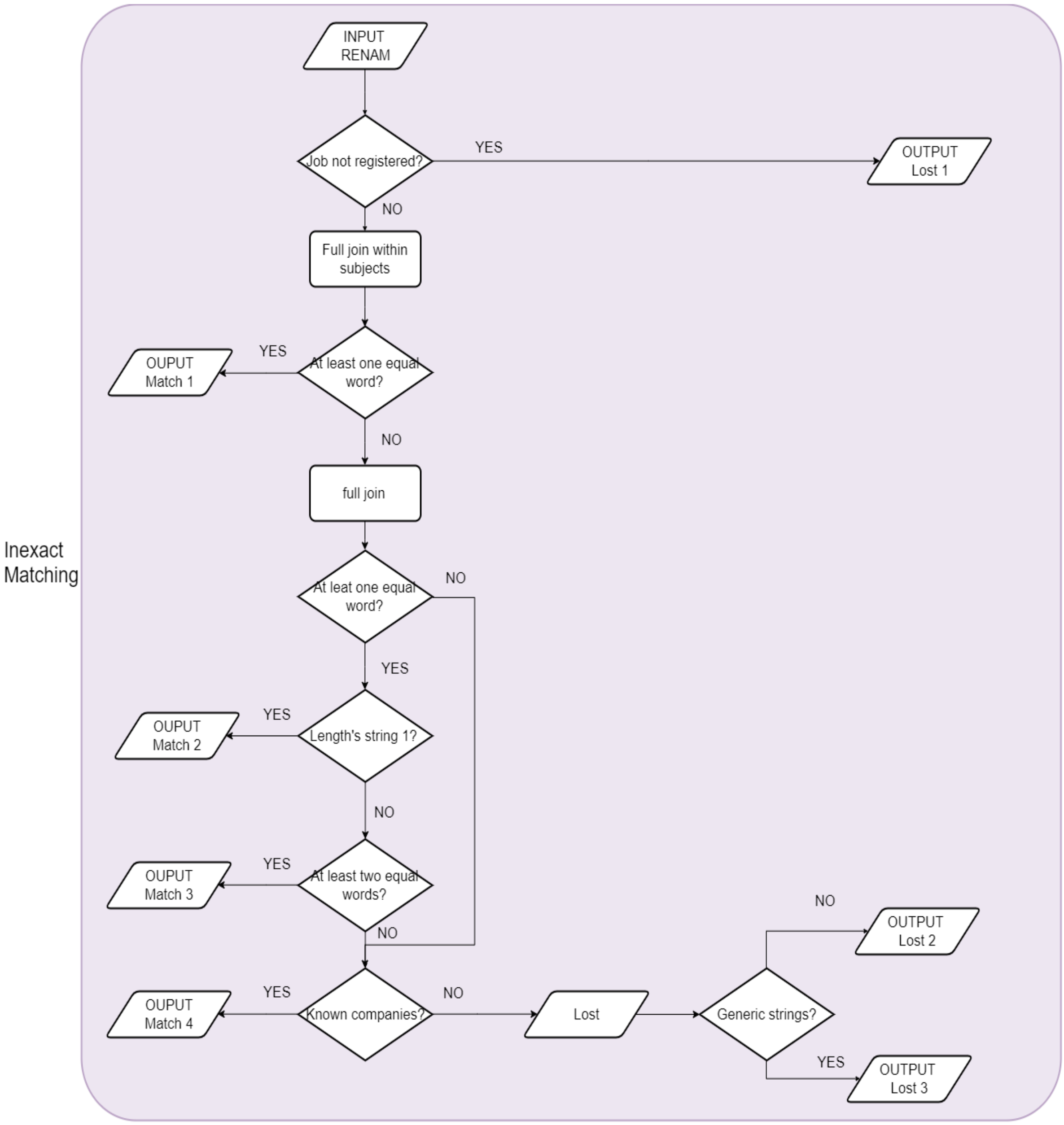
### ***Step 2.3 Inexact Matching Linkage***

The record linkage between the firm names was used to add the ReNaM evaluation of asbestos exposure to the INPS information. The term “inexact matching” refers to the fact that we used as linkage keys the single words (tokens) of the firm’s name and not the whole string to maximize the number of records linked within the regions.

The procedure was firstly created using Piedmont data and then it was tested on Tuscany data. Then, other improvements and modifications were done based on the other regions included in the project.

Before the matching, some categories, such as military, firemen or the agricultural sector were deleted from ReNaM as they were not recorded in INPS database. Then, the linkage algorithm was designed starting from the ReNaM dataset and performing different steps as shown in figure 3 and explained below.

*Figure 3. Record linkage flowchart: "inexact matching steps".*




First, we performed an inexact matching within-subject: we considered a correct match when the two datasets shared the cases code and at least one significant word of the firm name. Other variables could be included in the procedure to improve the success of linkage procedure, as for example the time period. However, time period was not used, as start and end date in ReNaM could be affected by recall bias.

Second, a full join was performed using the records unmatched in the previous step: all the remaining rows from ReNaM were combined with all records from INPS. The procedure was time-consuming and, to apply to our big databases, we split ReNaM into subgroups of 10 rows, selecting only the records that shared at least one word. The remaining records were used and re-controlled in the following step. We decided to consider a successful match when 2 words of the firm name were equal. The choice to use 2 words as a threshold was related to empiric criteria: 1 was too little (excess of sensitivity), and 3 was too big (excess of specificity). In some cases, after the preprocessing phase, the strings were reduced, and they were composed of only one term. In this case, a match was considered successful when the single word present in ReNaM was found in an INPS record.

Third, we searched into the remaining not linked records, the known companies that could be registered with more than one name or with a noun accompanied by other terms. Examples were "ENEL" or "PIRELLI", big firms with a large number of departments and subgroups. To avoid the possible loss of the records, we forced a list of the company's names known 'a priori'.

The ReNaM remaining unmatched records were reorganized into 3 groups: 'a priori' (lost 1), 'non generic terms' (lost 2), and 'generic term' (lost 3). The latter referred to a strings without specific nouns or element to identify a society in a unique way. For example, "COMUNE" was considered generic while "COMUNE DI CATANIA" specific. These matrices were manually controlled by the regional experts to find some recurrent mistakes and to improve the ability of algorithm to match the higher number of records in correct way. Some examples of modification done in this step are the "SIA" that is the "SOCIETA' ITALIANO AMIANTO" or "ORSI MANGELLI" named also "SAOM", "SIDAC" and so on.

The other outputs (match1-4), composed by the "successful" matches, were not controlled by the regions due to the great amount of work; it is time-consuming and impractical when applied to large data sources. As the accuracy assessment is a key component of record linkage, we extracted a sample of records and we manually verified and validated matched pairs to further refine linkage parameters and increase its overall effectiveness.

## Step 3. Dictionary

The result of the merge procedures among ReNaM and INPS was a general table ("dictionary"), including all firms with evidence of asbestos exposure. Each firm record included: the official name of the firm from INPS, its official administrative registration number, the rating of exposure to asbestos (certain, probable, possible) according to ReNaM experts and the firm name as reported at the ReNaM interviews. Duplicates were reduced to a single record, keeping the higher asbestos exposure rating.

The firm addresses were geocoded [Weber and Pèclat, 2018], and the latitude and longitude coordinates were used to design an Italian map of asbestos exposure industries.

## Step 4. OCM and dictionary

The application of the dictionary of asbestos exposing firms to the working histories of cancer cases from the OCM project was conducted for the 2 Italian regions (Lazio and Sicily), according to the OCM data available at the time of the research. All cancer types associated or possibly associated with asbestos according to IARC [IARC, 2012], were included. The information recorded for each case were: the diagnosis, the date of diagnosis, and the occupational history as provided by the INPS (firm names and codes). The output was the list of cases with at least a firm registration number present in the dictionary, i.e., the subjects who worked in a probably asbestos-exposed industry and who are eligible for a reimbursement based on Italian Law.

## Statistical analysis

Descriptive statistics were conducted reporting the absolute and relative number of cases and records for each database. Moreover, results of linkage procedure separated for the different steps were indicated. These results were calculated considering the overall Italian data available and separated for each region.

To evaluate the spatial distribution of firms with at least possible asbestos exposure, an Italian map was drawn using different colors based on a firm's region provenience.

All analyses were performed using SAS 9.3 (SAS Institute Inc., Cary, NC, USA.), R version 3.4.1 (R Core Team, Vienna, Austria), STATA 11 (StataCorp LLC, College Station, TX, USA).

# LATENT

## Study design

The present analyses used the data collected in a case-control study on pleural MM conducted in Casale Monferrato Local Health Authority (LHA) between 1 January 2001 and 30 June 2006 [Ferrante et al., 2016]. The LHA is the basic administrative unit of the Italian National Health Service. At the time of the study, Casale Monferrato LHA included the town of Casale Monferrato and the surrounding area, for a total of 60 towns and villages of different size; this area was little wider than Casale Monferrato SIN. The total resident population on December 31, 2001 was 117,680, (of whom 35,238 lived in Casale Monferrato). Population rosters of residents, maintained by the LHA for administrative purposes, were used to assess eligibility of cases and to sample controls. The map with names and boundaries of the towns included in the study is reported in appendix 2.

Cases of histologically verified MM of the pleura were actively searched in the departments of pathology, respiratory medicine, oncology, internal medicine, thoracic surgery, and radiotherapy of the hospitals serving the study area. Controls were randomly selected from the population rosters of the LHA of Casale Monferrato. Two hundred twenty-three incident cases and 552 age (± 18 months) and gender- individually matched controls were identified. To increase power in the younger age classes, the case control ratio was 1:2 for cases 60 years and older, and 1:4 for younger cases.

## Questionnaire and exposure assessment

Each subject (or the next-of-kin for decedents or subjects unable to answer) was interviewed using a structured questionnaire administrated by trained interviewers. Demographic characteristics, lifelong occupational and residential histories, selected leisure time activities and characteristics of the home environment possibly relevant for asbestos exposure were recorded. A short clinical history with focus on occupational chest diseases was also included. A lifelong occupational history was elicited, including, for each job, the job title, industry, and dates of beginning and ending. Further details on occupational activities were collected using job specific modules. Information collected for each dwelling included: address, a set of questions about the use of asbestos material in the house and its presence around it (roof coverings, courtyards, gardens) and a description of the neighborhood environment. A checklist was also used to investigate the presence of selected industrial activities nearby. Information was also gathered on family members and cohabitants. This section started with an abridged occupational history for each family member, supplemented by a checklist of occupational activities known to have entailed asbestos exposure. Positive answers were further investigated by using a particular specific module, including questions about work clothes and other materials possibly brought home.

The evaluation of asbestos exposure in term of frequency, intensity, and duration, separated for occupational, environmental, domestic sector, was then conducted blindly to the case/control status by an experienced rater, on the basis of interview data. As individuals usually had multiple exposure circumstance (work and not-work related), the assessment took into account their whole exposure history and computed cumulative exposure indexes, reflecting the contribution of all sources. Cumulative exposure was expressed in fibres/mL-year. More details on data collection procedures and exposure assessment have been published elsewhere [Ferrante et al., 2016].

## Geographical classification

All residential addresses reported at interview, generally more than one for each subject, were geocoded as Universal Transverse Mercator (UTM) geographic coordinates using a Global Positioning System (GPS) accessed from Google Maps. Addresses missed at first instance were manually checked, amended whenever appropriate, and resubmitted to geocoding process. The geographic coordinates of the AC factory location were determined in a similar manner. Moreover, geographical coordinates of Casale Monferrato and towns and villages of LHA's boundaries were downloaded from the Italian National Institute of Statistics [Istat, 2021]. The factory is in the NW area of Casale Monferrato town, at about 1500 m from the center and 250 m from the closest residential areas.

The dwelling address was chosen as a proxy to residential exposure. The dwelling with the longest residency duration (main residence) after excluding those outside the LHA boundaries was used as the location of study subjects for the analysis. The time window of interest was defined excluding the last 20 years before diagnosis to exclude any exposure occurring during the preclinical phase of the disease. The 20 years period was chosen as an extreme value, following the previous analyses in the same setting [Maule et al., 2017; Magnani et al., 2001]. As the duration of that interval is not known with precision for MM we also conducted analyses with consideration of shorter periods, in particular using the 10 years period customarily used in other studies [Newhouse and Berry, 1976; Health Effects Institute, 1991]. A further sensitivity analysis was carried out using as exposure proxy the closest residence, with reference to the distance from the Eternit factory.

## Statistical analysis

Categorical variables were summarized using absolute and percentage frequencies, while mean and standard deviation or median and interquartile range were used for numeric variables. Results were presented for all subjects and separately for cases and controls.

The geodetic distances of each subject location to the AC industry (pollution source - foci) were calculated. Maps with point locations, AC plant and Casale Monferrato boundaries were produced. Analyses were focused on the Casale Monferrato LHA, and residential addresses outside it were disregarded.

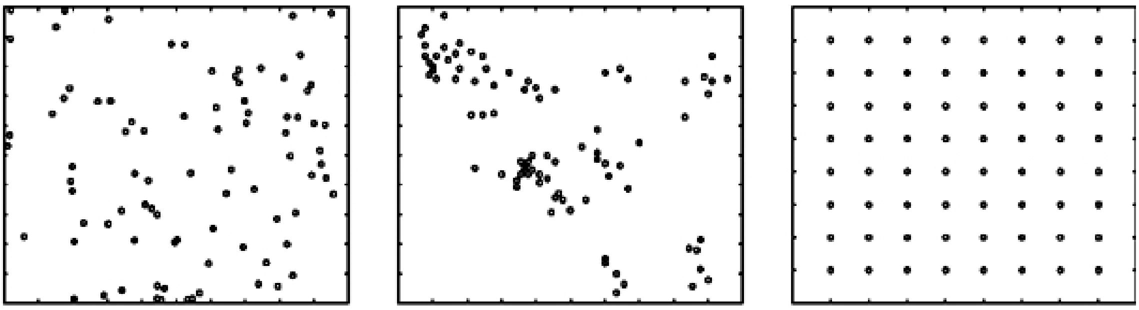
Unconditional multivariable logistic regression models, as suggested by Pearce [Pearce, 2016], were used to assess the relationship between distance from AC plant and the probability of being a case, adjusted for gender, age at diagnosis and type of interview, following Ferrante et al. [Ferrante et al., 2016]. Distance was considered in continuous and categorical way, choosing different thresholds (500 m, 3 Km, 5 Km). Moreover, to adjust for occupational and domestic asbestos exposure, the individual estimate of cumulative exposure previously calculated by experts using interview information were included in the models as either continuous covariates (in fibre/mL-year) or dichotomous categories. Odds ratios (OR) and 95% confidence intervals (95% CI) were reported.

First of all, we conducted a preliminary exploratory spatial analysis separately for cases and controls. Then, to explore the spatial distribution of cases' and controls' locations, in search of spatial aggregations, first- and second- order properties were assessed. The former investigate intensity and spatial density measuring the distribution of events in the study region. The latter give information on the interaction between cases and controls in terms of clustering ability. Finally, scans of local case/control ratios and nearest-neighbor statistics were calculated.

## ***Preliminary exploratory analysis***

A point process, based on Diggle definition [Diggle, 2003], is as stochastic mechanism which generates a countable set of events. The locations of the events generated by a point process in the area of study is called point pattern. Events, in our case, are of two different types: set of individuals with a disease (cases) and set of those without (controls) which usually reflect the spatial distribution of the population. The simplest theoretical model for a spatial point pattern is that of "complete spatial randomness" (CSR) in which the events are distributed randomly and independently according to uniform probability distribution over the region as show in figure 4.a. This implies that there are no regions where the events are more (or less) likely to occur, and that the presence of a given event does not modify the probability of other events appearing nearby. However, usually, the observed events display any systematic spatial pattern or departure from randomness either in the direction of clustering (4.b) or regularity (4.c) indicating presence of attraction and inhibition, respectively.

*Figure 4. Three spatial point patterns: random (4.a), clustered (4.b) and regularly dispersed (4.c).*



(a) (b) (c)

A primary goal of our exploratory analysis was to determine, separately for cases and controls, the degree of accomplishment of the CSR. To do this we had to examine the plots of empirical and theoretical G, F and J functions that represent the distribution of the number of pairs of nearest neighbors within particular distance ranges. Depending on the statistic, the distances can be evaluated with respect to other events as the nearest neighbor (G function) or to predefined grid indicating the average space left between events (F function). The G function represents the probability of finding an event of the point pattern with distance r to a random point, while the F function shows the probability of finding an event of the point pattern with distance r to a fixed arbitrary point. Then, the third statistic (J function) combines the F- and G- functions: it is the ratio of the probabilities of not finding any points within distance r from any random point and a fixed arbitrary point. Generally, to compare empirical and theoretical function, pointwise envelopes under CSR are computed by repeatedly simulating a CSR point process with the same estimated intensity in the study region and check whether the empirical function is contained inside. Particularly, suggestion of clustering (regularity) among events is defined when the empirical function is above (below) the theoretical for the G-function, is below (above) for the F function and when J function assumed values less (higher) than 1.

Based on the results obtained implementing these functions we can understand if cases and controls were random located in Casale Monferrato LHA or instead some different patterns could be identified.

## ***First-order property***

The first-order property investigates intensity and spatial density measuring the distribution of events in the study region. With respect to a spatial point process, a density function defines the probability of observing an event at a location, while the intensity calculates the number of events expected per unit area at a location. As a result, the density and the intensity functions differ by a constant of proportionality, which can result in visual stretching or shrinking of the modes (peaks) in perspective or contour plots of the intensity compared to similar plots of the densities. However, the relative spatial pattern (e.g., locations of peaks and valleys) in densities and intensities will be the same.

The primary goal for comparisons of intensity functions (first-order property) was to detect local difference between the spatial pattern in disease incidence observed in the cases from the spatial pattern observed in the controls, and to assign statistical significance to any differences observed. To do this, we assumed that, based on results obtained in preliminary analysis rejecting CSR hypothesis, the collection of cases and controls followed two inhomogeneous Poisson processes (IPP); so the events occurred independently and were distributed according to given intensities that vary spatially.

To evaluate the spatial distribution of cases and controls and assess the variation of the risk, different parametric and semi-parametric approaches were considered.

### *Kernel density ratio*

The first exploratory tool (Kernel density ratio), used for examining the first-order property, considers the estimator of the disease risk given by the ratio between the intensity of the cases and controls. Under the null hypothesis of equal spatial distribution, the ratio is constant. Alternatively, a risk can be estimated by working with the logarithm of the ratio of the intensities of cases and controls. Moreover, we know that probability densities are proportional to intensities so we worked with density functions. Kelsall and Diggle [Kelsall and Diggle, 1995a; Kelsall and Diggle, 1995b] proposed the use of a bivariate kernel smoothing to estimate each density, after choosing the bandwidth that best fit with the data. Despite the choice of the Kernel had relatively little effect, the bandwidth must be chosen more carefully since its value can have large effect on the smoothness of the resulting estimates. The implementation of this first step produces surface/contour plots of the ratio of cases to controls useful to explore a region and identify an area where cases are more or less likely than controls. Moreover, to assess the presence of a significant deviations from the null hypothesis of constant risk, a global test of clustering can be performed. Kelsall and Diggle proposed a Monte Carlo approach, under the random labelling hypothesis, to investigate the department from the situation where the spatial densities (and intensities) of cases and controls vary across the study area but are always in the same relative proportion. [Kelsall and Diggle, 1995a; Kelsall and Diggle, 1995b

In our study, we considered a bivariate Gaussian Kernel that is a kernel with infinite tails so the estimate of the control density is nonzero for all locations and bandwidths avoiding division by zero in the estimate of risk surface. As the bandwidth selection was more critical, we decided to repeat the analysis considering different combinations of distance and choosing the parameters that best perform our data among 1, 2, 3, 4, 5, 10, 20, 40 Km.

### *Generalized additive model*

Another approach used to evaluate the spatial variation in risk of a disease is the implementation of generalized additive models (GAM), that are more flexible than the density ratio as the regression analysis could include covariate terms. A GAM can be throught as a generalized linear model which has been extended to include arbitrary smooth functions in addition to linear terms in the linear predictor. Different smooth function could be considered but, because of non-homogeneity of population densities, we used the locally weighted regression smoother (LOESS). The optimal degree of smoothing was chosen minimizing Akaike's Information Criterion (AIC). [Kelsall and Diggle, 1998]

In our study, we considered a binary response variable that assume value 1 if a subject is a case and 0 if control. GAMs were firstly performed, for all study subjects and for only non-occupationally exposed, without covariates adjustment to directly compared the results with the consideration done using the kernel density ratio approach. Then, we adjusted the models for gender, age, type of interview and gender, age, type of interview, occupational and domestic asbestos exposure in continuous way and we evaluated the impact of inclusion of distance from the AC plant (in categorical and continuous way). This approach helps us to i) verify the significant presence of a spatial component in the distribution of cases and controls in Casale Monferrato area, adjusting for some important covariates, and ii) understand if the Eternit alone could explain the different spatial distribution of cases and controls.

### *Raised incidence model*

Finally, Diggle-Rowlingson raised incidence models were implemented. Similar to the GAM used before they were based on "focus approach" as we hypothesized the presence of a source of excess risk (AC plant). However, these models allow for a more flexible parametric modelling of the exposure according to the distance to the Eternit [Diggle, 1990; Diggle and Rowlingson, 1994].

The models were based on a multiplicative decomposition of the intensity function, with separate terms to describe natural spatial variation in intensity and possible raised incidence model around a prespecified point. A non-parametric kernel smoothing approach was used to describe natural spatial variation, while a parametric maximum likelihood approach was implemented to describe raised incidence near the source.

Through this approach we could estimate the alpha parameter that indicates the residential excess risk at the AC plant and the beta coefficient that represents the risk decay rate per unit squared distances (Km) moving away from the AC plant.

## ***Second-order property***

Secondly, we evaluated the general tendency for cases to occur more closely together than expected from random sampling. The second-order property measures the strength and the type of interactions between events of the point process, so it is more focused on clustering than on clusters.

To assess clustering of the cases as compared to the controls, Diggle and Chetwynd proposed a test based on K-functions, defined as the ratios between the expected number of events within distance *r* of a randomly chosen event and the intensity, for any positive distance $r$ [Diggle and Chetwynd, 1991]. Intuitively, the definition of the K function implies that under CSR, the value of K(r) is equal to $\pi r^2$ (the area of a circle of radius r). For processes more regular than CSR, we would expect fewer events within distance r of a randomly chosen event than under CSR, so K(r) would tend to be less than $\pi r^2$. Conversely, for processes more clustered than CSR, we would expect more events within a given distance than under CSR (K(r)> $\pi r^2$). Diggle and Chetwynd [Diggle and Chetwynd, 1991] considered the difference of the two K-functions to evaluate whether the cases tend to cluster after considering the inhomogeneous distribution of the population: $D(s) = K_1(s) - K_0(s)$ where $K_1(s)$ and $K_0(s)$ are the homogeneous K-functions of cases and control, respectively. Under the null hypothesis the expected value of the test statistic D is zero. Significant departure from 0 means that there is a difference in the distribution of cases and controls, with clustering occurring at the range of those distance for which D(s)>0. To evaluate the statistical significant of the test, a re-labelling hypothesis and Monte Carlo approach was used.

In our analysis we drawn the D function both with its pointwise envelopes and we explored at what distance any observed clustering tend to occur, average over the entire study area.

## ***Scans of local case/control ratios***

To determine any local areas where the observed ratio of cases/controls (number of cases compared to the number of controls) appeared inconsistent with the ratio observed in the rest of the study area, the Kulldorf test was used. Particularly, this scanning local rate approach was performed to find the most unusual aggregations of cases adjusted for multiple testing considering circular windows with variable radii [Kulldorff, 1997]. The test provides a single p-value for the study area, suggesting that it is a test of clustering. However, the spatial scan statistics observed is generated by a well-defined set of cases, and we can map and identity the potential cluster.

## ***Nearest-neighbor statistics***

The nearest-neighbor statistics examine local patterns of cases in the vicinity of other cases and evidence for clustering involves observing more cases among the nearest neighbors of cases than one would expect under the random labelling hypothesis.

The test developed by Cuzick and Edward [Cuzick and Edward, 1990] represents the number of the q nearest neighbors of cases that are also cases. Different values of q may generate different results, possibly indicating the scale (in the sense of the number of nearest neighbors, not necessarily geographic distance) of any clustering observed.

In our work, evidence for clustering was evaluated for 3, 5, 7, 9, 11, 13 and 15 nearest neighbors, values chosen ‘a priori.

## ***Sensitivity analyses***

The main analyses excluded the time window of 20 years before diagnosis and used the dwelling with longest residence. Sensitivity analyses excluding the time window of 10 years before diagnosis and using the dwelling at the nearest distance from the AC plant were performed to assess for different lag time periods and different selection of exposed dwellings. Unconditional logistic models were further adjusted for the individual estimate of cumulative asbestos due to environmental exposures, in addition to that stemming from occupational and domestic assessment, already included in our main analyses. The estimated environmental exposure level only partially overlapped with exposure index based on distance from the Eternit factory, as the quantitative assessment protocol took into account all sources of exposure. Moreover, only residential distances up to 2 Km from the Eternit factory, 1 Km from the warehouse/rail stock or 100 m from the transport path, whichever the closest, were considered. More details on the how the exposure was calculated are available elsewhere [Ferrante et al., 2016].

## ***Analysis***

Statistical significance was set at 0.05; 999 simulations were performed when Monte Carlo approach was used. SAS 9.4 was used to manage the database, while the models implementation, the spatial analysis and the tests were performed with STATA and R (packages spatstat, mgcv, MapGAM, splancs).

# RESULTS

## BRIC 59

For clarity, the results of BRIC 59 project were presented separately for the different steps described in the Material and Methods.

### Step 1

In our project, we included 6057 cases of MM from the seven Italian regions (Piedmont, Lombardy, Emilia-Romagna, Tuscany, Lazio, Apulia and Sicily): the remaining cases were excluded 'a priori' because they did not report complete identification data, or names of firms, or lacked any evidence of occupational asbestos exposure. Of these 6057, 4134 (68.25%) were successfully linked with INPS.

Table 1 presents the number of records, corresponding for ReNaM to the number of employments reported at the interview and for INPS to the number of different firms recorded in the database for the linked subjects. The number of records was higher than the cases because a subject could have worked in more than one firm. The regions with the higher number of cases where Lombardy (2276) and Piedmont (1104), while for Sicily and Lazio fewer subjects with MM were observed: 256 and 289, respectively, corresponding to the different incidence of MM and prevalence of occupation in industries contributing to INPS. The percentage of successful linkage ranged from 62% (Tuscany/Apulia) to 77% (Sicily). Finally, we had 12801 records from ReNaM and 11602 records from INPS.

***Table 1.*** *The number of malignant mesothelioma (MM) cases included in The Italian National Mesothelioma Registry (ReNaM) and the number of subjects linked (n, %) to the Social Security Institute (INPS) database, and the corresponding number of occupational records, by region. Total population at 2012 was also reported.*

| Region | Total population (2012) | ReNaM Cases | INPS Linked | ReNaM Records | INPS Records |
|---|---|---|---|---|---|
| Piedmont | 4416745 | 1104 | 742 (67.21%) | 2490 | 1885 |
| Lombardy | 9811011 | 2276 | 1595 (70.08%) | 3965 | 4083 |
| Emilia-Romagna | 4391314 | 891 | 634 (71.16%) | 1367 | 1906 |
| Tuscany | 3733535 | 906 | 568 (62.69%) | 3465 | 1600 |
| Lazio | 5605706 | 289 | 189 (65.40%) | 634 | 613 |
| Apulia | 4102797 | 335 | 208 (62.09%) | 440 | 642 |
| Sicily | 5061946 | 256 | 198 (77.34%) | 440 | 873 |
| **Total** | **37123054** | **6057** | **4134 (68.25%)** | **12801** | **11602** |

Table 2 describes the occupational records of the linked MM cases according to the level of occupational asbestos exposure rated by ReNaM, expanding the corresponding column of Table 1.

***Table 2.*** *Distribution of occupational asbestos exposure levels by region.*

| Region | Occupational Asbestos Exposure | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Total** | **Certain** | | **Probable** | | **Possible** | |
| | ***n*** | ***n*** | **%** | ***n*** | **%** | ***n*** | **%** |
| Piedmont | 2490 | 1480 | 59.44 | 208 | 8.35 | 802 | 32.21 |
| Lombardy | 3965 | 2797 | 70.54 | 274 | 6.91 | 894 | 22.55 |
| Emilia-Romagna | 1367 | 883 | 64.59 | 239 | 17.48 | 245 | 17.92 |
| Tuscany | 3465 | 2464 | 71.11 | 388 | 11.2 | 613 | 17.69 |
| Lazio | 634 | 152 | 23.97 | 176 | 27.76 | 306 | 48.26 |
| Apulia | 440 | 204 | 46.36 | 96 | 21.82 | 140 | 31.82 |
| Sicily | 440 | 177 | 40.23 | 119 | 27.05 | 144 | 32.73 |
| **Total** | **12801** | **8157** | **63.72** | **1500** | **11.72** | **3144** | **24.56** |

Overall, of 12801 records form ReNaM, 8157 (63.72%) were defined as certain asbestos exposure, 1500 (11.72%) as probable and 3144 (24.56%) as possible. Difference among regions could be observed. In Tuscany, Lombardy, Emilia Romagna and Piedmont the majority of firms were identified with certain occupational asbestos exposure while for the other regions higher proportions were observed for probable or possible exposure. These could be justified by different approaches used by the COR for identify asbestos exposure both with different knowledge of the firms.

## Step 2

Step 2 summarized the results obtained by text mining and text cleaning both with the inexact matching procedure.

As the first two approaches (text mining and text cleaning) were iterative and recursive it was impossible to report the results of them analytically. However, in appendix part, we indicated all the modifications done and stop words used in the algorithms implemented (appendix 1).

The results of the linkage procedure are instead presented in Table 3. ReNaM records reporting agricultural, military, and firemen sectors were ‘a priori’ excluded because these occupational sectors were not in the INPS database and, therefore, linkage was impossible. These excluded records ranged from 3% to 25% of ReNaM records, depending on the region considered. The remaining records were analyzed, and we considered a successful match when the records shared (1) at least one word into the same subjects, (2) at least two words, (3) one word in length’s string equal to one. Then, in (4), we selected the companies in the ‘a priori list’ of known companies.

The total matched records were 5327 (41.61%), but a huge variation was observed between regions. Better performance was observed in Piedmont (52.37%) followed by Emilia Romagna (46.23%), Sicily (45.91%), and Lombardy (45.30%) while the percentage of records linked was lower in Tuscany (29.99%), Lazio (31.39%), and Apulia (35.23%). If we do not consider the ‘a priori’ exclusions in the denominator, the percentage of successful matches was higher, particularly for Apulia (47.26%), Emilia Romagna (52.67%), and Lazio (35.99%) (results not reported).

***Table 3.*** *The number of ReNaM records and results of the record linkage, by region. The results of linkage procedures were reported to distinguish between excluded (‘a priori’), non-matched, and matched: (1) Within subjects with more than one equal word in firm’s name, (2) two equal words, (3) one equal word in a string of length equal to one, (4) known firms.*

| Region | ReNaM *n* | Excluded *n* (%) | Matched (By Matching Step) | | | | Matched Total *n* (%) | Non-Matched *n* (%) |
|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | | |
| | | | *n* (%) | *n* (%) | *n* (%) | *n* (%) | | |
| Piedmont | 2490 | 185 (7.43) | 861 (34.58) | 156 (6.27) | 168 (6.75) | 119 (4.78) | 1304 (52.37) | 1001 (40.2) |
| Lombardy | 3965 | 128 (3.23) | 1324 (33.39) | 271 (6.83) | 106 (2.67) | 95 (2.4) | 1796 (45.3) | 2041 (51.48) |
| Emilia-Romagna | 1367 | 167 (12.22) | 452 (33.07) | 71 (5.19) | 106 (7.75) | 3 (0.22) | 632 (46.23) | 568 (41.55) |
| Tuscany | 3465 | 364 (10.51) | 755 (21.79) | 117 (3.38) | 146 (4.21) | 21 (0.61) | 1039 (29.99) | 2062 (59.51) |
| Lazio | 634 | 81 (12.78) | 161 (25.39) | 12 (1.89) | 2 (0.32) | 24 (3.79) | 199 (31.39) | 354 (55.84) |
| Apulia | 440 | 112 (25.45) | 117 (26.59) | 7 (1.59) | 22 (5) | 9 (2.05) | 155 (35.23) | 173 (39.32) |
| Sicily | 440 | 32 (7.27) | 163 (37.05) | 23 (5.23) | 8 (1.82) | 8 (1.82) | 202 (45.91) | 206 (46.82) |
| **Total** | **12801** | **1069 (8.35)** | **3833 (29.94)** | **657 (5.13)** | **558 (4.36)** | **279 (2.18)** | **5327 (41.61)** | **6405 (50.04)** |

## Step 3

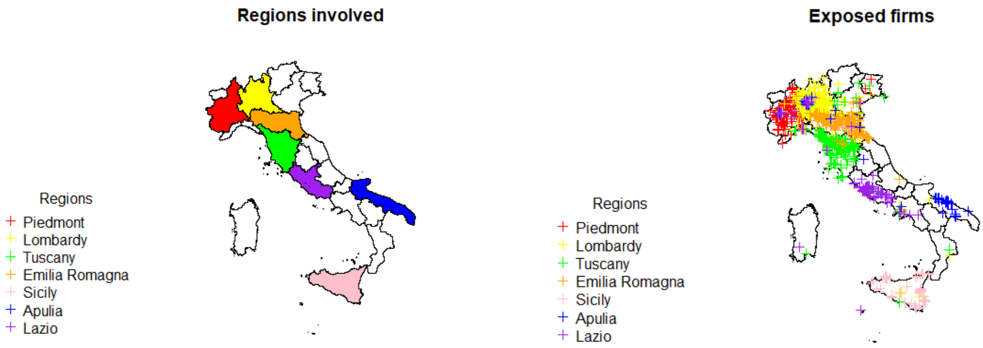
The numbers of different firms obtained from the linkage procedure after removing the duplicates are shown in Table 4, by region and level of asbestos exposure. As explained before, whenever different ratings were present for the same firm, we considered the highest rating, and when the same firm was shown in different regions, we considered it only once. Regions with more MM cases contributed with more firms, and particularly, Lombardy contributed for 40% (1080 firms) of the national total. The total number of unique firms identified was 2606: 1826 (70.07%) with certain, 222 (8.52%) with probable, and 558 (21.41%) with possible asbestos exposure. The list of official INPS firm names and codes formed the thus called ‘dictionary,’ to be used for the evaluation of asbestos exposure in the next step of the process.

***Table 4.*** *The number of firms listed for possible asbestos exposure after the linkage, and distribution of occupational exposure, by region. The total number is smaller than the sum of the number of firms in each region because the firm’s names observed in two or more regions were counted only once.*

| **Regions** | **Matched Records *n*** | **Firms *N*** | **Asbestos Exposure Rating of the Firm** | | |
|---|---|---|---|---|---|
| | | | **Certain *n* (%)** | **Probable *n* (%)** | **Possible *n* (%)** |
| Piedmont | 1304 | 429 | 267 (62.24) | 25 (5.83) | 137 (31.93) |
| Lombardy | 1796 | 1080 | 872 (80.74) | 41 (3.80) | 167 (15.46) |
| Emilia-Romagna | 632 | 370 | 259 (70.00) | 38 (10.27) | 73 (19.73) |
| Tuscany | 1039 | 537 | 420 (78.21) | 45 (8.38) | 72 (13.41) |
| Lazio | 199 | 129 | 33 (25.58) | 50 (38.76) | 46 (35.66) |
| Apulia | 155 | 99 | 50 (50.51) | 10 (10.10) | 39 (39.39) |
| Sicily | 202 | 142 | 57 (40.14) | 34 (23.94) | 51 (35.92) |
| **Total** | **5208** | **2606** | **1826 (70.07)** | **222 (8.52)** | **558 (21.41)** |

One of the advantages of this process is the provision of a list of ‘asbestos exposing firms’ that is not limited to the experience of regional experts. In Figure 5, we report the two Italian maps with the regional boundaries. In the left panel (Figure 5a), we highlight the regions involved, while in the right (Figure 5b), the spatial distribution of exposed firms, separately the region of origin of the MM cases, is indicated.

*Figure 5.* (*a*) *Italian map of regions involved in the study.* (*b*) *Italian map with the spatial distribution of exposed firms.*




(**a**) (**b**)

## Step 4

The cases of cancer included on the preliminary identification of asbestos exposing firms in the OCM system were 35010, 9848 (28.13%) from Sicily and 25162 (71.87%) from Lazio, selected from the local population cancer registry (Sicily) or the hospital discharge (Lazio) databases. In the case of multiple malignancies in the same subject, only the first malignant neoplasm was selected, in chronological order. Table 5 reports the distribution by tumor type. We can observe that the most common neoplasms among those considered where lung and colorectal (37.82% and 37.66%, respectively), followed by stomach (9.35%), ovary (5.58%) and larynx (5.32%).

*Table 5. Distribution of malignant neoplasm in occupational cancer monitoring (OCM) data for Lazio and Sicily.*

| **Malignant Neoplasm** | **Region** | | | | **All** | |
|---|---|---|---|---|---|---|
| | **Lazio (*n* = 25162)** | | **Sicily (*n* = 9848)** | | **(n=35010)** | |
| | ***n* cases** | **%** | ***n* cases** | **%** | ***n* cases** | **%** |
| Colorectal | 9373 | 37.25 | 3812 | 38.71 | 13185 | 37.66 |
| Pharynx | 605 | 2.40 | 137 | 1.39 | 742 | 2.12 |
| Larynx | 1288 | 5.12 | 573 | 5.82 | 1861 | 5.32 |
| Ovary | 1490 | 5.92 | 464 | 4.71 | 1954 | 5.58 |
| Pleura | 221 | 0.88 | 196 | 1.99 | 417 | 1.19 |
| Lung | 9590 | 38.11 | 3650 | 37.06 | 13240 | 37.82 |
| Nasopharynx | 196 | 0.78 | 140 | 1.42 | 336 | 0.96 |
| Stomach | 2399 | 9.53 | 876 | 8.90 | 3275 | 9.35 |

The ‘dictionary’ obtained from Step 3 to the list of INPS records was then applied to these cases. The subjects with malignant neoplasm who had worked in firms selected for asbestos exposure were 1454 (14.76%) and 5092 (20.24%) from Sicily and Lazio, respectively (6546, 18.7% overall). It is interestingly observed that the percentage were generally less than 15%, except for pleural neoplasms for which it reached a 32.61% of subjects who worked in a firm with asbestos exposure. Moreover, excluding pleural malignancies, 18.7% (6410) of the cancer cases extracted for the OCM system could be related to occupational asbestos exposure (Table 6).

*Table 6. Distribution of malignant neoplasm in OCM data (all and matched cases) for Lazio and Sicily.*

| **Malignant Neoplasm** | **Region** | | | | | |
|---|---|---|---|---|---|---|
| | **Lazio Cases** | | **Sicily Cases** | | **All** | |
| | **n** | **Matched (%)** | **n** | **Matched (%)** | **n** | **Matched (%)** |
| Colorectal | 9373 | 2020 (21.55) | 3812 | 579 (15.19) | 13185 | 2599 (19.71) |
| Pharynx | 605 | 113 (18.68) | 137 | 17 (12.41) | 742 | 130 (17.52) |
| Larynx | 1288 | 248 (19.25) | 573 | 81 (14.14) | 1861 | 329 (17.68) |
| Ovary | 1490 | 169 (11.34) | 464 | 28 (6.03) | 1954 | 197 (10.08) |
| Pleura | 221 | 83 (37.56) | 196 | 53 (27.04) | 417 | 136 (32.61) |
| Lung | 9590 | 1953 (20.36) | 3650 | 561 (15.37) | 13240 | 2514 (18.99) |
| Nasopharynx | 196 | 35 (17.86) | 140 | 12 (8.57) | 336 | 47 (13.99) |
| Stomach | 2399 | 471 (9.63) | 876 | 123 (14.04) | 3275 | 594 (18.14) |
| **Total** | **25162** | **5092 (20.24)** | **9848** | **1454 (14.76)** | **35010** | **6546 (18.7)** |

## LATENT

The study base included 223 eligible cases (mean age: 68.4; SD: 11.3; males: 62%) and 552 controls (mean age: 65.4; SD: 12.1; males: 61%). Two hundred cases (89.7%) and 348 (63.0%) controls accepted the invitation and were interviewed (Figure 6). Distribution by sex of participating cases and controls was similar (male: 63.5% vs 63.2% respectively), while cases were older than controls because of the oversampling of controls matched to cases under 60 and the different age distribution of non-participating controls (68.3±11.4 vs 63.3±11.6 years). Interviews took slightly longer among cases than among controls (74.2±18.3 vs 70.7±17.1 minutes). Interviews were face to face in 54% of the cases (due to death or poor health condition, 46% of interviews were with a close relative) and for nearly all controls (92.8%) [Ferrante et al, 2016]. Due to the higher number of controls who were not interviewed, comparison between responders and non-responders in terms of socio-demographic and clinical characteristics were done. Significant differences among responders and non-responders were observed for age, occupation, and education while gender and civil status were similar: non-responders were older, with less education level and more frequently unemployed or retired [Airoldi et al., 2020b].

Two thousand seven hundred sixty-one dwellings were reported at interview and were automatically geocoded (953 for cases and 1808 for controls; average number of dwellings per person: 4.8 for cases and 5.2 for controls); for about 5% (127/2761) of them a manual check was necessary to amend errors and imprecisions that precluded automatic georeferentiation. After excluding the few dwellings for which full address was not available (20/953=2.1% for cases and 40/1808=2.2% for controls) and those outside the LHA area (156/933=16.7% for cases and 526/1768=29.8% for controls), geographic coordinates were available for 777 and 1242 addresses, corresponding to 199 cases and 348 controls. The application of the 10-years and 20-years lags reduced the number of subjects eligible for the analyses to 198 cases and 331 controls, and to 196 cases and 322 controls, respectively (Figure 6).

*Figure 6. Case control study on MM in Casale Monferrato area. Flow chart of subjects included in the analysis. N indicates the number of subjects, Nr the number of dwellings. LHA=Local Health Authority.*

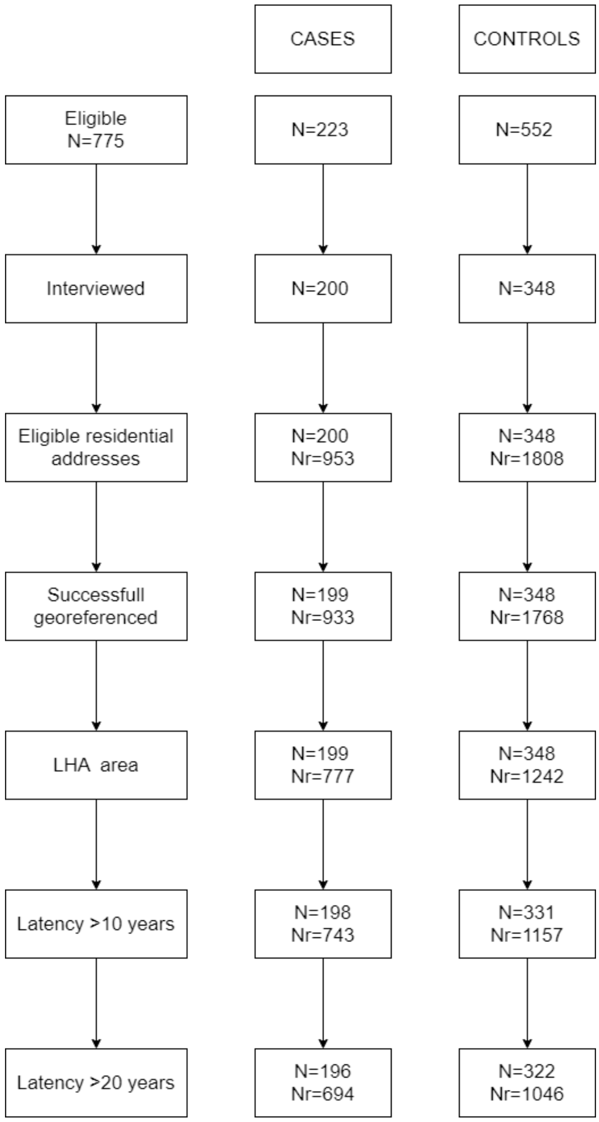


Figure 7 presents the distribution of index residences and the borders of the municipalities forming the Casale Monferrato LHA. The spatial distribution of the controls on the map (circles) represents the population density. The square covers a 2500 $km^2$ surface (50x50 Km).

*Figure 7. Case control study on MM in Casale Monferrato area. Spatial distribution of the residences of cases (triangles) and controls (circles) in the geographic area of LHA (solid line), approximately 2500 $km^2$ around Casale Monferrato (bold solid line). Longest-held residences after excluding 20 years before the date of diagnosis of the index case. The location of the AC plant is also indicated (red triangle). [R Spatstat]*

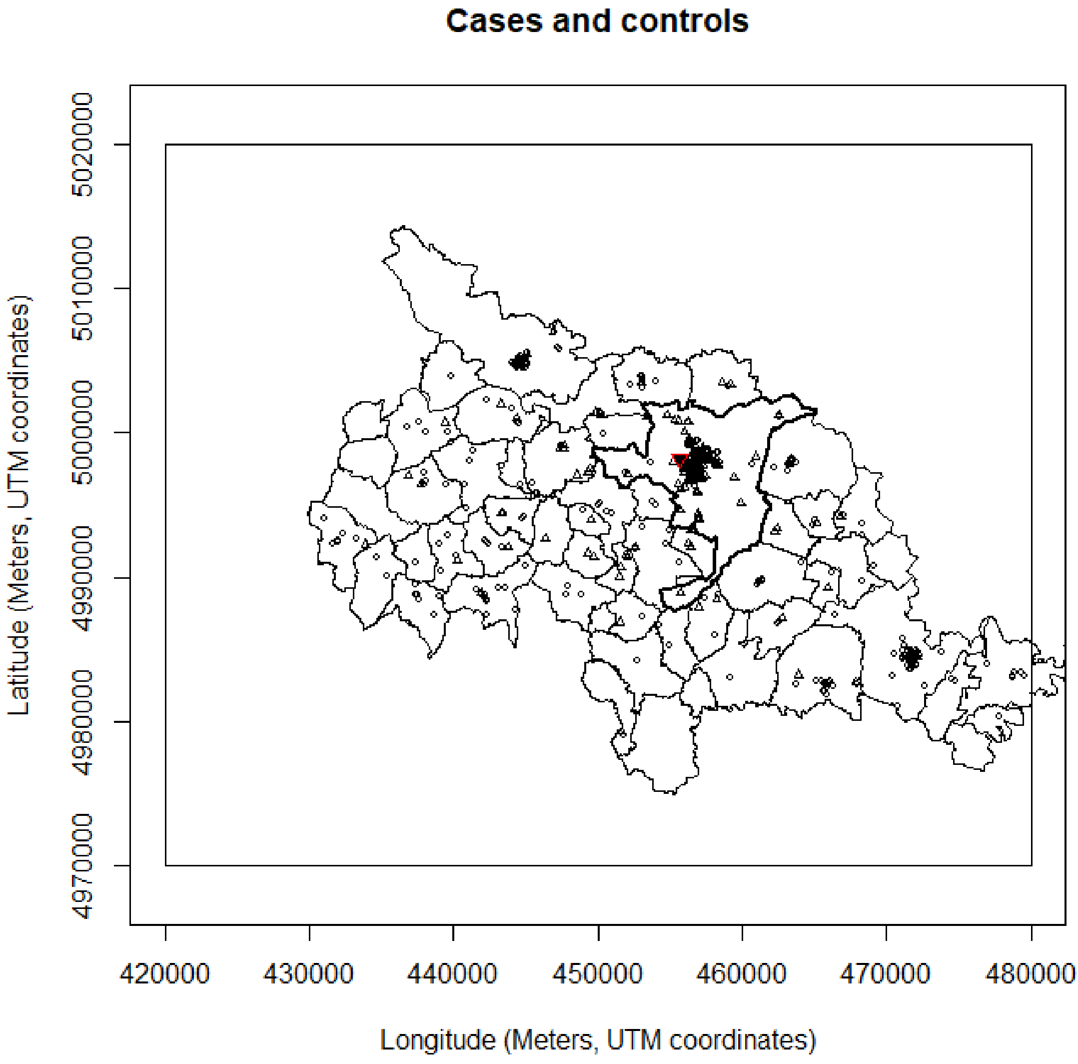


Table 7 presents results on MM risk by classes of residential distance, both considering all subjects and excluding subjects occupationally exposed. For ‘all subjects’ category, the median distance from residence to AC factory was 2.10 [IQR 5.92] and 11.32 [IQR 15.77] km for cases and controls, respectively. Cases lived more frequently near the AC plant and 64.3% of them had the longest-held residence within 5 km from the plant, compared to 34.5% of controls. Considering only non-occupationally exposed subjects (Table 7 bottom section), closely similar results were obtained: distances from the AC factory were lower in cases than controls (2.03 [IQR 4.59] vs 10.36 [IQR 15.86]), 71.95% of cases and 38.12% of controls had the index dwelling within 5 km from the factory. The analyses considering all subjects showed a constant trend of decreasing OR with increasing distance from the factory, both using distance as categorical or numerical variable. In distance classes up to 10 km from the factory (reference >15 km) a tenfold increase of risk was observed, statistically significant. The model adjusted for age, sex and type of interview showed similar ORs compared to that adjusted also for occupational and domestic exposure but AIC index was smaller. OR for 1 km increasing distance were 0.87 (95% CI 0.84-0.91) and 0.88 (95% CI 0.84-0.91), respectively. Similar results were observed when the analyses were restricted excluding subjects with occupational exposure (lower part of table 7), confirming that the risk relation with distance to the factory was not a spurious result and cannot be explained by the effect of occupational exposure.

*Table 7. Case control study on MM in Casale Monferrato area. Risk of MM of the pleura in relation to the distance of longest-held residence (after exclusion of 20 years before the date of diagnosis) from the AC plant. Absolute and relative frequencies of distance categories and median [interquartile range] of distance. Logistic models adjusted by age, sex, type of interview (*) and age, sex, type of interview and occupational and domestic asbestos exposure as continuous covariate (**) or age, sex, type of interview and domestic asbestos exposure as continuous covariate (***); odds ratios (OR), 95% confidence intervals (in brackets) and Akaike Information Criterion (AIC).*

| *All subjects* | | | | | |
|---|---|---|---|---|---|
| **Distance from the AC plant (km)** | **All N=518** | **Cases N=196** | **Controls N=322** | **OR adjusted*** | **OR adjusted**** |
| 0-5 | 237 (45.75) | 126 (64.29) | 111 (34.47) | 10.91 (5.32-22.38) | 10.81 (5.26-22.21) |
| 5-10 | 75 (14.48) | 40 (20.41) | 35 (10.87) | 10.48 (4.54-24.2) | 10.42 (4.51-24.07) |
| 10-15 | 85 (16.41) | 18 (9.18) | 67 (20.81) | 2.2 (0.9-5.33) | 2.2 (0.9-5.34) |
| >15 | 121 (23.36) | 12 (6.12) | 109 (33.85) | Ref | Ref |
| **AIC** | | | | 513.71 | 515.43 |
| Distance, Km | 6.06 [12.25] | 2.10 [5.9] | 11.32 [15.77] | 0.87 (0.84-0.91) | 0.88 (0.84-0.91) |
| **AIC** | | | | 514.91 | 516.61 |

| *Non-occupationally exposed subjects* | | | | | |
|---|---|---|---|---|---|
| **Distance from the AC plant (km)** | **All N=263** | **Cases N=82** | **Controls N=181** | **OR adjusted*** | **OR adjusted***** |
| 0-5 | 128 (48.67) | 59 (71.95) | 69 (38.12) | 13.73 (4.57-41.32) | 12.66 (4.19-38.28) |
| 5-10 | 30 (11.41) | 12 (14.63) | 18 (9.94) | 8.5 (2.29-31.61) | 7.15 (1.87-27.36) |
| 10-15 | 39 (14.83) | 6 (7.32) | 33 (18.23) | 2.81 (0.7-11.29) | 2.47 (0.61-10.04) |
| >15 | 66 (25.10) | 5 (6.10) | 61 (33.70) | Ref | Ref |
| **AIC** | | | | 251.83 | 249.96 |
| Distance, Km | 5.29 [13.56] | 2.03 [4.59] | 10.36 [15.86] | 0.86 (0.81-0.91) | 0.86 (0.81-0.91) |
| **AIC** | | | | 247.18 | 244.89 |

Results using different distance categories (0-3, 3-5, 5-7, 7-9, 9-11, 11-13, 13-15, >15 km and 0-500, 500-1000, 1000-1500, 1500-2000, >2000 meters) were reported in Table 8 and 9. Considering classes of 2 km (Table 8), the OR observed are consistent with the main analyses, although the trend was less regular and CIs were wider, as expected because of the smaller numbers of subjects in these smaller categories. When classes were further restricted, problems of scattered data occurred. No controls lived in a radius of 500 meters while 8 cases did, including 3 without occupational exposure. Subjects with residences in the 500-1000 meters band showed an increased risk: OR was about 14 (all subjects analyses) or 18 (exclusion of subjects with occupational exposure) (Table 8). In order to further control for the possible role of occupational, domestic and total (sum of occupational, domestic and environmental) asbestos exposure, specific analyses were performed by including exposure estimates in the model, using different metrics (Table 10). Results remained very close to those from the main analysis.

*Table 8. Case control study on MM in Casale Monferrato area. Risk of MM of the pleura in relation to the distance of longest-held residence (after exclusion of 20 years before the date of diagnosis) from the AC plant. Absolute and relative frequencies of distance categories. Logistic models adjusted by age, sex, type of interview (*) and age, sex, type of interview and occupational and domestic asbestos exposure as continuous covariate (**), and age, sex, type of interview and domestic asbestos exposure as continuous covariate (***); odds ratios (OR), 95% confidence intervals (in brackets) and Akaike Information Criterion (AIC).*

| | | *All subjects* | | | |
|---|---|---|---|---|---|
| **Distance from the AC plant (km)** | **All N=518** | **Cases N=196** | **Controls N=322** | **OR adjusted*** | **OR adjusted**** |
| 0-3 | 216 (41.70) | 115 (58.67) | 101 (31.37) | 11.46 (5.53-23.73) | 11.36 (5.48-23.55) |
| 3-5 | 21 (4.05) | 11 (5.61) | 10 (3.11) | 7.05 (2.11-23.57) | 7.04 (2.11-23.56) |
| 5-7 | 40 (7.72) | 20 (10.20) | 20 (6.21) | 10.09 (3.87-26.33) | 10.1 (3.87-26.35) |
| 7-9 | 30 (5.79) | 17 (8.67) | 13 (4.04) | 12.22 (4.28-34.84) | 12.04 (4.21-34.4) |
| 9-11 | 20 (3.86) | 7 (3.57) | 13 (4.04) | 3.25 (0.9-11.72) | 3.24 (0.9-11.72) |
| 11-13 | 43 (8.30) | 10 (5.10) | 33 (10.25) | 2.44 (0.85-6.99) | 2.45 (0.85-7.01) |
| 13-15 | 27 (5.21) | 4 (2.04) | 23 (7.14) | 1.7 (0.44-6.62) | 1.7 (0.44-6.61) |
| >15 | 121 (23.36) | 12 (6.12) | 109 (33.85) | Ref | Ref |
| **AIC** | | | | 520.93 | 522.69 |

| | | *Non occupationally exposed subjects* | | | |
|---|---|---|---|---|---|
| **Distance from the AC plant (km)** | **All N=263** | **Cases N=82** | **Controls N=181** | **OR adjusted*** | **OR adjusted***** |
| 0-3 | 115 (43.73) | 53 (64.63) | 62 (34.25) | 14.39 (4.71-43.91) | 13.35 (4.35-40.93) |
| 3-5 | 13 (4.94) | 6 (7.32) | 7 (3.87) | 9.91 (1.9-51.8) | 8.94 (1.73-46.29) |
| 5-7 | 18 (6.84) | 6 (7.32) | 12 (6.63) | 5.23 (1.14-24.03) | 4.54 (0.95-21.66) |
| 7-9 | 9 (3.42) | 5 (6.10) | 4 (2.21) | 23.23 (3.96-136.36) | 18.76 (3.05-115.58) |
| 9-11 | 9 (3.42) | 3 (3.66) | 6 (3.31) | 3.86 (0.58-25.6) | 3.72 (0.56-24.71) |
| 11-13 | 22 (8.37) | 4 (4.88) | 18 (9.94) | 3.93 (0.82-18.94) | 3.21 (0.65-15.95) |
| 13-15 | 11 (4.18) | | 11 (6.08) | - | - |
| >15 | 66 (25.10) | 5 (6.10) | 61 (33.70) | Ref | Ref |
| | | | | 252.52 | 251.28 |

*Table 9. Case control study on MM in Casale Monferrato area. Risk of MM of the pleura in relation to the distance of longest-held residence (after exclusion of 20 years before the date of diagnosis) from the AC plant. Absolute and relative frequencies of distance categories. Logistic models adjusted by age, sex, type of interview (*); age, sex, type of interview and occupational and domestic asbestos exposure as continuous covariate (**) and age, sex, type of interview and domestic asbestos exposure as continuous covariate (***); odds ratios (OR), 95% confidence intervals (in brackets) and Akaike Information Criterion (AIC).*

| | | *All subjects* | | | |
|---|---|---|---|---|---|
| **Distance from the AC plant (meters)** | **All N=518** | **Cases N=196** | **Controls N=322** | **OR adjusted*** | **OR adjusted**** |
| 0-500 | 8 (1.54) | 8 (4.08) | 0 (0.00) | - | - |
| 500-1000 | 18 (3.47) | 14 (7.14) | 4 (1.24) | 13.58 (4.22-43.73) | 13.7 (4.25-44.11) |
| 1000-1500 | 84 (16.22) | 38 (19.39) | 46 (14.29) | 2.08 (1.19-3.63) | 1.99 (1.14-3.5) |
| 1500-2000 | 57 (11.00) | 31 (15.82) | 26 (8.07) | 2.94 (1.55-5.57) | 2.92 (1.54-5.54) |
| >2000 | 351 (67.76) | 105 (53.57) | 246 (76.40) | Ref | Ref |
| **AIC** | | | | 545.30 | 546.19 |

| | *Non occupationally exposed subjects* | | | | |
|---|---|---|---|---|---|
| **Distance from the AC plant (km)** | **All N=263** | **Cases N=82** | **Controls N=181** | **OR adjusted*** | **OR adjusted***** |
| 0-500 | 3 (1.14) | 3 (3.66) | 0 (0.00) | | |
| 500-1000 | 13 (4.94) | 10 (12.20) | 3 (1.66) | 17.82 (4.38-72.44) | 18.45 (4.5-75.62) |
| 1000-1500 | 44 (16.73) | 14 (17.07) | 30 (16.57) | 1.15 (0.49-2.7) | 1.25 (0.53-2.95) |
| 1500-2000 | 31 (11.79) | 12 (14.63) | 19 (10.50) | 2.25 (0.89-5.7) | 2.24 (0.86-5.78) |
| >2000 | 172 (65.40) | 43 (52.44) | 129 (71.27) | Ref | Ref |
| **AIC** | | | | 258.11 | 255.07 |

*Table 10. Case control study on MM in Casale Monferrato area. Risk of MM of the pleura in relation to the distance of longest-held residence (after exclusion of 20 years before the date of diagnosis) from the AC plant. Absolute and relative frequencies of distance categories and median [interquartile range] of distance. Logistic models adjusted by age, sex, type of interview, occupational exposure in a dichotomous way (*) and age, sex, type of interview and domestic exposure in a dichotomous way (**) and age, sex, type of interview and asbestos exposure (occupational, domestic and environmental) as continuous covariate (***) and age, sex, type of interview and asbestos exposure (domestic and environmental) as continuous covariate (****); odds ratios (OR), 95% confidence intervals (in brackets) and Akaike Information Criterion (AIC).*

| | *All subjects* | | |
|---|---|---|---|
| **Distance from the AC plant (km)** | **OR adjusted*** | **OR adjusted**** | **OR adjusted***** |
| 0-5 | 11.75 (5.67-24.37) | 10.11 (4.9-20.85) | 10.8 (5.26-22.19) |
| 5-10 | 10.39 (4.47-24.16) | 9.54 (4.08-22.32) | 10.41 (4.51-24.06) |
| 10-15 | 2.22 (0.91-5.41) | 2.22 (0.91-5.38) | 2.2 (0.9-5.34) |
| >15 | Ref | Ref | Ref |
| **AIC** | 504.68 | 505.98 | 515.39 |
| Distance, Km | 0.87 (0.84-0.9) | 0.88 (0.85-0.91) | 0.88 (0.84-0.91) |
| **AIC** | 514.39 | 505.53 | 516.58 |

| | *Non-occupationally exposed subjects* | | |
|---|---|---|---|
| **Distance from the AC plant (km)** | | **OR adjusted**** | **OR adjusted****** |
| 0-5 | | 11.66 (3.74-36.37) | 7.39 (2.34-23.39) |
| 5-10 | | 7.57 (1.92-29.85) | 5.72 (1.46-22.38) |
| 10-15 | | 2.74 (0.67-11.28) | 2.31 (0.56-9.49) |
| >15 | | Ref | Ref |
| **AIC** | | 242.79 | 237.55 |
| Distance, Km | | 0.87 (0.82-0.92) | 0.89 (0.84-0.95) |
| **AIC** | | 237.13 | 233.77 |

## ***Preliminary spatial exploratory analysis***

Before the main analyses, already presented, some preliminary spatial analyses had been conducted and are briefly presented. First, we plotted the locations of cases and controls separately (Figure 8). As observed, events were not random allocated: the majority of subjects lived in the center of Casale Monferrato or in other urban areas, while rural areas were less inhabited. These suggestions were supported by the results of nearest-neighbor distance functions that measures the distribution of distances from events. In our case, the observed curves were outside of the expected under complete spatial randomness hypothesis, suggesting clustering (Figure 9, 10, 11): cases and controls tend to aggregate more than expected by chance. In details, for G and F functions the observed values were respectively above and below the envelope area (expected) while values of the J functions were lower than 1 (reference value).

*Figure 8. Case control study on MM in Casale Monferrato area. Spatial distribution of the residences of cases (left panel) and controls (right panel) in the geographic area of LHA, approximately 2500 km$^2$ around Casale Monferrato (solid line). Longest-held residences after excluding 20 years before the date of diagnosis of the index case. The location of the AC plant is also indicated (red triangle).*

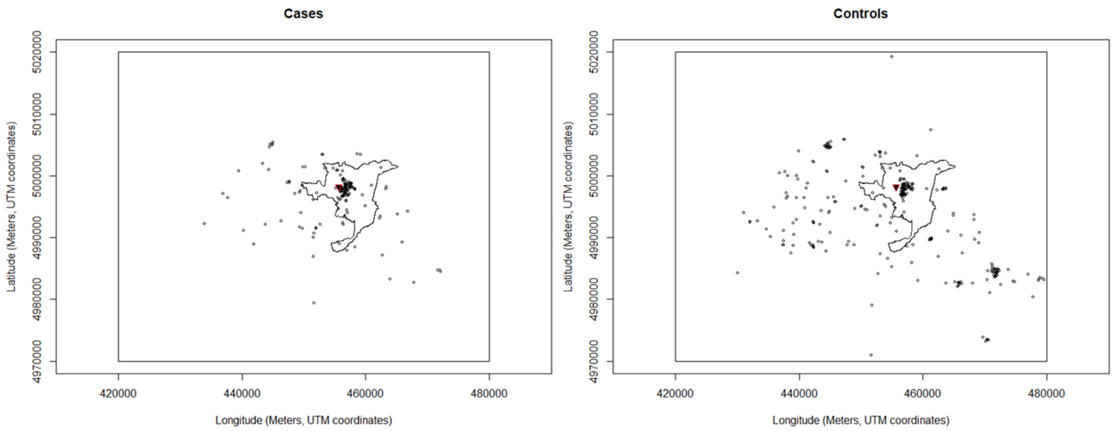


*Figure 9. G functions for cases and controls, separately. The solid black line represents the observed value, while the red dashed the expected. The grey area reports the envelopes (n simulation=999).*

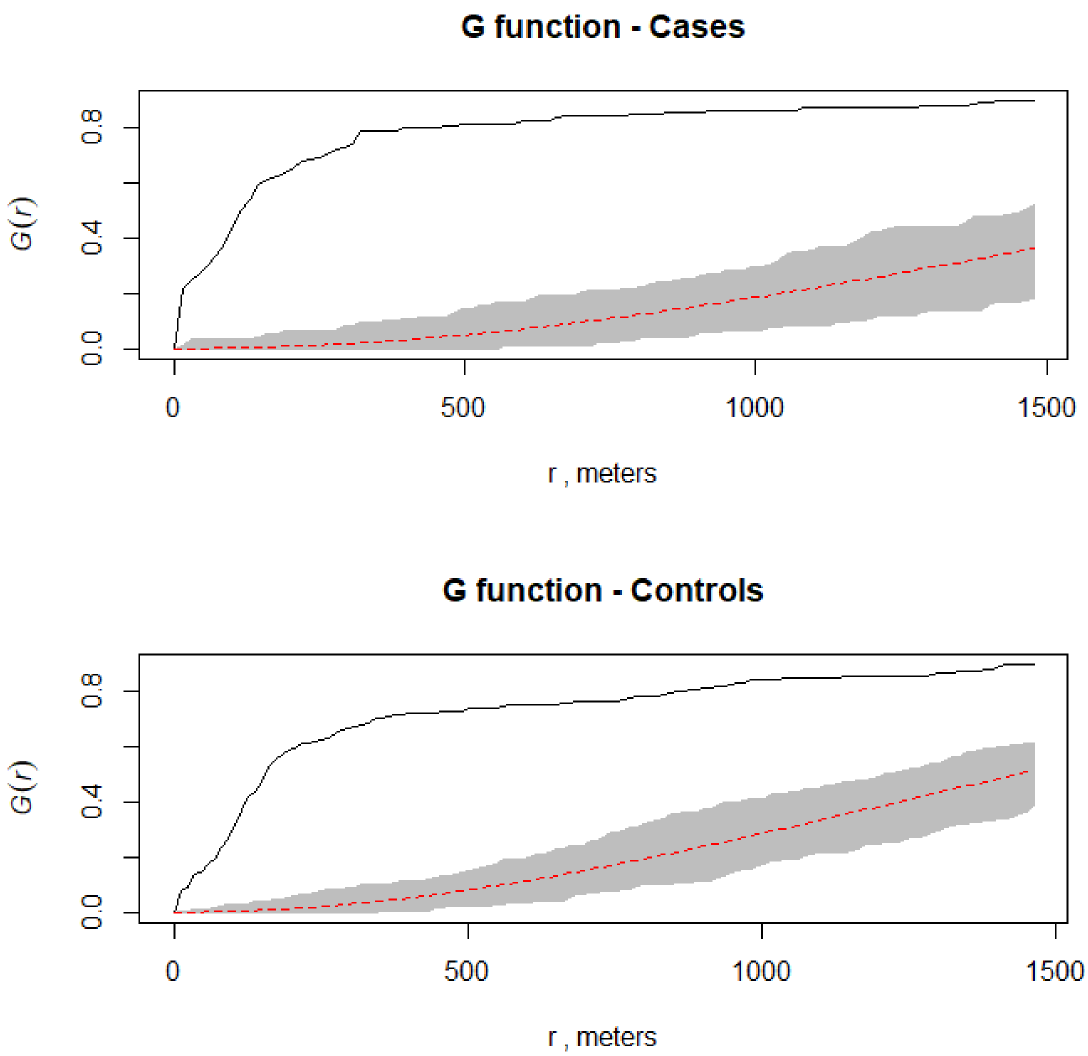
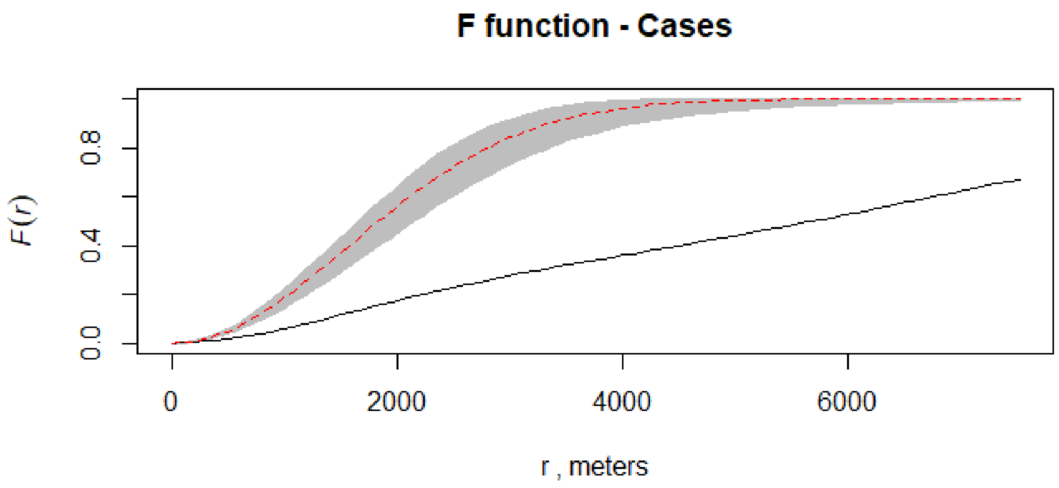


*Figure 10. F functions for cases and controls, separately. The solid black line represents the observed value, while the red dashed the expected. The grey area reports the envelopes (n simulation=999).*



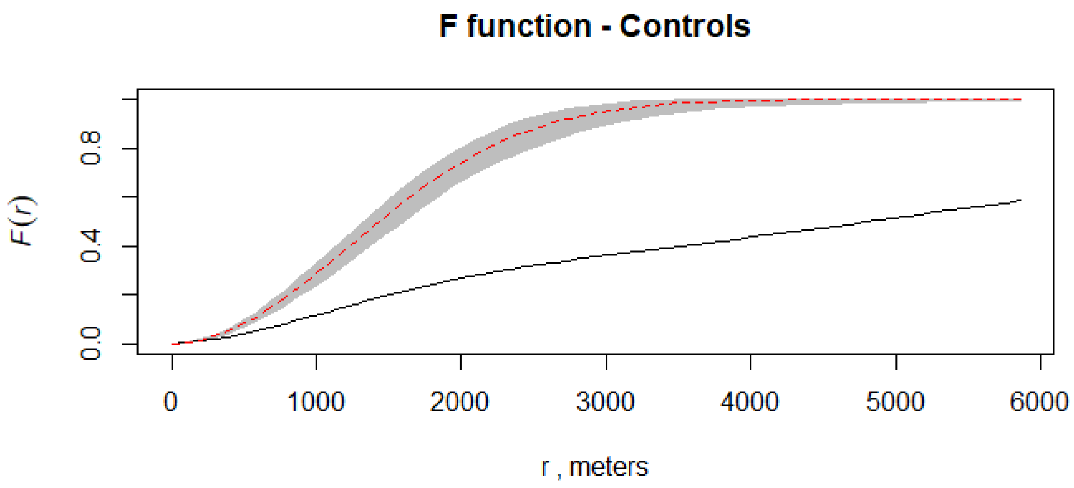
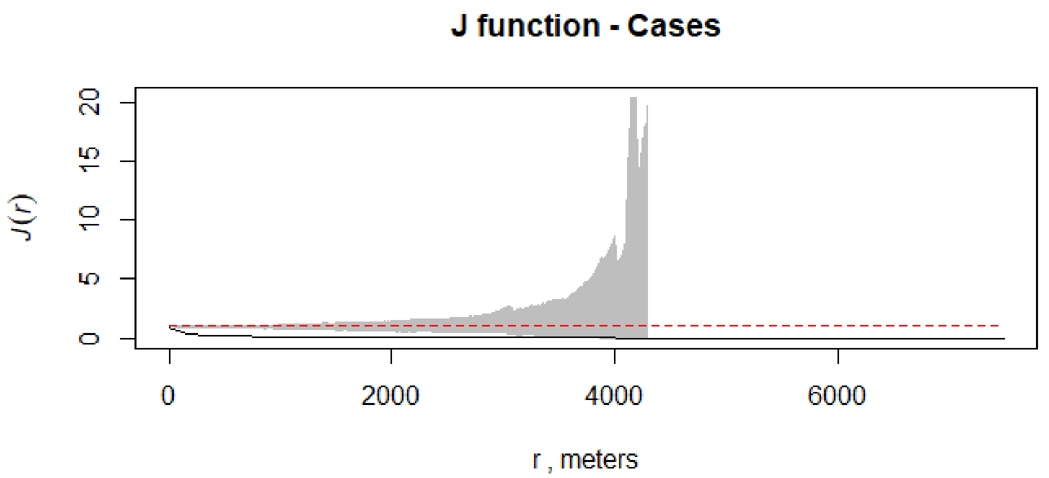
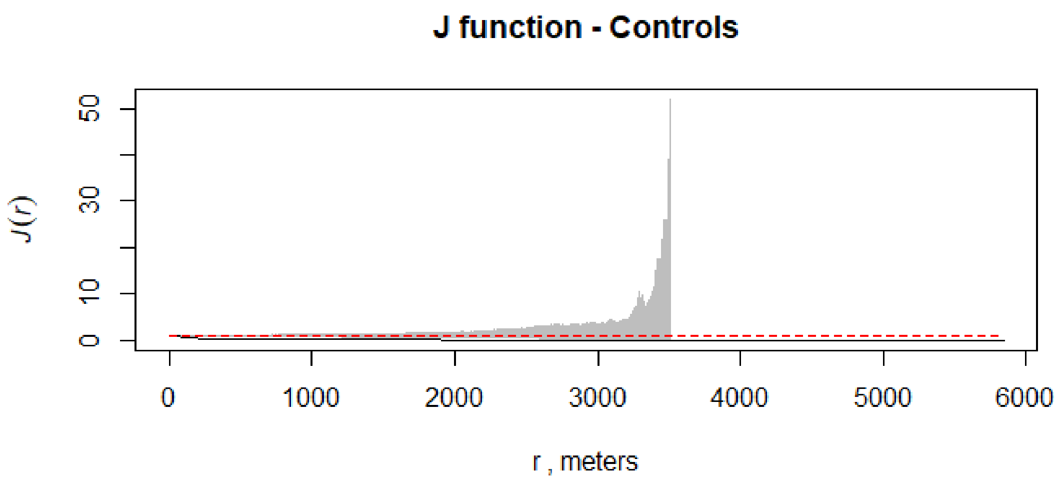



*Figure 11. J functions for cases and controls, separately. The solid black line represents the observed value, while the red dashed the expected. The grey area reports the envelopes (n simulation=999).*

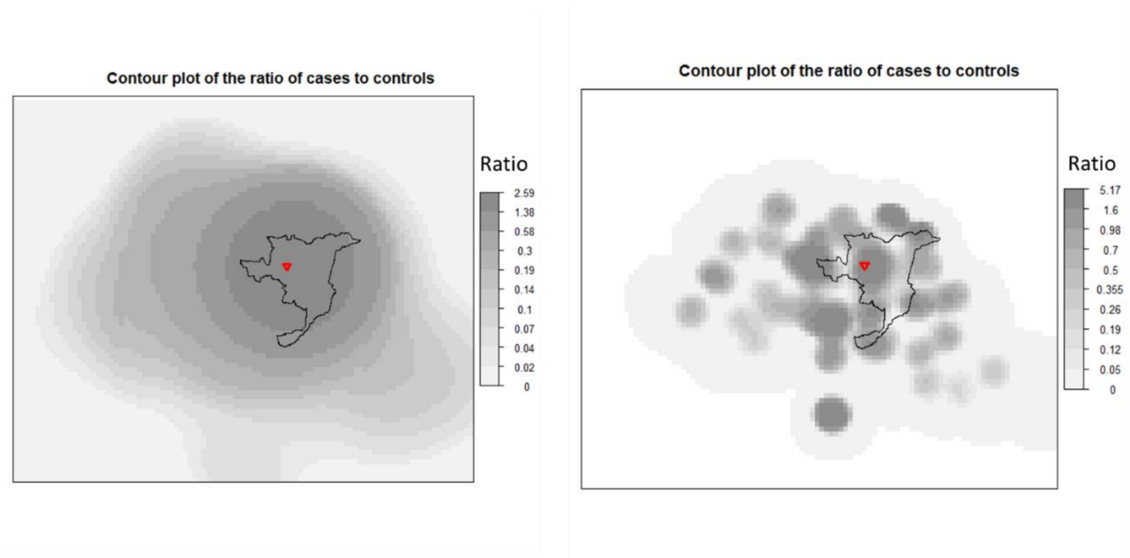


## ***First order property***

The contour plot of the ratio of Kernel density functions for cases and controls based on the Kelsall and Diggle method was obtained (figure 12). We observed (left panel) a well-defined peak near the AC factory, and risk seems to decrease monotonically in all directions when bandwidths were set at 10 and 20 km for cases and controls. The area around the Eternit was associated with a significantly increased risk, and the global test yielded a p-value of 0.001, indicating that the overall difference between intensities supports the idea of a global pattern of clustering. However, when we reduced the bandwidth values (2 and 4 km for cases and controls, respectively), large effect on the smoothness was observed, resulting in a bumpy surface (figure 12, right panel). This result must be taken with caution as statistical significance was not reached (p=0.656) and most local clusters were based on few subjects with little ‘a priori’ evidence of exposure hot points. Despite other values of bandwidths were considered, results were not reported as no significant contour plots were obtained.

*Figure 12. Case control study on MM in Casale Monferrato area. Contour plot of the kernel density surfaces of the case to control ratio, including occupationally and non-occupationally exposed individuals in an area of approximately 2500 km2 around Casale Monferrato (solid line). Bivariate Gaussian Kernels with smoothing parameters set to 10 and 20 km for cases and controls, respectively, in the left panel and 2 and 4 km in the right panel. The location of the AC plant (red triangle) is indicated. The legend reports the value of ratio and the corresponding grey shades.*




Then, generalized additive models were performed including a smooth spatial function. We started fitting the models without covariates and then we added potential confounder factors and distance variable. More details on covariates included, span choice, AIC and p-value of global test of spatial effect are reported in table 11. The models that did not take into account for the distance between residence and AC plant had a statistically significant spatial departure from the hypothesis of flat surface, which suggests that there was some residual spatial variation unexplained in the generalized linear model. When we added the distance from the AC as covariate (in categorical and continuous way), the global test did not reach the statistical significance indicating that all spatial variation of MM risk can be explained by the presence of the AC plant. Similar results were obtained considering only subjects non-occupationally exposed.

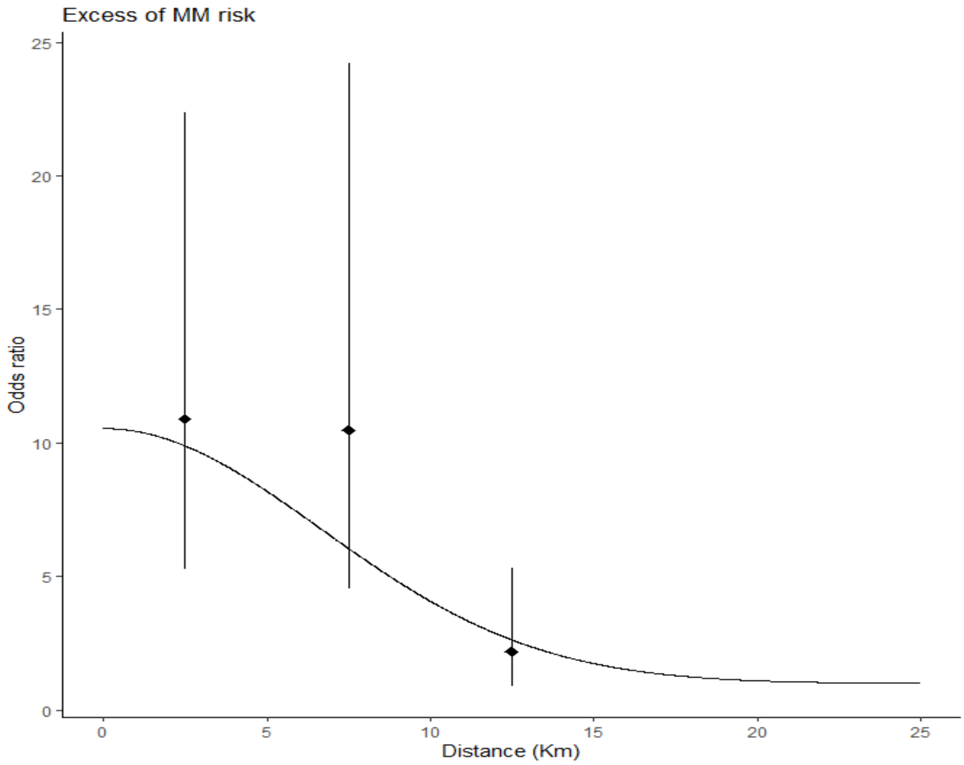
*Table 11. Risk of malignant mesothelioma of the pleura in Casale Monferrato in relation to the distance of each individual's longest-held residence (after excluding of 20 years from the date of the diagnosis) from the AC plant. Generalized additive models were performed including different covariates. Optimal span size and AIC, both with global test p-values of spatial effect based on 999 permutations were reported.*

| *All subjects* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **No distance** | | | **Distance (Categorical)** | | | **Distance (continuous)** | | |
| **Covariates adjustment** | **Span** | **AIC** | **p-values** | **Span** | **AIC** | **p-values** | **Span** | **AIC** | **p-values** |
| - | 0.3 | 608.64 | <0.001 | 0.95 | 608.83 | 0.550 | 0.95 | 612.63 | 0.815 |
| Gender, age, type of interview | 0.3 | 515.64 | <0.001 | 0.3 | 518.27 | 0.062 | 0.95 | 521.05 | 0.969 |
| Gender, age, type of interview, categorical asbestos exposure | 0.3 | 468.02 | <0.001 | 0.95 | 470.95 | 0.689 | 0.3 | 469.35 | 0.062 |
| Gender, age, type of interview, continuous asbestos exposure | 0.3 | 517.41 | <0.001 | 0.3 | 520.13 | 0.027 | 0.95 | 522.74 | 0.873 |

| *Non occupationally exposed subjects* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | No distance | | | Distance (Categorical) | | | Distance (continuous) | | |
| **Covariates adjustment** | **Span** | **AIC** | **p-values** | **Span** | **AIC** | **p-values** | **Span** | **AIC** | **p-values** |
| - | 0.85 | 291.25 | <0.001 | 0.95 | 296.17 | 0.361 | 0.95 | 292.11 | 0.461 |
| Gender, age, type of interview | 0.15 | 246.81 | <0.001 | 0.95 | 256.53 | 0.761 | 0.95 | 252.48 | 0.873 |
| Gender, age, type of interview, categorical asbestos exposure | 0.15 | 240.15 | 0.002 | 0.95 | 249.01 | 0.600 | 0.95 | 252.38 | 0.749 |
| Gender, age, type of interview, continuous asbestos exposure | 0.95 | 237.65 | 0.003 | 0.95 | 242.51 | 0.809 | 0.95 | 239.23 | 0.888 |

Finally, raised incidence models were implemented to investigate the possible elevation in disease risk around the putative source. Estimates of the point source models without covariates were alpha=14.49 and beta=0.012, where alpha represents residential excess risk at the source and beta represents risk decay rate per unit squared distances (km) moving away from the source. Considering the adjusted models, the alpha and beta parameters were respectively 9.54 and 0.011 (age, sex, type of interview) and 0.90 and 0.002 (age, sex, type of interview, occupational and domestic asbestos exposure). In terms of deviance (or log-likelihood), the best model is the second (adjusted for age, sex, type of interview). When the analysis was restricted to non-occupationally exposed cases and controls, the model without covariates had an excess risk at the source of 11.64 and the risk decay rate was 0.013 for km. The model with covariates did not converge for collinearity. Comparison between ORs estimated using categorical distance from the source and the raised incidence model, is shown in figure 13.

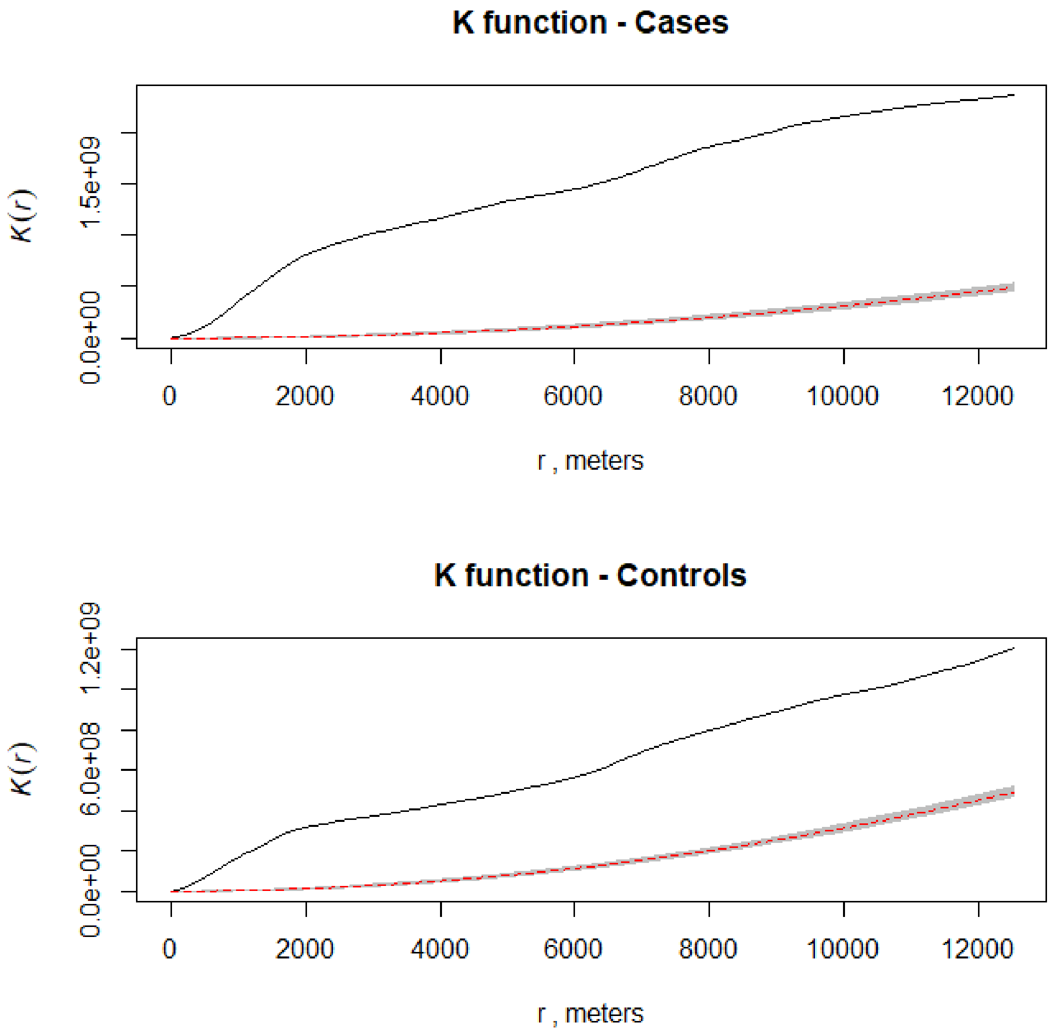
*Figure 13. Case control study on MM in Casale Monferrato area. Risk of MM of the pleura in Casale Monferrato in relation to the distance of individuals' longest-held residence (after exclusion of 20 years before the date of diagnosis) from the AC plant. Risk estimates are adjusted for age, sex, type of interview. Odds ratios and 95% confidence intervals estimated through the logistic model are represented by error bars while those estimated through the model with exponential decay of the risk by distance are shown as a smooth line.*



## ***Second order property***

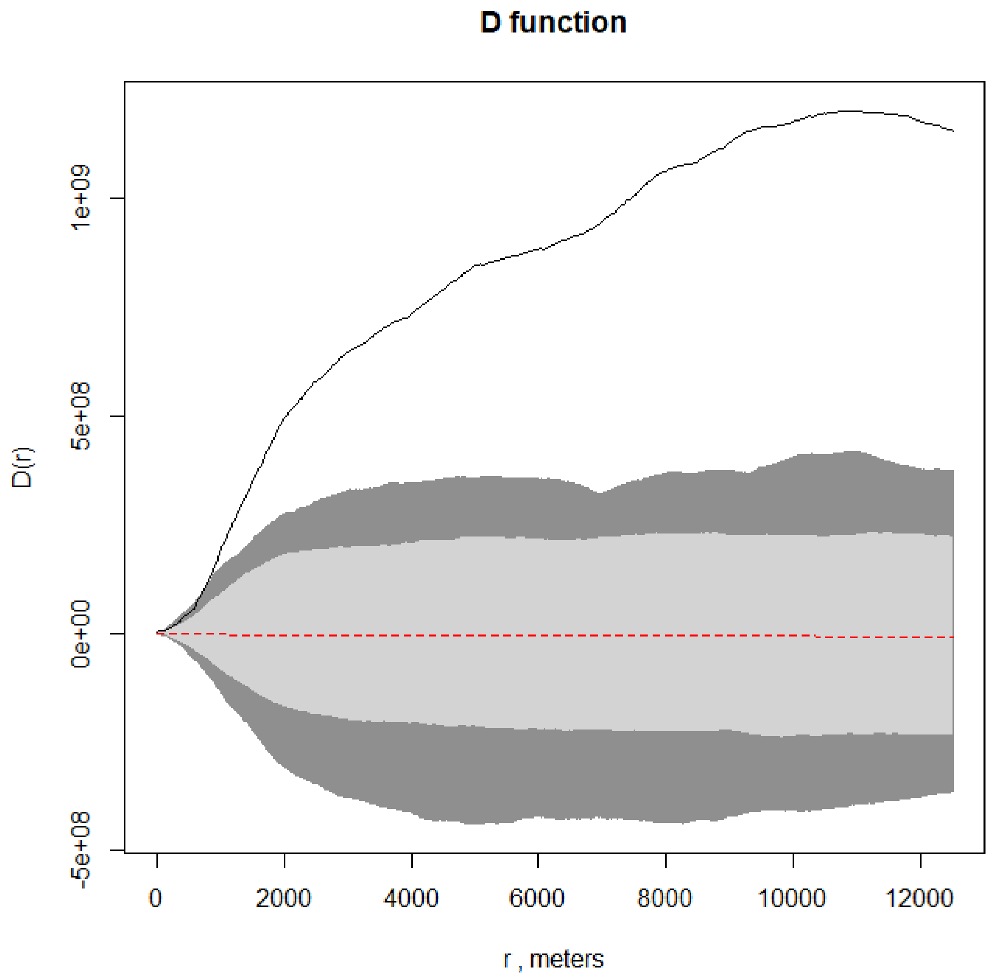
To assess clustering, second-order properties were assessed. Values of K functions were higher than $\pi r^2$, where r are the possible distances from 0 to 12 Km around an arbitrary event, suggesting that cases and controls were clustered processes (Figure 14).

*Figure 14. K functions for cases and controls, separately. The solid black line represents the observed value, while the red dashed the expected. The grey area reports the envelopes (n simulation=999).*



Moreover, the D function, that compared the distributions of cases and controls, provided evidence that their spatial distributions were different as the empirical D function was above the upper limit of the tolerance envelope (figure 15). In details, significant evidence of clustering occurred approximately after 700 meters, so for any distance greater than this threshold, cases tend to be more aggregated than controls. This conclusion is heavily consistent with the Diggle test results (p=0.001) as it supports the hypothesis that spatially close groups of cases occur more frequently than is consistent with completely random occurrence of the disease among members of the population at risk.

*Figure 15. Case control study on MM in Casale Monferrato area. D function. The solid black line represents the observed value, while the red dashed the expected. The grey area reports the envelopes (n simulation=999) and the lighter grey part the 95% envelopes.*



## ***Scans of local case/control ratios***

The Kulldorf test identified an area with a significant excess of cases (p-value = 0.001) centered at 1.71 Km South from the Eternit plant, with a radius of 8.64 km that included 168 cases and 146 controls (Supplementary material, figure 16). The cluster covered almost the whole city of Casale Monferrato and this area was very similar to the area with increased risk obtained using the Kelsall and Diggle method as reported in Figure 12 (left panel).

*Figure 16. Case control study on MM in Casale Monferrato area. Spatial distribution of the residences of cases (triangles) and controls (circles) in a geographic area of approximately 2500 km2 around Casale Monferrato (solid line). Residences are the longest-held among all residences of each individual after excluding 20 years before the date of diagnosis of the index case. The location of the AC plant (red triangle) and the center of the cluster found using the Kulldorf test (green triangle) in the town of Casale Monferrato are also indicated. [R Spatstat]*

## ***Nearest-neighbor statistics***

Finally, we applied the Cuzick and Edward statistics that calculated for each q nearest-neighbor the number of those that are cases. Table of test statistics and associate p-values both with contrasts for a variety of values of q is presented in supplementary materials (Table 12). Since all the contrasts, except the last (15 vs 13), suggested significant clustering, we conclude that the significant clustering observed among collections of more than thirteen nearest neighbors is driven primarily by clustering among the first thirteen nearest neighbors.

*Table 12. Case control study on MM in Casale Monferrato area. Cuzick and Edward test statistics ($T_q$) and associate p-values based on 999 random labelling simulations for a variety of q values (3, 5, 7, 9, 11, 13, 15).*

| **Number of nearest neighbors (q)** | **Tq** | **p-value** |
|---|---|---|
| 3 | 312 | <0.001 |
| 5 | 511 | <0.001 |
| 7 | 697 | <0.001 |
| 9 | 884 | <0.001 |
| 11 | 1073 | <0.001 |
| 13 | 1255 | <0.001 |
| 15 | 1414 | <0.001 |

| **Contrast** | **Value** | **p-value** |
|---|---|---|
| $T_5$-$T_3$ | 199 | <0.001 |
| $T_7$-$T_3$ | 385 | <0.001 |
| $T_9$-$T_3$ | 572 | <0.001 |
| $T_{11}$-$T_3$ | 761 | <0.001 |
| $T_{13}$-$T_3$ | 943 | <0.001 |
| $T_{15}$-$T_3$ | 1102 | <0.001 |
| $T_7$-$T_5$ | 186 | 0.004 |
| $T_9$-$T_5$ | 373 | <0.001 |
| $T_{11}$-$T_5$ | 562 | <0.001 |
| $T_{13}$-$T_5$ | 744 | <0.001 |
| $T_{15}$-$T_5$ | 903 | <0.001 |
| $T_9$-$T_7$ | 187 | <0.001 |
| $T_{11}$-$T_7$ | 376 | <0.001 |
| $T_{13}$-$T_7$ | 558 | <0.001 |
| $T_{15}$-$T_7$ | 717 | <0.001 |
| $T_{11}$-$T_9$ | 189 | <0.001 |
| $T_{13}$-$T_9$ | 371 | <0.001 |
| $T_{15}$-$T_9$ | 530 | <0.001 |
| $T_{13}$-$T_{11}$ | 182 | 0.005 |
| $T_{15}$-$T_{11}$ | 341 | 0.008 |
| $T_{13}$-$T_{15}$ | 159 | 0.210 |

## ***Sensitivity analyses***

All the analyses were repeated considering a 10 years' lag period, including 198 cases and 331 controls. The results of the analyses using distance, either categorical or continuous, were reported in table 13 and no relevant differences were observed compared to the main analyses (table 7). Odds Ratio of about 10 were observed for classes up to 10 km and they were statistically significant including all subjects. When we considered only non-occupationally exposed, confidence intervals were wider but the same trend was estimated. Also the cluster analyses using Kulldorf, and Cuzick and Edward tests provided unchanged results: we identified a big cluster that covered the area surrounded the AC plant and clustering hypothesis was supported by the data (data not tabulated).

*Table 13. Case control study on MM in Casale Monferrato area. Risk of MM of the pleura in relation to the distance of longest-held residence (after exclusion of 10 years before the date of diagnosis) from the AC plant. Absolute and relative frequencies of distance categories and median [interquartile range] of distance. Logistic models adjusted by age, sex, type of interview (*) and age, sex, type of interview and occupational and domestic asbestos exposure as continuous covariate (**) or age, sex, type of interview and domestic asbestos exposure as continuous covariate (***); odds ratios (OR), 95% confidence intervals (in brackets) and Akaike Information Criterion (AIC).*

| | | *All subjects* | | | |
|---|---|---|---|---|---|
| **Distance from the AC plant (km)** | **All N=529** | **Cases N=198** | **Controls N=331** | **OR adjusted*** | **OR adjusted**** |
| 0-5 | 237 (44.80) | 124 (62.63) | 113 (34.14) | 10.26 (5.09-20.67) | 10.17 (5.04-20.5) |
| 5-10 | 79 (14.93) | 42 (21.21) | 37 (11.18) | 10.72 (4.76-24.12) | 10.65 (4.73-23.99) |
| 10-15 | 86 (16.26) | 19 (9.60) | 67 (20.24) | 2.34 (0.99-5.54) | 2.34 (0.99-5.54) |
| >15 | 127 (24.01) | 13 (6.57) | 114 (34.44) | Ref | Ref |
| **AIC** | | | | 523.76 | 525.45 |
| Distance, Km | 6.18 [12.63] | 2.13 [6.15] | 11.38 [16.09] | 0.88 (0.85-0.91) | 0.88 (0.85-0.91) |
| **AIC** | | | | 526.03 | 527.71 |

| | | *Non-occupationally exposed subjects* | | | |
|---|---|---|---|---|---|
| **Distance from the AC plant (km)** | **All N=270** | **Cases N=83** | **Controls N=187** | **OR adjusted*** | **OR adjusted***** |
| 0-5 | 127 (47.04) | 57 (68.67) | 70 (37.43) | 13.01 (4.36-38.87) | 12 (4-36.03) |
| 5-10 | 34 (12.59) | 14 (16.87) | 20 (10.70) | 10 (2.81-35.58) | 8.66 (2.38-31.51) |
| 10-15 | 41 (15.19) | 7 (8.43) | 34 (18.18) | 3.38 (0.88-12.93) | 2.97 (0.77-11.52) |
| >15 | 68 (25.19) | 5 (6.02) | 63 (33.69) | Ref | Ref |
| **AIC** | | | | 259.67 | 257.85 |
| Distance, Km | 5.89 [13.56] | 2.07 [4.91] | 10.36 [15.98] | 0.87 (0.82-0.92) | 0.87 (0.82-0.92) |
| **AIC** | | | | 255.85 | 253.37 |

The analyses using the dwellings at the shortest distance from the factory are presented in Table 14 and again no relevant differences from the main analyses were observed, although a smoother decline of ORs with distance was noticed.

*Table 14. Case control study on MM in Casale Monferrato area. Risk of MM of the pleura in relation to the distance of shorter residence (after exclusion of 20 years before the date of diagnosis) from the AC plant. Absolute and relative frequencies of distance categories and median [interquartile range] of distance. Logistic models adjusted by age, sex, type of interview (*) and age, sex, type of interview and occupational and domestic asbestos exposure as continuous covariate (**) or age, sex, type of interview and domestic asbestos exposure as continuous covariate (***); odds ratios (OR), 95% confidence intervals (in brackets) and Akaike Information Criterion (AIC).*

| | | *All subjects* | | | |
|---|---|---|---|---|---|
| **Distance from the AC plant (km)** | **All N=518** | **Cases N=196** | **Controls N=322** | **OR adjusted*** | **OR adjusted**** |
| 0-5 | 282 (54.44) | 144 (73.47) | 138 (42.86) | 11.59 (5.36-25.08) | 11.48 (5.3-24.86) |
| 5-10 | 56 (10.81) | 29 (14.80) | 27 (8.39) | 9.04 (3.52-23.22) | 9.06 (3.53-23.27) |
| 10-15 | 73 (14.09) | 13 (6.63) | 60 (18.63) | 2.03 (0.76-5.47) | 2.04 (0.76-5.48) |
| >15 | 107 (20.66) | 10 (5.10) | 97 (30.12) | Ref | Ref |
| **AIC** | | | | 517.33 | 519.00 |
| Distance, Km | 2.70 [11.70] | 1.38 [4.35] | 8.12 [15.00] | 0.87 (0.83-0.9) | 0.87 (0.83-0.9) |
| **AIC** | | | | 512.52 | 514.27 |
| | | *Non-occupationally exposed subjects* | | | |
| **Distance from the AC plant (km)** | **All N=263** | **Cases N=82** | **Controls N=181** | **OR adjusted*** | **OR adjusted***** |
| 0-5 | 148 (56.27) | 63 (76.83) | 85 (46.96) | 14.04 (4.21-46.88) | 12.87 (3.83-43.2) |
| 5-10 | 22 (8.37) | 9 (10.98) | 13 (7.18) | 8.55 (1.9-38.34) | 7.1 (1.53-33.05) |
| 10-15 | 37 (14.07) | 6 (7.32) | 31 (17.13) | 3.4 (0.77-15.07) | 2.98 (0.66-13.41) |
| >15 | 56 (21.29) | 4 (4.88) | 52 (28.73) | Ref | Ref |
| **AIC** | | | | 256.53 | 254.45 |
| Distance, Km | 2.53 [11.70] | 1.37 [3.10] | 6.03 [14.77] | 0.86 (0.81-0.92) | 0.86 (0.81-0.92) |
| **AIC** | | | | 251.69 | 249.11 |

# DISCUSSION

The project has allowed to deepen the study of the effect of two types of asbestos exposure: the occupational (BRIC 59) and environmental (LATENT) ones.

Firstly, through the use of available databases (ReNaM and OCM) we improved the knowledge about the firms with at least possible asbestos exposure and the identification of cancer cases related to possible occupational exposure of this carcinogenic. Secondly, we analyzed the Italian National Priority Contaminated site of Casale Monferrato in terms of spatial variation of MM risk to better clarify the public health impact of the Eternit and asbestos related products.

## BRIC 59

The BRIC 59 aims at contributing to Italian surveillance on occupational cancer with a specific focus on asbestos exposure. In particular, it aims at supporting the identification of asbestos using factories when analyzing the occupational history of cases with information limited to industry names. Our exploratory project was limited to five regions, with reference to this area we have i) prepared a list of 2606 firms ("dictionary") with possible asbestos exposure, ii) transferred this information to the cancer cases identified in OCM in two regions selecting 6546 patients who probably had an occupational exposure to asbestos.

The creation of a dictionary [i)] that reported the official name of firms as recorded in INPS, the official administrative registration number, the rating of exposure to asbestos (certain, probable, possible) and the name reported in ReNaM is a tool that can be used for other projects. Particularly, this list can be applied to cases from OCM in a more extended selection of regions and can be provided to Regional Operating Centers for their routine activity of research. In our idea, the list and the map reporting the geographical position of firms, can help the trained interviewer to better describe the occupational histories of MM cases and the industrial hygienist to identify their asbestos exposure. Moreover, in a future we hypothesize that an automatic linkage among ReNaM and INPS may be performed: the firms identified with asbestos exposure according to interview information will be flagged directly on the INPS database. This idea was recently introduced in the ReNaLOCCAM report.

Another important aim reached in the project was the selection of cancer cases from OCM who worked in firms that we identified at risk of asbestos exposure [ii)]. Despite a more accurate and detailed evaluation must be done to verify the real exposure, a first screening was done. Particularly, the exploration of the asbestos-related etiology of cases requires enormous efforts, given the relatively low attributable fraction and the high numbers of patients. Our methodology allows to shrink the population of cases to investigate and maximize the resources available. In addition, cases with verified occupational exposure could ask to INPS an economical compensation and other personal support services so our work may have also an impact in term of reimbursements.

The analysis conducted was based on registries recorded for routinely surveillance system (ReNaM), administrative databases for pension contributions (INPS) or a hybrid system between them (OCM). The use of available databases not created for the specific purpose of the research, reduces the duration and cost of the project. Moreover, it allows to replicate the procedure performed as data are generally recorded in standardized forms in term of disease classification, questions on occupational histories, and exposure estimation. Finally, large amount of information can be evaluated, and statistics can be performed at national level. However, the databases not created for scientific purposes often lack critical information. In particular, the INPS database does not include details on occupational exposure as contact with specific material, use of personal protective equipment, job description and so on. Other typical problems of these type of databases are a not total overlap of observations recorded and the absence of a linkage key, that limits the connections between them. In ReNaM we considered only MM cases occupationally exposed, and we included in the analysis only the records/firms with a certain, probable, and possible asbestos exposure. On the other hand, the INPS database excluded some important occupational categories (agriculture, public employment, armed forces) and self-employees, mainly artisans and shopkeepers. Only contributions from 1974 were recorded [Crosignani et al., 2006]. Despite these limits, the main challenge was the definition of an algorithm to match the names of the firms from ReNaM to INPS. The different accuracy level on reporting the information of industries (free text in ReNaM and a drop-down list in INPS) causes problems on the use of classic deterministic linkage procedures, so a probabilistic approach was implemented. As infinite solutions could be applied, we decided to proceed using a step-by-step procedure and building an ad hoc algorithm. The quality of the work done was difficult to determine and we do not exclude that a better performing solution can be considered. We tried to minimize the “false negatives” corresponding to the absence of “signals” of asbestos exposure: we wanted to reduce the firms considered exposed in ReNaM but not included in the “dictionary”. On the contrary, the firms not exposed that were included in the final list (“false positives”) were considered a less serious mistake as they were manually checked by experts in the final step of the analysis.

## LATENT

The results obtained in the LATENT project increases the knowledge on the impact of environmental asbestos exposure in an Italian National Priority Contamined site (SIN), the city of Casale Monferrato and the surrounding area. We found a pattern of increased MM risk in the area around a large AC factory and we detected secondary clusters of cases, of smaller size and possibly associated with the local use of asbestos waste materials. Overall, Kernel estimates with wide bandwidths (10-20 Km), Kulldorf test and the models implemented were able to identify a large area with increased risk in the city of Casale Monferrato, around the AC plant. However, when narrow bandwidths (2-4 Km) were used and the Cuzick Edward test was performed, we observed a different pattern of cases and controls distributions: several peaks of MM risk appear to be spread over the area, not only around the AC plant, and cases tend to aggregate in smaller groups.

The first aim of this study was to explore the spatial variation of MM risk in the area of Casale Monferrato. Using different graphical and statistical methods, we found an increased risk of MM near the AC plant; models that included the distance as covariate showed that risk decreased as distance increased, with a significant excess of cases up to 10 km from the AC plant. A similar inverse relation between distance from a recognized asbestos pollution source and MM risk was observed in two other Italian spatial case-control studies, respectively in the same area of Casale Monferrato (Eternit)[Maule et al., 2007; Magnani et al., 2001] and Bari (Fibronit) [Musti et al., 2009]). In our previous study in Casale Monferrato, that included 103 cases and 272 controls, bands of approximately 2 Km had been used and significant ORs within 11 km had been estimated for occupationally and non-occupationally exposed subjects [Maule et al., 2007; Magnani et al., 2001]. A mixed additive-multiplicative model had been used estimating a RR of 10.5

for residence at zero distance from the source, and a risk decay rate of 0.11 x 10-7 per unit squared-distances (meters) moving away from the source. Results had been, thus, very similar to the present study, despite some differences in statistical modelling: in the previous study [Maule et al., 2007] we had adjusted for occupation in the AC industry and indoor or outdoor domestic exposure to asbestos, while in the current analyses we chose, based on deviance indicator, the model without asbestos exposure covariates. The Bari study involved the geographical area where the Fibronit AC factory had been active for about 50 years. It included 48 cases and 273 controls without occupational, domestic, and household asbestos exposure. A significant OR of 5.29 (95%CI 1.18-23.74) was observed for residential distance < 500 meters from the AC plant (reference: distance > 2000 m), while for intermediate distance bands the confidence intervals were very wide and included unity [Musti et al., 2009]. Our results present higher ORs: in our analysis no controls and 8 cases lived in the first band (infinite risk) and subjects within 500-1000 m (14 cases and 4 controls) had a high and significant OR. These differences might be partly explained by the different selection of the reference category (> 2000 m in Bari, > 15 km in the present study). However, the two areas are different: for the Fibronit AC factory in Bari the distance from the closest residential areas was less than 50 m, compared with about 200 m in Casale Monferrato, but the Fibronit AC factory was much smaller than the Eternit plant, employing a maximum of 400 workers against about 1500. Our contour plot is in line with the map proposed by Maule et al. [Maule et al., 2007] when the smoothing parameters are the same.

After decreasing the bandwidth values different findings were observed, as the results are strongly dependent on the parameters selection. Using narrow bandwidths, we were able to decompose the global peak observed in the city of Casale Monferrato in four areas with an increased risk of 4 or more. This observation is consistent with the 'a priori' information on the occurrence of local point sources of exposure, either associated to the main AC plant or to the use of production residues.

Other studies supported the view than environmental asbestos exposure entails an increased MM risk, despite using a different methodology - as generally they were population surveys where the standardized incidence or mortality ratio (SIR and SMR respectively) was calculated. In Broni, another Italian town where a large AC factory operated, an overall excess of MM cases above the regional average was observed for non-occupationally exposed people and MM incidence was similar to that observed in Casale Monferrato [Mensi et al., 2015; Consonni et al., 2020]. In Amagasaki (Japan) Kurumatani and Kumagai [Kurumatani and Kumagai, 2008] observed that among persons who had lived within a 300 m radius of an asbestos products factory, the MM SMR was 13.9 (95% CI 5.6-28.7) for men and 41.1 (95% CI 15.2-90.1) for women; the town districts with a significantly elevated SMR reached 2,200 meters from the plant.

It is worth noting that the Third report from the (Italian) National Mesothelioma Registry (ReNaM) included an historical outline of the Italian asbestos-cement industry, describing 37 plants, whereas reports on the occurrence of MM after residential exposures have been published only for those of Casale Monferrato, Bari and Broni. These three factories had some distinctive features: all used crocidolite in all formulations, produced – among the other articles – pressure pipes, whose formulation required larger amounts of crocidolite, were old (the three oldest in Italy), large (two of the three largest), close to urban areas with a large resident population and are known to have dumped asbestos-cement waste materials at uncontrolled sites accessible to the population and even distributed them for free, to be used by residents for paving roads, yards and, in general, private estates. We consider them, therefore, the most likely ones to have large impact on residents, but also the best suited to be the object of formal epidemiological studies. However, by no means the lack of comparable studies for the other plants can be considered as evidence of no effect on nearby people.

The second purpose of our study was to detect clustering and to find unusual aggregations of cases. Overall, we were able to demonstrate that cases tend to aggregate more than controls, and to observe a significant local cluster of 168 cases, in the area surrounding the AC plant, consistent with analyses by distance bands. Coherent with the other Italian studies, significant p-values were obtained testing the alternative hypothesis of clustering [Maule et al., 2007; Musti et al., 2009]. Despite the fact that the Kulldorf test was not sensitive enough to detect other clusters, we posit that minor aggregations of cases are incorporated into this major cluster, as already supposed by Maule et al. [Maule et al., 2007]. Indeed, the results of the Cuzick Edward test indicated than cases tend to significant aggregate up to 13 elements, which suggests the existence of multiple environmental point sources of exposure. It is realistic to suppose that the nature of such large aggregations is domestic and environmental; it is likely to result either from the diffusion of asbestos fibres when the factory was in operation or to be caused by proximity to its transport pathways, as it was centered on the Casale Monferrato main residential area and its extension was such to include the factory, rail-yard, warehouse and dumping site located West, East and North, respectively, of the central residential area. The use of asbestos tailings in buildings, paving of squares, courtyards and roads and other similar instances is also likely to have contributed, as access to waste materials was arguably easier for the closest residents [Hansen et al., 1998; Howel et al., 1997; Chiappino et al., 1991; Marconi et al., 1989, Driece et al., 2010]. We may offer a tentative explanation for two further clusters. The first is East of Casale Monferrato and appears to be centered on the residential area of Frassineto Po, a small town where from the nineteen-fifties to the nineteen-seventies a family-based enterprise recycled Eternit’s used jute bags. The second one is South-West of Casale Monferrato, centered on the town of Cella Monte, where interviewed cases reported the widespread use of finely broken asbestos-cement waste materials for paving private roads, paths and courtyards. We have no similar explanation for the other clusters and indeed our analysis is intended to stimulate further investigations to identify and remediate possible hitherto unreported asbestos deposits.

As we know that models appeared to be sensible to distance threshold, selection of covariates and other modelling choices [Dreassi et al., 2008], different statistical methods were applied and compared. Changing the threshold of distance from the putative source and adjusting the models for different sets of covariates yielded consistent results, except for the raised incidence model.

Sensitivity analysis using a shorter lag time period (10 years) included a similar number of subjects and produced similar results. This is consistent with the most commonly adopted estimate for the lag time in studies on MM [Health Effects Institute, 1991]. It should however also be considered that the factory ceased its activity in 1986, that is 15 to 20 years before the diagnosis of our study cases, which may decrease any possible difference between using 20 or 10 years lag.

One limitation of our study is the higher level of non-response in controls: it affects the study power and, more importantly, may have led to selection bias. However, the difference between respondents and non-respondents was investigated recording socio-demographic and health-related data from administrative source and applying logistic regression models to investigate the variable associated with nonresponse. We observed that age and education were inversely associated with the response status, while gender was unrelated. Then, the association between asbestos exposure and MM was re-estimated and it remained unchanged, suggesting that no bias actually occurred [Airoldi et al., 2020b]. Secondly, in the original analysis [Ferrante et al., 2016] asbestos exposure was considered as the sum of occupational, domestic and environmental component while in our study we adjusted only for occupational and domestic fibre concentration. Although, when we adjusted the models also for environmental component the distance estimates were similar and coherent.

# REFERENCES

Airoldi C, Ferrante D, Miligi L, Piro S, Stoppa G, Migliore E, Chellini E, Romanelli A, Sciacchitano C, Mensi C, Cavone D, Romeo E, Massari S, Marinaccio A, Magnani C. Estimation of Occupational Exposure to Asbestos in Italy by the Linkage of Mesothelioma Registry (ReNaM) and National Insurance Archives. Methodology and Results. Int J Environ Res Public Health. 2020 Feb 6;17(3).

Airoldi C, Ferrante D, Mirabelli D, Azzolina D, Magnani C. Evaluation of Nonresponse Bias in a Case-Control Study of Pleural Mesothelioma. Int J Environ Res Public Health. 2020 Aug 24;17(17):6146.

Airoldi C, Magnani C, Lazzarato F, Mirabelli D, Tunesi S, Ferrante D. Environmental asbestos exposure and clustering of malignant mesothelioma in community: a spatial analysis in a population-based case-control study. Environ Health. 2021 Sep 15;20(1):103.

Arpa. 2020. Mappatura amianto in Piemonte. Https://Webgis.Arpa.Piemonte.It/Amianto_storymap_webapp/.

Berry M. Mesothelioma incidence and community asbestos exposure. Environ Res. 1997 Oct;75(1):34-40.

Binazzi A, Miligi L, Giovannetti L, Piro S, Franchi A, Mensi A, Calisti R, Galli P, Camagni A, Romeo E, Zajacovà J, Sorasio D, D'Errico A, Di Marzio D, Bonafede M, Marinaccio A, and the ReNaTuNS working group. ReNaTuNS sorveglianza epidemiologica dei tumori naso-sinusali. Rome, 2020

Chiappino G, Sebastien P, Todaro A. L'inquinamento atmosferico da amianto nell'ambiente urbano: Milano, Casale Monferrato, Brescia, Ancona, Bologna, Firenze [Atmospheric asbestos pollution in the urban environment: Milan, Casale Monferrato, Brescia, Ancona, Bologna and Florence]. Med Lav. 1991 Sep-Oct;82(5):424-38. Italian.

Consonni D, De Matteis S, Dallari B, Pesatori AC, Riboldi L, Mensi C. Impact of an asbestos cement factory on mesothelioma incidence in a community in Italy. Environ Res. 2020 Apr;183:108968Epub 2019 Nov 29.

CPO (Piedmont Reference Center for Epidemiology and Cancer Prevention). Malignant

Mesothelioma Registry. https://www.cpo.it/It/Articles/Show/Incidenza-Sopravvivenza-Ed-Esposizione-Dei-Mesoteliomi-1990-2016/.

Crosignani P, Massari S, Audisio R, Amendola P, Cavuto S, Scaburri A, Zambon P, Nedoclan G, Stracci F, Pannelli F, Vercelli M, Miligi L, Imbriani M, Berrino F. The Italian surveillance system for occupational cancers: characteristics, initial results, and future prospects. Am J Ind Med. 2006 Sep;49(9):791-8.

Cuzick J. and Edwards R. 1990. Spatial Clustering for Inhomogeneous Populations. J R Stat Soc. Series B (Methodological), 52 (1), 73-104.

Driece HA, Siesling S, Swuste PH, Burdorf A. Assessment of cancer risks due to environmental exposure to asbestos. J Expo Sci Environ Epidemiol. 2010 Jul;20(5):478-85. Epub 2009 Oct 28.

Diggle PJ. A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point, J R Stat Soc Ser A Stat Soc, 1990, 153, 349-62

Diggle PJ. Chetwynd AG. Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics. 1991 Sep;47(3):1155-63.

Diggle PJ and Rowlingson BS. 1994. A Conditional Approach to Point Process Modelling of Elevated Risk. J R Statist Soc 157-part 3:433-40

Diggle PJ. Statistical Analysis of Spatial and Spatio-Temporal Points Patterns, Third Edition, Monographs on Statistics and Applied Probability, 2003, CRS press

Dreassi E, Lagazio C, Maule MM, Magnani C, Biggeri A. Sensitivity analysis of the relationship between disease occurrence and distance from a putative source of pollution. Geospat Health. 2008 May;2(2):263-71

Ferrante D, Bertolotti M, Todesco A, Mirabelli D, Terracini B, Magnani C. Cancer mortality and incidence of mesothelioma in a cohort of wives of asbestos workers in Casale Monferrato, Italy. Environ Health Perspect. 2007 Oct;115(10):1401-5.

Ferrante D, Mirabelli D, Tunesi S, Terracini B, Magnani C. Pleural mesothelioma and occupational and non-occupational asbestos exposure: a case-control study with quantitative risk assessment. Occup Environ Med. 2016 Mar;73(3):147-53.

Furuya S, Chimed-Ochir O, Takahashi K, David A, Takala J. Global Asbestos Disaster. Int J Environ Res Public Health. 2018 May 16;15(5):1000.

Goswami E, Craven V, Dahlstrom DL, Alexander D, Mowat F. Domestic asbestos exposure: a review of epidemiologic and exposure data. Int J Environ Res Public Health. 2013 Oct 31;10(11):5629-70.

Hansen J, de Klerk NH, Musk AW, Hobbs MS. Environmental exposure to crocidolite and mesothelioma: exposure-response relationships. Am J Respir Crit Care Med. 1998 Jan;157(1):69-75.

Health Effects Institute. 1991. Asbestos in public and commercial buildings: A literature review and synthesis of current knowledge.

Hodgson JT, Darnton A. The quantitative risks of mesothelioma and lung cancer in relation to asbestos exposure. Ann Occup Hyg. 2000 Dec;44(8):565-601.

Howel D, Arblaster L, Swinburne L, Schweiger M, Renvoize E, Hatton P. Routes of asbestos exposure and the development of mesothelioma in an English region. Occup Environ Med. 1997 Jun;54(6):403-9.

Istat. 2021. Demografia in cifre. Http://Demo.Istat.It/.

Kelsall JE, Diggle PJ  Kernel estimation of relative risk. Bernoulli 1995a. 1(1/2):3-16

Kelsall JE, Diggle PJ. Non-parametric estimation of spatial variation in relative risk. Stat Med. 1995b Nov 15-30;14(21-22):2335-42.

Kelsall, J, Diggle, PSpatial variation in risk of disease: a nonparametric binary regression approach. Appl Stat 1998; 47(4): 559–573.

Kielkowski D, Nelson G, Rees D. Risk of mesothelioma from exposure to crocidolite asbestos: a 1995 update of a South African mortality study. Occup Environ Med. 2000 Aug;57(8):563-7.

Kulldorff M. A spatial scan statistic. Communication in Statistics - Theory and Methods 1997. 26:1481-96.

Kurumatani N, Kumagai S. Mapping the risk of mesothelioma due to neighborhood asbestos exposure. Am J Respir Crit Care Med, 2008. 178(6):624-9.

IARC Monogr Eval Carcinog Risks Hum. Arsenic, metals, fibres, and dusts. IARC

Working Group on the Evaluation of Carcinogenic Risks to Humans. 2012;100(Pt C):11-465.

ISPESL. Linee Guida per la rilevazione e la definizione dei casi di mesotelioma maligno e la trasmissione delle informazioni all'ISPESL da parte dei Centri Operativi Regionali. Seconda edizione. Roma, ISPESL, 2003

Istat. 2021. Demografia in cifre. Http://Demo.Istat.It/.

Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. Soviet physics. Doklady 1965.10: 707-710.

Magnani C, Terracini B, Ivaldi C, Botta M, Mancini A, Andrion A. Pleural malignant mesothelioma and non-occupational exposure to asbestos in Casale Monferrato, Italy. Occup Environ Med 1995. 52 (6):362-7.

Magnani C, Terracini B, Ivaldi C, Mancini A, Botta M. Mortalita per tumori e altre cause tra i lavoratori del cemento-amianto a Casale Monferrato. Med Lav 1996. 87(2):133-476,

Magnani C, Agudo A, González CA, Andrion A, Calleja A, Chellini E, Dalmasso P, Escolar A, Hernandez S, Ivaldi C, Mirabelli D, Ramirez J, Turuguet D, Usel M, Terracini B. Multicentric study on malignant pleural mesothelioma and non-occupational exposure to asbestos. Br J Cancer 2000. 83(1):104-11.

Magnani C, Dalmasso P, Biggeri A, Ivaldi C, Mirabelli D, Terracini B. Increased risk of malignant mesothelioma of the pleura after residential or domestic exposure to asbestos: A case-control study in Casale Monferrato, Italy. Environ Health Perspect 2001. 109(9):915-9.

Magnani C, Ferrante D, Barone-Adesi F, Bertolotti M, Todesco A, Mirabelli D, Terracini B. Cancer risk after cessation of asbestos exposure: A cohort study of Italian asbestos cement workers. Occup Environ Med 2008. 65(3):164-70.

Magnani C, Fubini B, Mirabelli D, Bertazzi PA, Bianchi C, Chellini E, Gennaro V, Marinaccio A, Menegozzo M, Merler E, Merletti F, Musti M, Pira E, Romanelli A, Terracini B, Zona A. Pleural mesothelioma: Epidemiological and public health issues.

Report from the second italian consensus conference on pleural mesothelioma. Med Lav 2013. 104 (3):191-202.

Marconi A, Cecchetti G, Barbieri M.. Airborne mineral fibre concentrations in an urban area near an asbestos-cement plant. IARC Scie Publ; 1989. (90):336-46.

Marinaccio A, Corfiati M, Binazzi A, Di Marzio D, Bonafede M, Verardo M, Migliore E, Gennaro V, Mensi C, Schallemberg G, Mazzoleni G, Fedeli U, Negro C, Romanelli A, Chellini E, Grappasonni I, Pascucci C, Madeo G, Romeo E, Trafficante L, Carrozza F, Angelillo IF, Cavone D, Cauzillo G, Tallarigo F, Tumino R, Melis M; ReNaM Working Group. The epidemiological surveillance of malignant mesothelioma in Italy (1993-2015): methods, findings, and research perspectives. Epidemiol Prev. 2020 Jan-Feb;44(1):23-30.

Marinaccio A, Binazzi A, Di Marzio D, Mensi C, Consonni D, Chellini E, Angelini A, Baldassarre A, Cavone D, Vimercati L, Migliore E, Verardo M, Benfatto L, Malacarne D, Eccher S, Mazzoleni G, Comiati V, Negro C, Romanelli A, Grappasonni I, Pascucci C, Madeo G, Stracci F, Romeo E, Di Giammarco A, Carrozza F, Angelillo IF, Labianca M, Tallarigo F, Tumino R, Melis M, Bonafede M, Massari S, Scarselli S, Iavicoli S, Signorini S, ReNaM Working Group. Il registro nazionale dei mesoteliomi. Settimo rapporto. INAIL, Rome 2021.

Massari S, Miligi L, Binazzi A, Bonafede M, Branchi C, Paris L, Calisti R, Crosignani P, De Girolamo G, Magnani C, Mensi C, Migliore E, Mirabelli D, Oddone E, Paredes Alpaca RI, Quattrini G, Marinaccio A. ReNaLOCCAM. Il sistema di monitoraggio delle neoplasie a bassa frazione eziologica. INAIL, Rome 2021

Maule MM, Magnani C, Dalmasso P, Mirabelli D, Merletti F, Biggeri A. Modeling mesothelioma risk associated with environmental asbestos exposure. Environ Health Perspect; 2007. 115 (7):1066-71.

Mensi C, Riboldi L, De Matteis S, Bertazzi PA, Consonni D. Impact of an asbestos cement factory on mesothelioma incidence: Global assessment of effects of occupational, familial, and environmental exposure. Environ Int; 2015. 74:191-9.

Musti M, Pollice A, Cavone D, Dragonieri S, Bilancia M. The relationship between malignant mesothelioma and an asbestos cement plant environmental risk: A spatial case-control study in the city of Bari (Italy). Int Arch Occup Environ Health; 2009. 82(4):489-97.

Newhouse ML, Berry G. Predictions of mortality from mesothelial tumours in asbestos factory workers. Br J Ind Med. 1976;33:147–51

Odell MK, "The profit in records management". Systems. New York, 1956, 20: 20.

Pearce, N. Analysis of matched case-control studies. *BMJ (Online)*, 2016, *352*, 1–4.

Pirastu R., Ancona C., Iavarone I., Mitis F., Zona A., Comba P., SENTIERI Project. Mortality study of residents in Italian polluted sites: evaluation of the epidemiological evidence. Epidemiol Prev. 2010; supplementary 3

Rai K., Combine Datasets using Inexact Character Variables in SAS, SAS Global Forum 2012, 134, Available online: https://support.sas.com/resources/papers/proceedings12/134-2012.pdf (last access on November, 2021).

Rees D, Myers JE, Goodman K, Fourie E, Blignaut C, Chapman R, Bachmann MO. Case-control study of mesothelioma in South Africa. Am J Ind Med; 1999. 35(3):213-22.

Robinson BW, Musk AW, Lake RA. Malignant mesothelioma. *Lancet* 2005;366:397-408.

Roesch A., Matching Data Using Sounds-Like Operators and SAS Compare Functions., SAS Global Forum 2012, 122, Available online: https://support.sas.com/resources/papers/proceedings12/122-2012.pdf (last access on November, 2021).

Silge, J.; Robinson, D. Text Mining with R, 1st ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2017.

Valenzuela M, Giraldo M, Gallo-Murcia S, Pineda J, Santos L, Ramos-Bonilla JP. Recent Scientific Evidence Regarding Asbestos Use and Health Consequences of Asbestos Exposure. Curr Environ Health Rep 2016;3:335-347. Review

Virta, R.L., 2006, Worldwide asbestos supply and consumption trends from 1900 through 2003: U.S. Geological Survey Circular 1298, available only online.

Wagner JC, Sleggs CA, Marchand P. Diffuse pleural mesothelioma and asbestos exposure in the North Western Cape Province. Br J Ind Med; 1960. 17(4):260-71

Weber S, Pèclat M. A simple command to calculate travel distance and travel time. Stata J. 2018.

World Health Organization. Chrysotile Asbestos. Geneva: WHO; 2014.

Xu R, Barg FK, Emmett EA, Wiebe DJ, Hwang WT. Association between mesothelioma and non-occupational asbestos exposure: Systematic review and meta-analysis. Environ Health; 2018. 17(1): 90.

Zona A, Fazzo L, Binazzi A, Bruno C, Corfiati M, Marinaccio A, SENTIERI – Epidemiological Study of Residents in National Priority Contamined Sites: incidence of mesothelioma. Epidemiol Prev. 2016; supplementary 1

# APPENDIX 1

## REMOVE WORDS

| remove records |
|---|
| IMPRESA INDIVIDUALE |
| IMPRESA FAMILIARE |
| UNSPECIFICIED |
| SPECIFICATA |
| NESSUNO |
| NESSUNA |
| UNSPECIFIED |
| IGNOTA |
| IGNOTO |
| IGNOTI |
| IGNOTE |
| UNSPECIFIIED |
| NON SPECIFICATA |
| NR |

## MISTAKE AND CORRECTIONS

| old | new |
|---|---|
| AINERI E CONTERNO | AIMERI E CONTERNO |
| FIATO AUTO | FIAT AUTO |
| FONDERIA LIMONE | FONDERIE LIMONE |
| MARZIO MAURO | MARZO MARIO |
| COMFER | CONFER |
| DEMANT | DEMONT |
| NAZI VERSOLATTO | NAZZI VERSOLATTO |
| TST | TINTORIA STAMPERIA TESSUTI |
| LIDT | LITD |
| ZORZANA | ZORZAN ARMANDO |
| BELOIT ITALIA | PMT ITALIA SPA |
| BENDER MARTINY | BENDER MARTINI |
| NOVI INDUSTRIA ACCIAIO | NOVINDUSTRIA ACCIAIO |
| IVECO | IVECO FIAT |
| RDB | RDB PREFABBRICATI |
| SIDERMAR | SIDERMAR NAVIGAZIONE |
| ATAP | ATAP PROVINCIE BIELLA VERCELLI |

| | |
|---|---|
| FINAFF | FINAFF CERUTTI GIAMPAOLO |
| SICMA | SICMA CHIMICA |
| FRENDO | FRENDO ABEX |
| AMBROSINO E DEAMBROGIO | AMBROSINO E DE AMBROGIO |
| LA PIEMONTESE | LA PIEMONTESE IMP GEN COSTRUZIONI |
| CANTIERE NAVALE ORLANDO | CANT NAV FLLI ORLANDO LCA COSTUDIO LAMANNA |
| ANSALDO CAMNTIERE NAVALE ORLANDO | CANT NAV FLLI ORLANDO LCA COSTUDIO LAMANNA |
| MONTEDISON | MONTECATINI EDISON |
| DEUTCHE BABCOCK | DEUTSCHE BABCOCK |
| SIA | SOCITALIANA PER L AMIANTO |
| TUBIGOMMA | TUBIGOMMA |
| LEUMANN | NLEUMANN |
| SOCIETÀ GENERALE PER LINDUSTRIA DELLA MAGNESIA | SOCGENERALE PER L INDDELLA MAGNESIA |
| ABRATE CO AMIANTIFERA | ABRATE CO AMIANTIFERA DI BALANGERO |
| MANDELLI | FONDERIE E ACCIAIERIE GIOVANNI MANDELLI |
| TORTA CO KELLER E CO AMIANTIFERA | TORTA CO KELLER E CO AMIANTIFERA DI BALANGERO |
| BORSALINO | BORSALINO GIUSEPPE E FLLO |
| MARCHINO UNICEM MERONE | UNICEM |
| CEMENTIFICIO BUZZI | BUZZI CEMENTI |
| CEMENTERIA MARCHINO | BUZZI CEMENTI |
| UNIONE CEMENTI MARCHINO | BUZZI CEMENTI |
| CEMENTERIA MARCHINO UNICEM | BUZZI CEMENTI |
| BUZZI SPA (POI CEMENTI BUZZI) | BUZZI UNICEM |
| ERMENEGILDO ZEGNA | ZEGNA BARUFFA LANE BORGOSESIA |
| MARIO ZEGNA | ZEGNA BARUFFA LANE BORGOSESIA |
| FIBRONIT | FINANZIARIA FIBRONIT |
| GABBA E MIGLIETTA | GABBA ELSA E C CONFEZIONI ELGA |
| ISTITUTO DONEGANI | ISTITUTO GUIDO DONEGANI |
| MAGIC PAGLIERI | MAGIC |
| MAGGIORA ERMENEGILDO | VENCHI UNICA |
| MAGGIORA | VENCHI UNICA |
| AGIP RAFFINERIA PETROLI | ERG AGIP RAFFINERIA PETROLI |
| ANTONVENETA | BANCA ANTONIANA POPOLARE VENETA |
| COTONIFICIO VALLE SUSA | VALLESUSA INDUSTRIE TESSILI |

| | |
|---|---|
| COTONIFICIO VALSUSA | VALLESUSA INDUSTRIE TESSILI |
| FACIS | FACIS GFT NET |
| FIAT=>IMP=>COMAU | COMAU DIPENDENTI |
| GFT | GFT NET |
| MANIFATTURA ROSSATI VARZI | MANIFATTURA ROSSARI VARZI |
| MOKADOR | MOKAOR |
| OFFICINE VIBERTI | OFFICINE VIBERTI VEICOLI INDUSTRIALI |
| PRINCIPI DI PIEMONTE | GRANDE ALBERGO SITEA |
| RIVOIRA PIETRO | RIVOIRA |
| RUMIANCA | RUMIANCA SIR ENI |
| GFT CONFEZIONI DI MATELICA | GFT NET CONFEZIONI DI MATELICA |
| SINGER | COMPAGNIA SINGER |
| SOCIETÀ DELLALLUMINIO ITALIANO | ALCAN ALLUMINIO |
| TALCO E GRAFITE VAL CHISONE | FILSETA VALCHISONE |
| TEKSIDAST TORINO (TO) | TEKSID AST TORINO (TO) |
| TRIONE FEDERALMOGUL | FEDERAL MOGUL OPERATIONS ITALY |
| VISCOSA | SNIA VISCOSA |
| BREDA | ANSALDOBREDA |
| BARBINI OTELLO CO CANTIERE ORLANDO | CANT NAV FLLI ORLANDO LCA COSTUDIO LAMANNA |
| IMPARATO LUCIANO | SOLMINE |
| LMI | LMI LAZZARI MONTAGGI INDUSTRIALI |
| MARZOTTO | MANIFATTURA LANE GAETANO MARZOTTO E FIGLI |
| OFFICINA MECCANICA S LUCIA | SANTA LUCIA OFFMECCANICHE PIATTELLI |
| OFFICINE SAN MARCO | OFFICINE S MARCO |
| SPINELLI CO CANTIERE ORLANDO | CANT NAV FLLI ORLANDO LCA COSTUDIO LAMANNA |
| SPINELLI CANTIERE NAVALE | CANT NAV FLLI ORLANDO LCA COSTUDIO LAMANNA |
| ANTINORI | AZIENDA AGRICOLA ANTINORI |
| LEBOLE | LEBOLEMODA |
| LA MAGONA | LA MAGONA LUCCHINI AMMINISTRAZIONE STRAORDINARIA |
| FIBRONIT | FIBRONIT |
| FIBRONIT (SAPIC) | FIBRONIT |
| FIBRONIT BARI | FIBRONIT |
| SAPICFIBRONIT | FIBRONIT |

| | |
|---|---|
| ILVA TARANTO | ILVA |
| ITALSIDERILVA | ILVA |
| AFP | ACCIAIERIE FERRERIE PUGLIESISTAB |
| ANSALDO TERMOSUD SPA | ANSALDOBREDA |
| PPTT | POSTE ITALIANE TELEFONI |
| OMI REGGIANE | OMIREGGIANE |
| ORSI MANGELLI | ORSI MANGELLI SAOM SIDAC |
| ITAMIANT | ITAMIANT LANDINI SPA DI LANDINI CAVMIRCO |
| ICAR | ICAR INDUSTRIA ETERNIT REGGIO EMILIA |
| AIA ZUCCHERIFICIO SAN PIETRO IN CASALE | ERIDANIA |
| AIE ZUCCHERIFICIO SAN PIETRO IN CASALE | ERIDANIA |
| ZUCCHERIFICIO MOLINELLA | ERIDANIA |
| ZUCCHERIFICIO OSTIGLIA | ERIDANIA |
| LOMBARDINI A SOCIO UNICO | LOMBARDINI A SOCIO UNICO |
| MONTECATINI | MONTECATINI EDISON |
| ROMANA ZUCCHERI | ROMANA ZUCCHERO |
| COPROB | COOP PRODUTTORI BIETICOLI COPROB |
| ZUCCHERIFICIO DI MINERBIO | COOP PRODUTTORI BIETICOLI COPROB |
| ZUCCHERIFICIO DI MIRANDOLA | ITALIANA INDUSTRIA ZUCCHERI |
| OGR FS | OGR OFFICINE GRANDI RIPARAZIONI FERROVIE DI STATO |
| MONTECATINI MONTEDISON | MONTECATINI EDISON |
| MONTECATINI EDISON SPA | MONTECATINI EDISON |
| ATB RONCADELLE | ATB RIVA CALZONI |
| ATB BRESCIA | ATB ACCIAIERIA E TUBIFICIO DI BRESCIA |
| CANTONI COTONIFICIO | TESSITURE CANTONI |
| CANTONI TESSITURA | TESSITURE CANTONI |
| ICAR INDUSTRIA ETERNIT REGGIO EMILIA | INDUSTRIA ETERNIT REGGIO EMILIA |
| OFFICINE FERROVIARIE GRANDI RIPARAZIONI | OGR OFFICINE GRANDI RIPARAZIONI FERROVIE DI STATO |
| TESSITURA BASSETTI | INDUSTRIA TESSILE GIOVANNI BASSETTI |
| BASSETTI SPA TELERIA | INDUSTRIA TESSILE GIOVANNI BASSETTI |
| BASSETTI VIMERCATE | INDUSTRIA TESSILE GIOVANNI BASSETTI |
| ATM MILANO | AZIENDA TRASPORTI MILANESI |
| GALBANI | EGIDIO GALBANI |
| TELERIE FRETTE | FRETTE |

| | |
|---|---|
| FINCANTIERI PALERMO | FINCANTIERI CANTIERI NAVALI ITALIANI |
| CANTIERI NAVALI PALERMO | FINCANTIERI CANTIERI NAVALI ITALIANI |
| SICIL TUBI | SICILTUBI |
| ANIC GELA | RAFFINERIA DI GELA |
| AGIP GELA | RAFFINERIA DI GELA |
| CANIERE SINCAT | EDISON *PERCHE I SINCAT FA PARTE DEL GRUPPO EDISON |
| SINCAT MONTEDISON | FINCANTIERI CANTIERI NAVALI ITALIANI SPA |
| SINGAT-EDISON | FINCANTIERI CANTIERI NAVALI ITALIANI SPA |
| CANTIERE NAVALE AUSTRALIA | FINCANTIERI CANTIERI NAVALI ITALIANI SPA |
| FINCANTIERI PALERMO | FINCANTIERI CANTIERI NAVALI ITALIANI SPA |
| CANTIERE RASIOM | EDISON |
| GRUPPO CELENE | EDISON |
| COGECO | COGECOCOMPAGNIA GENERALE COSTRUZIONI |
| FABBRICA CEMENTO AMIANTO | ETERNIT |
| ENI SPA PIAZZALE ENRICO MATTEI ROMA VARIE SEDI MILANO RA ( ) PISTICCI (MT) GHANA ( ) PAVIA MOSCA ROMA | AGIP PETROLI |
| ACOTRAL | COTRAL CONS TRASP PUBBL LAZIO |
| ACOTRALCOTRAL METRO | COTRAL CONS TRASP PUBBL LAZIO |
| MONDIAL DIAMOND | MONDIAL PULIMENTO |
| CANTIERI VARI TRA CUI IANNARILLI A ROMA | CANTIERI IANNARILLI |
| CGE ASSIGEN IN VARIE LOCALITÀ ITALIANE | CGE ASSIGEN |
| DITTE EDILI VARIE GENIO CIVILEDITTA FAMILIARE | DITTE EDILI GENIO CIVILE |
| DITTE VARIE ANCHE PRESSO LA CENTRALE DI TVS ( ANNO) | CENTRALE TVS |
| INDUSTRIA ALIMENTARE VARIE TRA CUI LA DITTA BLASI | INDUSTRIA ALIMENTARE BLASI |
| SAIP (SOCIETÀ AUTOMAZIONE IMPIANTI PRODUTTIVI SRL) PIEMONTE PRESSO OLEODOTTI METANODOTTI PIATTAFORME MARINE IN VARIE SEDI | SAIP PIEMONTE PRESSO OLEODOTTI METANODOTTI |
| VARIE DITTE EDILI TRA CUI SILVERIO MINOTTI | DITTA EDILE SILVERIO MINOTTI |
| BPD | BPD SNIA |
| BPD (BOMBRINI PARODI DELFINO) | BPD SNIA |
| BPD (COLLEFERRO) | BPD SNIA |
| BPD COLLEFERRO | BPD SNIA |
| CFC | CFC SNIA |

| | |
|---|---|
| CFC CARROZZE FERROVIARIE COLLEFERRO | CFC SNIA |
| DOMUS BPD E SUCCESSIVE DENOMINAZIONI | DOMUS BPD SNIA |
| DOMUS PRESSO BPD | DOMUS BPD SNIA |
| COLGATE PALMOLIVE ANZIO | COLGATE PALMOLIVE ANZIO |
| SIT SOCIETÀ IMPIANTI TELEFONICI | SITE SOCIMPTELEFONICI ELETTRICI |
| ASGEN | ASGEN ANSALDOBREDA |
| ATAC | ATAC AZIENDA TRANVIE E AUTOBUS COMUNE DI ROMA |
| ATAC SPA | ATAC AZIENDA TRANVIE E AUTOBUS COMUNE DI ROMA |
| AZIENDA OSPEDALIERA S CAMILLO FORLANINI | OSPEDALIERO CAMILLO FORLANINI |
| CANTIERI NAVALI RIUNITI PALERMO | CANTIERI NAVALI RIUNITI PALERMO MILITARE |
| CAR A CASERTA E SERVIZIO A CIVITAVECCHIA | CAR A CASERTA E SERVIZIO MILITARE A CIVITAVECCHIA |
| CASERMA AUTOGRUPPO X | CASERMA AUTOGRUPPO X DIFESA NAZIONALE |
| CASERMA DI ALMONE IN ROMA | CASERMA DI ALMONE IN ROMA DIFESA NAZIONALE |
| CECCHIGNOLA | CECCHIGNOLA DIFESA NAZIONAL |
| CENTRALE TERMICA STAZ TERMINI ROMA | CENTRALE TERMICA STAZ TERMINI ROMA FFSS |
| CNEN ENEA | ENEA ENTE PER LE NUOVE TECNOLOGIE LENERGIA E AMBIENTE |
| COMECI CIVITAVECCHIA | COSTRUZIONI MECCANICHE CIVITAVECCHIA |
| COMECI | COSTRUZIONI MECCANICHE CIVITAVECCHIA |
| FONDERIA MIRAFIORI TORINO (TO) | FONDERIA MIRAFIORI TORINO FIAT |
| FRATELLI ROSCIOLI | FRATELLI ROSCIOLI FFSS |
| FS GRANDI OFFICINE (TO) | FFSS GRANDI OFFICINE (TO) |
| FS NUOVO SALARIO (SMISTAMENTO) ROMA | FFSS NUOVO SALARIO (SMISTAMENTO) ROMA |
| FS OFFICINA SEDE DI ROMA | FFSS OFFICINA SEDE ROMA |
| FS OFFICINA SEDE ROMA | FFSS OFFICINA SEDE ROMA |
| FS STAB PRESSO CAMPO DI MARTE FIRENZE | FFSS STAB PRESSO CAMPO DI MARTE FIRENZE |
| FS STAB PRESSO SLORENZO ROMA | FFSS STAB PRESSO SLORENZO ROMA |
| GRANDI OFFICINE FS TORINO | GRANDI OFFICINE FFSS TORINO |
| OFFICINA GRANDI RIPARAZIONI DELLE FS VICENZA | OFFICINA GRANDI RIPARAZIONI DELLE FFSS VICENZA |
| HASSEY FERGUSSON | MASSEY FERGUSON |
| IPZS | POLIGRAFICO E ZECCA DELLO STATO |
| PPTT BELSITO | POSTE ITALIANE TELEFONI BELSITO |

# STOP WORDS

| | | | | | |
|---|---|---|---|---|---|
| UNSPECIFIED | VARIO | AZIENDA | IMPRESA | COOPERATIVA | PRIVATI |
| UNSPECIFICIED | FAMILIARE | GIÀ | IMPRESE | COOPERATIVE | PUBBLICA |
| NESSUNO | FAMILIARI | FALLIMENTO | INDUSTRIA | ITALIA | POLITECNICO |
| NESSUNA | PROPRIO | FALL | INDUSTRIE | ITALY | UNIVERSITA |
| IGNOTA | DEGLI | FALLIMENTI | SOCIETA | ITALIANA | MULINO |
| IGNOTO | DELLA | LIQUIDAZIONE | SOCIETÀ | ITALIANI | OSPEDALE |
| IGNOTEIGNOTI | DEI | LIQUIDAZIONI | NEGOZIO | ITALIANO | ELIMINATA |
| DELLO | DELLE | NUOVA | NEGOZI | ISTITUTO | ELIM |
| DEL | VIAPER | NUOVO | DITTA | ISTITUTI | FALLITA |
| E | DOMICILIO | NUOVI | DITTe | STATALE | FALLITO |
| SNC | IN | NUOVE | OFFICINA | AGENZIA | FALLITI |
| SRL | DI | AZIENDA | OFFICINE | BOTTEGA | POI |
| SPA | SAS | AZIENDE | FLLI | BOTTEGHE | |
| VARIE | SCRL | IMPIANTO | FRATELLI | PRIVATA | |
| VARI | SOC | IMPIANTI | ENTE | PRIVATO | |

# BI-GRAMS

| |
|---|
| CANTIERE NAVALE |
| CASALE MONFERRATO |
| COMPAGNIA NAVIGAZIONE |
| SCUOLA MEDIA |
| SCUOLA SUPERIORE |
| SCUOLA ELEMENTARE |
| SCUOLA PRIMARIA |
| SCUOLA SECONDARIA |
| ASILO NIDO |
| SCUOLA PROFESSIONALE |
| FORZE ARMATE |
| GUARDIA FINANZA |
| DIFESA NAZIONALE |
| MARINA MERCANTILE |

# SPECIAL WORDS

| | | |
|---|---|---|
| PIRELLI | OLIVETTI | ANSALDOBREDA |
| ENEL | CEAT | MOTOFIDES |
| FIAT | TELECOM | SIA |
| ETERNIT | CMF | EDISON |
| MONTEFIBRE | SAIPEM | MONTEDISON |

| | | |
|---|---|---|
| FARMITALIA | SOLVAY | OGR |
| RIV | TOREMAR | FALCK |
| NOVINDUSTRIA | FINTECNA | SNIA |
| AERITALIA | ILVA | ENI |
| SKF | BREDA | |

# NON-SPECIFIC WORDS

| | | | | |
|---|---|---|---|---|
| IDRAULICO | AUTOFFICINA | ORAFO | AUTONOMO | VERDURA |
| IDRAULICA | COIBENTAZIONE | LAVANDERIA | IDRAULICI | STAB |
| BAR | FORNACE | EDILE | PORTO | BALNEARE |
| CALZATURIFICIO | FORNACI | EDILI | ARSENALE | VIAGGIO |
| CALZATURIFICI | SARTORIA | NAVALE | AEREOPORTO | VIAGGI |
| RISTORANTE | SPORT | CARROZZIERI | ACCIAIERIE | IMPORT |
| RISTORANTI | TESSITURA | TRASPORTI | VETRERIA | EXPORT |
| CALZOLAIO | TESSITURE | TRASPORTO | PRIGIONIERO | CERAMICA |
| PASTICCERIA | TESSILE | ZUCCHERIFICIO | GUERRA | STRADALE |
| PASTICCERIE | PANETTIERE | ZUCCHERIFICI | FONDERIA | STABILIMENTO |
| PASTICCERE | PANETTERIA | HOTEL | AUTOMOBILE | SORVEGLIANZA |
| BARBIERE | PANIFICIO | VILLA | FORNAIO | ROSTICCERIA |
| LATTERIA | CARBONAIO | LANIFICIO | AMMINISTRATORE | PESCHERIA |
| LATTERIE | FORNACE | MERCERIA | AUTONOMO | AMBULANTE |
| CERNITA | STUDIO | MERCERIE | FAMILIARE | PARTITO |
| CERNITE | GEOMETRA | RIFINITURA | PETROLIERE | POLITICO |
| CEMENTIFICIO | NEGOZIO | OSPEDALE | CASA | FIORI |
| CEMENTIFICI | ABBAGLIAMENTO | COMUNE | IMBIANCATURE | ABBIGLIAMENTO |
| ALIMENTARI | ELETTRODOMESTICI | CONSORZIO | MANIFATTURA | MISSIONE |
| TAPPEZZERIA | ORTOFRUTTA | COMPAGNIA | OFFICINE | INGROSSO |
| TAPPEZZERIE | ALIMENTARI | NAVIGAZIONE | OFFICINA | CEREALI |
| AUTOLAVAGGIO | MINIERA | NAVALE | EDILIZIA | IMBIANCHINO |
| AUTOLAVAGGI | CARBONE | NAVIGAZIONE | FALEGNAMERIA | IMPAGLIATURA |
| MACELLERIA | MERCURIO | MARITTIMA | FALEGNAMERIE | SEDIE |
| MACELLERIE | DISCARICA | STRADALE | CARPENTERIA | COSTRUZIONE |
| MACCELLAIO | RIFIUTI | MERCI | FABBRICA | COSTRUZIONI |
| MACELLAI | DISTRIBUTORE | CANTIERE | BICICLETTE | ALBERGO |
| ORTOFRUTTA | CARBURANTE | COMPAGNIA | FABBRICHE | ALBERGHI |
| CARROZZERIA | BARBIERE | ELETTROMECCANICA | MOTO | INDUSTRIALI |
| CARROZZIERE | FABBRO | LAVORATORE | FRUTTA | INDUSTRIALE |
| MURATORI | | | | |

# APPENDIX 2

Trino
Palazzolo Vercellese
Balzola
Villanova Monferrato
Morano sul Po
Camino
Moncestino
Gabiano
Coniolo
Villamiroglio
Pontestura
Solonghello
Mombello Monferrato
Casale Monferrato
Frassineto Po
Cerrina Monferrato
San Giorgio Monferrato
Ozzano Monferrato
Odalengo Grande
Serralunga di Crea
Treville
Ticineto
Valmacca
Cereseto
Borgo San Martino
Murisengo
Ponzano Monferrato
Odalengo Piccolo
Castelletto Merli
Sala Monferrato
Rosignano Monferrato
Terruggia
Bozzole
Pomaro Monferrato
Cella Monte
Occimiano
Villadeati
Giarole
Moncalvo
Ottiglio
Alfiano Natta
Frassinello Monferrato
Olivola
Conzano
Mirabello Monferrato
Camagna Monferrato
Valenza
Vignale Monferrato
Lu
San Salvatore Monferrato
Bassignana
Cuccaro Monferrato
Altavilla Monferrato
Pecetto di Valenza
Rivarone
Fubine Monferrato
Montecastello

# FUNDING

The BRIC 59 was conducted thanks to a grant from INAIL (Piano Ricerca 2016–2018, “Programma Speciale Amianto”, Ricerca BRIC id 59).

The Casale case-control study was supported by a grant from Regione Piemonte, Ricerca Sanitaria Finalizzata 2004 and by the Regione Piemonte, Ricerca Scientifica Applicata 2003 (grant no.258/2003). Analyses were supported by the Italian National Institute of Health—ISS, ‘Progetto Amianto’.

# ACKNOWLEDGEMENT

*Grazie. Il primo vero grazie è alla mia "nuova" e "vecchia" famiglia.*

*A te Marco, che mi hai preso per mano e hai deciso di camminare con me per tutta la vita. Grazie per starmi accanto, non davanti, non dietro, ma semplicemente accanto..mano nella mano. Grazie per accompagnarmi in tutte le gioie e le sofferenze che la vita ci ha messo, ci mette e ci metterà davanti. Grazie per il supporto e la stima incondizionata che hai di me e che mi permette di seguire i miei sogni, sempre. Grazie perché quando il mare era in tempesta sei stato un porto sicuro dove attraccare e ora un nuovo faro di luce illumina la nostra vita.*

*Alla mia mamma e al mio papà, che ci sono sempre e che nonostante qualche acciacco sono il mio pilastro e la mia forza. A voi che siete un rifugio sicuro, un nido caldo e accogliente dove non mancano parole per scaldare il cuore e cibi per riempire la pancia, grazie. A mia sorella e mio fratello, sempre presenti in ogni momento della vita. A voi, alle volte così diversi da me ma così simili, grazie per indicarmi sempre la via, per accompagnarmi e guidarmi passo dopo passo. Grazie anche ai vostri compagni di vita che hanno reso la famiglia ancora più allegra ma soprattutto alla piccola Ilaria, che non smette di insegnarmi cosa conta davvero nella vita. A lei che mi dà l'energia per credere e combattere sempre per un mondo migliore, grazie. Grazie anche al piccolo in arrivo, non ti conosciamo ancora ma ti amiamo già tanto! Non sarei la persona che sono oggi senza di voi e non vi cambierei per nulla al mondo.*

*Ma questo dottorato, non sarebbero mai stato possibile senza la presenza di tante altre persone speciali.*

*Grazie in primis al prof Corrado Magnani, che dopo avermi introdotto nel mondo della ricerca sugli effetti dell'amianto ha continuato a darmi fiducia e mi ha incentivato ad iniziare questo dottorato. Grazie poi a Daniela, che dopo avermi sopportato per la tesi magistrale ha continuato a seguirmi in questi anni consigliandomi come approcciarmi a questo mondo. Grazie a Dario, sempre scrupoloso e preciso nel definire ogni aspetto relativo all'esposizione ad amianto. Grazie di cuore anche al prof Faggiano, senza il quale non avrei avuto l'opportunità di continuare a lavorare in università. Da voi, oltre* *alle conoscenze tecniche, quello che ho davvero imparato, è la passione nel fare ricerca, nel porsi continuamente domande e dubbi e cercare le risposte.*

*Grazie ai miei colleghi di ufficio, Silvano e Lorenza, che in questi anni hanno non solo sopportato i miei borbottii e lamentele ma si sono mostrati ottimi maestri. A voi che risolvete i miei dubbi statistici e sollevate sempre nuovi problemi metodologici da affrontare, un grazie davvero speciale…siete poco normali ma sicuramente statisticamente significativi per me! Grazie anche a tutti gli altri colleghi che ho incontrato in questi anni e che hanno reso sempre speciale la permanenza in UNIUPO. Erica, Silvia, Sara, Anil, Andrea, Alessandro, Roberto, i colleghi di PHD e tutti voi che avete condiviso un caffè, una chiacchierata …grazie.*

*Ma la vita non è fatta di solo lavoro, anzi. Un grande grazie agli amici di sempre che magari non possono darmi qualche consiglio su come scrivere una tesi o fare un modello ma che sicuramente sanno come riempire di gioia le mie giornate e soprattutto brindare ai successi! Nicole, Viviana, Isabel voi siete le mie ochette, le mie amichette con cui crescere insieme, ridere della vita e piangere delle sfighe (sempre tante)! Grazie perchè ormai siete così presenti nella mia vita che in fondo, vi voglio bene! Alle mie compagne di viaggio, Ari, Ale e Tati che hanno sempre un biglietto pronto per evadere dalla vita e una meta dove trascorrere momenti speciali. A Fede e Vale, così distanti fisicamente ma così unite nel cuore. Grazie anche al gruppo "Sfiga", con voi è sempre bello condividere passioni e disagi.*

*E l'ultimo grazie va ancora a te, piccolo mio. Per te ho posticipato questo dottorato, per te sono stata 40 giorni in ospedale, per te ho stravolto le mie priorità…ma lo rifarei mille e ancora mille volte perché quello che ho dato non è nulla rispetto a quanto ho ricevuto. Grazie di avermi dato la possibilità di essere la tua mamma!*