Master of Public Health

Master international de Santé Publique

Spatial distribution of environmental exposures relationship with socio-economic characteristics

Presented by:
Thanh Huyen NGUYEN

EHESP MPH 2011

Location of the practicum: EHESP

Professional and academic advisor
Dr. Séverine DÉGUEN, EHESP

Acknowledgments

I would like to take this opportunity to thank all those who contributed to the accomplishment of this work.

My deepest thanks go to my academic advisor, Dr. Séverine Déguen for her guidance and Benoît Lalloué for his thorough support during the period of my practicum. I would also like to thank Dr. Denis Zmirou and Cindy Padilla for their contribution to this work.

I would like to express my special and sincere gratitude to all members of Infobiostat Department of EHESP for their assistances and facilitation for my work and life in Rennes.

I would like to thank Dr Alan O'Rourke, my supervisor during the first year in the University of Sheffield as a Europubhealth student.

I am indebted to the Europubhealth Consortium, the University of Sheffield, the MPH program and the EHESP for their organizing, co-ordinating and support during my two-year master course.

Last but not least, I also thank to my family and friends who always give me encouragement and motivation during my study.

TABLE OF CONTENT

List of Figures

List of Tables

List of Acronyms

CI: Confidence interval

CO_2: Carbon dioxide

EHESP: The French school of Public Health (École des Hautes Études en Santé Publique)

INSEE: National Institute of Statistics and Economic Studies (Institut national de la statistique et des études économiques)

IQR: Interquartile range

IREP: French Register of Pollutant Emissions (Registre Français des Emissions Polluantes)

IRIS: French Census Block (Ilots Regroupés pour l'Information Statistique)

NO_2: Nitrogen dioxide

NOx: Nitrogen oxides

OR: Odds ratio

$PM_{2.5}$: Particulate Matter with 2.5 micrometers or less

PM_{10}: Particulate Matter, 10 micrometers or less

PRTR: Pollutant Release and Transfer Register

SES: Socioeconomic status

SOA: Super output area

SO_2: Sulfur dioxide

VOC: Volatile organic compound

WHO: World health organization

Spatial distribution of environmental exposures relationship with socio-economic characteristics

Abstract in English

Background: In the recent past decades, researches suggested that deprived population are more likely to have greater environmental exposure than the wealthier. However, literature in environmental justice research revealed mixed findings and has not been well documented in France.

Objectives: It aims to explore the spatial distribution of environmental exposures in relation with socio-economic characteristics in the two French metropolitan areas Lille and Lyon.

Methods: Industrial proximity index was produced by using two distance-based approaches: spatial coincidence and buffer. Logistic regression was employed to test the association between industrial proximities and socioeconomic conditions at French census block (IRIS) level. Odds ratios and their CI 95% were estimated as the measure of association strength between the proximity to industrial facilities and SES.

Results: There were important differences of data between the two metropolitan areas Lille and Lyon, for most of index in both SES and industry (p-value <0.001), for example the SES index of foreigners, unemployed or all industrial proximity indices with buffer. Analyses also revealed the existence of significant associations between certain socioeconomic characteristics of populations and proximity to polluting industries especially in the metropolitan area of Lyon, e.g. p-value <0.001 for all associations between indices of education and the proximity index of industry presence. Overall, socio-economically disadvantaged populations live closer to polluting industrial sites.

Conclusions: This study demonstrated that the deprived population are more likely to suffer from higher industrial exposure. These disadvantaged populations are also known to be more vulnerable to the effects of pollution. Additional analysis is underway to take into account the emissions of each industry in the construction of the index of proximity to polluting industries.

Distribution spatiale des relations entre les expositions environnementales et les caractéristiques socio-économiques-une étude écologique menée à fine échelle

Abstract in French

Contexte: Les recherches menées depuis plusieurs années ont suggéré que les populations socioéconomiquement défavorisées vivraient dans des zones plus polluées que les populations plus riches. Cependant, des etudes de la justice environnementale ont révélé plus récemment qu'en Europe les résultats étaient plus discutés et cela nécessitait d'être documentées en France.

Objectifs: Explorer la distribution spatiale des expositions environnementales avec des caractéristiques socio-économiques dans deux agglomérations françaises, Lille et Lyon.

Méthodes: L'indice de proximité industrielle est calculé en utilisant deux approches basées sur la distance: la méthode de « coïncidence spatiale » et celle du buffer. Un modèle de régression logistique a été utilisé pour tester l'association entre la proximité des populations aux industries polluantes et leurs caractéristiques socio-économiques à l'échelle de l'IRIS (Ilots Regroupés pour l'Information Statistique). La force d'association et sa significativité ont été estimés par l'odds ratio et son intervalle de confiance à 95%.

Résultats: Des différences significatives entre les deux agglomérations étudiées ont été révélées pour de nombreuses caractéristiques socioéconomiques et différents indicateurs de proximité aux industries polluantes (p-valeur <0.001) (par exemple pour étranger, chômeur ou les indices avec buffer). Les analyses mettent également en évidence l'existence d'associations significatives entre certaines caractéristiques socioéconomiques et la proximité des populations aux industries polluantes notamment sur l'agglomération de Lyon, par exemple, p <0.001 pour les associations entre tout les indices de l'éducation à Lyon et l'indice de proximité de la présence de l'industrie. Globalement, les populations socio-économiquement défavorisées vivraient à plus proximité des industries polluantes.

Conclusions: Cette étude a révélé que les populations défavorisées sont plus exposées aux expositions industrielles. Ces populations, connues pour être plus vulnérables aux effets de la pollution, supporteraient également un différentiel d'exposition. Des analyses complémentaires sont en cours de réalisation afin notamment de prendre en compte les émissions de chaque industrie dans la construction de l'index de proximité aux industries polluantes.

1. Introduction

1.1. Environmental inequity and its relation to vulnerability of human health

In recent past decades, findings from researches in public health suggested the likely location of hazardous industrial sites in areas populated by ethnic minorities (Brown, 1995, Perlin et al., 1999) , a situation that contradicts the principle of environmental justice which refers to “the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation and enforcement of environmental laws, regulations and policies” (Environmental Protection Agency, 2010).

To examine this growing issue, researches have been further conducted to investigate distributional dimension of environmental justice, which focuses on the uneven distribution of environmental burden in different social, economic, racial and ethnic groups. Obviously, evidence for such inequities, referred to as environmental injustice, has a critical suggestion that marginalized and socially disadvantaged groups are more likely to suffer from higher polluted living environment (Brown, 1995, Morello-Frosch, 2002a, Morello-Frosch, 2002b). In other words, deprived population becomes subject to farther vulnerability of health with hazardous environmental burden, in addition to the health impact of being disadvantaged in socio-economic status.

1.1.1. Health impact of environmental nuisances

Over the last three decades there has been increasing global concern over the public health impacts attributed to environmental pollution, in particular, the global burden of disease. The World Health Organization (WHO) estimates that about a quarter of the diseases facing mankind today occurs due to prolonged exposure to environmental pollution. For example, Kunzli and co-workers in their study (2010) reported an significant association between exposure to air pollution ($PM_{2.5}$ and traffic proximity) and the progression of atherosclerosis in human, indicated with changes in common carotid artery intima-media thickness. Their findings suggest that air pollution may contribute to the acceleration of cardiovascular disease development, the main causes of morbidity and mortality in many countries (Künzli N, 2010). To further support the statistical results, biological mechanisms have also been suggested to explain the important effect of air pollution on cardiovascular morbidity and mortality in a study focusing on effects of particulate air pollution on cardiovascular risk (Franchini M, 2009).

It is well known that exposures to environmental pollution represent important risk factors for many diseases. Environmental hazards could come from diverse chemical, physical or biological agents, which have been proved to be related with a wide range of adverse health effects as cancer development, endocrine disruption, respiratory diseases, reproductive and developmental effects, cardiovascular effects, mortality, mental disorders, etc (Iqbal S, 2010, Peek et al., 2009). A large and growing body of literature has investigated the effect of hazardous environmental exposure on human health, for instance smoking substances, water disinfection by products, pesticides or air pollution, of which air pollution is one of the widely investigated agents.

Air pollution comes from many different sources, including natural processes and man-made activities of which emissions from industrial, power plants and vehicles play important roles. Including several types of substances as particulate matter, carbon monoxide, nitrogen, sulfur oxides, metals, volatile organic compounds, pesticides, radiation and bioaerosols, etc, air pollution constitutes one of the major threats to human in terms of public health. The health effect of air pollution has been recognized for recent past decades, with different levels in human health and a wide range of organ systems such as cardiovascular, respiratory, immunological, neurological, reproductive and developmental systems (Curtis, 2006). In a study conducted in 1989, Gorham and colleagues found statistically significant positive associations between these two measures of air pollution (sulfur dioxide and ultraviolet-light-blocking aerosols) and age-adjusted mortality rates for colon cancer in both sexes and breast cancer in women (Gorham ED, 1989), while significant associations between NO_2, PM_{10} and ischemic stroke admission and in Hospital mortality were also found in a more recent study (Vidale S, 2010).

Many studies have revealed the potential health impacts of living close to industrial sites, eg waste sites or incinerators, which constitutes a public health issue. Along with the diversity of types of hazardous industrial sites, there are also a wide range of toxicological substances released from these facilities into the environment. Researches are ongoing on investigating health effects of different variables of environmental hazard domain, including not only the emissions of environmental pollutants but also the proximity to hazardous sources.

Generally speaking, ambient concentrations of some certain air pollutants (eg, aromatic hydrocarbons) were inversely associated with distances to hazardous sources, such as major roadways with high traffic densities and gasoline stations (Kwon J, 2006). The risk of adverse health events tended to increase when people lived closer to waste sites, and became more important in the urban areas than in rural areas (Kuehn, 2007). Kuehn and co-workers reported an increased risk of heart malformations, skin, musculoskeletal, reproductive and urogenital alterations related to the proximity to waste sites. Also, another outstanding study in Japan suggested positive associations between proximity to major roads and preterm births at all gestational ages (Yorifuji T, 2011). For instance, living within 200 m increased the risk of births before 37 weeks by 1.5 times (95% CI = 1.2-1.8), birth before 32 weeks by 1.6 times (1.1-2.4), and births before 28 weeks by 1.8 times (1.0-3.2). It is interesting to note that exposure to traffic-related air pollution increases even the risk of preterm births of less than 30 weeks' gestational age and proposes a possible mechanism (Yorifuji T, 2011). To conclude, not only air pollutants that is already well-known for its relationship with certain adverse events, but also industrial proximity can be considered as predictors of health effects.

1.1.2. Health impact of being disadvantaged socioeconomic status

A considerable amount of literature has been published on the association between social deprivation and the rate of mortality as well as morbidity, across a wide range of diseases and areas (Carstairs, 2000, Davey Smith, 2002). Davey Smith and colleagues in a study conducted in the UK (1998) pointed that manual social class and early termination of full time education were associated with higher blood pressure, shorter height and poorer lung function. Also, in almost European countries, the rates of death were substantially higher in groups of lower socioeconomic status (Mackenbach JP, 2008). Across the Atlantic Ocean, Heo S and colleagues in the United States (2010) revealed the findings that better economic status, along with other social factors, is related to better health-related quality of life (Heo S, 2010). In conclusion, socioeconomic conditions are one of predictors for health outcome.

1.1.3. Complex relationships among the 3 domains

When taking into account the 3 domains (including SES, living environment and health status), and combining the 2 associations above together, people who are socially disadvantaged or marginalized have to suffer from “triple jeopardy” of environmental inequity (Jerrett, 2001). This “multiple deprivation” of poor socio-economic status, poor environmental condition and poor health was also mentioned in Carstairs’ study (1981).

According to Kantrowitz and O’Niell, there are 2 dominant mechanisms that explain how environmental exposure may contribute to social health inequalities (Kantrowitz, 2002, O'Neill et al., 2003). Firstly, disadvantaged groups are recognized as being more frequently or more intensely exposed to pollution than affluent population (**differential exposure**). This potential mechanism constitutes precisely the objective of our work which aims to test the association between socio-economic status and environmental condition. And details of the hypothesis will be given in the section of rational and objective. Secondly, exposure to environmental nuisances might cause greater health effects among population with low socioeconomic status than among those with higher socioeconomic level (**differential**

susceptibility). Thus, some factors that are more common in the disadvantaged populations can modify the relationship between pollution and health outcomes.

1.2. Association between SES and industrial proximity

With the objective of testing the differential exposure hypothesis, this part we reviewed the literature that focuses on the association between socio-economic condition and the industrial proximity index. The searching method and summary of article review will be reported, respectively.

1.2.1. Searching methods

Articles relating to research field were obtained by using literature search tools in Medline database of National Library of Medicine. Only articles written in English were selected.

Three principle terms were used for the literature search: '(Social OR economic) AND (industrial OR pollution) AND (proximity OR distance).' A number of synonymous expressions were also used for the two key terms, such as 'deprivation, education, income, poverty' for socioeconomic status and 'facility, hazardous sites, landfill, waste site' for industrial sites. To maximize the effectiveness of search tools, some more general expressions, such as environment (in)equity, environment (in)justice were also included.

Using these keys words, there were 141 articles available in search results, however, after reading the abstracts, 135 papers were excluded. Those were researches which did not deal precisely with the topic of our literature review, either out of the field of the topic or dealing with the effect modification of SES or environmental factors.

One paradigm of irrelevant articles achieved by this search was a study conducted in Mediterranean countries with title "Industrial harvesting of olive tree pruning residue for energy biomass"(Spinelli R, 2010) There was no information of the relationship between industrial variables and socio-economic status. Instead, the research aimed to test two industrial pruning harvesters, capable of overcoming the limits of lighter units appeared in the past years. In another study, the authors investigated the variations of ground-level ozone and its precursors in Pearl River Delta. The variations of concentrations of several air pollutants (eg nitrogen oxides (NOx), volatile organic compounds (VOCs)) were mentioned but no relation to socio-economic conditions were either given in the results or discussed in the discussion section. (Shao M, 2009)

Using these above key words, there were also a number of papers investigating the modification of socioeconomic conditions or industrial indicators to the relationship between the other two out of three domains, including SES, environment and health. For example, a study in Canada was conducted to evaluate whether proximity to highway interacts with individual and neighbourhood socioeconomic status (SES) to influence birth outcomes (Généreux M, 2008). And counterintuitively, high SES mothers was found to be more likely than low SES mothers to experience adverse births associated with residential proximity to highway. This finding is important in studying of differential susceptibility; however, details of SES characteristics of the population living near highway were not reported in the results or discussion sections. In another study in Brazil, Martins et al also suggested that socioeconomic deprivation represents an effect modifier of the association between air pollution and respiratory deaths. Out of line with suggestions from other researches, the effect of PM_{10} in this study was found to be negatively correlated with both percentage of people with college education and high family income, and it was positively associated with the percentage of people living in slums (Martins MC, 2004). Being extremely useful when doing research in environmental inequality, however, no information about differential exposure was mentioned in the results or discussed in the discussion part of the paper.

Besides 6 articles selected by using this search method, additional 5 papers from different sources were reviewed to assess the direct associations between environmental pollution and the socio-economic status. Apart from some articles dealing with systematic review, total 11 papers were summarized in table 1 with critical information on the geographical origin, study designs, different variables and key results.

1.2.2. Article review

A. Geographical origin

Most of studies were conducted in America, eg (Anderton DL. , 1994, Kearney and Kiros, 2009), with only a few conducted in Europe such as the UK (Anderton DL. , 1994, Brender JD, 2008) or France (Viel et al., 2011).

B. Variables and geographical unit

To generate the association between environmental factors and social aspects, different types of variables were used in different studies depending on settings, approaches and capacity. Environmental variables was proximity to hazardous facilities or high traffic roads (Kearney and Kiros, 2009, Perlin SA, 2001, Viel et al., 2011), while as for socio-economic status variables, a wide range of social indicators were applied in environmental equity studies, from Scottish index in researches in UK (Fairburn, 2005), to variables of singular social domain (eg maternal occupation in US (Brender JD, 2008)) or combination of several options in numerous social aspects: income, education, employment status, race, ethnic, etc (Downey, 2008, Finkelstein, 2005). Data on air pollution and proximity were mostly achieved by using specific models based on industry data (Fairburn, 2005), whereas information on socio-economic were obtained from census or individuals for personal details (Brender JD, 2008)

Different geographical units were investigated in different studies. At crude unit levels or a finer geographical scales, the studied units could be at district scale of 139.000 persons, ward scale of 6200 persons, super output scale of 1500 persons in England (Briggs et al., 2008), or an enumeration area of 750 population in Canada (Finkelstein, 2005).

C. Study design

Majority of researches were conducted at geographical level, whereas only a few dealing with individual data and multilevel approach taking into account both individual and geographical information.

- Ecological studies

Among the group of ecological papers, most of studies conclude that population with low SES are recognised as being more often exposed to air pollutants and toxicants (Perlin et al., 1999, Briggs et al., 2008). This association occurs in both environmental variables: the mean concentration of pollutants and the proximity of people's living places to pollution source, including traffic roads, industrial facilities and waste disposal sites (Mohai et al., 2009, Perlin SA, 2001, Viel et al., 2011). Briggs strongly pointed out positive correlation with most of air pollutants (VOCs, r=0.32; PM_{10}=0.26) at super output area (SOA) level, although the association between deprivation index and proximity to hazardous sites are rather weak (r=0.09-0.16) (Briggs et al., 2008). Especially, both race and SES predict a disproportionate spatial distribution of environmental burdens. In a study investigating the association/relation between race, education, income and the proximity to hazardous sites, Mohai and colleagues showed that blacks and residents at lower educational levels were significantly more likely to live within a mile of a polluting facility. Also, low income level is one of the predictors of living closer to polluted site, though to a lesser degree (Mohai et al., 2009). In addition, results were consistent in Perlin's study which focused on race and poverty (Perlin SA, 2001), and Viel's with income and ethnic minority status as the major SES variables (Viel et al., 2011).

Nevertheless, there was a study found non-linear relationships between SES and proximity to noxious facilities (Anderton DL. , 1994). In Anderton's study, no statistically significant differences were obtained between the racial or ethnic composition of tracts which contain commercial facilities for treatment, storage & disposal of hazardous wastes.

Although inverse relationship existed in some studies of SES and air pollution concentraion, where a greater burden is borne by the affluent (Perlin et al., 1995, Forastiere et al., 2007, McLeod H. et al., 2000), no inverse association was found in the literature review of association between SES and industrial proximity by using this search method. It is, however, important to note that there was still the evidence for that inverted direction in environmental inequity studies. For instance, revealing the suggestion that households of higher social class were more likely to be located in areas with high traffic emissions, Forastiere gave the results at block income level, of which concentrations of all investigated pollutants increase from low to high income categories (eg 10.4 vs 26.7 µg/m^3 for NO_x). Moreover, when investigating relationship of SES and common pollutants (NO_2, SO_2, PM_{10}) in Wales and England, McLeod also found results that were not in line with majority studies in environmental inequity and indicated that the 3 pollutants have significantly positive relationships with not only social class but also ethnicity. The existed inverse results in studies dealing with social condition and air pollution might raise the need for further research on the association between SES and proximity index, including clear description of methods and diverse settings.

- Multilevel (semi-individual)

Majority of studies with multilevel approach has their results in line with suggestions from previous works that deprived population are more likely to have greater environmental exposure than the wealthier. A multilevel study in Texas, US aimed to test the relationship between maternal occupation and residential proximity to industrial sources also brought the evidence of environmental inequity, suggested that women in production occupations were twice as likely (95% confidence interval 1.3 to 3.0) to live near industrial facilities than women in management/professional occupations (Brender JD, 2008).

While there are a certain amount of purely individual studies investigating association between SES and pollutant concentration, no research at individual level was found in literature dealing with industrial or traffic proximity. Nevertheless, to give more information on the individual studies in environmental justice dealing with air pollutants, there are several researches confirmed the same implication that deprived population were prone to environmental exposure (Kohlhuber et al., 2006). For example in a study in Finland, personal 48h $PM_{2.5}$ exposure was suggested to be strongly associated with occupational status, with lower exposures for white-collar employees compared to other category (mean $PM_{2.5}$ levels are 11.97 and 20.46 µg/m3 respectively) (Rotko T, 2000). However, weak association and mixed findings were also found in individual researches, which implied that the assumption that more disadvantage populations have higher levels of exposure to environmental population does not always be applied and requires further explanation in different international settings (Vrijheid, 2010)

In summary, the strength and direction of association varied at observations at different approach and different settings. Most of researches on the relationship of social condition and proximity index come from America and were conducted at geographical level rather than individual degree.

2. Rationale and objectives

Basically, socio-economic status is the result of multiple interactions of factors where environmental index plays an important role. And the association between environmental exposures and SES has been described in previous studies.

In environmental justice research, especially in the field of differential exposure, proximity to industrial sites has been recognized to be negatively related to social conditions (Brender JD, 2008). However, the results still vary depending on the settings and approach. Therefore, there is a burning need for further studies.

In epidemiological studies, socio-economic status may act as modifying factors when an association between SES and environmental pollution occurs in epidemiological studies (Blakely, 2004, Blakely, 2000). In several recent researches, socioeconomic status were revealed to modify the association between air pollution and mortality (Forastiere et al., 2007, Jerrett et al., 2004, Zanobetti A, 2000). Obviously, the lack of control for such effects may exist. And this shortcoming has to be taken into account when caution is required to interpret results of epidemiological studies. Therefore, the importance of investigating relationship between SES and environmental conditions needs to be concerned for the sake of epidemiological research.

This study was part of a project in Environmental Health and Health Inequalities named “Equit’Area” conducted by the Environmental Health Department of the French School of Public Health (EHESP).

The objective of the research is to study the spatial distribution of environmental nuisances and pollution in the two French metropolitan areas Lille and Lyon. At French census block, it aims to explore the association between proximity to industrial facilities and the socio-economic characteristics.

3. Materials and methods

3.1. Study design

An ecological study was carried out in Lille and Lyon Metropolitan Areas, with a study period ranged from 2000 to 2009. The analysis was conducted at French Census Block (named IRIS in French for Ilots Regroupés pour l'Information Statistique that is Housing Blocks Regrouped for Statistical Information). Each IRIS represents on average a neighbourhood of residence of 2000 inhabitants and this geographical unit is the smallest geographic area in France for which demographic and socioeconomic data from the national census is available.

In our study, the IRIS geographical unit used is comparable to the US census block group and the US census tract which are, according to Krieger et al. (2003), the most pertinent geographic units for measuring socio-economic inequality and highlighting social inequalities in health.

This study is part of a research project previously approved by the French data protection authority.

3.2. Study area

The study was conducted in 2 metropolitan areas Lille and Lyon in France.

Lille is the fourth-largest metropolitan area in France, of which Lille city is the principal city of the metropolitan area, the capital of the Nord-Pas-de-Calais Region and the Prefecture of the North Department. In 2007, the National Institute of Statistics and Economic Studies (INSEE) reported a population of 1 106 885 inhabitants in Lille metropolitan area. It is composed of 506 French Census Blocks (IRIS) during the 1999 French National Census.

Lyon, the third-largest metropolitan area in France, has Lyon City as the capital of the Rhône-Alpes region. In 2007, the National Institute of Statistics and Economic Studies (INSEE) reported a population of 1 757 180 inhabitants in Lyon metropolitan area. And it is divided into 510 IRIS during the 1999 French National Census.

3.3. Data collection

3.3.1. Socioeconomic characteristics data

Socioeconomic and demographic data at census block level were extracted from the 1999 National French Census conducted by the INSEE, France.

21 variables were withdrawn, which represent diverse aspects of socioeconomic deprivation as income, education level, employment, housing characteristics, family structure, and immigration status.

3.3.2. Industrial facilities data

A. General information

The industrial facilities data used in our study were extracted from the French Register of Pollutant Emissions (Registre Français des Emissions Polluantes - IREP) from 2003 to 2009. The collection of this data meets the requirements of the international protocol on Pollutant Release and Transfer Register (PRTR) in Kiev 2003, and the requirements of the European regulation concerning the establishment of a European Pollutant Release and Transfer Register (E-PRTR n°166/2006, 18 January 2006).

This register consisted of data reported annually through an e-filing site (Déclaration annuelle des emissions pollutantes), of which the requirements for reporting by the operators of industrial facilities, municipal sewage treatment plants of more than 100,000 population equivalents and livestock is fixed with the decree of 31 Jan 2008. The register concerns 92 pollutants released into water, 81 for air emissions, 65 for soil emissions and 400 categories of hazardous waste. Total pollutants are classified into 10 groups as general parameters of water pollution, general air pollutants, metals and their compounds, organochlorines, pesticides, other organic compounds, carbon dioxide (CO2), other greenhouse gas emissions, toxic or carcinogenic and other compounds.

The study focused on the proximity to industrial facilities and data concerning different emitted pollutants in the metropolitan areas was not taken into account in the analysis.

B. Proximity data at census block level

Industrial proximity index was produced by using two distance-based approaches: spatial coincidence and buffer.

<u>Spatial coincidence</u>

This method is considered as a proxy of exposure to environmental hazards particularly to industrial facilities assuming that exposure depends directly on distance to the pollutant source. It involves selecting a predefined geographic unit (such as county, municipality or census block in this case), identifying the presence (host unit) or absence (non-host unit) of the hazard within each unit, and then comparing, between the host and non host units, the socio-demographic characteristics of the population aggregated over the unit of analysis (See Annex). This approach is also known as the “unit-hazard coincidence” method (Mohai P, 2006). Alternatively, the spatial coincidence method consists in counting the number of hazard sources (such as industrial facilities) within each geographic unit.

Buffer

This method consists in constructing at each potential source of pollution a buffer to delineate the zone and populations at risk. There are three types of buffers: point, area and line buffer (Sheppard E, 1999). In this study, the point buffer is used to identify population living within a given radius (500 metres and 1 kilometre) of a point-source emission (See Annex).

To investigate environmental inequities at proximity of an industrial facility, a circle is drawn with a predefined radius around the industry after its localisation on a map. Populations living in the geographic area within IRIS overlapped (or “touched”) by the circular buffer are considered as exposed. Their characteristics are then compared with those of unexposed populations residing in geographic units not intersected by the buffer.

3.3.3. Statistical methods

At French census block level, the statistical analysis was performed using STATA 11 statistical software.

Firstly, Shapiro-Wilk test was used to test the normality of distribution of the data. Secondly, the differences between the distributions of the socioeconomic and demographic variables in the two cities were then described and tested using Wilkoxon rank-sum test; this non-parametric statistical test was applied because most variables showed a non-normal distribution. Differences were considered significant at $p < 0.05$.

Thirdly, univariate logistic regression was employed to test the association between each industrial proximity index and each of the 21 socioeconomic variables. For each proximity index, p-value was classified after achieving 21 results, and those with p-value > 0.2 were excluded. Fourthly, all of SES variables with p-value ≤ 0.2 were included in multivariate regression analysis. The SES indicator with smallest p-value was tested first in a model, and then included one by one SES indicator with larger p-value. Suppressed were the SES variables which made one of the variables in the model become non-statistically significant. Level of statistical significance was set at $p < 0.05$. Odds ratios were used as the measure of association strength between the proximity to industrial facilities and socioeconomic conditions.

Last but not least, stepwise forward regression was used to find the best-fit model. However, a significant lack of data can show importantly different results.

4. Results

4.1. Data description

4.1.1. Socioeconomic characteristics

A. Socioeconomic characteristics in Lille

Using Shapiro-Wilk test to test the normality of the distributions, all socioeconomic variables are non-normally distributed. Descriptive analysis showed the variability of the socioeconomic characteristics of the study population in the two areas Lille and Lyon. In Lille, substantial gaps were observed in several variables related to unemployment, manual work, housing, poverty (assessed by the median of income) and education attainment aspects (See Table 1, Annex). For example, percentage of persons aged 15 years and older without further elementary education approximately doubles in the third quartile compared to that in the first quartile of the population. As for the poverty aspect, the median income also varies significantly among different IRIS; the average income in the third quartile is 1,5 time greater than that in the first quartile (IQR: 16992.5-25774).

Table 1: Socioeconomic characteristics in Lille metropolitan area

Variables	**1st quartile**	**Median**	**3rd quartile**	**Min**	**Max**
People aged 25 years or younger (%)	32.76	36.05	40.59	21.45	77.78
Foreigners (%)	1.88	3.87	7.93	0.00	33.23
Unemployed people (%)	9.16	13.61	19.86	0.00	57.24
Self-employed people (%)	3.96	5.57	8.26	0.00	33.33
People with insecure job (%)	9.08	11.62	14.71	0.00	40.31
People with stable job (%)	59.65	67.40	72.31	26.25	100.00
Managers in labour force (%)	4.78	9.73	17.81	0.00	61.54
Blue-collar workers (%)	13.72	21.03	28.26	0.00	80.00
Single-parent families (%)	10.11	13.87	19.84	0.00	39.31
People aged 15 years or older who did not go beyond an elementary education (%)	10.64	16.98	23.56	0.00	46.19
People aged 15 years or older with a general or vocational maturity certificates (%)	8.33	9.83	11.89	0.00	40.00
People aged 15 years or older with at least a tertiary education (%)	5.68	8.03	10.33	0.00	40.00
People aged 15 years or older with a higher educational degree (%)	3.58	6.37	11.61	0.00	42.86
Primary residence with a garage or other parking space (%)	26.5	43.79	64.37	0.00	100.00
Non-owner occupied primary residence (%)	25.94	44.2	69.73	4.59	100.00
Households without a car (%)	14.53	24.10	33.62	0.00	86.96
Households with 2 or more cars (%)	14.66	22.53	37.23	0.00	80.61
Subsidized housing among all primary residences (%)	0.58	10.30	29.00	0.00	133.48
*Median income per consumption unit (euro)**	16992.5	21165	25774	8162	55608
Primary residence with more than 1 person per room (%)	4.09	5.97	9.19	0.00	48.53
*Mean number of people per room**	0.61	0.65	0.72	0.51	1.13

All variables are expressed in percentage (0-100%) except (*).

B. Socioeconomic characteristics in Lyon

Similarly, all socioeconomic variables in Lyon are non-normally distributed. Descriptive analysis also highlighted that there were big differences among IRIS such as in the percentage of managers in labour force (IQR: 6-22%), people with no further elementary education (9-21%), primary residence with a garage or other parking space (34-72%), or non-owner occupied primary residence (39-72%). Important gap was also noticed in the percentage of primary residence with subsidized housing, with 16 times higher in the third quartile compared to that in the first quartile.

Table 2: Socioeconomic characteristics in Lyon metropolitan area

Variables	1st quartile	Median	3rd quartile	Min	Max
People aged 25 years or younger (%)	28.83	32.12	35.75	0.00	72.86
Foreigners (%)	4.30	6.55	11.11	0.00	64.76
Unemployed people (%)	8.30	10.98	14.96	0.00	51.09
Self-employed people (%)	5.26	7.31	10.46	0.00	47.83
People with insecure job (%)	8.97	11.19	13.64	0.00	100.00
People with stable job (%)	63.13	68.22	72.50	0.00	100.00
Managers in labour force (%)	6.36	13.39	22.95	0.00	63.64
Blue-collar workers (%)	10.25	16.74	25.12	0.00	85.71
Single-parent families (%)	10.79	13.88	16.92	0.00	100.00
People aged 15 years or older who did not go beyond an elementary education (%)	9.86	14.10	21.68	0.00	72.50
People aged 15 years or older with a general or vocational maturity certificates (%)	9.42	11.23	12.69	0.00	24.56
People aged 15 years or older with at least a tertiary education (%)	6.43	9.69	12.23	0.00	17.03
People aged 15 years or older with a higher educational degree (%)	4.36	9.67	17.63	0.00	66.67
Primary residence with a garage or other parking space (%)	34.03	52.88	72.73	0.00	100.00
Non-owner occupied primary residence (%)	39.93	58.27	72.64	5.99	100.00
Households without a car (%)	11.71	23.06	32.30	0.00	100.00
Households with 2 or more cars (%)	14.53	21.88	34.72	0.00	71.43
Subsidized housing among all primary residences (%)	2.16	13.37	34.66	0.00	138.23
*Median income per consumption unit (euro)**	18982.5	22655.5	27056	11339	47909
Primary residence with more than 1 person per room (%)	5.39	7.08	9.71	0.00	50.00
*Mean number of people per room**	0.66	0.7	0.76	0.28	1.10

All variables are expressed in percentage (0-100%) except (*).

C. Differences of SES data between Lille and Lyon

Wilkcoxon rank-sum test was used to test differences between the underlying distributions of socioeconomic variables in the two metropolitan areas. Statistical analysis revealed that there were statistically significant differences between most of SES variables in the two areas, with the level of statistical significance set at p<0.05. No difference, however, was found in persons with insecure job (p=0.08), single-parent families (p=0.15), households without a car (p=0.07) and households with two or more cars (p=0.70).

Table 3: Differences of SES data between Lille and Lyon

Variables	P-value (statistical significance : p<0.05)	Conclusion
People aged 25 years or younger (%)	0.00	
Foreigners (%)	0.00	
Unemployed people (%)	0.00	
Self-employed people (%)	0.00	
People with insecure job (%)	**0.08**	No difference
People with stable job (%)	0.03	
Managers (%)	0.00	
Blue-collar workers (%)	0.00	
Single-parent families (%)	**0.15**	No difference
People aged 15 years or older who did not go beyond an elementary education (%)	0.03	
People aged 15 years or older with a general or vocational maturity certificates (%)	0.00	
People aged 15 years or older with at least a tertiary education (%)	0.00	
People aged 15 years or older with a higher educational degree (%)	0.00	
Primary residence with a garage or other parking space (%)	0.00	
Non-owner occupied primary residence (%)	0.00	
Households without a car (%)	**0.07**	No difference
Households with 2 or more cars (%)	**0.70**	No difference
Subsidized housing among all primary residences (%)	0.01	
*Median income per consumption unit (euro)**	0.00	
Primary residence with more than 1 person per room (%)	0.00	
*Mean number of people per room**	0.00	

All variables are expressed in percentage (0-100%) except (*).

4.1.2. Industrial proximity data

A. Industrial proximity data in Lille

There were six industrial proximity indexes which were constructed differently. In the simplest way, presence and numbers of industrial facilities were identified in each IRIS. These two variables showed quite similar results, though the second one demonstrated a clearer composition of the presence of industries in each IRIS compared to the first one (Absence =0, Presence= 1 or more facilities). The similarity was given because of the small number of industries in comparison with the number of IRIS.

More sophisticatedly, the other four variables were established by using buffers with different radii of 500m and 1km. When populations living in the geographic area of IRIS overlapped by the circular buffer are considered as exposed, the presence of exposed IRIS increased with the increase of radius. With a 1km radius, the presence of buffers (number of exposed IRIS) nearly doubled in comparison with that within a 500m radius, and is 8 times higher than the presence of facility in the IRIS (radius is equal to 0) (See Figure 1, Annex).

Also as expected, the increase of numbers of buffers in an IRIS was matched by the corresponding increase of the radius. For example, the percentage of IRIS with 2 buffers tripled with 1 km radius compared to that with a 500m radius (17.98% and 5.53% respectively). Especially, the number of IRIS with 3 buffers with 1km radius is significantly 12 times higher than that with 500m radius (7.11% and 0.60% respectively) (See Figure 2, Annex).

B. Industrial proximity data in Lyon

Similarly, six industrial variables were constructed for Lyon metropolitan area. The index showing numbers of facilities in each IRIS gave us more detailed information on the composition of the presence of industries in each IRIS (Absence =0, Presence= 1 or more facilities). The two indexes of presence and numbers of facilities revealed similar results because the number of industries is not high enough in comparison with the number of IRIS.

Regarding the variables established by using buffers, the presence and numbers of buffers in an IRIS also increased corresponding to the increase of radii (500m and 1km). With a 1km radius, the presence of buffers (number of exposed IRIS) nearly doubled in comparison with that in a 500 radius, and is 7.4 times higher than the presence of facility in the IRIS - when radius is equal to 0 (29.02%, 15.69% and 3.92% respectively) (See Figure 3, Annex).

As can be seen from Figure 4 (See Annex), the number of buffers in an IRIS increased likewise when increasing the radius of buffers. For instance, the percentage of IRIS with 2 buffers within 1 km radius is 2.6 times higher than that within a 500m radius; 8.43% and 3.14% respectively.

C. Differences in industrial proximity data between Lille and Lyon

Chi-square test is employed to test differences between variables of the two metropolitan areas. The results revealed that there were differences in most of variables when comparing the two metropolitan areas (See Table 4). Nevertheless, no difference between the presence of industry in each IRIS in the two areas was found (at p=0.062). Besides, when number of industries in the IRIS was divided into 3 categories (0, 1, 2), the rule of Chi-square test was not respected due to small expected number. As for the two variables that counting numbers of buffers within 500m and 1km radius, some groups of the value were reorganized (e.g. 2 or more, 3 or more). The way of assembling and creating smaller number of categories allowed the results to have sufficient expected number - an important requirement when applying Chi-square test (Expected number =5).

Furthermore, attention should also be paid to the constitution of proximity index. From the first to the last industrial index in the following table, the results of association changed when variables were more sophisticatedly constructed. The four variables that were constructed by using buffers gave significantly low p-value, showing the statistically significant differences in proximity data between the two areas.

Table 4: Differences in industrial proximity data between Lille and Lyon (Chi-square test)

Variables	**Categories**	**Smallest expected number (P>5)**	**p-value (statistical significance : p<0.05)**	**Conclusion**
Presence of 1 or more industry in the IRIS	Presence, absence	26	**0.062**	No difference
Presence of 1 or more buffer with a 500m radius	Presence, absence	114	0.000	
Presence of 1 or more buffer with a 1km radius	Presence, absence	211	0.000	
Number of industries in the IRIS	0, 1, 2	**04**	0.013	
Number of buffer with a 500m radius which touch the Iris	0, 1, 2 or more	21	0.000	
Number of buffer with a 1km radius which touch the Iris	0, 1, 2, 3 or more	16	0.000	

4.2. Statistical analysis

Regarding industrial variables tested in the study, the logistic regression analysis was conducted only with two out of six industrial proximity indexes in the database: the presence of facilities in each IRIS and the presence of buffers with a 1km radius.

4.2.1 Statistical analysis in Lille

A. Presence of industry in Lille

Logistic regression analysis revealed no statistically significant association between the presence of industry in IRIS and a single SES indicator (p-value > 0.05) in Lille.

B. Presence of buffers with a 1km radius

Logistic regression

Univariate logistic regression analysis demonstrated 4 SES indicators having statistically significant association with industry index of 1km-buffer presence (See Table 5). However, the associations were rather weak between the industrial proximity indicator and either unemployed people or foreigners. There is 1.03 times more risk for the IRIS located within buffers with a 1km radius (OR =1.03, CI [1.00, 1.05]) for 1% increase in the percentage of the unemployed. The same implication could be interpreted for the foreigner variable (OR=1.03, CI [1.00, 1.07]).

As expected, people with stable job and median income showed relationships with the proximity index but this correlation was completely weak for the latter indicator. The odds that an IRIS located in one or more buffers with 1km radius decrease 2% (OR=0.98, CI [0.96, 0.99]) compared to that of an IRIS with 1% increased of people with stable job. Also, the odds of the presence of buffers decrease 0.29% (OR=0.9971, CI [0.9944, 0.9998]) compared to that in an IRIS with 100 euro increased in median income.

Table 5: Univariate regression analysis in Lille for industry index of 1km-buffer presence

SES variables	OR	95% CI	P-value
Unemployed people (%)	1.0285	[1.0087;1.0486]	0.005
People with stable job (%)	.9780	[.9620;.9944]	0.009
*Median income per consumption unit (euro)**	.9971	[.9944;.9998]	0.034
Foreigners (%)	1.0348	[1.0000;1.0707]	0.050

OR: Increased risk for additional 0.01unit except (*).
OR for median income (*) is the increased risk for additional 100 unit.

Stepwise forward

Stepwise forward analysis revealed that there was a significant association ($p<0.05$) between being in industrial buffers within 1 km radius and status of being unemployed and foreigners. For 1% increase in the percentage of the unemployed, there is 1.09 times more risk for the IRIS located within buffers with a 1km radius (OR=1.09; CI 95%=[1.04,1.14]). Foreigners, however, had an inverse correlation with proximity index in the model. The odds that an IRIS located in one or more buffers with 1km radius decrease 9% (OR=0.91; CI 95%=[0.84-0.99]) compared to that of an IRIS with 1% increase of foreigners (See Table 10, Annex).

The association between the proximity index and the two SES variables above were verified with multivariate logistic regression. However, no statistical significance was found for foreigners in the model (p-value =0.422 > 0.05). The difference between the two results could be explained by the lack of data, especially for median income variable. While the multivariate logistic regression took into account data of 504 IRIS, only 387 IRIS were included in stepwise regression (See Table 10, Annex).

4.2.2 Statistical analysis in Lyon

A. Presence of industry in Lyon

Univariate logistic regression

Different results were observed between the 2 metropolitan areas for the industrial index of the presence of facilities in an IRIS. While no SES predictors was found for Lille, univariate regression revealed significant associations ($p<0.05$) between the presence of industry in an IRIS and each of the 12 SES variables in Lyon (See Table 6). In term of education attainment, the results showed that the further the population in the IRIS achieved in education, the less likelihood for the IRIS located in an area with facilities. For example, the odds that an IRIS has one or more industrial facilities within itself decrease 20% (OR=0.80, CI 95%=[0.71, 0.91]) compared to that of an IRIS with 1% increase of people aged 15 or older with at least a tertiary education.

Also as expected, people who did not go beyond an elementary education were one of the predictors for the increased risk of the presence of industry in an IRIS. The odds of having facilities in an IRIS are 1.04 times (OR=1.04, CI 95%=[1.00, 1.08]) as large than that in an IRIS with 1% decrease in percentage of people aged 15 years or older who did not go beyond an elementary education.

Not in line with previous empirical researches, results concerning housing composition domain and property owning demonstrated positive associations between the proximity indicator and either one of the two variables: the primary residence with garage and households with 2 or more cars (OR=1.02, CI 95%= [1.00, 1.05] and OR=1.03, CI 95% [1.00, 1.06], respectively). However, these associations were rather weak.

Table 6: Univariate regression analysis in Lyon for industry presence index

Variables	OR	95% CI	P-value
People aged 25 years or younger (%)	1.1029	[1.0494,1.1592]	0.000
People aged 15 years or older with at least a tertiary education (%)	.8045	[.7100,.9117]	0.001
People aged 15 years or older with a higher educational degree (%)	.8298	[.7431,.9267]	0.001
Blue-collar workers (%)	1.0456	[1.0135,1.0786]	0.005
Households without a car (%)	.9419	[.9025,.9830]	0.006
Primary residence with more than 1 person per room (%)	1.0785	[1.0223,1.1378]	0.006
Managers in labour force (%)	.9254	[.8734,.9804]	0.008
Foreigners (%)	1.0490	[1.0075,1.0922]	0.020
People aged 15 years or older who did not go beyond an elementary education (%)	1.0420	[1.0060,1.0794]	0.022
Primary residence with a garage or other parking space (%)	1.0245	[1.0034,1.0460]	0.023
Households with 2 or more cars (%)	1.0291	[1.0036,1.0553]	0.025
Single-parent families (%)	.8982	[.8264,.9763]	0.012

(OR: Increased risk for additional 0.01unit)

Multivariate logistic regression

Multivariate logistic regression highlighted a model with 3 independent SES variables and the presence of industry in the IRIS (See table 7). The directions of relationships between each independent SES indicator and the proximity index were still unchanged in comparison with the results in univariate regression. The education attainment again expressed significant association with proximity index, the odds that an IRIS has one or more facilities decrease 18% (OR=0.82, CI 95%= [0.72, 0.94]) compared to that of an IRIS with 1% increase in people aged 15 or older with at least a tertiary education.

Table 7: Regression analysis in Lyon for industry presence index

	Univariate regression (Obs: 503)			**Multivariate regression** (Obs: 503)		
	OR	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**
People aged 25 years or younger (%)	1.1029	1.0494 1.1592	0.000	1.0765	1.0187 1.1376	0.009
People aged 15 years or older with at least a tertiary education (%)	.8045	.7100 .9117	0.001	.8211	.7176 .9396	0.004
Households without a car (%)	.9419	.9025 .9830	0.006	.9335	.8949 .9737	0.001

(OR: Increased risk for additional 0.01unit)

Stepwise forward

To verify the best-fit model, stepwise forward was processed for all SES indicators which had the univariate regression results of p-value ≤ 2. The new model, however, included only 2 SES variables of people with a general or vocational maturity certificates and households without a car. Regarding the people with a general or vocational maturity certificates, this social indicator became statistically significant in the new model (p-value close to 0) while it was non-significant in univariate regression (p-value=0.029) (See Table 11, Annex).

To explain the different results between multivariate regression processes and stepwise, the reason of missing data could be given. Whereas the multivariate logistic regression took into account data of 503 IRIS, only 415 IRIS were included in stepwise regression.

B. Presence of buffers with a 1km radius

Univariate logistic regression

There were 15 out of 21 SES variables giving statistically significant results when testing with the presence of buffers with a 1km radius. The significantly low p-value suggested important associations between these SES and the proximity index, e.g. p-value <0.001 for all of indicators in education attainment domain and most of index in employment or poverty (assessed by the median of income).

Table 8: Univariate regression analysis in Lyon for industry index of 1km-buffer presence

Variables	P-value
People aged 25 years or younger (%)	0.000
Foreigners (%)	0.000
Unemployed people (%)	0.000
Self-employed people (%)	0.000
Managers in labour force (%)	0.000
Blue-collar workers (%)	0.000
People aged 15 years or older without further elementary education (%)	0.000
People aged 15 years or older with a general or vocational maturity certificates (%)	0.000
People aged 15 years or older with at least a tertiary education (%)	0.000
People aged 15 years or older with a higher educational degree (%)	0.000
Subsidized housing among all primary residences (%)	0.000
*Median income per consumption unit (euro)**	0.000
Primary residence with more than 1 person per room (%)	0.000
*Mean number of people per room**	0.000
People with stable job (%)	0.003

Multivariate logistic regression

Multivariate logistic regression suggested several models. For example, one of these included 4 SES indicators of people aged 25 years or younger, foreigners, self-employed people and primary residence with parking space (See Table 9).

Table 9: Logistic regression analysis in Lyon for industry index of 1km-buffer presence

	Univariate regression (obs: 503)			**Multivariate regression** (obs: 503)		
	OR	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**
People aged 25 years or younger (%)	1.0842	1.0498 1.1196	0.000	1.0495	1.0136 1.0868	0.007
Foreigners (%)	1.0806	1.0511 1.1110	0.000	1.0635	1.0268 1.1016	0.001
Self-employed people (%)	.8926	.8477 .9399	0.000	.9353	.8845 .9891	0.019
Primary residence with a garage or other parking space (%)	1.0058	.9980 1.0137	0.146	1.0168	1.0071 1.0266	0.001

(OR: Increased risk for additional 0.01unit)

Stepwise forward

The stepwise forward regression, however, reported the best-fit model with only 2 SES indicators. The variable of primary residence with a garage or other parking space became significant in the new model (p-value close to 0), significantly associated with the industry index (OR=1.02, CI 95%= [1.01, 1.04]) in comparison with the non-significant result in univariate regression (p-value=0.146). In contrast, people aged 15 years or older with at least a tertiary education revealed an inverse relationship with the dependent industrial variable. The odds that an IRIS had one or more buffers with 1km radius decreased 33% (OR=0.67, CI 95%= [0.62, 0.73]) in comparison with that of an IRIS with 1% increase in people aged 15 or older with at least a tertiary education (See Table 12, Annex)

And again, the difference between results of multivariate regression and stepwise can be explained by the lack of data. In STATA, whereas data in 503 IRIS was included in multivariate logistic regression, only 415 IRIS were taken into account in stepwise forward process.

5. Interpretation and discussion

5.1. Interpretation

5.1.1. Existence of environmental inequality

The study revealed that different approaches in the context of ecological studies such as the use of a specific unit of analysis and a different statistical method to conduct the analysis could highlight different results.

The significance of the associations between industrial proximity and social gradients changed in different local settings with different construction of proximity index. However, the results generally highlighted patterns of concentration of sites in deprived communities. Especially in Lyon, significant associations were observed between the presence of buffers with 1km radius and all of indicators in educational level, most of index in employment status and poverty (p-value close to 0). In other words, these social conditions above were demonstrated as significant determinants for the industrial exposure. The existence of environmental inequality was also confirmed using the multivariate analysis.

5.1.2. Proximity index composition and its relationship with SES

In both metropolitan areas, the industrial index which was more sophisticated (buffer) gave more frequently significant results compared to that from the simply constructed indicator.

In Lille, while no association was observed between the industry presence index and SES, statistically significant associations were found between the index of 1km-buffer presence and 4 SES variables. These correlations were especially strong with unemployed and people with stable jobs (p-value=0.005 and 0.009 respectively), though it was not the case for median income and foreigners (p-value= 0.034 and 0.050, respectively).

In Lyon, the index of industry presence showed significant associations with 12 SES indicators, remarkably in demographics (people less than 25 years old, p-value close to 0, OR=1.10, 95% CI=[1.05,1.16]) and education achievement (people with tertiary education p-value=0.001, OR=0.80, 95%CI=[0.71,0.91]; and people with a higher educational degree, p-value=0.001, OR=0.83, 95%CI=[0.74,0.93]). As for the 1km-buffer index, important associations were observed with 15 social indicators. P-values were significantly low (close to 0) for all variables in education attainment and some other indices in employment and poverty, which largely contributed to the suggestion that the deprived population are prone to hazardous exposure.

Also, more multivariate models were additionally given for the index of 1km-buffer presence compared to the simple proximity indicator of industry presence. Despite differences in the types of SES variables and the strength of correlations between the two metropolitan areas, general conclusion from multivariate analysis was still suggested that socially disadvantaged groups are more likely to suffer from higher industrial exposure.

5.1.3. Comparison between the two cities

Significant differences between data of the two metropolitan areas were found in most of indicators, for both SES and industrial proximity (most of p-values were significantly low, close to 0). Only 4 SES variables of people with insecure jobs, single-parent family, household without a car and household with two or more cars have no significant difference between the two areas p>0.05). As for proximity, no difference between the two areas was observed for the simplest constructed proximity index- the presence of industry in an IRIS.

Also, there were differences between Lille and Lyon in the associations between proximity and SES, in both aspects of models and strength of associations. For both proximity variables tested in this study, data in Lyon showed more frequently significant associations with social conditions compared to that in Lille.

For the purpose of trying to understand the reasons behind these differences, information on industries in the two cites Lille and Lyon was studied. While Lille is well-known for its textile sector and other heavy industries including metallurgy, chemicals, engineering, and distilling, Lyon is famous for its high-tech industries such as chemicals, biotechnology, pharmaceutical and software development. Besides, the development of medical research and technology, non-profit institutions and universities in Lyon makes crucial difference in industries between the 2 cities. The dissimilarity in dominant industrial patterns in the two areas could, therefore, partly explain the difference in social conditions.

5.2. Comparison with other studies

Some certain findings in this research are in line with previous studies. Briggs and colleagues, for example, determined source proximity by buffering around the source, intersecting the resulting areas with postcode locations, and selecting the postcodes falling within the buffer zone. And in this study, they found significant correlations (increasing deprivation associated with higher potential exposure) for most of proximity index though the associations for the proximity to point sources and industrial land are rather weak (Pearson r=0.09-0.16).

Besides, in Fairburn's report, there was a clear relationship between proximity and deprivation for all of buffer distances used to estimate numbers of people living near to industrial sites. For the 500m buffer there are 6 times as many people in the most deprived decile living near to a facility when compared to the least deprived. For the 1km buffer nearly 20% of the population in the most deprived decile live near to a site, compared to only 6% in the least deprived i.e. the most deprived are three times more likely to be living near to an industrial site than the least deprived. Statistic results revealed that the relationship with deprivation for 500m and 1km buffers is strongly linear, whilst a curvilinear relationship emerges for 2km and 4km. This proportional difference becomes lower as the buffer size increases (approximately 4 times for 1km, 2 times for 2 km) and the CI values accordingly decrease from 0.26 to 0.08. The relationship with deprivation for 500m and 1km buffers is strongly linear,

However, literature yields mixed findings regarding exposure disparities. The association between air pollution and SES has been often translated into poorer populations or areas being at greater exposure in most of studies, though inverse correlation was also found when richer population have been reported at greater exposure in other instances.

5.3. Advantages and limitations

5.3.1. Advantages

This is one of the few studies in environmental justice conducted in France. To avoid the disadvantages resulting of crude unit levels, a fine geographical scale (IRIS) could partly help to respect the heterogeneous characteristic of the population.

This study has its own strengths inherited in ecological design. Firstly, the ecological study can investigate differences between populations when there are greater differences between populations (IRIS) than within them. The analysis could identify the IRIS with accumulative social and environmental inequalities. As an advantage for public health, the study aimed at group level. Programmes and actions to mitigate the social and environmental disparities can be implemented on census block scale, which can benefit a group of population rather than the individuals. Thirdly, it takes advantages of the convenience and availability of group-level data when we only have data available at group level. In our study, socioeconomic indices of each IRIS have already been published by INSEE and industrial facilities data has been updated on IREP website. .Also, the uses of routine data already available can be less costly and time-consuming in comparison to studies at individual-level, especially when finance and potential for lost to follow-up are always the main short-comings of individual studies.

The socioeconomic indicators ranged in different domains from demographics, employment, types of work, family status, housing, poverty to education attainment, which presented multi-aspect of social status in each IRIS. Moreover, statistical analysis run for each of SES variables with industrial proximity helped to find the most determining SES for industrial exposure. In this study, demographics, education level, employment status and poverty were found to have the most significant correlations with proximity. This research may, therefore, avoid the limitation of other studies which used only one deprivation index constructed from different social domains (eg (Briggs et al, 2008) in England). In our study, different models with different compositions of variables were applied for the two areas because the importance of each SES aspect in association with proximity indicators varied in different metropolitan areas.

5.3.2. Limitations

The interpretation of our findings considers some weakness.

Study design

This is an ecological study; therefore, it owns its inherited limitation. Firstly, ecological studies may be unable to measure information of other important risk factors that may also be associated with the outcome under study (because the data are collected already and were collected for other purposes). Secondly, data on exposure and outcome may be collected in different ways with different definitions which may bias the results. Differences in the way data are collected over time may also differ systematically. Thirdly, geographical comparisons may suffer from migration of populations between groups over the period of the study, which may dilute the difference between groups. Finally, the index is based on aggregated data corresponding to a geographic approach and the associations observed at the census block level cannot be extrapolated to the individual level. Indeed, in this study, the results found at the census block level cannot be applied to the individual level (ecological fallacy).

Data sources

The study relied essentially on the quality of the socioeconomic and demographic data recorded by the INSEE for the 1999 French national. This source of data registered thorough information. So far, socioeconomic data were extracted from published data in 1999 by INSEE. We plan to re-explore the existence of social -environmental inequalities using the more recent data.

Industrial facilities data were extracted from the French Pollutant Emission Registry (IREP) following the E-PRTR criteria. A possible limitation of these data could be that only the facilities with excess of threshold limit are reported by the IREP. Other industrial sites which released several pollutants below the threshold limit, consequently, may not be included in the report. In addition, the data reported by the IREP did not provide information about other important polluted industries, such as energy industries, etc. The weakness of the industrial data could possibly bias the results.

Construction of data

There are several limitations which come from the method used to characterize the proximity of the population to facilities, which requires a careful consideration of the association between proximity index and SES at French census block level. The method applied in this study (spatial coincidence method and buffers) use proximity as a proxy for pollutant exposure related to the industrial facilities located in French census block (IRIS).

Regarding the indicator of the presence or absence of facilities within each IRIS, this definition implied that the industrial exposure was equally distributed in the entire geographical unit (IRIS). All individuals residing there, then, were assumed of being equally exposed, regardless the distance to pollution sources and time-activities pattern which might vary among individuals and according to SES.

Another limitation can be given that the proximity data did not consider the exact location of the industrial facilities. The hazardous sites could be located in the limit border of an IRIS and its pollution emissions could affect the neighbour IRIS. Our results are valid only at the geographical level initially defined (IRIS) due to the IRIS border limits which were fixed by administrative process. It is important, however, to note that the dispersion of pollutants released by industrial facilities might not respect the administrative boundaries. Or in other words, there might be important difference between the affected geographical zones and the administrative boundaries limited for the census blocks.

The other proximity index using buffers can avoid some other limitations above, but still have its own drawbacks. The consideration of exposure homogeneity is still a problem, not only within each IRIS but also within the entire buffer. When proximity constructed in this way, the pollution level is consider equally distributed within the buffers either in the centre or on the line.

Moreover, both indicators of industry presence and buffer presence do not provide other characteristics of industrial sites (e.g. size, complexity, type of industrial activities) and the pollutants (e.g. different chemical and physical properties of the pollutants or their specific pathways of exposure), which could affect in a different manner the SES characteristics of the population. In addition, meteorological conditions which modify the properties and dispersion of pollutants were also not taken into account.

Socioeconomic and demographic data extracted from the 1999 French National census were presented as percentages at IRIS level, and we assumed that these percentages characterize the whole population of the IRIS, assuming a certain degree of homogeneity. However, studies conducted at individual level could be an alternative approach to study environmental inequity and the spatial distribution.

Among the set of industrial index, there are industrial indicators reporting the number of facilities or number of buffers in IRIS. However, these variables were not analysed due to limitation of time.

Other limitations

Missing data, especially in the indicator of median income, was proved to have important effects on the results.

This study does not consider indoor air quality or air pollution from traffic road - another important pollution source. Availability of green areas was also not included in this research, which was studied to have significant associations with income of the surrounding residents (Kruize et al, 2007). Moreover, the data used in the study did not allow establishing differences between urban and rural areas that could have an impact in the results.

In addition, information on the characteristics of industry (types of industrial sector, number of facilities, etc) in each metropolitan should be given, which could possibly explain the differences between the two areas.

6. Recommendations and proposals for actions

Further researches need to be conducted for more recent data, especially data on socioeconomic characteristics in 2006 which has been available in INSEE website. Also, the emission of the industries should be taken into account to construct the industrial index. Besides, there is a need to construct a common SES index for Lille and Lyon, which is not only to better compare the results of the associations with the proximity index but also to avoid the collinearity in the multivariate model. Studies in future should also take into account the spatial autocorrelation between the different IRISs.

References

Anderton DL. , A. A., Oakes JM., Fraser MR. (1994) Environmental Equity: The Demographics of Dumping *Demography,* 31**,** 229-248.

Blakely, T., Kawachi, I., Atkinson, J., & Fawcett, J. (2004) Income and mortality: the shape of the association and confounding New Zealand census-mortality study, 1981-1999. *International Journal of Epidemiology,* 33**,** 874-883.

Blakely, T. A., & Woodward, A. J. (2000) Ecological effects in multi-level studies. *Journal of Epidemiology and Community Health,* 54 367-374.

Brender JD, S. L., Langlois PH, Steck M, Zhan FB, Moody K. (2008) Are maternal occupation and residential proximity to industrial sources of pollution related? *J Occup Environ Med,* 50**,** 834-839.

Briggs, D., Abellan, J. J. & Fecht, D. (2008) Environmental inequity in England: Small area associations between socio-economic status and environmental pollution. *Social Science & Medicine,* 67**,** 1612-1629.

Brown, P. (1995) Race, Class, and Environmental Health: A Review and Systematization of the Literature. *Environmental Research,* 69**,** 15-30.

Carstairs, V. (1981) Multiple deprivation and health state. *Journal of Public Health,* 3**,** 4-13.

Carstairs, V. I. P. E., J. Wakefield, N. Best, & D. Briggs (Eds.), (2000) Spatial epidemiology. Methods and applications *Socio-economic factors at areal level and their relationship with health.*, Oxford: Oxford University Press.

Curtis, L., Rea, W., Smith-Willis, P., Fenyves, E., & Pan, Y. (2006) Adverse health effects of outdoor air pollutants. *Environ.Int.,* 32**,** 815-830.

Davey Smith, G., Dorling, D., Mitchell, R., & Shaw, M. (2002) Health inequalities in Britain: continuing increases up to the end of the 20th century. *Journal of Epidemiology and Community Health,* 56**,** 434-435.

Downey, B. H. L. (2008) Race, Income, and Environmental Inequality in the United States. *Sociological Perspectives,* 51**,** 759-781.

Environmental Protection Agency, U. (2010) Environmental Justice.

Fairburn, J., Walker, G., Smith, G. (2005) Investigating environmental justice in Scotland: links between measures of environmental quality and social deprivation Scotland.

Finkelstein, M., Jerrett, M., Sears M., (2005) Environmental inequality and circulatory disease mortality gradients. *J Epidemiol Community Health,* 59**,** 481-487.

Forastiere, F., Stafoggia, M., Tasco, C., Picciotto, S., Agabiti, N., Cesaroni, G. & Perucci, C. A. (2007) Socioeconomic status, particulate air pollution, and daily mortality: Differential exposure or differential susceptibility. *American Journal of Industrial Medicine,* 50**,** 208-216.

Franchini M, M. P. (2009) Particulate air pollution and cardiovascular risk: short-term and long-term effects. . *Semin Thromb Hemost,* 35**,** 665-70.

Généreux M, A. N., Goneau M, Daniel M. (2008) Neighbourhood socioeconomic status, maternal education and adverse birth outcomes among mothers living near highways. *J Epidemiol Community Health,* 62**,** 695-700.

Gorham ED, G. C., Garland FC. (1989) Acid haze air pollution and breast and colon cancer mortality in 20 Canadian cities. *Can J Public Health,* 80**,** 96-100.

Heo S, M. D., Chung ML, Lennie TA. (2010) Social status, health-related quality of life, and event-free survival in patients with heart failure. . *Eur J Cardiovasc Nurs,* In Press, Corrected Proof.

Institut national de la statistique et des études économiques (INSEE). Available at http://www.insee.fr/fr/default.asp (last accessed on 03 May 2011)

Iqbal S, C. J., Boehmer TK, Yip FY, Garbe P. (2010) Carbon monoxide-related hospitalizations in the U.S.: evaluation of a web-based query system for public health surveillance. . *Public Health Rep. ,* 125**,** 423-32.

Jerrett, M., Burnett, R. T., Brook, J., Kanaroglou, P., Giovis, C., Finkelstein, N. & Hutchison, B. (2004) Do socioeconomic characteristics modify the short term association between air pollution and mortality? Evidence from a zonal time series in Hamilton, Canada. *Journal of Epidemiology and Community Health,* 58**,** 31-40.

Jerrett, M., Burnett,R.T., Kanaroglou,P., Eyles, J., Finkelstein,N., Giovis,C., et al. (2001) A GIS-environmental justice analysis of particulate air pollution in Hamilton, Canada. . *Environment and Planning A,* 33**,** 955-973.

Kantrowitz, G. E. a. E. (2002) Socioeconomic status and health: The Potential Role of Environmental Risk Exposure. *Annual Review of Public Health,* 23**,** 303-331.

Kearney, G. & Kiros, G.-E. (2009) A spatial evaluation of socio demographics surrounding National Priorities List sites in Florida using a distance-based approach. *International Journal of Health Geographics,* 8**,** 33.

Kohlhuber, M., Mielck, A., Weiland, S. K. & Bolte, G. (2006) Social inequality in perceived environmental exposures in relation to housing conditions in Germany. *Environmental Research,* 101**,** 246-255.

Kuehn, C. M., Mueller, B. A., Checkoway, H., & Williams, M. (2007) Risk of malformations associated with residential proximity to hazardous waste sites in Washington State *Environ.Res,* 103**,** 405-412.

Künzli N, J. M., Garcia-Esteban R, Basagaña X, Beckermann B, Gilliland F, Medina M, Peters J, Hodis HN, Mack WJ. . (2010) Ambient air pollution and the progression of atherosclerosis in adults *PLoS One,* 5**,** e9096.

Kwon J, W. C., Turpin BJ, Zhang J, Korn LR, Morandi MT, Stock TH, Colome S. (2006) Source proximity and outdoor-residential VOC concentrations: results from the RIOPA study. *Environ Sci Technol,* 40**,** 4074-82.

Mackenbach JP, e. a. (2008) Socioeconomic Inequalities in Health in 22 European Countries.European Union Working Group on Socioeconomic Inequalities in Health in 22 European countries. . *N Engl J Med,* 358**,** 2468-81.

Martins MC, F. F., Véspoli TC, Martins LC, Pereira LA, Martins MA, Saldiva PH, Braga AL. (2004) Influence of socioeconomic conditions on air pollution adverse health effects in elderly people: an analysis of six regions in São Paulo, Brazil. . *J Epidemiol Community Health. ,* 58**,** 41-46.

McLeod H., Langford I. H., Jones A. P., Stedman J. R., Day R. J., Lorenzoni I. & J., B. I. (2000) The relationship between socio-economic indicators and air pollution in England and Wales: implications for environmental justice. *Regional Environmental Change,* 1**,** 78-85.

Mohai, P., Lantz, P. M., Morenoff, J., House, J. S. & Mero, R. P. (2009) Racial and Socioeconomic Disparities in Residential Proximity to Polluting Industrial Facilities: Evidence From the Americans' Changing Lives Study. *Am J Public Health,* 99**,** S649-656.

Mohai P, S. R. (2006) Reassessing racial and socioeconomic disparities in environmental justice research. *Demography.,* 43**,** 383-399.

Morello-Frosch, R., Pastor, M., Porras, C., & Sadd, J. (2002a) Environmental justice and regional inequality in Southern California: implications for future research. *Environmental Health Perspectives,* 110.

Morello-Frosch, R. A. (2002b) Discrimination and the political economy of environmental inequality. *Environment and Planning C: Government and Policy,* 20**,** 477-496.

O'Neill, M. S., Jerrett, M., Kawachi, I., Levy, J. I., Cohen, A. J., Gouveia, N., Wilkinson, P., Fletcher, T., Cifuentes, L., Schwartz, J., Pollution, W. o. A. & Conditions, S. (2003) Health, Wealth, and Air Pollution: Advancing Theory and Methods. *Environ Health Perspect,* 111.

Peek, M. K., Cutchin, M. P., Freeman, D., Stowe, R. P. & Goodwin, J. S. (2009) Environmental hazards and stress: evidence from the Texas City Stress and Health Study. *Journal of Epidemiology and Community Health,* 63**,** 792-798.

Perlin, S. A., Setzer, R. W., Creason, J. & Sexton, K. (1995) Distribution of Industrial Air Emissions by Income and Race in The United States: An Approach Using the Toxic Release Inventory. *Environmental Science & Technology,* 29**,** 69-80.

Perlin, S. A., Sexton, K. E. N. & Wong, D. W. S. (1999) An examination of race and poverty for populations living near industrial sources of air pollution. *J Expo Anal Environ Epidemiol,* 9**,** 29-48.

Perlin SA, W. D., Sexton K. (2001) Residential proximity to industrial sources of air pollution: interrelationships among race, poverty, and age. *J Air Waste Manag Assoc. ,* 51**,** 406-421.

Registre Français des Emissions Polluantes (IREP). Available at: http://www.pollutionsindustrielles.ecologie.gouv.fr/IREP/index.php

Rotko T, K. K., Hänninen O, Jantunen M. (2000) Sociodemographic descriptors of personal exposure to fine particles (PM2.5) in EXPOLIS Helsinki. *J Expo Anal Environ Epidemiol.,* 10**,** 385-393.

Shao M, Z. Y., Zeng L, Tang X, Zhang J, Zhong L, Wang B. (2009) Ground-level ozone in the Pearl River Delta and the roles of VOC and NO(x) in its production. *J Environ Manage,* 90**,** 512-8.

Sheppard E, L. H., McMaster RB, Tian H. (1999) GIS-based measures of environmental equity: exploring their sensitivity and significance. *J Expo Anal Environ Epidemiol,* 9**,** 18-28.

Spinelli R, P. G. (2010) Industrial harvesting of olive tree pruning residue for energy biomass. *Bioresour Technol.,* 101**,** 730-5.

Vidale S, B. A., Guidotti M, Arnaboldi M, Sterzi R. (2010) Air pollution positively correlates with daily stroke admission and in hospital mortality: a study in the urban area of Como, Italy. . *Neurol Sci. ,* 31**,** 179-82.

Viel, J.-F., Hägi, M., Upegui, E. & Laurian, L. (2011) Environmental justice in a French industrial region: Are polluting industrial facilities equally distributed? *Health & Place,* 17**,** 257-262.

Vrijheid, M., Martinez, D. et al. (2010) Socioeconomic status and exposure to multiple environmental pollutants during pregnancy: evidence for environmental inequity? *J Epidemiol Community Health,* 10.1136/jech.2010.117408.

Yorifuji T, N. H., Kashima S, Ohki S, Murakoshi T, Takao S, Tsuda T, Doi H. (2011) Residential proximity to major roads and preterm births. *Epidemiology,* 22**,** 74-80.

Zanobetti A, S. J. (2000) Race, gender, and social status as modifiers of the effects of PM10 on mortality. *Journal of occupational and environmental medicine American College of Occupational and Environmental Medicine,* 42**,** 469-474.

Annexes

Table 10 : Logistic regression analysis in Lille for industry index of 1km-buffer presence

	Univariate regression (Obs : 504)			**Stepwise** (Obs : 387)			**Multivariate logistic regression** (Obs: 504)		
	OR	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**
Unemployed people (%)	1.0285	[1.0087; 1.0486]	0.005	1.0865	[1.0359, 1.1395]	0.001	1.0408	[1.0049, 1.0780]	0.026
Foreigners (%)	1.0348	[1.0000; 1.0707]	0.050	.9074	[.8350, .9862]	0.022	.9747	[.9157, 1.0376]	**0.422**

(OR: Increased risk for additional 0.01unit)

Table 11 : Logistic regression analysis in Lyon for industry presence index

	Univariate regression (Obs : 503)			**Stepwise** (Obs : 415)			**Multivariate logistic regression** (Obs: 503)		
	OR	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**
People aged 15 years or older with a general or vocational maturity certificates (%)	.8980	.7799 1.0340	0.135	.5349	.3998 .7158	0.000	.8621	.7550 .9845	0.029
Households without a car (%)	.9419	.9025 .9830	0.006	.8994	.8388 .9643	0.003	.9371	.8990 .9769	0.002

(OR: Increased risk for additional 0.01unit)

Table 12 : Logistic regression analysis in Lyon for industry index of 1km-buffer presence

	Univariate regression (Obs : 503)			**Stepwise** (Obs : 415)			**Multivariate logistic regression** (Obs: 503)		
	OR	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**
People aged 15 years or older with at least a tertiary education (%)	.7630	.7170 .8119	0.000	.6730	.6172 .7339	0.000	.7412	.6929 .7929	0.000
Primary residence with a garage or other parking space (%)	1.0058	.9980 1.0137	0.146	1.0226	1.0103 1.0350	0.000	1.0162	1.0065 1.0261	0.001

(OR: Increased risk for additional 0.01unit)

FIGURES

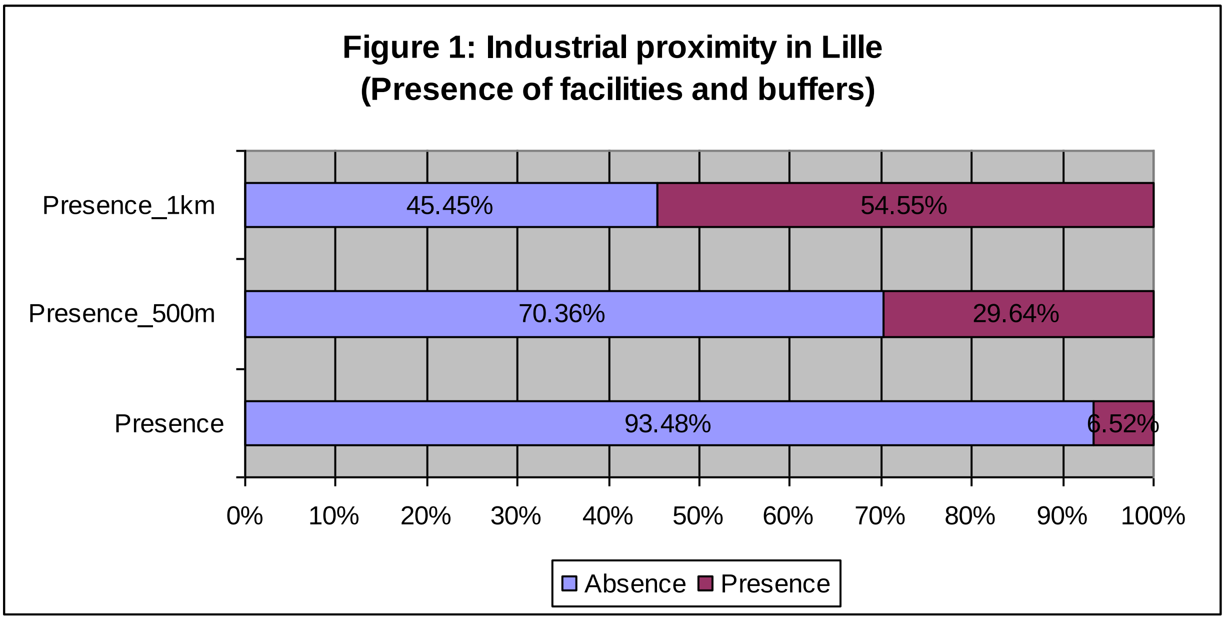


Figure 1: Industrial proximity in Lille
(Presence of facilities and buffers)
Presence_1km
Presence_500m
Presence
45.45%
54.55%
70.36%
29.64%
93.48%
6.52%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Absence
Presence

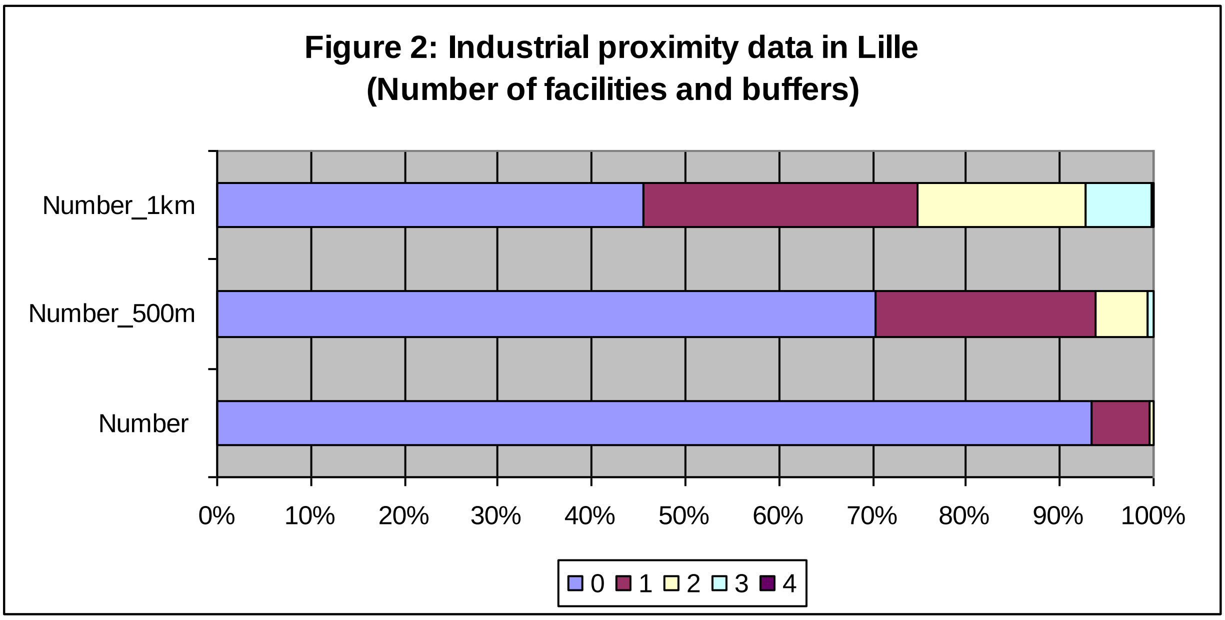
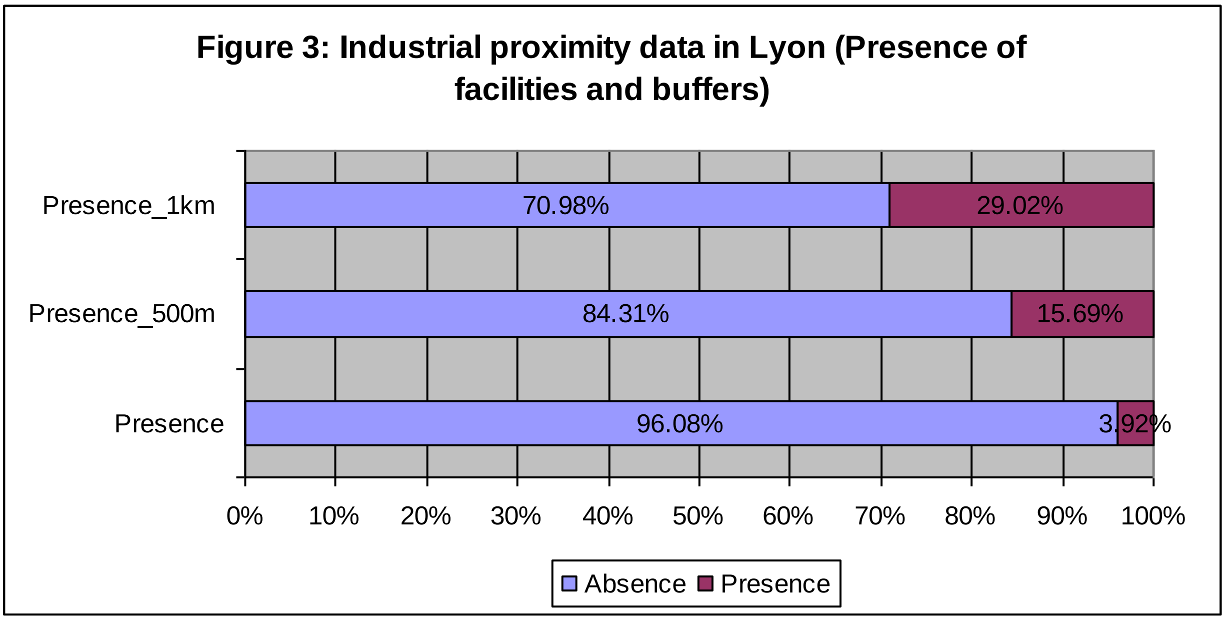
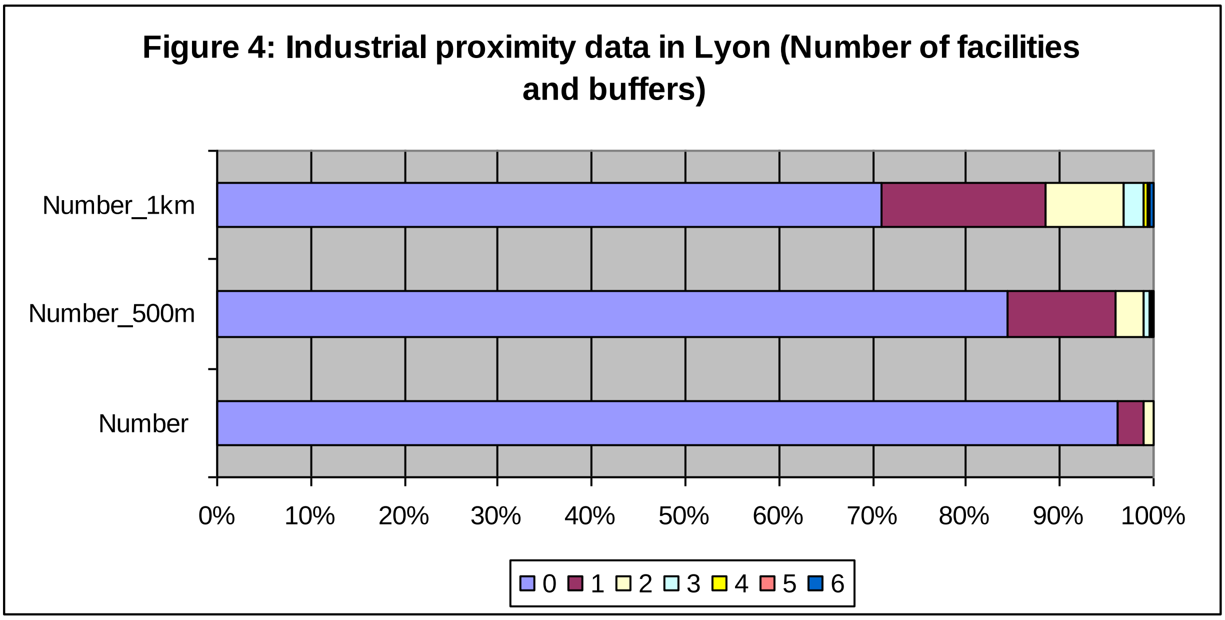


Figure 2: Industrial proximity data in Lille
(Number of facilities and buffers)
Number_1km
Number_500m
Number
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
0
1
2
3
4





Presence: the presence of one industry or more in the IRIS

Presence_500m: the presence of one or more buffer with a 500m radius

Presence_1km: the presence of one or more buffer with a 1km radius

Number: the number of industries in the IRIS

Number_ 500m: the number of buffers within a 500m radius which touch the IRIS

Number_1km: the number of buffers within a 1km radius which touch the IRIS

Figure 5: Spatial Coincidence Method

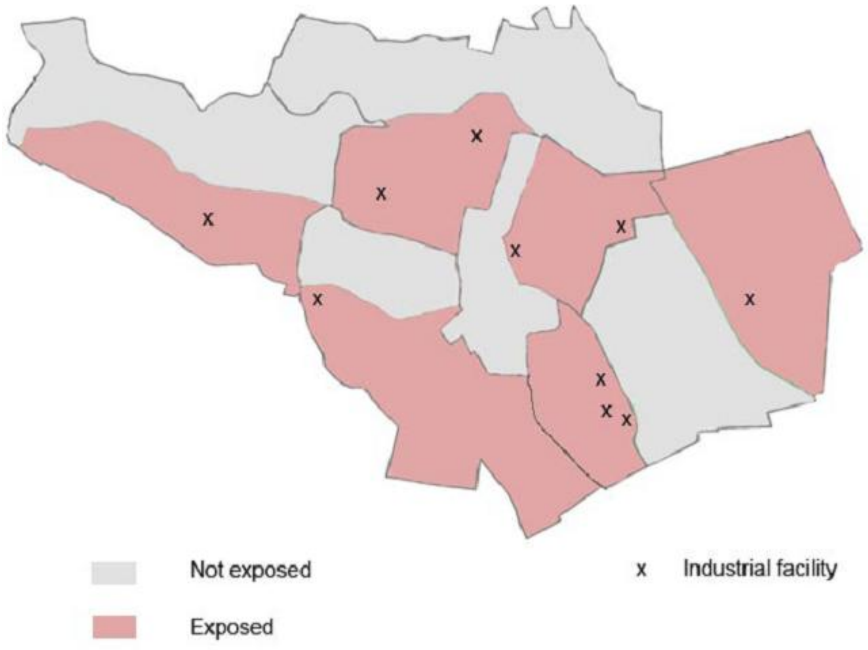


Figure 6: Buffer Method