



HAL
open science

Analyse des variations spatiales d'incidence des cas de cardiopathies ischémiques dans le cadre de l'étude MONICA au sein de la Métropole Européenne de Lille, entre 2008 et 2011

Antoine Longo

► **To cite this version:**

Antoine Longo. Analyse des variations spatiales d'incidence des cas de cardiopathies ischémiques dans le cadre de l'étude MONICA au sein de la Métropole Européenne de Lille, entre 2008 et 2011. Sciences de l'Homme et Société. 2018. dumas-03441033

HAL Id: dumas-03441033

<https://dumas.ccsd.cnrs.fr/dumas-03441033v1>

Submitted on 22 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Antoine Longo

Analyse des variations spatiales
d'incidence des cas de cardiopathies
ischémiques

Dans le cadre de l'étude MONICA au sein de la
Métropole Européenne de Lille, entre 2008 et 2011

Mémoire soutenu le 5 septembre 2018

En vue de l'obtention du master

Mention Géographie, Aménagement, Environnement et Développement

Parcours Territoires, Villes et Santé

Université Paris Nanterre / Université Paris Est Créteil

Année Universitaire 2017/2018

Sous la direction de M. Michaël Genin (tuteur de stage)

Et M. Stéphane Rican (tuteur universitaire)

Table des matières

REMERCIEMENTS	4
MOTS-CLES / KEY WORDS	5
RESUME	5
ABSTRACT	5
TABLE DES SIGLES	6
PARTIE INTRODUCTIVE – Géographie et Maladie Cardio-vasculaire : Contexte et description de l'étude MONICA	8
Chapitre introductif : Contexte épidémiologique	8
1. Les Maladies Cardiovasculaires (MCV).....	8
1.1 <i>Définition</i>	8
1.2 <i>Epidémiologie à travers le monde</i>	9
2. Le Projet Monica.....	11
2.1 <i>Étude générale</i>	11
2.2 <i>Le registre</i>	12
3. Problématique et objectifs.....	12
PARTIE I – MATERIEL ET OUTILS D'ANALYSE	14
Chapitre 1 : Territoire d'étude et cadre de recherche	14
1. Territoire d'étude et échelle d'analyse	14
1.1 <i>Site et situation de la MEL</i>	14
1.2 <i>Histoire du développement de la MEL</i>	15
1.3 <i>La Métropole aujourd'hui</i>	17
1.4 <i>Description de l'échelle d'étude : l'IRIS</i>	19
2. Revue de la littérature.....	20
2.1 <i>Méthodologie de la revue de la littérature</i>	20
2.2 <i>Principaux enseignements</i>	23
Chapitre 2 : Cadre d'étude	23
1. Le registre MONICA Lillois	23
1.1 <i>Description du registre</i>	23
1.2 <i>Critères d'inclusion</i>	24
2. Le géocodage.....	25
2.1 <i>Nécessité</i>	25
2.2 <i>Méthode</i>	26
Chapitre 3 : Outils de caractérisation	30
1. Bases de données INSEE.....	30
1.1 <i>Critère de jugement principal</i>	30

1.2	<i>Les bases de données</i>	30
2.	Outils de mesure de la défaveur sociale : les indices.....	31
2.1	<i>Description des indices</i>	31
2.2	<i>Le French EDI</i>	35
2.3	<i>Le Fdep</i>	36
2.4	<i>Le choix de l'indice</i>	37
3.	Mesurer la pollution atmosphérique	39
3.1	<i>Le choix des polluants</i>	40
3.2	<i>Méthode de captation</i>	40
3.3	<i>Principales sources d'émissions</i>	42
3.4	<i>Les indicateurs de pollution</i>	45
PARTIE II – ANALYSE SPATIALE D'INCIDENCE ET CARACTERISATION DES PRINCIPAUX FACTEURS DE RISQUES		46
Chapitre 1 : Description du registre MONICA		46
1.	Profil des cas incidents	46
2.	Les taux d'incidence : bruts et standardisés	48
2.1	<i>Taux d'incidence brute</i>	48
2.2	<i>Les différents types de standardisation</i>	50
2.3	<i>Taux d'incidence standardisés par âge</i>	51
2.4	<i>Taux d'incidence standardisés par âge et par sexe</i>	52
Chapitre 2 : Distribution spatiale de l'incidence et détection de clusters spatiaux		53
1.	Distribution spatiale des rapports standardisés d'incidence	53
1.1	<i>Méthode de calcul</i>	53
1.2	<i>Limites des SIR bruts</i>	55
1.3	<i>Lissage des SIR par modèle bayésien</i>	56
2.	Détection de clusters spatiaux	59
2.1	<i>Objectif</i>	59
2.2	<i>Méthodes</i>	59
2.2.1	Choix du type de cluster	60
2.2.2	Phase de détection du cluster	60
2.2.3	Phase d'inférence	61
2.2.4	Choix du type de scan.....	62
2.3	<i>Résultats des détections</i>	63
Chapitre 3 : Caractérisation des clusters spatiaux et influence des facteurs de risque dans les variations spatiales d'incidence		67
1.	Objectif	67
2.	Les stratégies de comparaisons de groupes d'IRIS.....	68

3.	Les méthodes statistiques de comparaison	68
3.1	<i>Les tests de Kruskal-Wallis</i>	69
4.	Résultats	70
4.1	<i>Première stratégie</i>	70
4.2	<i>Deuxième stratégie</i>	72
5.	Limites des stratégies d'analyse de la caractérisation des clusters	73
PARTIE III – La défaveur sociale : unique facteur de risque ?		75
Chapitre 1 : Ajustement des SIR lissés sur la défaveur sociale et distribution spatiale de l'incidence		75
1.	Objectif de l'ajustement des nouveaux SIR lissés	75
2.	Ajustement et lissage des SIR ajustés sur la défaveur sociale par régression écologique.....	76
3.	Résultats des SIR lissés ajustés sur la défaveur sociale	76
Chapitre 2 : Détection de clusters spatiaux ajustés sur la défaveur sociale.....		78
1.	Objectif de la détection de nouveaux clusters spatiaux	78
2.	Technique d'ajustement par régression de Poisson	79
3.	Résultats	79
Chapitre 3 : Régressions écologiques ajustées sur la défaveur sociale		80
1.	Objectif de la régression écologique ajustée	80
2.	Méthodologie de l'analyse bivariée	80
3.	Résultat des régressions écologiques ajustées sur la défaveur sociale	82
3.1	<i>Analyse bivariée</i>	82
3.2	<i>Analyse multivariée</i>	85
DISCUSSION.....		87
CONCLUSION.....		94
BIBLIOGRAPHIE		95
SITOGRAFIE		97
TABLE DES ILLUSTRATIONS.....		98
ANNEXES		99
ANNEXE 1 : GEOCODAGE D'ADRESSE A L'IRIS		99
ANNEXE 2 : DETECTION DE CLUSTERS ISOTONIQUES PAR LE LOGICIEL SATSCAN		109
ANNEXE 3 : LES TESTS POST HOC DES TESTS KW SUR FDEP, POLLUTION ET MOS		116

REMERCIEMENTS

Je souhaite tout d'abord remercier mon maître de stage, Michaël Génin, qui a toujours été très enthousiaste, patient et pédagogue lors des nombreuses leçons qu'il m'a donné et de ses conseils toujours pertinents. Je le remercie aussi d'avoir été aussi ouvert aux discussions et aux échanges. Cette sympathie et cette considération marqueront l'un de mes stages les plus formateurs, et je peux dire que j'ai mûri à son contact.

Ensuite, je souhaite plus généralement remercier l'ensemble de l'équipe du CERIM ainsi que les stagiaires qui ont participé à ce que ce stage et ces 6 mois à Lille soient une très belle expérience pour moi.

Je remercie également Florent Occelli et Caroline Lanier, qui m'ont transmis leurs connaissances dans le domaine de la biologie, de la santé et de l'environnement. Ce mémoire et cette étude n'auraient pas été réalisables sans leur travail en amont.

Je souhaite aussi remercier Julie Dumont, qui m'a accueilli pendant près de trois semaines dans son bureau à l'Institut Pasteur pour la réalisation du géocodage. Son enthousiasme et sa bonne humeur ainsi que sa réactivité et son aide précieuse m'ont permis d'accomplir ma tâche dans les meilleures conditions.

De plus, je remercie Stéphane Rican pour ses conseils avisés et sa confiance en moi. Ce stage n'a été possible que grâce à la confiance portée sur ma capacité à réaliser cette étude, et je mesure l'assurance que l'on m'a confié. J'espère que ce mémoire sera à la hauteur de cette espérance. Je remercie d'ailleurs l'ensemble de mes professeurs du master Territoires, Villes et Santé des Université de Paris Est Créteil et Paris Ouest Nanterre pour m'avoir transmis leurs savoirs sans lesquelles je n'aurais pu mener ce stage à bien.

Enfin, je remercie mes proches qui ont su m'éclairer et m'épauler lors de la rédaction de ce mémoire.

MOTS-CLES / KEY WORDS

Maladies cardiovasculaires, cardiopathies ischémiques, incidence, défaveur sociale, pollution de l'air, analyse spatiale, clusters. / Cardiovascular diseases, ischemic heart diseases, incidence, social deprivation, air pollution, spatial analysis, clusters.

RESUME

Les maladies cardiovasculaires, et plus précisément les cardiopathies ischémiques, représentent aujourd'hui l'un des plus grands enjeux de santé publique mondiale. En France, près de 140 000 personnes décèdent chaque année des complications liées à ces maladies. Si les facteurs individuels comme le tabac ou l'alcool sont aujourd'hui largement reconnus dans la survenue d'un événement aigu de cardiopathie, les facteurs environnementaux, et plus particulièrement la défaveur sociale et la pollution atmosphérique restent peu explorés. Le projet de recherche MONICA a été développé dans les années 1980 pour connaître la relation entre mortalité liée aux maladies cardiovasculaires en général et les facteurs individuels. Aujourd'hui, cette étude est toujours en cours mais s'intéresse désormais à faire la lumière sur l'impact de l'ensemble des facteurs environnementaux dans le risque de survenue d'un événement aigu. Ce mémoire se concentre sur l'analyse des facteurs environnementaux. L'originalité de cette étude basée sur un registre de personnes victimes de ces événements aigus à Lille est double puisque i) la question de la défaveur sociale comme facteur de risque d'incidence en santé cardiaque n'est apparue que très récemment dans la littérature scientifique et ii) la pollution atmosphérique pourrait jouer un rôle dans les variations spatiales d'incidence de ces maladies.

ABSTRACT

Cardiovascular diseases are one of the most important public health issues. Ischemic heart diseases represent a large part of cardiovascular diseases, especially in France where they cause the death of about 140.000 people every year. If individual factors such as tobacco or alcohol are well recognized as factors of ischemic heart diseases, environmental factors – especially social deprivation and air pollution – remain neglected in academic studies. The MONICA Project has been developed first to analyse these individual factors, then to analyse the environmental factors on ischemic heart disease. Our study, based on a patient register in Lille, France from the MONICA Project, focuses on environmental dimensions. The originality of the study is double because i) social deprivation is a very recent question among ischemic heart diseases studies, and ii) air pollution may have an impact on the spatial variations of incidence of those diseases.

TABLE DES SIGLES

AVC : Accident Vasculaire Cérébral

BAN : Base Adresse Nationale

CERIM : Centre d'Etudes et de Recherches en Informatique Médicale

CI : Cardiopathies Ischémiques

CNIL : Commission Nationale de l'Informatique et des Libertés

IDM : Infarctus du Myocarde

IGN : Institut Nationale de l'Information Géographique

INSEE : Institut Nationale de la Statistique et des Etudes Economiques

INVS : Institut Nationale de Veille Sanitaire

IRIS : Ilots Regroupés pour l'Information Statistique

MCV : Maladies Cardiovasculaires

MEL : Métropole Européenne de Lille

MONICA : *Multinational mONItoring of trends and determinants of Cardiovascular diseases*

OMS : Organisation Mondiale de la Santé

RR : Risque Relatif

SIG : Système d'Information Géographique

PARTIE INTRODUCTIVE – Géographie et Maladie Cardio-vasculaire : Contexte et description de l'étude MONICA

SOMMAIRE

PARTIE INTRODUCTIVE – Géographie et Maladie Cardio-vasculaire : Contexte et description de l'étude MONICA.....	8
Chapitre introductif : Contexte épidémiologique	8
1. Les Maladies Cardiovasculaires (MCV).....	8
2. Le Projet Monica.....	11
3. Problématique et objectifs.....	12

Chapitre introductif : Contexte épidémiologique

1. Les Maladies Cardiovasculaires (MCV)

1.1 Définition

L'Organisation Mondiale de la Santé (OMS) distingue deux types de maladies afin de mieux cibler ses luttes : les maladies transmissibles et les maladies non-transmissibles (MNT). En 2010, elle rappelle que les « *maladies non-transmissibles sont, à l'échelle mondiale, la principale cause de mortalité* ». (OMS, 2017). L'incidence de ces dernières représente environ 70% des décès dans le monde chaque année, soit environ 40 millions de décès. Parmi ces maladies non-transmissibles, les maladies cardiovasculaires (MCV) sont au premier rang des causes de décès dans le monde avec 17,7 millions en 2017, soit près d'un décès par MNT sur deux.

Les « maladies cardiovasculaires » (MCV) désignent l'ensemble des troubles qui affectent l'activité cardiaque et la circulation sanguine (INSERM, 2014). En 2015, l'OMS établit un état des lieux des différentes MCV :

- Les cardiopathies coronariennes (touchant les vaisseaux sanguins qui alimentent le muscle cardiaque) ;
- Les maladies cérébro-vasculaires (touchant les vaisseaux sanguins qui alimentent le cerveau) ;
- Les artériopathies périphériques (touchant les vaisseaux sanguins qui alimentent les bras et les jambes) ;
- Les cardiopathies rhumatismales, affectant le muscle et les valves cardiaques ;

- Les malformations cardiaques congénitales (malformations de la structure du cœur déjà présentes à la naissance) ;
- Les thromboses veineuses profondes et les embolies pulmonaires (obstruction des veines des jambes par un caillot sanguin, susceptible de se libérer et de migrer vers le cœur ou les poumons). (OMS, 2015).

Parmi ces maladies, les plus courantes restent les cardiopathies coronariennes (ou « maladie coronarienne » ou encore « cardiopathie ischémique ») qui se caractérisent par l'obstruction de l'artère coronaire (celle qui irrigue le muscle cardiaque) par de multiples dépôts formant une plaque appelée « athérome ». Cette maladie est la première cause de mortalité dans le monde depuis le début du XXI^{ème} siècle (OMS, 2017).

La conséquence de ces maladies est le plus souvent la survenue d'un événement aigu¹ comme un Infarctus Du Myocarde (IDM), plus communément appelé « crise cardiaque », ou un Accident Vasculaire Cérébral (AVC).

Le myocarde est le muscle qui génère les contractions cardiaques et qui permet au cœur de pomper le sang à travers les artères et les veines. Si l'artère coronaire se bouche à cause d'un athérome, cela s'appelle une « ischémie ». Le myocarde se retrouve alors peu irrigué en sang, et donc en oxygène, ce qui va engendrer la mort de plusieurs cellules du muscle (aussi appelée « hypoxie ») et ainsi son dysfonctionnement voire son arrêt total : il est alors question d'un infarctus du myocarde.

1.2 Epidémiologie à travers le monde

Les IDM se localisent à 80% dans les pays à faibles ou moyens revenus, ainsi que dans les populations les plus pauvres dans chaque pays (Yusuf et *al.*, 2004 ; OMS, 2010). Cela s'explique par des modes de vie plus à risque chez les populations les plus défavorisées. En effet, certains facteurs de risques dits « modifiables » – c'est-à-dire qu'ils dépendent du mode de vie de la personne – sont notamment observés chez les populations les plus défavorisées. Dans son « Rapport sur la situation mondiale des maladies non transmissibles » établi en 2010 (*Global Status Report on Noncommunicable Diseases*, 2010) l'OMS classe ces facteurs de risque du plus au moins morbide :

¹ En santé, le terme « aigu » désigne un événement ponctuel comme une crise ou un accident, dont la douleur a un début, une fin et est de courte durée, contrairement à « chronique » qui désigne une douleur ou un trouble continu dans le temps, même si la cause qui l'a engendré a disparu (INCa, 2017).

- Le tabagisme (actif et passif) ;
- Le manque d'exercice physique / la sédentarité (moins de 30 minutes d'exercice par jour) ;
- L'usage nocif de l'alcool ;
- L'alimentation déséquilibrée (pas assez de fruits et légumes, excès de sel et de graisses saturées) ;
- L'hypertension artérielle ;
- L'obésité (d'après l'IMC) ;
- L'hypercholestérolémie (un taux trop élevé de mauvais cholestérol dans le sang) ;
- Les infections associées aux cancers (HPV², Hépatites B et C).

Ces facteurs sont néanmoins de plus en plus en plus fréquents chez les populations favorisées et dans les pays les plus riches, notamment à propos de l'hypercholestérolémie et de l'alimentation déséquilibrée. Dans une étude publiée en 2004 à propos des effets des potentiels facteurs de risque modifiable (Yusuf et *al.*, 2004), les auteurs rapportent que certains facteurs sont plus à risques dans certaines zones géographiques que d'autres. Effectivement, bien que les facteurs observés (tabagisme actif, diabète, hypertension, obésité abdominale, index psychosocial³, consommation de fruits et légumes, consommation d'alcool, exercice, ratio d'apolipoprotéines ApoB/ApoA1⁴) soient tous significativement démontrés comme « à risque » toutes populations confondues, on lit notamment que l'Europe de l'Ouest, l'Amérique du Nord, l'Australie et la Nouvelle-Zélande (pays à hauts revenus) ainsi que l'Asie du Sud (région de pays à moyens revenus) ont une part de risque associée à l'obésité abdominale plus forte qu'au tabagisme, alors que l'on observe le contraire en Chine par exemple, où le tabagisme est plus fréquent au sein de la population que l'obésité, ce qui influence le niveau de risque.

D'après le rapport de l'OMS de 2010, les prévisions concernant les décès engendrés par les maladies non-transmissibles devraient avoir augmenté de 15% dans le monde entre 2010 et 2020, notamment en Afrique, Asie de l'Est et l'Est de la Méditerranée (Proche-Orient) où cette augmentation dépassera les 20%. Il est donc possible de constater ici que les maladies

² HPV : Papillomavirus humain (*Human PapillomaVirus*), responsable notamment des cancers du col de l'utérus.

³ Index calculé à partir de données issues des réponses des participants à l'études sur : dépression ou non, stress au travail ou à la maison ou non, stress modéré ou sévère à propos de sa situation financière ou stress minimal ou non, une naissance ou plus ou non, et un score de situation dans les trois quartiles inférieurs par rapport au quartile supérieur de la distribution issu de ce calcul.

⁴ Les apolipoprotéines sont des protéines ayant un rôle sur l'action des graisses dans le corps. Les ApoB sont celles qui peuvent être à risque pour le foie et les intestins, les ApoA1 sont celles qui régulent le niveau de graisses. Le risque survient donc lorsque le niveau d'ApoB est plus élevé que les ApoA1 dans le corps.

cardiovasculaires, et en particulier les infarctus du myocarde, sont des enjeux de santé publique majeurs et globaux, et que leur prépondérance n'est pas prête de s'estomper.

En France, les maladies cardiovasculaires sont la deuxième cause de décès annuelle avec environ 140 000 décès par an, derrière les cancers (InVS, 2017). Précisément pour les maladies coronariennes, la prévalence en 2008 était estimée à 3,7% de la population adulte (de Peretti *et al.*, 2014). En 2008, ce sont 360 000 nouveaux cas qui ont été déclarés. Le taux d'incidence – c'est-à-dire la proportion de nouveaux cas par an sur une population donnée – tend à diminuer mais on observe qu'elle est plus forte chez les hommes que chez les femmes. En effet, d'après une étude menée grâce à une observation de la population française dans trois villes (Lille, Strasbourg et Toulouse) entre 2000 et 2007, le taux d'incidence serait passé de 292 à 237 pour 100 000 habitants chez les hommes, et de 69 à 56 pour 100 000 habitants chez les femmes (Wagner *et al.*, 2011). En ce qui concerne la France dans sa globalité, le taux d'incidence moyen est de 216 cas pour 100 000 habitants (Dujardin & Cambou, 2005). En d'autres termes, la tendance en France serait, à première vue, à la baisse si l'on compare aux estimations mondiales de l'OMS pour les années à venir, mais les maladies coronariennes et les autres maladies cardiovasculaires en général restent néanmoins des causes de mortalité importantes en France, et par conséquent, un enjeu de santé publique majeur.

2. Le Projet Monica

2.1 Étude générale

Dans ce contexte d'augmentation du risque de maladie cardiovasculaire, l'OMS a mis en place le projet MONICA (*Multinational mONItoring of trends and determinants of Cardiovascular diseases*⁵). Lancé d'abord au début des années 80, ce projet réunissait 37 régions réparties dans 21 pays du monde, dont la grande majorité en Europe (Ducimetière, Actualité et Dossier en Santé Publique, Haut Conseil de la Santé Publique, 2004) et consistait à surveiller quotidiennement plus de 10 millions de personnes à travers le monde (OMS, 1987). Les objectifs étaient alors de caractériser les facteurs de survenue de problèmes de Cardiopathies Ischémiques (ou « CI », infarctus du myocarde notamment), de mesurer les différentes tendances de mode de vie à risque dans les populations et d'observer l'évolution de la mortalité par MCV.

En France, le projet MONICA s'est mis en place au sein de deux régions et d'une communauté urbaine : le Bas-Rhin, la Haute-Garonne et la Métropole Européenne de Lille

⁵ En français : Surveillance internationale des tendances et déterminants des maladies cardiovasculaires.

(MEL)⁶. Cette étude a permis de montrer que la France s'inscrivait dans un gradient Nord/Sud de mortalité par événement coronaire. En effet, d'après un article publié dans la revue *The Lancet* au début des années 1990, les données issues de l'étude MONICA ont montré que les pays du nord de l'Europe avaient tendance à avoir une mortalité beaucoup plus élevée de causes cardiovasculaires que les pays méditerranéens (Renaud & De Lorgeril, 1992). Les données MONICA ont permis de mettre fin au mythe du « *French Paradox* », qui consistait à penser que les Français décédaient moins des causes de MCV que leurs voisins du Nord malgré une consommation soutenue d'alcool et de matières grasses. Grâce aux résultats des trois laboratoires d'études situés dans les régions françaises concernées par MONICA (Strasbourg, Toulouse et Lille), cette idée a pu être déconstruite car il a pu être observé que la mortalité par MCV était décroissante de Lille à Strasbourg puis à Toulouse, résultat d'un mode de vie plus à risque au Nord qu'au Sud, notamment à cause d'une plus forte consommation de matière grasse, de tabac, et une moins forte consommation de fruits et légumes. On note également que la consommation d'alcool est sensiblement la même entre les trois régions, mais les Toulousains ayant une forte consommation de vin, il semblerait que ce dernier ait un effet « protecteur » d'après les auteurs de l'article, en comparaison aux autres alcools plus consommés dans les régions du Nord et de l'Est comme la bière ou les spiritueux.

2.2 Le registre

Afin de pouvoir observer les premières données de mortalité et de morbidité, trois registres ont été créés dans les villes d'études françaises, dès le début de l'étude en 1985 (OMS, 1987). L'objectif était de continuer les observations cette fois-ci sur la base des admissions à l'hôpital pour cause d'événement de cardiopathie ischémique, afin d'étudier les profils de personnes qui étaient victimes de ces attaques, ainsi que d'analyser leur état de santé 28 jours après l'événement (survie ou non, et si non, durée de survie en jours, heures et minutes).

A Lille, le registre créé permet d'enregistrer l'ensemble des cas de cardiopathie ischémique de la Métropole. Chaque personne déclenchant un accident cardiovasculaire est inscrite dans la base et renseigne de son sexe, son âge, mais aussi de son adresse afin de connaître son quartier d'habitation qui sera nécessaire pour l'étude, comme il le sera décrit plus tard dans ce mémoire.

3. Problématique et objectifs

⁶ Au moment de la création, la collectivité s'appelait « Communauté Urbaine de Lille (CUDL) ». Elle est devenue « Métropole » au 1^{er} janvier 2015.

L'incidence des maladies cardiovasculaire à Lille est donc observée depuis maintenant plus de quarante ans. **Mais comment se caractérise, spatialement et étiologiquement, cette incidence ?**

En utilisant les données du registre lillois entre 2008 et 2011, ce travail de recherche, effectué au Centre d'Études et de Recherche en Informatique Médicale (CERIM) à Lille, va permettre de répondre à cette question, qui soulève deux objectifs.

Dans un premier temps, il s'agira d'évaluer la distribution spatiale de l'incidence des évènements cardiovasculaires sur l'ensemble de Métropole Européenne de Lille grâce à des outils statistiques et cartographiques.

Dans un second temps, l'objectif sera de mettre en lien ces variations spatiales d'incidence avec des facteurs socio-économiques et environnementaux, afin de déterminer l'influence de ces facteurs dans la survenue d'un accident cardiovasculaire.

PARTIE I – MATERIEL ET OUTILS D’ANALYSE

SOMMAIRE

PARTIE I – MATERIEL ET OUTILS D’ANALYSE	14
Chapitre 1 : Territoire d’étude et cadre de recherche	14
1. Territoire d’étude et échelle d’analyse	14
2. Revue de la littérature.....	20
Chapitre 2 : Cadre d’étude	23
1. Le registre MONICA Lillois	23
2. Le géocodage.....	25
Chapitre 3 : Outils de caractérisation.....	30
1. Bases de données INSEE.....	30
2. Outils de mesure de la défaveur sociale : les indices.....	31
3. Mesurer la pollution atmosphérique	39

Chapitre 1 : Territoire d’étude et cadre de recherche

1. Territoire d’étude et échelle d’analyse

1.1 Site et situation de la MEL

A l’heure actuelle, en 2018, la MEL regroupe 90 communes des Hauts-de-France et près de 1,2 millions d’habitants, ce qui fait d’elle la quatrième métropole la plus peuplée en France, derrière le Grand Paris, La Métropole de Lyon et l’Agglomération Aix-Marseille. De plus, elle possède un territoire s’étendant sur 611,45 km², faisant d’elle la cinquième plus grande métropole, derrière les métropoles d’Aix-Marseille, Nice, Paris et Rennes.

Elle est située en plein cœur du département du Nord qui la borde à l’est et à l’ouest. Au nord, la Métropole partage 84 km de frontière commune avec la Belgique, tandis qu’au sud elle s’ouvre sur le département du Pas-de-Calais.

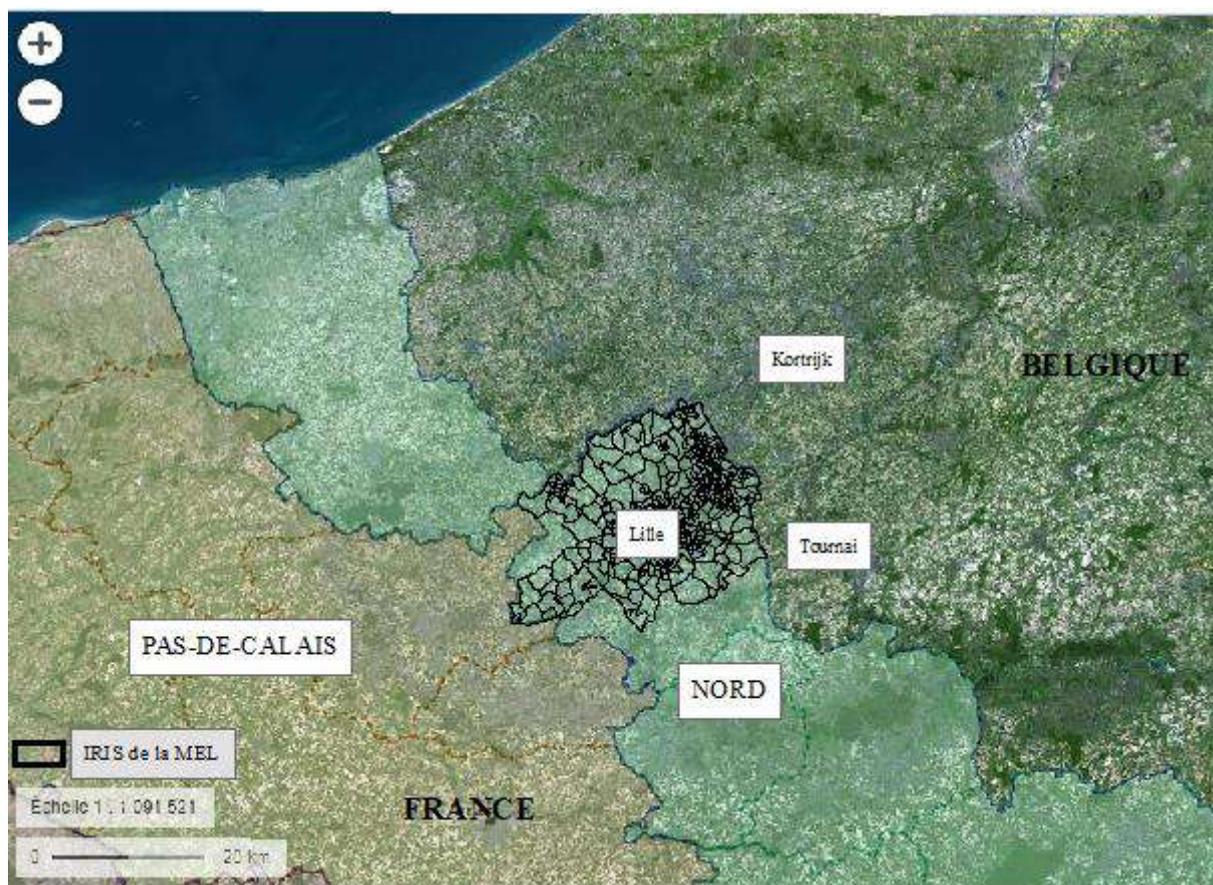


Figure 1 : Carte de situation de la MEL. Sources : IGN, Géoportail.

1.2 Histoire du développement de la MEL

La partie de l'étude MONICA auquel ce mémoire s'intéresse se localise au sein de la Métropole Européenne de Lille. Cette entité voit le jour en 1966 lors de la première loi à propos de la création des Etablissement Publics de Coopération Intercommunale (EPCI) qui permet de créer, à l'époque, les communautés urbaines de Lyon, Bordeaux, Strasbourg et Lille. Cette dernière prend alors le nom de Communauté Urbaine De Lille (CUDL) et regroupe alors 36 communes.

Ancien territoire industriel, et notamment textile grâce aux pôles Roubaix-Tourcoing-Lille pendant la révolution industrielle, la Métropole lilloise est aujourd'hui un espace majeur du développement économique et urbain des territoires du Nord de la France. A l'image des premiers centres commerciaux Auchan, ouverts à Roubaix en 1961, puis Roncq dans la banlieue de Tourcoing en 1967 avant de devenir le premier employeur de la région des Hauts-de-France (La Voix du Nord, 2015), la MEL est devenue un centre économique de premier choix.

En 1996, après des travaux d'ampleur régionales et nationales comme la création du centre commercial et d'activités d'Euralille (figure 2) ou l'agrandissement de l'aéroport de Lille-Lesquin (figure 3), la communauté décide de marquer ce changement d'échelle d'influence par une évolution du nom de la communauté en y ajoutant le terme de « métropole », devenant alors Lille Métropole Communauté Urbaine (LMCU).



Figure 2 : Aéroport Lille-Lesquin. Source : France 3 Région Hauts-de-France (à gauche)

Figure 3 : Centre Commercial et d'activités Euralille. Source : Site internet Euralille (à droite)

Jusqu'en 1970, les grands travaux de transports permettent d'offrir à la MEL plusieurs voies de dessertes autoroutières nationales et européennes. Tout d'abord, l'A1 ouvre la route vers Paris qui n'est plus qu'à 2h30 en voiture. Puis, l'A25 permet de rejoindre la côte avec Dunkerque en seulement 1h afin d'ouvrir les bords de mers aux touristes nordistes plus retranchés dans les terres du Nord mais aussi pour le commerce vers la Grande-Bretagne. Enfin, l'A23 permet de rejoindre Bruxelles en 1h30 de transport routier. Plus récemment, la MEL organise des grands travaux de transports ferroviaires, qui accroît sa capacité de liaison internationale et sa position centrale en Europe. En 1994, la gare de Lille-Europe ouvre ses premiers rails, à quelques mètres seulement de la gare de Lille-Flandres, qui permet d'ouvrir la voie aux Lignes à Grande Vitesse (LGV). La plus populaire des LGV est l'Eurostar, ouvrant l'accès à Londres en 1h20. En 2014, l'entreprise Thalys ouvre la MEL à une autre capitale européenne à 2h30 de train : Amsterdam. Lille et son agglomération se retrouvent donc à moins de 3h de transports de quatre capitales d'Europe occidentale, ce qui conforte sa situation de pôle économique majeur.

Dans cet élan de modernisation de ses transports, la MEL construit deux lignes de métro, le Transpole, en 1983 et en 1989. Ces lignes desservent la MEL de l'Ouest de Lille au Nord de Tourcoing, près de la frontière belge, par la ligne 2 (en rouge sur la figure 4), la plus récente, tandis que la ligne 1 dessert le Sud de Lille jusqu'à l'Est de la Métropole (en jaune). De plus, en 1994, la MEL modernise ses deux lignes de tramway, qui furent les premiers transports en commun lillois en 1909, allant jusqu'à Tourcoing et Roubaix (en bleu et vert).

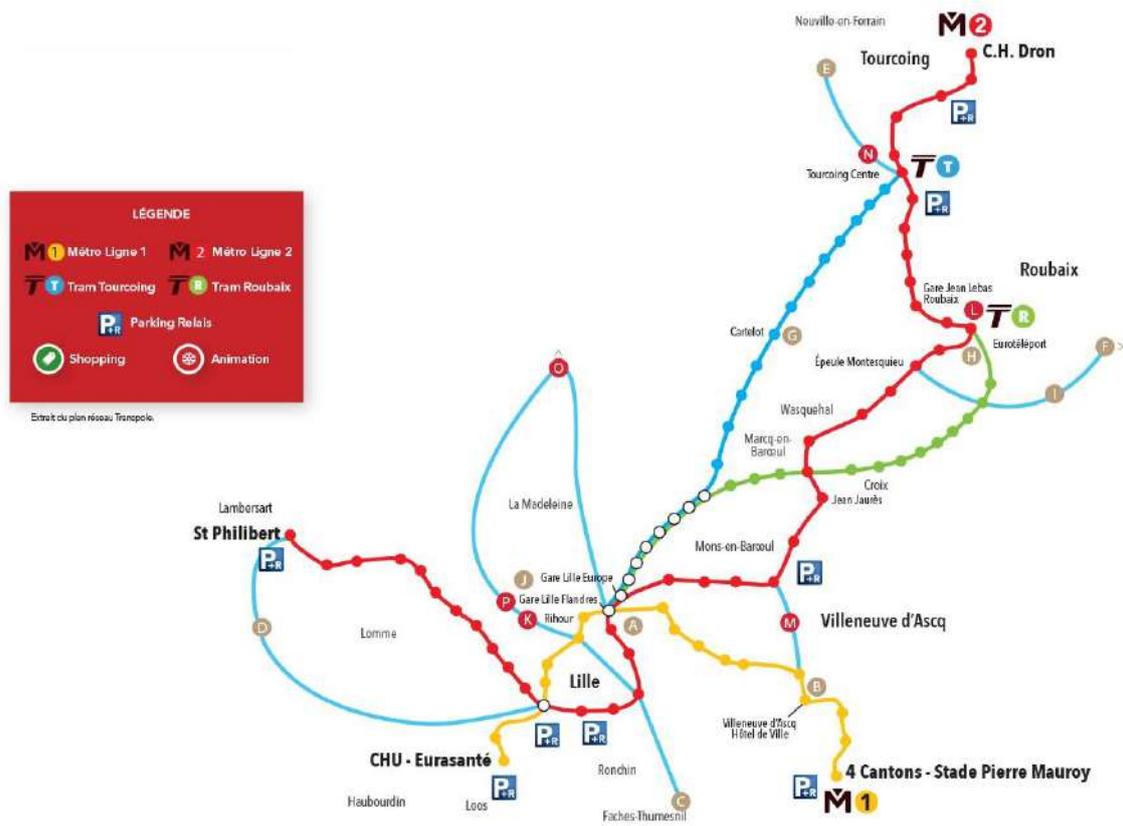


Figure 4 : Tracé des lignes du Transpole de la MEL en 2018. Source : Transpole.

En 2014, la loi sur la Modernisation de l'Action Publique Territoriale et d’Affirmation des Métropoles (MAPTAM) permet la création d’un statut particulier de métropole aux communautés urbaines de plus de 400 000 habitants. La LMCU change alors de nom et devient le Métropole Européenne de Lille (MEL), affichant donc son importance sur le territoire national (« métropole ») mais aussi son attachement à sa zone d’influence dépassant les frontières du territoire national (« européenne »), qui est déjà confirmé depuis 2008 avec la création de l’Eurométropole Lille-Kortrijk-Tournai (figure 1). Cette entité est le premier Groupement Européen de Coopération Territoriale (CEGT) à avoir vu le jour, et regroupe 127 communes et 2,1 millions d’habitants répartis entre le nord de la France et la Belgique.

En 2017, la loi portant sur la Nouvelle Organisation Territoriale de la République (NOTRe) fait évoluer le découpage des régions et incite les communautés urbaines de moins de 15 000 habitants à se regrouper avec des communautés plus importantes. Cette loi est la dernière à l’heure actuelle à avoir fait évoluer le territoire de la MEL, attirant la communauté urbaine de Weppes, au sud de la métropole.

1.3 La Métropole aujourd’hui

La Métropole Européenne de Lille présente aujourd’hui un territoire hétérogène et diversifié. En effet, les grands bouleversements économiques récents ont réussi à faire rayonner les pôles urbains de l’agglomération, tandis que la grande majorité du territoire reste néanmoins à caractère rural.

Aujourd’hui, 65% du territoire de la MEL est composé de terres agricoles (Base de données SIGALE⁷, 2009). La MEL est à l’image de sa région, les Hauts-de-France, un espace à majorité rural, où l’agroalimentaire est le premier secteur d’emplois industriels avec (Préfecture et services de l’Etat en région Hauts-de-France, 2013). Toutefois, les principaux pôles urbains sont visibles, en violet sur la carte ci-dessous (figure 5). Il s’agit de la conurbation Lille-Roubaix-Tourcoing, ainsi que, dans une moindre mesure, la commune d’Armentières à l’Ouest de la MEL.

La figure 5 montre l’occupation du sol de la MEL.

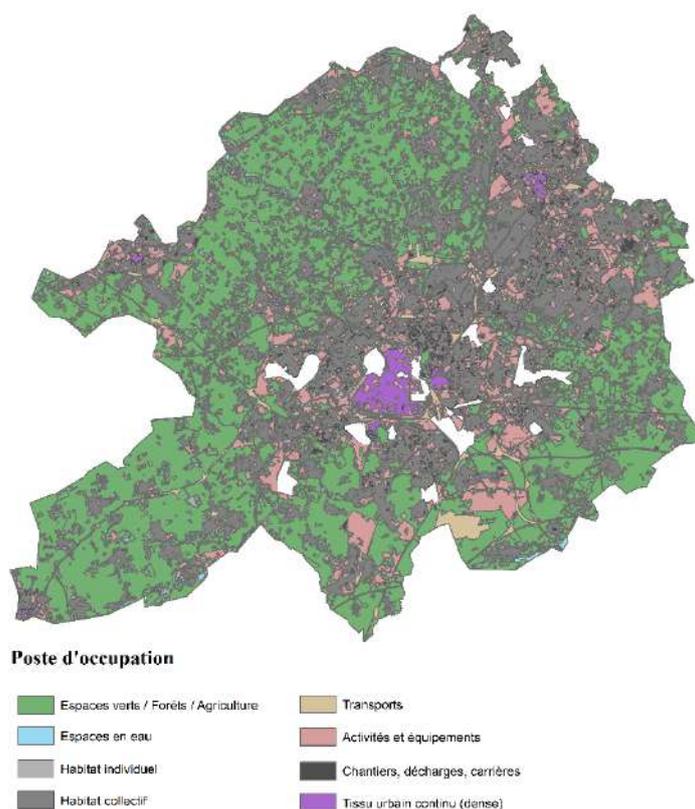


Figure 5 : Occupation du sol dans la MEL en 2009. Sources : Base de données SIGALE, Région Hauts-de-France

⁷ SIGALE : Système d’Information Géographique et Analyse de L’Environnement. Base de données créée en 1996 par la Direction du Développement Durable, de la Prospective et de l’Evaluation du Conseil Régional du Nord-Pas de Calais, puis des Hauts-de-France depuis 2015.

D'autre part, certains aspects du passé de la MEL sont toujours présents sur le territoire, notamment les logements d'ouvriers en briques rouges (figure 6) ou les industries fermées pour causes de délocalisations (figure 7), stigmates du passé industriel de ces communes de l'agglomération.



Figure 6 : Ancienne industrie métallurgique fermée, à Roubaix. Source : Antoine Longo (à gauche)

Figure 7 : Anciens logements ouvriers en briques rouges à Roubaix. Source : Antoine Longo (à droite)

La Métropole Européenne de Lille est donc un territoire diversifié et hétérogène, tant sur l'occupation de son sol et de ses activités que sur la vitesse de sa modernisation.

1.4 Description de l'échelle d'étude : l'IRIS

La MEL concentre un effectif de 90 communes. Comme le reste du territoire français, ces communes sont divisées en IRIS.

Les IRIS sont définis comme des Ilots Regroupés pour l'Information Statistique par l'INSEE (INSEE, 2016). Avec 2000 habitants en moyenne par secteurs, les IRIS représentent le découpage le plus fin et le plus précis disponible pour traiter des données du recensement de la population INSEE. Ce découpage a été créé en 1999. Il en existe trois différents types : les IRIS d'habitat (les plus fréquents, 92% du territoire français), les IRIS d'activités (qui regroupent au moins 1000 habitats et au minimum deux fois plus de salariés que de résidents) et les IRIS divers (qui regroupent les zones de faible densité d'habitat comme les zones agricoles ou forestières).

En ce qui concerne le territoire de la MEL, les 90 communes sont divisées en 504 IRIS. Le choix de réaliser des recherches à l'échelle des IRIS s'inscrit dans l'idée de réaliser une étude à l'échelle la plus fine possible. Les objectifs de l'étude concernant l'influence de facteur

environnementaux, et notamment les variations d'incidence suivant la pollution atmosphérique, une étude la plus fine possible permettre de se rapprocher au plus près de variations de ces facteurs. Comme évoqué lors de l'introduction, chaque personne ayant un cas de CI renseigne son adresse lors de son admission. Or, depuis 1978 et la loi relative à l'informatique, aux fichiers et aux libertés, la Commission nationale de l'informatique et des libertés (CNIL, créée pour garantir cette loi), les données de santé ne peuvent pas être détenues par d'autres personnes que le personnel soignant, lui-même devant garantir le secret professionnel. Les analyses à l'échelle de l'adresse étant impossible, l'échelle la plus fine se trouvait donc être l'IRIS, soit le quartier de résidence. La méthode d'attribution d'un quartier de résidence à un cas grâce à la connaissance de son adresse s'appelle un « géocodage », qui sera décrit plus tard dans ce mémoire (cf. Partie I, chapitre 2, 2.)

2. Revue de la littérature

L'objectif de ce travail de revue de la littérature est d'identifier le plus exhaustivement possible les travaux sur les déterminants des cas de cardiopathies ischémiques, dans le but de repérer les relations (ou l'absence de relation) entre les déterminants environnementaux et la survenue de ces cas.

2.1 Méthodologie de la revue de la littérature

Afin de produire une revue de la littérature la plus riche et complète possible, plusieurs ressources documentaires : *Google Scholar*, *Sciences Direct* ou encore le réseau social de chercheurs *Research Gate*, ont été utilisées avec les mots clés « *myocardial infarctus* », « *cardiovascular disease and environmental factor* » ou encore « *ischemic heart disease* ». Ces mots-clés ont aussi été cherché en français. La constitution de cette revue de la littérature s'est déroulée de la mi-mars jusqu'à mi-mai, puis une veille a été effectuée entre mi-mai et début juin. Au total, 14 articles ont été retenus. Parmi eux, 8 étudient des facteurs environnementaux tandis que les 6 autres s'intéressent soient aux facteurs individuels, soient à l'évolution des cas de l'étude MONICA depuis les années 90.

Le tableau 1 ci-dessous résume la description de l'ensemble des 8 documents discutant de facteurs environnementaux en précisant les caractéristiques de l'étude (nombre de sujets, pays, type d'étude et échelle d'analyse), la pathologie observée (maladie cardiovasculaire, cardiopathie ischémique, infarctus...) et les déterminants environnementaux étudiés.

Tableau 1 : Résumé de la revue de la littérature

N° article	Auteur	Date	Pays	Echelle	Nombre de sujets	Population	Nombre d'articles	Pathologie observée	Facteur observé
1	Carlos Mena et al.	2018	/	Multiples (Quartiers, Ville, Province/Comté, Région, Pays)	Revue de la littérature : pas de données disponibles	Revue de la littérature : pas de données disponibles	21 études	Maladies cardiovasculaires	Température Pollution de l'air Bruit Qualité de l'eau Type d'espace (rural/urbain) Défaveur sociale
2	Cai et al.	2016	Multiples (14 pays)	Région, Pays, Continent	« Plus de 4 millions de patients »	Pas de précision	33 études	Infarctus du myocarde	Pollution de l'air (PM10 et PM2,5)
3	Barry et al.	2015	Monde	Pays, continent	Revue de la littérature, pas de données disponibles	Revue de la littérature, pas de données disponibles	125 articles	Maladies cardiovasculaires Maladies coronariennes	Pollution de l'air
4	Stockfelt et al.	2017	Suède, ville de Göteborg	836 Small Areas for Market Statistics (SAMS, équivalent d'IRIS)	10 350 cohortes : 5850 + 4500	(2 25-75 ans	Bibliographie de 61 articles	Maladies cardiovasculaires	Pollution de l'air (PM10, PM2,5, Dioxyde d'azote, Noir de carbon ⁸)

⁸ Le Noir de Carbon, ou carbone suie, est un élément de la taille des PM2,5, émis par combustion de biomasse et d'énergies fossiles et qui fait partie de la classe des « forceurs climatiques » (Centre Interprofessionnel d'Etudes de la Pollution Atmosphérique, 2013)

N° article	Auteur	Date	Pays	Echelle	Nombre de sujets	Population	Nombre d'articles	Pathologie observée	Facteur observé
5	Vienneau et al.	2015	Multiples (8 pays)	Pas de précision	Méta-analyse : pas de précision	Tous les âges	11 études	Cardiopathies ischémiques	Bruit
6	Backholer et al.	2016	Multiples (4 continents)	Monde	Environ 22 millions	Entre 20 et 107 ans	44 articles	Maladies cardiovasculaires	Défaveur sociale
7	Ford & Highfield	2016	Etats-Unis, comté de Harris, Texas	649 census tracts (équivalent IRIS)	4,2 millions	5 classes d'âges entre 0 et plus de 65 ans.	Bibliographie de 59 articles	Maladies cardiovasculaires	Défaveur sociale
8	Bagheri et al.	2015	Australie, commune d'Adélaïde	148 Statistical Area One (SA1, équivalent Australien)	4 740 personnes	Entre 30 et 74 ans	Bibliographie de 30 articles	Maladies cardiovasculaires	Défaveur sociale

2.2 Principaux enseignements

La défaveur sociale est un facteur environnemental qui semble être de plus en plus à l'étude, puisqu'il est retrouvé dans 4 articles sur 8. La pollution atmosphérique est aussi présente comme un facteur d'étude dans 4 articles, ce qui l'ancre comme une piste d'étude importante puisqu'elle constitue l'un des axes de recherche avancé par l'équipe de recherche MONICA avant l'étude.

De plus, un facteur semble se dégager comme étant l'une des pistes des futures recherches à mener : il s'agit du bruit. Ce facteur sera abordé plus tard, lors de la discussion.

D'après la revue de la littérature et les objectifs de recherche de l'équipe d'étude de MONICA, les facteurs qui seront mis à l'étude seront donc la défaveur sociale (comme facteur socio-économique) et la pollution atmosphérique (comme facteur environnemental). Le lien entre ces facteurs et les variations spatiales d'incidence dans les IRIS de la MEL constitue le cadre de cette recherche.

Chapitre 2 : Cadre d'étude

1. Le registre MONICA Lillois

1.1 Description du registre

Le registre MONICA de Lille recense l'ensemble des événements aigües liés à des maladies cardiovasculaires subis par des habitants de la MEL depuis 1985. A l'origine, sa création avait pour objectif de caractériser les effets des facteurs individuels liés à la mortalité par MCV (pour rappel, il s'agit de la consommation de tabac et d'alcool, d'hyperglycémie...). Il constitue une base de données précieuse puisque chaque cas est caractérisé suivant plusieurs variables. Les dates et les adresses des patients ayant subi une attaque liée à une MCV sont aussi renseignés afin de réaliser des études de variations spatiales et temporelles d'incidence.

Le registre MONICA de Lille compte 15 variables décrivant les caractéristiques du patient ayant eu un cas de CI. Chaque ligne du registre correspond à un cas de CI. Un patient peut avoir été victime de plusieurs accidents, et par conséquent se retrouver plusieurs fois dans le registre.

Chaque cas se voit attribuer un identifiant numérique appelé « IDNUM » dans le registre. Ces identifiants sont donc uniques.

Chaque patient se voit attribuer un identifiant appelé « ID_SUJET » dans le registre.

Les quartiers d'habitation qui ont été géolocalisés grâce aux adresses ont un identifiant spatial appelé « SP_ID », soit le code que l'INSEE attribue à chaque IRIS français.

L'âge et le sexe du « cas » est inscrit dans le registre. Puis la présence d'antécédent de cardiopathie est aussi inscrite : « 1 » si le cas a déjà été victime d'une attaque, « 2 » dans le cas contraire, « 9 » pour les cas non-renseignés.

Le jour, le mois et l'année de l'épisode de CI est inscrit dans le registre. La survie après 28 jours est aussi notée (« 1 » en cas de survie, « 2 » en cas de décès, « 9 » pour les cas non-renseignés) tout comme le nombre de jours, d'heures et de minutes de survie après un événement. De plus, la survie estimée après un cas constitue aussi une variable : 1 pour une survie inférieure à 1h (à partir du début des symptômes), 2 pour une survie de 1 à 24h, 3 pour une survie supérieure à 24h et où l'on connaît le nombre de jours de survie, 4 pour une survie inférieure à 24h, 5 pour une survie supérieure à 24h mais on ne connaît pas le nombre de jours de survie (car la date de début de l'événement est inconnue), 8 lorsque le cas est vivant au 28ème jour et 9 pour les cas non-renseignés.

Enfin, les différentes catégories d'épisodes sont inscrites dans le registre. Il en existe huit notées ainsi : 1 pour les infarctus du myocarde (vivant ou décédé), 2 pour les décès coronaire (décès avec des symptômes et/ou des antécédents coronaires non diagnostiqués en catégorie 1, 5, 6 ou 7), 3 pour les morts subites en moins de 24h sans cause identifiée, 4 pour les sujets décédés pour une raison autre que d'origine coronaire, 5 pour les syndromes coronariens aigus (vivant ou décédé), 6 pour les angors instables (vivant ou décédé), 7 pour les autres épisodes d'insuffisance coronaire aigue (vivant ou décédé), et 9 pour les sujets décédés d'une cause inconnue.

Néanmoins, pour l'étude MONICA de Lille, tous les cas n'ont pas été conservés. L'étude a mis en place des critères d'inclusion, afin de cibler une certaine partie de la population : celle qui est considérée le plus « à risque » dans la MEL.

1.2 Critères d'inclusion

Tout d'abord, ne sont conservés que les cas ayant été répertoriés entre le 1^{er} janvier 2008 et le 31 décembre 2011, dates fixées par l'équipe de recherche.

Ensuite, seuls les cas âgés de 35 à 74 ans sont intégrés dans le registre. Ici, l'objectif est de réunir les cas par tranche d'âge (35-44 ans, 45-59 ans, 60-74 ans) et que celles-ci soient pertinentes d'après la revue de la littérature (Tunstall-Pedoe et *al.*, 1994 ; Richard, 1988).

Puis, tous les cas ayant eu une attaque telles qu'un infarctus du myocarde, un décès coronaire, une mort subite en moins de 24H, un syndrome coronaire aigu (vivant ou décédé), un angor instable (vivant ou décédé) ou tout autre forme d'insuffisance coronaire aiguë (vivant ou décédé) a été inscrit dans la base. Sont donc exclus, les catégories de cas 4 et 9.

De plus, n'ont été conservés que les cas n'ayant pas d'antécédents de cardiopathie ischémique avant leur événement. L'objectif étant de travailler sur une base d'incidence, seuls les premiers cas ont de l'intérêt pour l'étude.

Enfin, parmi les cas, seules les personnes résidant dans la Métropole lilloise sont intégrées. Ce qui exclut les touristes, mais aussi les résidant ayant eu une attaque hors de la MEL. De plus, il arrive que certains cas se localisent dans des IRIS que l'INSEE considère comme non-peuplés. Les cas se trouvant dans ces IRIS sont donc exclus.

Afin donc de connaître l'IRIS d'habitation des cas pour savoir s'il peut être inclus dans le registre ou non, un géocodage a été effectué. Cette étape va ainsi pouvoir permettre de mettre en application les critères d'inclusion de connaître le nombre final de cas incidents sur lequel l'étude va porter.

2. Le géocodage

2.1 Nécessité

Lors du début du stage, les cas retenus pour les analyses étaient les cas incidents sur deux années, entre le 1^{er} janvier 2008 et le 31 décembre 2009. Après un géocodage réalisé préalablement aux débuts des analyses que ce stage devait permettre d'effectuer, le nombre de cas étaient de 2 727. Un tri sur la présence d'antécédents ou non a ensuite pu permettre de constater une base d'incidence de 1694 cas, soit 1694 personnes ayant contracté une cardiopathie ischémique au cours de ces deux années.

Cependant, au cours de ces analyses, il a été constaté que certains résultats n'étaient pas significatifs mais que leurs *p-value* étaient proches de la significativité⁹, ce qui n'indique pas forcément une invalidité de l'hypothèse, mais plutôt un manque de puissance statistique. Aussi, après concertation avec les principaux chercheurs impliqués dans cette étude, il a été décidé d'inclure deux années d'incidence supplémentaires afin de d'augmenter la puissance statistique. L'idée ayant été validée, les cas datant du 1^{er} janvier 2010 au 31 décembre 2011 ont été ajoutés aux analyses de départ.

2.2 Méthode

L'exercice de géocodage constitue le fait d'attribuer des coordonnées spatiales (X,Y ou latitude et longitude, ce qu'on appelle un système de coordonnées) à des points dans un espace. Travaillant sur une base d'adresses, le but était de connaître à quel IRIS appartenait chaque adresse, afin de réaliser dans un premier temps des analyses statistiques, mais aussi dans un souci d'éthique et de conservation d'anonymat des cas dont les adresses étaient entre les mains des chercheurs de l'étude, comme expliqué précédemment.

Les adresses de cas de ces deux dernières années d'études n'étaient pas géocodées à l'IRIS, ce qu'il a donc fallu effectuer avant de réaliser quelque analyse statistique ou spatiale. Mais ce géocodage connaissait une contrainte : ne pas utiliser internet. Comme expliqué précédemment, l'anonymat des personnes ayant eu cas de CI lors de ces deux années ne devaient en aucun cas avoir le risque d'être diffusé ou rendu public. Le géocodage sans accès à internet a donc été recommandé (Annexe 1). Pour ce faire, le choix a été de télécharger deux bases réunissant la quasi-totalité des adresses françaises : la Base d'Adresses Nationales (BAN) et la BD Adresses de l'IGN.

Dans ces deux bases de données, les adresses sont rangées par numéro de voie, type de voie (rue, allée, boulevard...), nom de voie, code postal et commune. Pour chaque adresse, les coordonnées spatiales sont connues. Le choix d'utiliser les deux bases étaient d'abord par prévention, car certaines adresses auraient pu se trouver au sein d'une des bases et pas dans l'autre.

⁹ Lors d'un test de significativité, la p-value est la probabilité d'observer une valeur de la statistique de test aussi ou plus extrême que la valeur observée sur l'échantillon de base. Une p-value de valeur inférieure ou égale à 0,05 indique une probabilité inférieure ou égale à 5% d'observer un résultat égal ou plus extrême.

Une fois ces deux bases acquises, il a fallu réaliser le géocodage à l'Institut Pasteur, sur l'un des ordinateurs fournis par l'équipe en charge de l'étude car il était sécurisé (pas de connexion internet). La base MONICA des années 2010-2011 comprenait 2992 adresses. Mais avant de commencer à attribuer des coordonnées spatiales à chacune de ces adresses, il a fallu « nettoyer » la base MONICA afin que les adresses qui étaient inscrites dans le registre correspondent exactement à la façon dont étaient écrites celles des bases nationales. La combinaison était ainsi : Numéro de voie, type de voie, nom de la voie, commune. Toutes les adresses étaient écrites en majuscule et sans accent. Le détail de cet exercice est rédigé en annexe 1.

Les coordonnées spatiales qui étaient recherchées dans ces bases de données correspondaient à un système de coordonnées classique appelé WGS 1984, basant sa latitude sur l'équateur et sa longitude sur le Méridien de Greenwich. Mais ces données ont ensuite été converties vers un système de coordonnées français appelé « Lambert 93 ». Différent d'un système basé sur des points de longitude et de latitude, le système de coordonnées « Lambert 93 » est organisé selon un repère dit « cartésien ». Il prend pour base de latitude le parallèle 44°N et pour base de longitude le Méridien de Greenwich pour base de longitude.

La fonction « RECHERCHEV » de l'outil Excel a permis de faire correspondre les coordonnées spatiales des adresses des bases nationales à celles de la base MONICA. Cette technique a permis de géocoder 2 196 cas sur les 2 992, soit environ 73%. Les 796 adresses n'ayant pas trouvé de correspondance ont donc dû être géocodées manuellement, c'est-à-dire en utilisant un autre ordinateur, cette-fois-ci connecté à internet, et tapées dans l'outil *Google Maps* afin de récupérer les coordonnées spatiales. Cette deuxième étape a pris environ 14h de travail et a permis de géocoder 761 adresses sur les 796 restantes. Les 35 dernières adresses ont été soumises à une vérification, qui a nécessité d'aller chercher les dossiers d'admissions des patients lors de leur événement de cardiopathie. Après vérification, seulement 7 adresses se sont retrouvées sans coordonnées spatiales : 5 patients sans domicile fixe, une « adresse inconnue » et un patient dont l'hypothèse est que son adresse n'a pas dû être bien comprise par le professionnel de santé qui l'a pris en charge lors de la rédaction de son dossier. Sur 2992, 2985 ont pu être géocodées. Puis, il a fallu exclure les cas ayant un diagnostic qui ne seraient pas utiles pour des études futures : la catégorie 4 (i.e. les sujets décédés pour une raison autre que d'origine coronaire, c.f figure 8) et la catégorie 9 (i.e. les sujets décédés d'une cause inconnue). Finalement, le registre d'étude MONICA, après

géocodage et jointures des cas des années 2008/2009 et 2010/2011, comptait 5448 cas entre le 1^{er} janvier 2008 et le 31 décembre 2011.

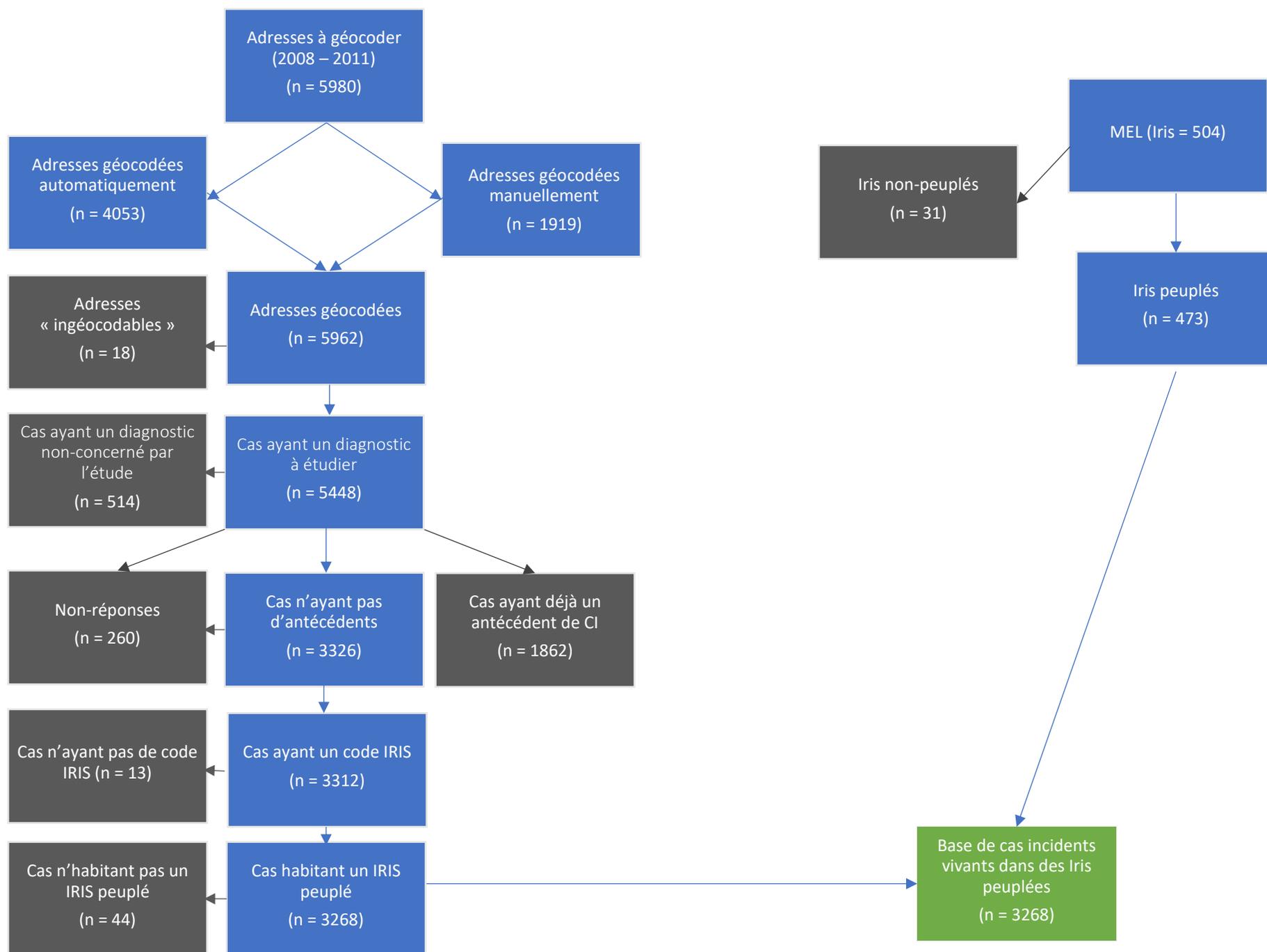
Le registre MONICA géocodé, sans l'ensemble des critères d'inclusion, comprend donc 5448 cas. Néanmoins, comme indiqué précédemment, seules les lignes comprenant des cas n'ayant pas d'antécédent ont été conservées. La base des cas incidents se réduit alors à 3 326 cas.

En ce qui concerne la sélection des IRIS, la base de travail était au départ les 504 IRIS qui composent la Métropole Européenne de Lille. Un premier tri a été effectué avec les données issues de la base « Population en 2009 » de l'INSEE afin de retirer les IRIS non-peuplés, soit 31 IRIS, ce qui porte le total d'IRIS d'étude à 473. Grâce à ce tri, la base des cas incidents a pu être modifiée une dernière fois, où seuls les cas localisés dans des IRIS « peuplés » ont été conservés, soit finalement 3 268 cas.

L'étude MONICA a donc utilisé les données de 3 268 patients localisés dans 473 IRIS de la Métropole Lilloise.

L'organigramme ci-dessous résume le processus d'inclusion/exclusion des cas du registre et des IRIS pour l'étude.

Figure 8 : Organigramme du processus d'inclusion des cas et des IRIS dans l'étude MONICA



Chapitre 3 : Outils de caractérisation

1. Bases de données INSEE

1.1 Critère de jugement principal

Le critère de jugement principal est la variable qui va aider à déterminer l'influence des facteurs étudiés sur la variabilité d'un phénomène. En ce qui concerne le registre MONICA, l'incidence est le critère de jugement principal car l'objectif est de savoir si celle-ci est significativement influencée par des facteurs environnementaux, ici la défaveur sociale et la pollution atmosphérique. Cette incidence peut se traduire de plusieurs manières. En effet, l'incidence est un indicateur qui peut être discret (nombre de nouveaux cas sur une période) mais peut aussi désigner une proportion : on parle alors de taux d'incidence. Celui-ci représente un rapport entre le nombre de nouveaux cas sur une période et la population « à risque » sur cette même période. Pour le cas de l'étude MONICA, le taux d'incidence sera mesuré grâce au nombre de cas du registre sans antécédents (3 268 cas) rapporté sur la population « à risque », c'est-à-dire ici le nombre d'habitants de la MEL âgés entre 35 et 74 ans. Pour connaître la taille de cette population, des bases de données INSEE ont été sélectionnées.

1.2 Les bases de données

Afin d'obtenir le nombre d'habitants dans la Métropole, les bases de données utilisées ont été téléchargé sur le site internet de l'INSEE. Néanmoins, la base la plus commune « Population en [Année recherchée] » n'avait pas les mêmes tranches d'âges que celles que la registre MONICA propose. Tandis que le registre a effectué un découpage en 3 tranches d'âges (35-44 ans, 45-59 ans, 60-74 ans), la base de l'INSEE disposait d'un découpage plus large pour la première tranche, la limite basse étant de 30 ans. Or pour ne pas fausser les résultats, une autre base a dû être spécialement demandée auprès de l'INSEE. Celle-ci possédait un découpage particulier qui indiquait l'effectif de chaque personne par âge et par IRIS, ce qui a permis de pouvoir réorganiser les tranches d'âge pour correspondre avec les données du registre.

De plus la base sélectionnée est celle datée de 2009. L'hypothèse a été faite que la population de la MEL entre 2008 et 2011 n'avait pas assez évoluée pour altérer les résultats. Pour vérifier cette hypothèse, un calcul du solde migratoire a été effectué (« *Le solde migratoire est la*

différence entre le nombre de personnes qui sont entrées sur le territoire et le nombre de personnes qui en sont sorties au cours de l'année. » INSEE, 2016). Le résultat de celui-ci a montré une augmentation d'environ 1% de population, avec une augmentation annuelle de 0,3% entre 2009 et 2010. Ces données étant calculées sur la population totale et ces résultats montrant une quasi-stabilité de la population, le choix a été fait de conserver uniquement la base de population en 2009 et de considérer la population de la MEL comme constante. De plus, à titre de comparaison, le solde migratoire français correspond à une augmentation annuelle d'environ 0,7% de sa population, ce qui renforce la décision de principe de stabilité de la population (INSEE, 2015). De surcroît, d'autres bases ont été téléchargées sur la plateforme internet de l'INSEE. Mais ces données, issues de recensements, ont cette fois pour vocation d'aider à mesurer la défaveur sociale. Pour ce faire, les bases d'Activités, de Logements, de Couple-Famille-Ménage et de Diplômes ont été extraites.

2. Outils de mesure de la défaveur sociale : les indices

2.1 Description des indices

Ces données de population n'ont pas seulement pour objectif de servir à calculer l'incidence. En effet, celles-ci seront utilisées aussi pour mesurer la défaveur sociale. En effet, la plupart des calculs pour mesurer cette défaveur nécessite de connaître l'effectif de population à l'intérieur des unités spatiales d'études (ici, les IRIS).

Pour mesurer la défaveur, il existe des outils qui permettent de rassembler un certain nombre de variables qui peuvent rendre compte d'une certaine défaveur. Ces outils s'appellent des « indices ». Ces indices sont « composites », c'est-à-dire qu'ils sont calculés à partir de plusieurs variables. Chaque indice à sa méthode de calcul, mais le résultat est toujours une valeur unique qui traduit une tendance ou non à la défaveur. Tous les indices se traduisent différemment.

En France, trois principaux indices se dégagent de la littérature : le *French Deprivation Index* (FDep), le *French European Deprivation Index* (EDI) et le *French Social Deprivation Index*.

Ces indices constituent une mesure écologique (c'est-à-dire qu'ils caractérisent une population dans son ensemble et non individuellement) sur une période donnée. La valeur de cet indice est donc distribuée sur l'ensemble de l'unité spatiale étudiée. Ils sont construits grâce à des variables individuelles agrégées à l'unité spatiale d'étude, et sont donc disponibles grâce à des données de recensement (taux de chômeurs, pourcentage d'ouvriers au sein de la

population active...), notamment les données INSEE récoltées plus tôt (Rapport de la DREES, 2017).

Le tableau 2 résume la construction de ces différents indices français.

Tableau 2 : Résumé des indices de défaveur sociale en France

Indice	Auteur et année	Variables	Source Variables
French Deprivation Index (FDep)	Grégoire Rey, 2009	4 variables : <ul style="list-style-type: none"> • Le revenu fiscal médian par unité de consommation ; • La part des diplômés de niveau baccalauréat (au minimum) dans la population de 15 ans ou plus non scolarisée ; • La part des ouvriers dans la population active de 15 à 64 ans ; • La part des chômeurs dans la population active de 15 à 64 ans. 	INSEE
French Ecological Deprivation Index (EDI)	Carole Pornet, 2012	Score calculé sur 10 variables = <ul style="list-style-type: none"> • 0.11 x "Surpeuplement" + • 0.34 x "Pas d'accès au chauffage central ou électrique" + • 0.55 x "Non propriétaires" + • 0.47 x "Chômage" + • 0.23 x "Nationalité étrangère" + • 0.52 x "Pas d'accès à une voiture" + • 0.37 x "Ouvriers non qualifiés-ouvriers agricoles" + • 0.45 x "Ménages avec au moins 6 personnes" + • 0.19 x "Faible niveau d'étude" + • 0.41 x "Familles monoparentales". 	Enquête Individuelle Européenne de défavorisation – EU-SILC + Recensement de la population (INSEE).

Indice	Auteur et année	Variable	Source Variable
French Social Deprivation Index	Sabrina Havard, 2008	19 variables <ul style="list-style-type: none"> • Ouvriers dans la main-d'oeuvre • Résidences primaires qui sont des maisons ou des fermes • Résidences primaires qui sont des logements collectifs • Ménages sans voiture • Ménages avec deux voitures ou plus • Ouvriers avec un travail précaire • Les personnes âgées de 15 ans ou plus ayant une licence ou diplôme professionnelle • Les personnes âgées de 15 ans ou plus avec au moins un faible niveau d'éducation dans le supérieur • Les personnes âgées de 15 ans ou plus qui ne sont pas allé au-delà de l'école primaire • Locataire de sa résidence primaire • Ouvriers avec un emploi stable • Logement subventionné parmi les résidences principales • Familles monoparentales • Revenu médian par unité de consommation (en euros/an) • Résidence primaire avec plus d'une personne par chambre • Etrangers dans la population totale • Nombre moyen de personne par chambre • Taux de chômage chez les ouvriers • Ouvriers au chômage depuis plus d'un an 	INSEE

Pour l'étude MONICA, un seul indicateur a dû être choisi pour déterminer l'importance de la défaveur sociale sur le territoire de la MEL. Un choix a donc dû être fait. Mais avant de commencer les comparaisons, un premier indicateur a été supprimé des options. Il s'agit du *French Social Deprivation Index* car le nombre de variables était trop important (19). Ce nombre important de variables peut potentiellement constituer un frein si l'étude doit être transposée dans d'autres pays, car tous les pays n'ont pas forcément les mêmes accès aux données. De plus, le French EDI ainsi que le Fdep ont aussi la légitimité de retranscrire une défaveur sociale proche de la réalité¹⁰, chacun s'étant basé soit sur des recueils de données concernant la santé perçue et la pauvreté objective (Pornet et al., 2012) ou bien sur une analyse sur l'association spatiale entre désavantage social et mortalité étudiée selon la tranche d'unité urbaine (Rey et al., 2013). Le choix s'est donc effectué entre le French EDI et le Fdep.

2.2 Le French EDI

Le French EDI est l'acronyme de *French European Deprivation Index*. Cet indice a été créé par une équipe de recherche autour de Carole Pornet, chercheuse en épidémiologie et santé publique à l'Université de Caen. L'objectif de base était de réaliser un indice qui était capable d'être transposé à l'échelle européenne, donc à toutes les nations de l'Union Européenne (UE). Pour ce faire, les échantillons sélectionnés pour construire l'indicateur n'étaient pas uniquement des données de recensement nationaux : ils étaient complétés par des données issues d'une enquête transversale européenne visant à mesurer la pauvreté au sein de l'UE : l'EU-SILC (European Union – Statistics on Income and Living Condition)¹¹. Cette enquête a notamment permis de créer un indice européen de défaveur social appelé *Individual Deprivation Indicator* qui sert essentiellement de base pour construire des indices à l'échelle nationale. Le French EDI est donc un indice issu d'une construction européenne, et chaque pays l'a donc adapté à son territoire en y rajoutant des données de recensement nationaux (en France, grâce à l'INSEE). L'équipe de construction du French EDI s'est servi des données françaises issues de l'enquête EU-SILC (Pornet, 2012). Cette enquête a été menée par

¹⁰ « Proche » car si ces indices prennent en compte des variables de recensement, ils ne tiennent pas compte de l'aspect « territorial » de l'unité spatiale, donc il n'y a aucune donnée sur l'enclavement des territoires, le temps de transport jusqu'à un service de santé, l'insalubrité des logements...

¹¹ En français : Union Européenne - Statistiques sur les Revenus et les Conditions de Vie.

Eurostat¹². Les données françaises issues de cette enquête sont datées de 2006 et concernent 10 036 ménages, 24 940 individus dont 19 253 âgés de 16 ans et plus (Pornet et *al.*, 2013).

Les 10 variables constituant l'indice ont été sélectionnées suivant une méthode consistant à mesurer la « pauvreté perçue » par les enquêtés. Une première partie des variables nommées « besoins fondamentaux » composée de 19 variables a donc été mise à l'épreuve, comme par exemple « posséder un réfrigérateur » ou bien « posséder une voiture » et les participants avaient six choix de réponse possible : 'Vous êtes vraiment à l'aise' ; 'Vous êtes plutôt à l'aise' ; 'ça va' ; 'C'est juste, il faut faire attention' ; 'Vous y arrivez difficilement' ; 'Vous ne pouvez pas y arriver sans faire de dettes'. Les données de défaveur individuelles ont été ajoutées par la suite en tant que « pauvreté objective ».

Dix variables ont donc été sélectionnées parmi celles qui étaient les plus représentatives à la fois de la pauvreté objective et de la pauvreté subjective, ou « perçue ».

Une régression logistique multivariée a permis de déterminer l'influence de ces 10 variables sur la variation de l'*Individual Deprivation Indicator* (IDI), qui, pour rappel, est issu de l'enquête SILC aussi composée de variables mesurant la « pauvreté objective » et la « pauvreté perçue ». Le coefficient de régression associé à chaque variable indique comment celles-ci vont influencer la variation de l'IDI. C'est ce coefficient qui servira de pondération à chaque variable dans le calcul du score du French EDI.

La valeur du French EDI sur une unité spatiale est comprise entre 0 et $+\infty$. Plus la valeur de l'indice est proche de 0, moins l'unité spatiale est défavorisée, et inversement.

2.3 Le Fdep

Le Fdep est l'acronyme du terme « French Deprivation » (« défaveur française » en français). Sa construction est le fruit du travail d'une équipe de recherche menée par Grégoire Rey, biostatisticien et épidémiologiste au Cépidec (Centre d'épidémiologie sur les causes médicales de décès). L'objectif de ce travail était de construire un indice capable à la fois de mesurer la défaveur sociale à l'échelle des IRIS, mais que ces données informent aussi sur la qualité des conditions socioéconomiques et résidentielles des populations observées (Temam et *al.*, 2017). Sa construction a été réalisée suivant une méthode statistique appelée

¹² Bureau des Statistiques de la Commission Européenne

« Analyse en Composante Principale » (ACP). Cet outil statistique permet de réduire la dimension d'un jeu de données caractérisé par un ensemble de variables par la construction de combinaisons linéaires des variables (appelées composantes principales) maximisant la variance expliquée. Pour ce faire, plusieurs variables ont dû être choisies en fonction de leur capacité à représenter la défaveur sociale, de leur corrélation avec le taux de mortalité dans chaque unité spatiale d'observation et leur corrélation avec d'autres variables du même profil¹³. Ainsi, 4 variables ont été choisies comme les plus représentatives selon l'équipe de recherche. Ces 4 variables sont issues des données de recensement de l'INSEE. Il s'agit du pourcentage d'ouvriers dans la population active, du pourcentage de bacheliers chez les 15 ans et plus, du pourcentage de chômeurs dans la population active et du revenu médian par foyer.

La valeur du Fdep est comprise entre $-\infty$ et $+\infty$ avec 0 comme valeur neutre. Plus la valeur de l'indice est supérieure à 0, plus l'unité spatiale est défavorisée, et inversement.

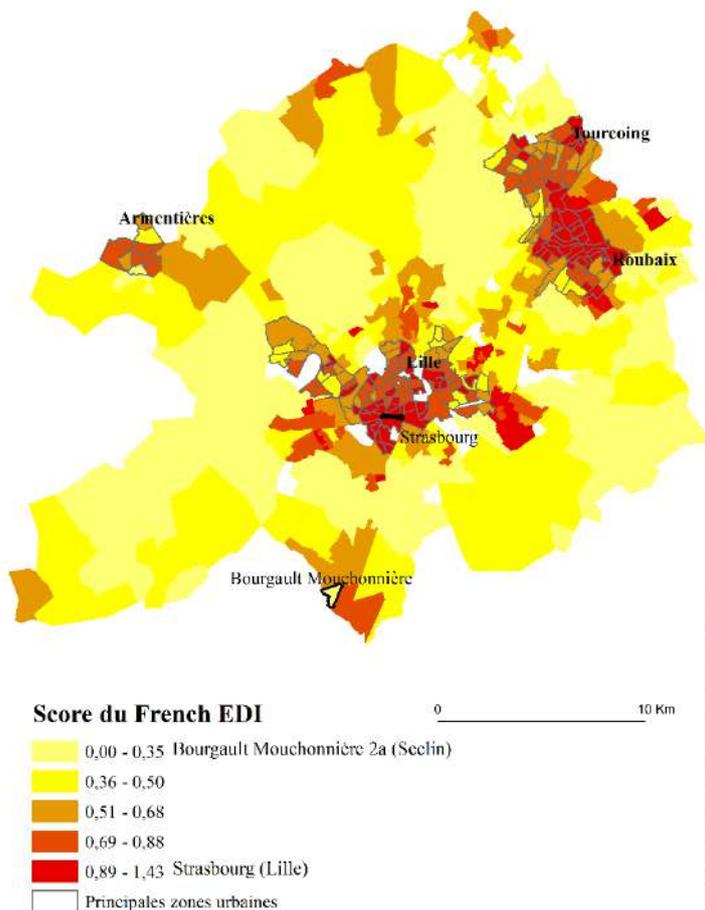
2.4 Le choix de l'indice

Pour choisir l'indice qui allait être utilisé pour représenter la défaveur sociale dans l'étude MONICA de Lille, plusieurs vérifications ont eu lieu. Premièrement, les valeurs du Fdep et du French EDI ont été calculées pour chaque IRIS. Puis un coefficient de corrélation de Pearson a été effectué afin de vérifier la corrélation entre les deux indicateurs pour être sûr que l'un et l'autre montraient sensiblement les mêmes variations spatiales de la défaveur sociale. Cet indicateur est compris entre -1 et 1. Plus il est proche de 1, plus les variables sont corrélées positivement, et plus il se rapproche de -1, plus les variables sont corrélées négativement. Enfin, plus le coefficient est proche de 0, moins les variables sont corrélées. Le coefficient de corrélation entre le Fdep et le French EDI sur la MEL est de 0,699, ce qui peut donc être considéré comme deux variables corrélées positivement. La significativité du coefficient de corrélation a été évaluée au moyen d'un test de nullité du coefficient. Dans le cas, présent, la corrélation entre les deux indices s'est avérée significative ($p = <0,001$).

Ces mêmes valeurs par IRIS ont ensuite été intégrées dans un Système d'Information Géographique (SIG) afin de constater si les IRIS de la MEL considérés comme défavorisés par un indicateur l'étaient aussi pour l'autre indicateur.

¹³ Indicateurs de suivi de l'évolution des inégalités sociales de santé dans les systèmes d'information en santé, Haut Conseil de la Santé Publique, 2013

Répartition de la défaveur sociale D'après le French EDI, dans le MEL en 2009



Répartition de la défaveur sociale d'après le Fdep, dans la MEL en 2009

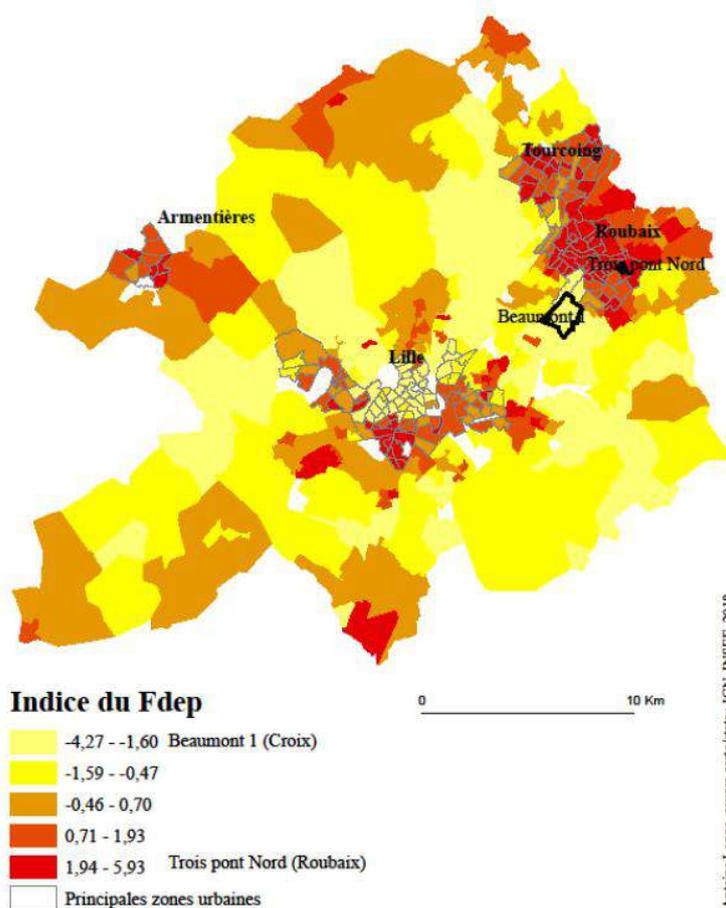


Figure 10 : Cartographie de la défaveur sociale suivant les valeurs du French EDI (gauche)
Figure 9 : Cartographie de la défaveur sociale suivant les valeurs du Fdep (droite).

D'après ces deux cartes, plusieurs constats sont à faire. Tout d'abord, sur la carte du French EDI, la majorité des zones urbaines ont été mises en valeur ; les zones périurbaines et rurales appartenant majoritairement aux classes les moins défavorisées. Les zones urbaines, quant à elles, montrent une appartenance aux mêmes classes les plus défavorisées, notamment la ville de Lille intra-muros qui semble entièrement défavorisée.

En ce qui concerne le Fdep, celui-ci montre plus d'hétérogénéité, que ce soit en milieu périurbain ou en zone urbaine. En effet, il semblerait que quelques communes du Sud de la MEL soient dans une situation plus défavorable que ce que montrait la carte du French EDI. De plus, si la conurbation Tourcoing-Roubaix reste parmi les communes ayant les IRIS les plus défavorisées de la MEL, les quartiers de Lille semblent eux moins défavorisés qu'avec la

vision du French EDI. Cela s'explique par le fait que les variables du French EDI, notamment celles définies comme les variables de « pauvreté perçue », ne reflète pas forcément une pauvreté réelle, surtout en milieu urbain. En effet, si ne pas « posséder une voiture » est considéré comme le deuxième coefficient le plus important de démonstration d'une potentielle défaveur dans l'indicateur, il n'en n'est pas réellement témoin dans une ville capitale de la quatrième métropole de France puisque les transports en communs sont multiples (métro, bus, tramway, vélo libre-service, 600 km de bande cyclables, deux gares européennes, un aéroport européen). De plus, la variable ayant le plus gros coefficient (donc qui influence le plus la valeur finale de l'indice de défaveur) est le fait d'être « propriétaire de son logement ». Or, si les zones péri-urbaines et rurales connaissent plus de propriétaires, le nombre de locataires en zone urbaine dense est élevé, notamment à cause du fait que ces zones sont emplies de logements collectifs et qu'il est plus commun de louer un appartement plutôt que de l'acheter. La ville de Lille est l'une des communes françaises connaissant le plus grand nombre d'étudiants à l'année avec 11% de sa population qui est étudiante. Cette population est donc plutôt encline à louer un appartement plutôt qu'à l'acheter et peut constituer l'une des explications au fait qu'en milieu urbain le nombre de propriétaires est plus faible qu'en milieu périurbain ou rural, ce qui, de fait, montre une potentielle défaveur plus importante à Lille par le French EDI que par le Fdep.

En conclusion, le choix s'est donc porté sur le Fdep qui comprend des variables plus représentatives de la défaveur dans un milieu qui connaît des zones d'urbanisation dense et qui semble plus révélateur d'une réelle défaveur sociale et économique. Autre atout, moins discriminant lors du choix mais néanmoins avantageux : il ne possède que 4 variables, toutes disponibles sur la plateforme de l'INSEE, ce qui est à la fois plus rapide à obtenir et moins contraignant.

3. Mesurer la pollution atmosphérique

L'étude MONICA au sein de la MEL a pour objectif d'étudier les variations spatiales d'incidences liées aux facteurs environnementaux. Avec la défaveur sociale, la pollution atmosphérique fait partie de ces facteurs environnementaux dont l'influence est à observer. En collaboration avec une équipe de chercheurs en environnement travaillant sur le projet SIGLES (Système d'Information Géographique et Liens Environnement et Santé), la répartition spatiale et le niveau de présence dans l'air de 19 polluants ont été analysés, mais

aussi sur 3 autres indicateurs de pollution : le RIM (indice multimétallique), l'eutrophie et le Score Environnemental.

Le principe de ces récoltes de polluants repose sur l'idée d'analyser un cumul d'exposition aux polluants des populations. Cette méthode correspond à un projet d'analyse sur l'exposition au long terme.

3.1 Le choix des polluants

Les polluants qui ont été récoltés et analysés sont issues d'un choix réalisé grâce à une revue de la littérature en amont de la recherche SIGLES. Ces polluants sont : l'aluminium, l'arsenic, le benzène, le cadmium, le monoxyde de carbone, le chrome, le cuivre, le mercure, le manganèse, le nickel, le plomb, le palladium, l'antimoine, le titane, le vanadium, le zinc et le dioxyde d'azote. À cela se rajoute une dénomination appelée « poussières » qui réunit les particules fines. Ces particules ont été récoltées sur des feuilles de peupliers qui ont ensuite été mélangées dans un produit aqueux pour déterminer une réaction montrant leur potentielle présence. Le mélange dans ce liquide a donc rassemblé les particules, rendant impossible l'observation sur leur taille (confusions entre les particules fines, les PM10, PM2,5, PM1¹⁴...).

3.2 Méthode de captation

La captation de ces polluants s'est effectuée grâce à deux méthodes : la surveillance physico-chimique et la biosurveillance.

La surveillance physico-chimique désigne une méthode qui consiste à déterminer la concentration en polluants dans un milieu. Pour effectuer ce type d'analyse, des institutions publiques organisent des observations via des capteurs de pollutions disposés sur le territoire étudié. Il s'agit notamment de l'organisme Atmo¹⁵ qui va réaliser ces prélèvements grâce à des capteurs.

¹⁴ PM10 : « Particulate Matter » ou « Matières Particulaires » en français désignent les particules en suspension. Le chiffre qui suit désigne leur taille en micromètres, notée 10µm.

¹⁵ Atmo : Réseau National des Associations Agréées pour la Surveillance de la Qualité de l'Air. Fédération d'association en charge de surveiller la qualité de l'air à l'échelle des régions sur le territoire français.



Figure 11 : Capteurs de surveillance physico-chimique.
Photos : Florent Occeci

L'organisme Atmo a ensuite publié plusieurs cartes ainsi que plusieurs données concernant la proportion de polluants de l'air (l'unité de ces polluants est en microgramme/mètre cube, notée $\mu\text{g}/\text{m}^3$). Ces données ont été ensuite utilisées pour réaliser un modèle d'interpolation spatiale.

L'interpolation spatiale désigne le principe d'attribuer une donnée à un point d'après les données connues des points voisins. Pour cette étude, cela veut dire qu'étant donné que les capteurs étaient situés à plusieurs points de la MEL, une interpolation a été réalisée pour donner une valeur de concentration de pollution à tous les endroits où il n'y avait pas de capteurs. Le calcul s'est effectué en faisant une moyenne pondérée des valeurs enregistrées par les capteurs, où la pondération était la distance entre la zone à calculer et les capteurs alentours. Le schéma ci-dessous résume cette explication.

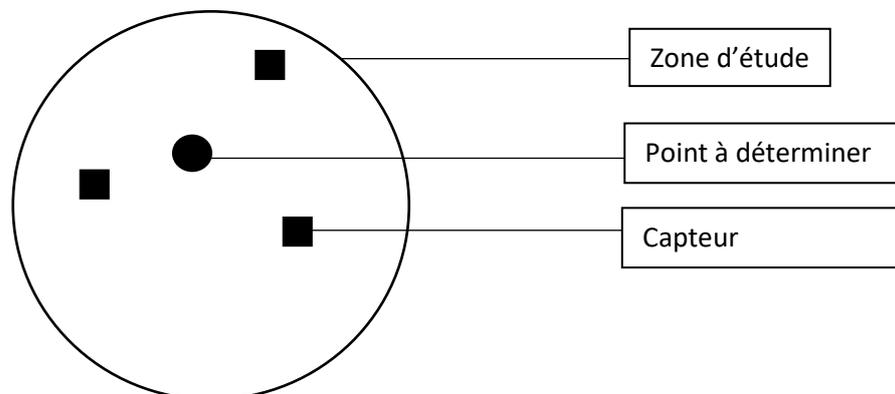


Figure 12 : Schéma d'une interpolation spatiale

*Données du point à déterminer = données du capteurs * distance entre le capteur et le point à déterminer.*

Mais au sein de l'équipe de recherche SIGLES, la méthode de captation et d'analyse des polluants n'a pas été exclusivement réalisée par surveillance physico-chimique. Une autre méthode a été utilisée : la biosurveillance.

Le mot « biosurveillance » est construit du préfixe « bio » désignant « la vie » en grec. Cette méthode est le fait d'observer l'impact d'un perturbateur, ici la pollution, sur le vivant (Santé Publique France, 2016). En d'autres termes, cela indique que la surveillance physico-chimique sert à mesurer la concentration d'un polluant pendant que la biosurveillance est utile pour connaître ses effets sur le vivant.

Ce qui est désigné comme « le vivant » concerne notamment des êtres végétaux et fongiques.



Figure 13 : Schéma des espèces végétales et fongiques utilisés pour de la biosurveillance. Schéma : Association pour la Prévention de la Pollution Atmosphérique

Pour l'étude SIGLES, une récolte des lichens a été effectuée sur 500 arbres répartis dans la MEL. Les lichens qui ont été récoltés devaient avoir une taille d'au moins 6cm. Cela s'explique par le fait qu'ils grandissent d'un centimètre tous les ans et que le choix a été fait d'étudier le cumul d'exposition sur les six années précédant la récolte (période 2003-2009). Les lichens ont été choisis car leur capacité à capter les métaux lourds a été jugée nécessaire dans une région au long passé industriel, qui est d'ailleurs leur principale source d'émissions de ces polluants.

3.3 Principales sources d'émissions

D'après les données issues de la plateforme publique Atmo, les polluants observés proviennent souvent des mêmes sources d'émissions. Parmi ces sources, on retrouve notamment les industries et la métallurgie pour 6 polluants et l'automobile pour 4 polluants,

ce qui en fait les deux sources les plus émettrices de pollution. Néanmoins, ces pollutions peuvent aussi être émises par d'autres sources. Le tableau 3 ci-dessous résume, par polluants, les sources principales.

Tableau 3 : Les principales sources des polluants étudiés

Polluant	Source principale (SP)	% Emis par la SP	Sources secondaires
Aluminium (Al)	Matériaux de construction	NR (Non-renseigné)	Métallurgie (non-ferreux)
Arsenic (As)	Habitat résidentiel	21%	Matériaux de construction ; Métallurgie (non-ferreux) ; Transport routier (Diesel)
Benzène (Be)	Résidentiel/tertiaire	58%	Transports ; Industries ; Agriculture
Cadmium (Cd)	Métallurgie (Ferreux)	31%	Agriculture ; Transport routier (Diesel) ; Matériaux de construction
Monyxyde de carbone (CO)	Résidentiel	42%	Métallurgie (ferreux) ; Transport routier (Essence et deux roues) ; Transport fluvial
Chrome (Cr)	Résidentiel	28%	Métallurgie (ferreux) ; Matériaux de construction ; Chauffage urbain ; Transport routier (Diesel)
Cuivre (Cu)	Transport routier (Diesel)	56%	Transport ferroviaire ; Transport routier (essence)
Mercure (Hg)	Industries chimiques Métallurgie (ferreux)	17%	Matériaux de construction ; Transformation énergétique ; Traitement des déchets
Manganèse (Mn)	Métallurgie (ferreux)	NR	Activité minière
Nickel (Ni)	Raffinerie	18%	Secteur tertiaire ; Métallurgie (ferreux) ; Industries chimiques ; Chauffage urbain ; Production électricité ; Papier, carton
Plomb (Pb)	Transport routier (Diesel)	37%	Métallurgie (ferreux) ; Résidentiel ; Transport routier (essence)
Palladium (Pd)	Métallurgie	NR	/
Antimoine (Sb)	Activité minière	NR	Industries ; Transport routier
Titane (Ti)	Industries chimiques	NR	Métallurgie
Vanadium (V)	Métallurgie	NR	Extraction minière
Zinc (Zn)	Transport routier (Diesel)	48,8%	Résidentiel ; Métallurgie (ferreux) ; Transport routier (essence)
Dioxyde d'azote (NO2)	Transport routier (Diesel)	54%	Agriculture ; Résidentiel
Particules fines	Agriculture + Elevage	55,5%	Construction ; Résidentiel ; Industries manufacturières

Mais en plus de ces polluants, 3 indicateurs cités précédemment viennent compléter les données de captation.

3.4 Les indicateurs de pollution

Parmi ces indicateurs, il a en premier lieu le Ratio d'Imprégnation Moyen (RIM) ou indicateur multimétallique. Cette méthode est utilisée grâce à de la biosurveillance par lichen. Elle consiste à observer les 18 polluants métalliques observés (sans les « poussières ») et va révéler la mesure de dépassement global des concentrations de polluant. Afin de connaître ce dépassement global par IRIS, une méthode dite de « krigage » est appliquée pour quadriller le territoire et affilier chaque point de relevé à un carreau, puis ces carreaux sont réagregés par IRIS. Un RIM inférieur à 1 désigne une absence de sur-présence de pollution métallique. Un RIM supérieur indique une présence de pollution. Par exemple, un indice RIM de 2 signifie qu'en moyenne les concentrations en métaux dans un IRIS sont deux fois plus élevés que sur le reste de la zone étudiée.

En second lieu, un indice d'eutrophie a été utilisé. L'eutrophisation désigne le niveau de couverture d'un arbre par des lichens. Il est notamment utilisé comme indicateur d'un milieu impacté par une pollution azoté (les NOx). Les données récoltées pour calculer ce niveau datent d'entre 2000 et 2012. Cet indice sert d'ailleurs habituellement à calculer un autre indice, l'Indice Biologique des Lichens Epiphytes (IBLE) mis en place pour connaître l'influence de l'azote sur un espace grâce au recouvrement des lichens sur les arbres ainsi que la diversité de ces lichens.

Enfin, un score appelé « Score environnemental » a été calculé. Ce score utilise toutes les variables environnementales utilisées afin de résumer le niveau de pollution en un seul indicateur. L'unité de calcul de ce score est en microgramme par mètre cube ($\mu\text{g}/\text{m}^3$).

Le matériel de recherche et les outils qui seront utilisés pour caractériser les risques ayant été décrit, les méthodes d'analyses qui vont être appliqués pour détecter les populations à risque vont être décrites, ainsi que les premiers résultats qui s'ensuivront.

PARTIE II – ANALYSE SPATIALE D’INCIDENCE ET CARACTERISATION DES PRINCIPAUX FACTEURS DE RISQUES

SOMMAIRE

PARTIE II – ANALYSE SPATIALE D’INCIDENCE ET CARACTERISATION DES PRINCIPAUX FACTEURS DE RISQUES.....	46
Chapitre 1 : Description du registre MONICA	46
1. Profil des cas incidents	46
2. Les taux d’incidence : bruts et standardisés	48
Chapitre 2 : Distribution spatiale de l’incidence et détection de clusters spatiaux.....	53
1. Distribution spatiale des rapports standardisés d’incidence	53
2. Détection de clusters spatiaux	59
Chapitre 3 : Caractérisation des clusters spatiaux et influence des facteurs de risque dans les variations spatiales d’incidence	67
1. Objectif	67
2. Les stratégies de comparaisons de groupes d’IRIS.....	68
3. Les méthodes statistiques de comparaison	68
4. Résultats	70
5. Limites des stratégies d’analyse de la caractérisation des clusters	73

Chapitre 1 : Description du registre MONICA

1. Profil des cas incidents

Pour rappel, les cas incidents désignent tous les cas n’ayant jamais eu d’antécédents de cardiopathie. Cela signifie que dans ce registre, chaque cas désigne un patient. Le registre MONICA sur lequel les analyses statistiques seront effectuées est donc composé de 3 268 patients. Le tableau 4 résume quelques données descriptives des patients.

Tableau 4 : Description du registre MONICA

Variables descriptives	Registre Monica (n = 3268)
Sexe*	
Hommes	71.38 (n = 2333)
Femmes	28.62 (n = 935)
Age**	
Age	58 [51 ; 66]
Année de l'épisode*	
2008	24.97 (n = 816)
2009	25.98 (n = 849)
2010	23.07 (n = 754)
2011	25.98 (n = 849)
Catégorie d'épisode*	
Infarctus	39.75 (n = 1299)
Syndrome coronarien aigu (sca)	29.22 (n = 955)
Mort subite	15.88 (n = 519)
Décès coronaire	6.58 (n = 215)
Autre épisode d'insuffisance coronaire aigue	4.68 (n = 153)
Angor instable	3.89 (n = 127)

* variables qualitative en pourcentage (effectif) ; ** Age médian [Intervalle interquartile].

D'après le tableau 4, plusieurs éléments sont à constater pour caractériser le registre. Premièrement, les hommes composent majoritairement la base, avec plus de 70% des cas.

Ensuite, l'âge médian est de 58 ans, et la moitié des patients enregistrés lors de leur premier évènement était âgée de 51 à 66 ans.

Les années où les patients ont été les plus nombreux étaient en 2009 et 2011 avec 849 cas, tandis qu'en 2010 le nombre de patients est à son niveau le plus bas avec 754 cas.

Enfin, l'infarctus est le type de cardiopathie le plus fréquent avec 39,75% des cas, suivi d'un syndrome coronarien aigu (29,22%) et des morts subites (15,88 %)

Ce que le registre révèle ici est que les hommes sont les personnes qui semblent être le plus vulnérables aux événements de cardiopathies ischémiques et plus précisément face aux

infarctus. Mais étant donné leur forte présence dans le registre, ceux-ci peuvent fortement influencer les proportions.

Cette description permet donc d'avoir un premier aperçu des profils de patients à risque en ce qui concerne les cas de CI. Afin de pouvoir constater comment se répartissent les risque au sein des différents profils de population, des taux d'incidence ainsi que des rapports d'incidence standardisés ont été effectués.

2. Les taux d'incidence : bruts et standardisés

2.1 Taux d'incidence brute

Comme expliqué précédemment (cf. Chapitre introductif, 1.2), le taux d'incidence représente un rapport entre le nombre de nouveaux cas sur une période et la population « à risque » sur cette même période. Ce taux est généralement exprimé pour 100 000 habitants pour qu'il ait plus d'impact à la lecture (50 cas sur 100 000 habitants est plus envisageable que 0,05% de malades).

Ici, les taux d'incidence bruts ont été calculés en plusieurs étapes. Tout d'abord, la base des cas incidents de MONICA (n=3268 cas) a été réorganisée suivant les âges et les sexes des personnes. 24 catégories ont donc été créées pour y classer les cas (2 sexes * 3 classes d'âge * 4 années = 24 catégories).

Pour calculer les incidences il a fallu ensuite connaître la population à risque, ici, la population des IRIS peuplés de la Métropole lilloise. La base de l'INSEE qui a été spécialement demandé pour l'étude a été utilisée, grâce à son séquençage spécial (cf. Chapitre 3, 1.2). Ces données de populations ont donc été réagrégées par la suite pour convenir aux classes d'âge nécessaires.

La formule pour calculer les incidences de chaque catégorie de personne est :

$$\text{Nombre de cas} / \text{population à risque}$$

Les résultats sont tous multipliés par 100 000 afin de permettre une représentation simplifiée. Pour l'exemple, dans le tableau 5 ci-dessous, l'un des résultats est interprétable ainsi : en 2011, l'incidence des cardiopathies ischémiques chez les femmes âgées de 45 à 59 dans la Métropole lilloise est de 89,91 nouveaux cas pour 100 000 habitantes.

Tableau 5 : Incidences brutes par années et classes de populations

Classe d'âge	Hommes			Femmes		
	35-44 ans	45-59 ans	60-74 ans	35-44 ans	45-59 ans	60-74 ans
2008	79,02	300,15	463,32	23,89	79,28	226,04
2009	89,92	306,51	495,34	33,18	85,08	204,36
2010	96,73	265,15	412,47	23,89	76,38	202,81
2011	94,00	317,12	493,46	19,91	89,91	191,98

Ces données ont été ensuite intégrées dans un graphique afin de comparer les évolutions des incidences plus distinctement.

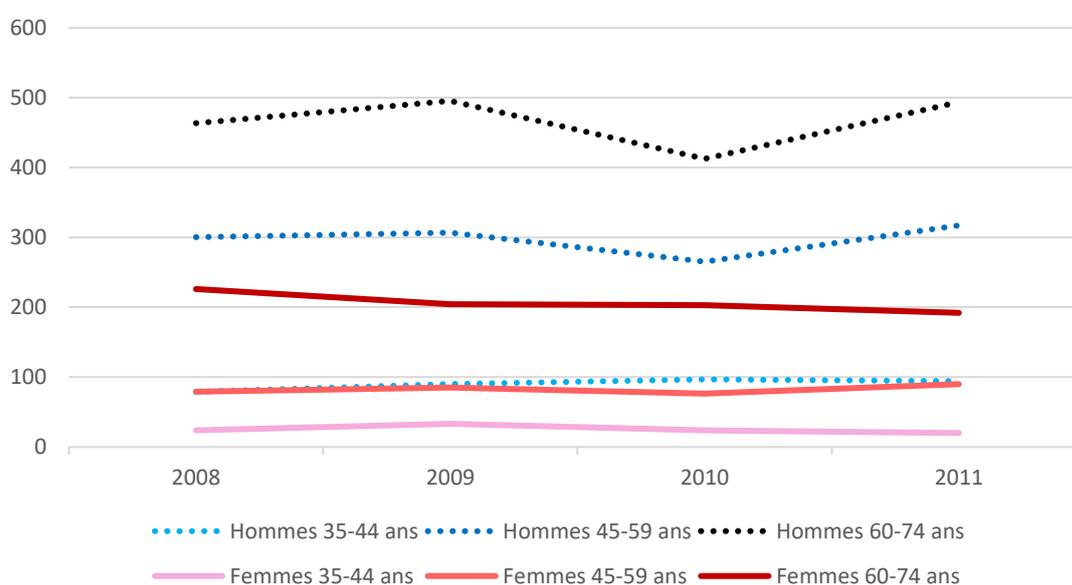


Figure 14 : Taux d'incidence brutes

La figure 14 ci-dessus est un graphique qui appuie ce que les données descriptives montraient précédemment, ainsi que les données issues de la littérature scientifique. En effet, les hommes ont tendance à avoir des incidences supérieures à celles des femmes, tous âges confondus. De plus, les deux classes les plus âgées chez les hommes montrent une tendance à une hausse de l'incidence (malgré un pic de diminution en 2010) sur les quatre années d'études, tandis que les hommes de 35-44 ans montrent une incidence plutôt constante depuis 2009. Le pic le plus élevé est atteint chez les hommes de 60-74 ans en 2009 où l'incidence culmine à 495,34 nouveaux cas pour 100 000 habitants. En 2011, l'incidence est d'ailleurs revenue à ce niveau.

Chez les femmes, la tendance est relativement à la baisse, notamment chez les femmes de 34-45 ans et celles de 60-74 ans. Ces dernières montrent une diminution constante sur les quatre

années d'études, seule catégorie dans cette situation, passant de 226,04 à 191 nouveaux cas pour 100 000 habitantes. Les femmes de 45-59 ans sont les seules montrant une hausse d'incidence, passant de 79,28 à 89,91 nouveaux cas/100 000 habitantes.

Le taux d'incidence brut sert ici de premier indicateur de répartition de la maladie à travers le temps, mais surtout à travers les profils de personnes. Ce premier calcul confirme la tendance à ce que ce soit les hommes qui sont les plus touchés, ainsi que les populations les plus âgées, ce qui est pour l'instant cohérent avec la littérature et les chiffres nationaux. Néanmoins, ce calcul d'incidence constitue une limite, et particulièrement lorsque l'on calcule ces taux pour chaque IRIS. En effet, un taux d'incidence brut élevé peut montrer une incidence réellement élevée sur un IRIS comme aussi révéler simplement un IRIS regroupant plus de personnes à risque au sein de celui-ci. Ce qui veut dire que pour l'étude MONICA, un IRIS regroupant plus d'hommes âgés de plus de 60 ans sera sûrement considéré plus à risque qu'un IRIS regroupant une majorité de femmes de 35 ans, ce qui n'est plus une question de différence de risque mais simplement un biais de calcul. Pour supprimer ce biais, une méthode consiste à calculer un taux d'incidence en appliquant à chaque IRIS la même structure par tranche d'âge et par sexe. Cette méthode s'appelle la standardisation, et elle peut être réalisée de différentes manières.

2.2 Les différents types de standardisation

Une standardisation est une méthode statistique consistant à atténuer l'impact des **facteurs de confusion** d'un calcul. Ces facteurs sont appelés ainsi car ils peuvent influencer l'interprétation d'un résultat statistique.

Dans le cas d'une étude sur les infarctus du myocarde, les personnes les plus touchées sont des hommes âgés de plus de plus de 60 ans. Les facteurs de confusions ici sont donc l'âge et le sexe, puisque, par exemple, une proportion de personnes plus âgées et/ou plus masculines peut influencer l'augmentation de l'incidence, et par conséquent le risque, sur un territoire. Une prise en compte de ces facteurs permettra de répondre à la question « Quel serait le taux d'incidence si tous les IRIS avaient la même répartition de population par âge et par sexe ? »

Pour atténuer ces facteurs de confusion, il existe deux types de standardisation : directe, et indirecte.

- La standardisation directe permet de savoir quel serait le taux d'incidence si toutes les unités (spatiales ou temporelles) avaient la même structure d'âge (le même effectif de personnes par tranche d'âge). Le résultat obtenu est un taux standardisé.
- La standardisation indirecte permet de savoir quels seraient les effectifs de chaque unité (spatiale ou temporelle) par tranche d'âge si toutes ces unités avaient le même taux d'incidence. Le résultat obtenu est un effectif attendu.

La standardisation indirecte a été préférée puisque les calculs d'effectifs attendus serviront pour calculer le taux d'incidence standardisé, puis des ratios d'incidence standardisé.

La standardisation indirecte s'effectuera alors en deux temps : d'abord une standardisation sur l'âge, puis sur le sexe.

2.3 Taux d'incidence standardisés par âge

Pour effectuer une standardisation indirecte, la méthode de calcul est en deux étapes. Il faut, pour chaque tranche d'âge :

- Calculer l'incidence brute
- Incidence brute sur une unité (spatiale ou temporelle) * effectif total de personnes de la classe d'âge sur l'ensemble des unités auxquelles une comparaison veut être faite (ici la France entière)

Les résultats de ces standardisations sont détaillés dans le tableau 6 ainsi que la figure 15 ci-dessous.

Tableau 6 : Taux d'incidence standardisé par âge, pour 100 000 habitants

Année	Hommes	Femmes
2008	279,5	112,7
2009	294,4	111,3
2010	255,8	104,2
2011	299,6	105,6

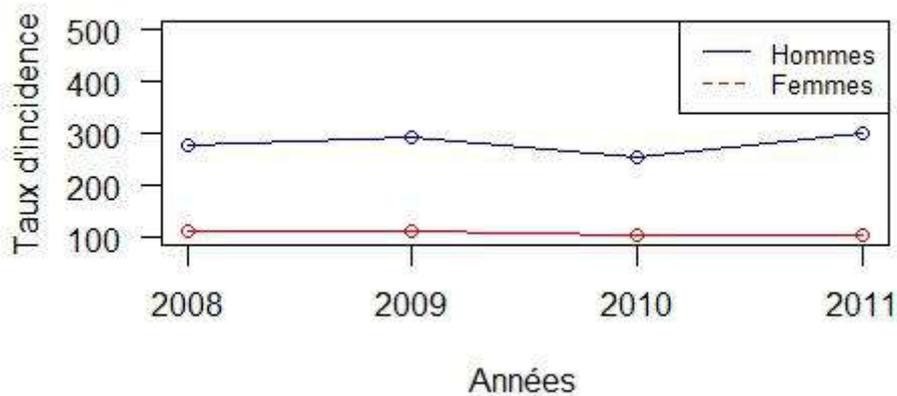


Figure 15 : Graphique d'incidence standardisée par âge

Sur la figure 15 la courbe de l'incidence pour 4 années confirme les tendances observées sur le graphique des incidences brutes. Premièrement, elle est plus élevée que chez les hommes, ce qui apparaît logique aux vues de la littérature explicitée et des calculs d'incidence bruts précédents. Mais ce sont les évolutions des courbes qui paraissent plus notables. En effet, malgré un pic de diminution en 2010, il apparaît que l'incidence chez les hommes de la MEL s'accroît tandis que celle des femmes montre une diminution, certes moins conséquente, mais qui est à souligner. En 4 ans l'incidence masculine a augmenté de 20 cas pour 100 000, passant de 279 à 299 cas pour 100 000 habitants. Chez les femmes, ce chiffre est passé de 112 à 105 cas pour 100 000 habitants, soit presque trois fois moins important que chez les hommes. Mais ces résultats sont à mettre en regard avec les incidences françaises car, si le taux d'incidence féminin apparaît comme faible dans les graphiques, il n'en n'est pas moins témoin d'une situation plus alarmante. En France, pour rappel, le taux d'incidence chez les femmes est situé entre 56 et 69 cas pour 100 000 habitantes (Wagner et *al.*, 2011).

2.4 Taux d'incidence standardisés par âge et par sexe

Une fois l'incidence standardisée par âge, une deuxième et ultime standardisation était à effectuer : celle par sexe. Le principe est le même : chercher à savoir quel serait le taux d'incidence global par année si la répartition par sexe était la même dans chaque unité spatiale.

Sur la figure 14 après une standardisation sur l'âge et le sexe, la courbe globale montre que l'incidence, malgré un pic de diminution en 2010, semble constante. En effet, entre 2008 et 2011, ce chiffre a augmenté de 7 points, mais n'a pas encore dépassé son plus haut taux atteint en 2009, à savoir 197,2 cas pour 100 000 habitants. Comparé à la France où le taux en 2008 n'est que de 78 cas globaux pour 100 000 habitants, cela met en lumière un risque de cardiopathies ischémique plus fort dans les Métropole Lilloise que dans le reste de la France.

Ces données annuelles montrent donc comment se répartit l'incidence dans le temps. Néanmoins, en parallèle de ces calculs, les effectifs attendus résultants de ces calculs ont été calculés pour chaque IRIS. Ce sont ces données qui seront utiles au calcul d'un Rapport d'Incidence Standardisé (*Standardized Incidence Ratio* : SIR) car cela servira à évaluer le niveau de risque de chaque IRIS dans la MEL.

Chapitre 2 : Distribution spatiale de l'incidence et détection de clusters spatiaux

1. Distribution spatiale des rapports standardisés d'incidence

1.1 Méthode de calcul

Le SIR, en anglais « *Standardized Incidence Ratio* » (SIR), vient établir un rapport entre un effectif observé et un effectif attendu. Il est un estimateur du risque relatif par IRIS, à savoir le risque de développer la maladie par rapport au reste du territoire. Par ailleurs, le SIR permet d'estimer un risque ajusté sur des facteurs de confusion. Il fréquemment utilisé dans le domaine de la santé, notamment pour des questions de calculs d'incidence de maladie ou d'effectif de populations dépistées, par exemple. En ce qui concerne l'étude MONICA, le calcul d'un SIR sera effectué grâce aux effectifs observés de cas incidents d'infarctus par IRIS, par rapport aux effectifs attendus par IRIS. Aussi, pour un IRIS i , le SIR est calculé de la manière suivante :

$$\frac{O_i}{E_i}$$

Où O_i correspond au nombre de cas observés au sein de l'IRIS i , et E_i au nombre de cas attendus issus d'une standardisation indirecte sur l'âge et le sexe. Par ailleurs, le nombre de cas attendus est calculé sous l'hypothèse d'une répartition spatiale homogène des cas de maladies cardiovasculaires sur l'ensemble du territoire couvert par la MEL.

Cet indicateur varie de 0 à $+\infty$, mais ce sera son éloignement à la valeur 1 qui sera observée.

Si le SIR associé à un IRIS est supérieur à 1, cela montre que plus de cas ont été observés qu'attendus. L'interprétation qui peut être faite est alors que l'IRIS présente un risque de CI plus élevé que sur l'ensemble du territoire.

Si le SIR est inférieur à 1, cela montre qu'il existe moins de cas observés qu'attendus. Le territoire en question peut donc être considéré comme un territoire où le risque est plus faible que sur les autres.

Si le SIR est égal à 1, alors le risque dans l'IRIS est identique à l'ensemble du territoire.

Le calcul de ces SIR est donc considéré comme un estimateur du Risque Relatif (RR). En effet, si le RR est un indicateur permettant de connaître le niveau de risque d'un facteur en épidémiologie, le SIR permet d'estimer la valeur du RR pour des études où toute la population observée est considérée comme étant « à risque ». Cela peut être utile lorsque des études complémentaires sur les différents facteurs de risques de la maladie observée sont misent en place.

La répartition spatiale des SIR bruts au sein de la MEL est représentée au sein de la Figure 16.

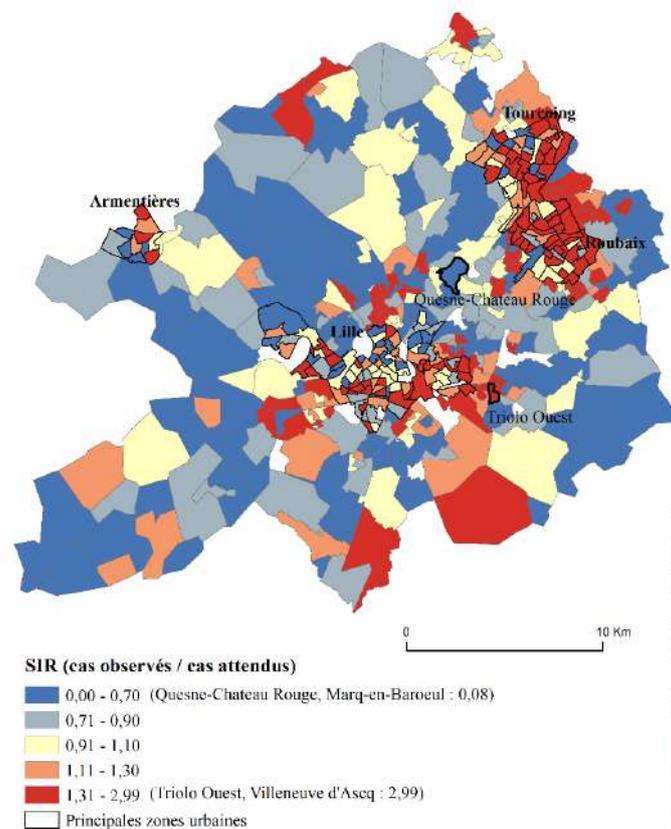


Figure 16: Répartition spatiales des SIR Bruts dans la MEL

Les IRIS ayant des SIR à 0 sont ceux où aucun cas n'a été enregistré. Ils sont 11 au total sur les 473 IRIS qui composent la Métropole. Grâce à ces SIR bruts, il est d'ores et déjà possible d'observer que le sud de Lille et les communes de Tourcoing et Roubaix semblent plus enclins à voir survenir des cas plutôt que les IRIS plus ruraux (cf. Carte du MOS, figure 5). Parmi les IRIS ayant enregistré au moins un cas, celui du Quesne-Château-Rouge situé à Marcq-en-Baroeul est celui ayant le SIR brut le plus faible avec 0,08, soit 92% de risque en moins que le reste de la MEL de voir survenir un cas de CI. A l'inverse, l'IRIS Triolo Ouest situé à Villeneuve d'Ascq est celui ayant le SIR le plus fort, enregistré est à 2,99. Ce SIR peut s'interpréter de la manière suivante : dans cet IRIS, le risque de voir survenir un cas de CI est presque 3 fois plus fort que dans le reste de la Métropole. Néanmoins, malgré cette première approche, les SIR bruts connaissent quelques limites.

1.2 Limites des SIR bruts

Le calcul des SIR, bien qu'utile à une première vision de l'étendu du risque sur un territoire lorsqu'il est calculé par IRIS, présente néanmoins plusieurs limites.

Premièrement, chaque SIR est calculé par IRIS, et par conséquent, se base sur la population présente dans un IRIS. Ce qui induit que la variance de chaque SIR est liée à la population de chaque IRIS¹⁶. Alors si 2 cas sont observés dans un IRIS habité par 4 personnes, le SIR sera de 2, mais la variance de celui-ci sera élevée. Et plus la variance d'une variable est élevée, plus son estimation est peu fiable et, par voie de conséquence, son interprétation est contestable.

Deuxièmement, il arrive qu'aucun cas ne soit observé dans les IRIS d'études. Leur SIR est donc de 0. Au moment d'avoir une vision globale des SIR – notamment lorsque l'on les cartographie – les SIR étant à 0 sont tous dans la même classe, ce qui semble logique à première vue. Néanmoins, il n'existe aucune différenciation des SIR qui sont à 0 dans un IRIS faiblement peuplés des SIR à 0 dans les IRIS fortement peuplés. Or, un SIR calculé à 0 dans un IRIS de 10 habitants ne peut pas forcément être considéré comme aussi faiblement à risque qu'un SIR calculé à 0 dans un IRIS de 1000 habitants.

¹⁶ La variance est un calcul permettant de mesurer les écarts de chaque valeur par rapport à la moyenne d'une distribution, et ainsi de caractériser la dispersion de ses valeurs. Plus sa valeur augmente, plus les valeurs sont proches de la moyenne, et inversement. Sa formule est : $V(X) = E[(X-E(X))^2]$

Enfin, chaque IRIS a son SIR propre à son territoire et à sa population. Or la MEL n'est pas un archipel constitué d'îles et de terres dissociées. Ses IRIS forment un territoire uni. Des voies de communications existent entre eux. De plus, pour rappel, les IRIS ne sont qu'un découpage statistique de l'INSEE, et non administratif. Par conséquent, rares sont les coupes franches et nettes entre deux IRIS qui marqueraient une quelconque séparation. Et comme le disait le géographe Waldo Tobler en 1970, et ce qui reste aujourd'hui comme la « première loi de la géographie » : « tout interagit avec tout, mais deux objets proches ont plus de chance de le faire que deux objets éloignés » (Tobler, 1970). Cela indique que les IRIS voisins peuvent avoir des comportements similaires, ce qui induit la notion d'autocorrélation spatiale. Calculer des SIR en ne prenant compte uniquement la population de l'IRIS étudié n'est donc pas forcément pertinent, tant les facteurs de risques ne sont pas délimités par des frontières terrestres et tant les personnes se déplacent au sein de la MEL.

Afin de pallier ces différentes limites, un modèle statistique hiérarchique bayésien a été utilisé afin de produire des SIR lissés plus robustes.

1.3 Lissage des SIR par modèle bayésien

Le modèle hiérarchique bayésien utilisé dans cette étude est celui développé par Besag, York et Mollié (Besag, J., York, J. and Mollie, A., 1991).

L'aspect hiérarchique du modèle réside dans le fait qu'il est constitué de trois niveaux différents. Au premier niveau, le nombre de cas observés dans un IRIS i est modélisé par une loi de Poisson (courant en épidémiologie pour modéliser le comptage des maladies rares) :

$$O_i \sim P(\theta_i e_i),$$

Où O_i correspond au nombre de cas observés, e_i au nombre de cas attendus issus d'une standardisation indirecte et θ_i correspond au risque relatif associé à l'IRIS. C'est ce paramètre que le modèle va estimer afin de produire ce qui sera appelé par la suite un SIR lissé.

Au deuxième niveau, le logarithme du risque relatif est considéré comme une combinaison linéaire de plusieurs effets :

$$\log(\theta_i) = \beta_0 + U_i + V_i - \log(e_i),$$

Où β_0 correspond au risque relatif moyen sur l'ensemble de la zone étudiée, U_i et V_i étant des effets aléatoires. L'effet aléatoire U_i capte la sur-dispersion du nombre de cas observés non

modélisée par la loi de Poisson et ne présentant pas de structure spatiale. U_i est distribué selon une loi normale de moyenne nulle et de variance σ_u^2 :

$$U_i \sim N(0, \sigma_u^2).$$

L'effet aléatoire V_i capte la sur-dispersion du nombre de cas qui est structurée spatialement. En d'autres termes, cet effet permet de prendre en compte l'autocorrélation spatiale potentielle entre les IRIS proches. Cet effet aléatoire est distribué selon le modèle conditionnel autorégressif suivant :

$$(V_i/V_j, i \neq j) \sim N\left(\frac{\sum_{j \neq i} w_{ij} v_j}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_v^2}{\sum_{j \neq i} w_{ij}}\right),$$

Où w_{ij} est égal à 1 si les deux IRIS i et j sont voisins et 0 sinon. En d'autres termes, l'effet d'une unité spatiale est la résultante d'une moyenne pondérée des effets de ces voisins.

Au troisième niveau, les variances des effets aléatoires sont considérées aléatoires et distribuées selon des lois Gamma dites « non-informatives » car nous ne disposons pas d'information *a priori* sur leur distribution.

L'estimation des différents paramètres du modèle est réalisée au moyen de techniques de simulations de type *Integrated Nested Laplace Approximation* qui, du fait de leur complexité, ne seront pas explicitées dans le présent mémoire.

Le résultat de ces SIR lissés indiquent comment le risque est évalué suivant les valeurs de risque des voisins et la population à risque au sein de l'IRIS observé. Cette technique permet donc de connaître plus précisément la valeur du risque auquel un IRIS est confronté. Dans la MEL, la répartition des SIR Lissés est établit sur la figure 17.

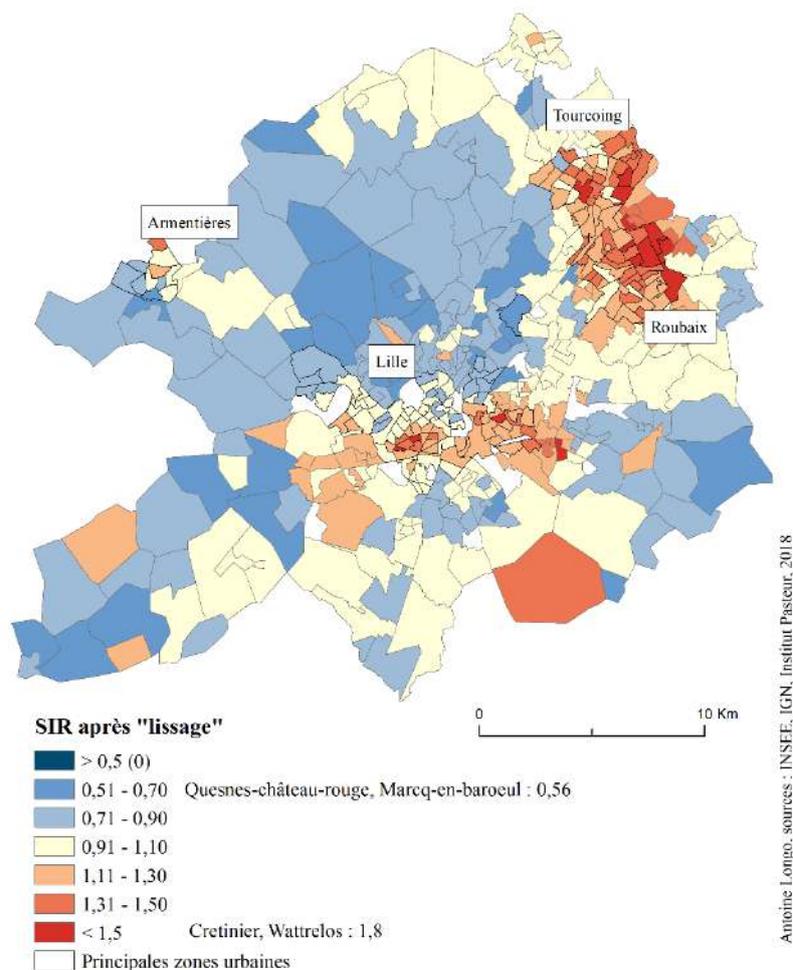


Figure 17 : Répartition des SIR Lissés à l'échelle des IRIS de la MEL

La répartition des SIR lissés montrent ici une autre vision du risque relatif dans la MEL. En effet, d'après la carte 17, il est possible d'affirmer qu'il existe une hétérogénéité spatiale de l'incidence des cardiopathies ischémiques dans le territoire de la métropole. De plus, il est désormais possible d'observer des agrégations de zones qui semblent avoir des profils de risque ressemblant. Le nord-ouest de la Métropole semble ici étendre une zone de faible risque d'incidence – ou sous-incidence – jusqu'à la pointe sud de la métropole. Parmi cette zone d'agrégation, l'IRIS Quesnes-Château-Rouge montre une nouvelle fois un SIR plus faible, évalué à 0,56. Cela peut s'interpréter de la façon suivante : dans cet IRIS, le risque de voir survenir un cas de CI est quasiment deux fois plus faible que dans le reste de la zone, ou 46% plus faible.

A l'inverse, les tendances de sur-risque observées grâce aux SIR bruts semblent se confirmer après lissage. En effet, les zones du Sud de Lille ainsi que les communes de Tourcoing et Roubaix montrent des SIR plus forts que le reste de la MEL. L'IRIS montrant le SIR lissés le plus élevé est l'IRIS Crétinier qui se situe à Wattrelos, commune limitrophe de Roubaix et Tourcoing. Son SIR s'élève à 1,8, soit 80% de risque en plus d'observer un cas de CI que dans le reste de la MEL.

Néanmoins, plusieurs questions restent en suspens après ces calculs : existe-t-il des zones regroupant des IRIS de même nature (montrant une sur-incidence ou une sous-incidence) et sont-elles significatives ? Et quelle est l'influence des facteurs de risques au sein de ces IRIS ? Ces réponses sont apportées par deux méthodes de statistiques spatiales : la détection de cluster et la régression écologique.

2. Détection de clusters spatiaux

2.1 Objectif

Les SIR calculés précédemment permettent donc d'avoir une première approche sur la répartition spatiale des risques relatifs sur la MEL. Cependant, les informations nécessaires pour l'étude sont de savoir s'il existe des zones où les risques seraient plus faibles ou plus forts que dans le reste du territoire ; des zones regroupant plusieurs IRIS ayant un risque relatif similaire : des clusters spatiaux. Dans le cadre d'une étude des variations de risques dans le domaine de la santé à une échelle fine, la détection de cluster spatiaux semble être l'outil le plus adapté pour repérer des zones où les incidences sont similaires et donc particulièrement notables. Ces clusters spatiaux peuvent être détectées suivant une méthode statistique particulière : le « scan ». Cette méthode est notamment utilisée via le logiciel SaTScan, qui permet de détecter rapidement des clusters en effectuant automatiquement les statistiques de scan (Annexe 2)

Le résultat de cette détection, si des clusters sont repérés, permettra par la suite d'établir une caractérisation de ces clusters, afin d'analyser les éventuels effets des variables environnementales de l'étude, à savoir la défaveur sociale et la pollution atmosphérique.

2.2 Méthodes

Les méthodes de statistiques de scan spatial ont pour objectifs la détection de clusters d'évènements (agrégation d'évènements dans l'espace) et le test de leur significativité

statistique. Les méthodes les plus utilisées sont celles développées par Martin Kulldorff de l'université de Harvard (Kulldorf, 1995 ; Kulldorf, 1997). Ce dernier a également développé le logiciel SaTScan (Kulldorf, 1998) qui vient automatiser les calculs. Ces méthodes sont décomposées en deux étapes principales : une étape de détection d'un cluster et l'étape d'inférence statistique, à savoir le test de sa significativité statistique.

2.2.1 Choix du type de cluster

Lors d'une étude à propos des variations spatiales de l'incidence comme celle du projet MONICA, deux types de clusters peuvent être recherchés : les clusters de sous-incidence et les clusters de sur-incidence. Un cluster de sur-incidence est un cluster qui, s'il est détecté, désigne une zone où le risque est plus élevé que dans le reste du territoire, et inversement pour les clusters de sous-incidence. Pour cette étude, les deux types de clusters seront recherchés, le but étant de caractériser ces deux différents types, suivant des variables choisies au préalable (défaveur sociale, pollution atmosphérique) afin de constater s'il existe des différences notables, ce qui constitue ici le cœur de cette recherche.

2.2.2 Phase de détection du cluster

Les méthodes de statistiques de scan spatiales reposent sur un test statistique dont les hypothèses sont les suivantes :

H₀ : les cas d'infarctus sont distribués de manière homogène sur le territoire de la MEL (absence de cluster)

H₁ : il existe un cluster spatial dans lequel le risque de développer un infarctus est plus élevé (ou plus faible) que dans le reste de la zone géographique étudiée.

La phase de détection consiste à venir balayer l'ensemble de zone géographique étudiée avec une fenêtre de balayage (appelée fenêtre de scan) de forme circulaire, de rayon variant de 0 à une certaine limite et prenant pour centre le centre de l'unité spatiale considérée (ici l'IRIS). La limite est souvent fixée à un rayon tel que le cluster ne couvre pas plus de 50% de la population à risque. Cette limite est justifiée par le fait qu'un cluster présentant un risque élevé et couvrant plus de 50% de la population à risque serait uniquement que le reflet d'un sous-risque à l'extérieur du cluster. La phase de détection, considérant chacun des centres des

unités spatiales et des cercles de rayon multiples conduit à une collection importante de clusters candidats (Figure 18).

Une solution pour évaluer les clusters significatifs pourrait consister à tester de manière statistique chacun des clusters candidats par rapport à l'ensemble du territoire. Cependant, à la vue du nombre très important de clusters candidats, cette approche conduirait inéluctablement au problème de multiplicité des tests statistiques et ainsi à un nombre de faux-positifs extrêmement élevé.

Aussi, au sein de ces méthodes, l'approche considérée est la suivante : à chacun de ces clusters candidats est associé un rapport de fonctions de vraisemblance où le numérateur correspond à la fonction de vraisemblance sous l'hypothèse nulle d'absence de cluster et le dénominateur à la fonction de vraisemblance sous l'hypothèse alternative de présence d'un cluster. Ces fonctions de vraisemblance sont des probabilités et, de ce fait, plus le rapport est élevé, plus le cluster candidat a peu de chances d'être observé par hasard. Le cluster maximisant le rapport de vraisemblance est appelé le *Most Likely Cluster* (MLC). C'est celui-ci qui va faire l'objet d'une inférence statistique. Le rapport de vraisemblance associé au MLC constitue la statistique de test λ du test explicité *supra*.

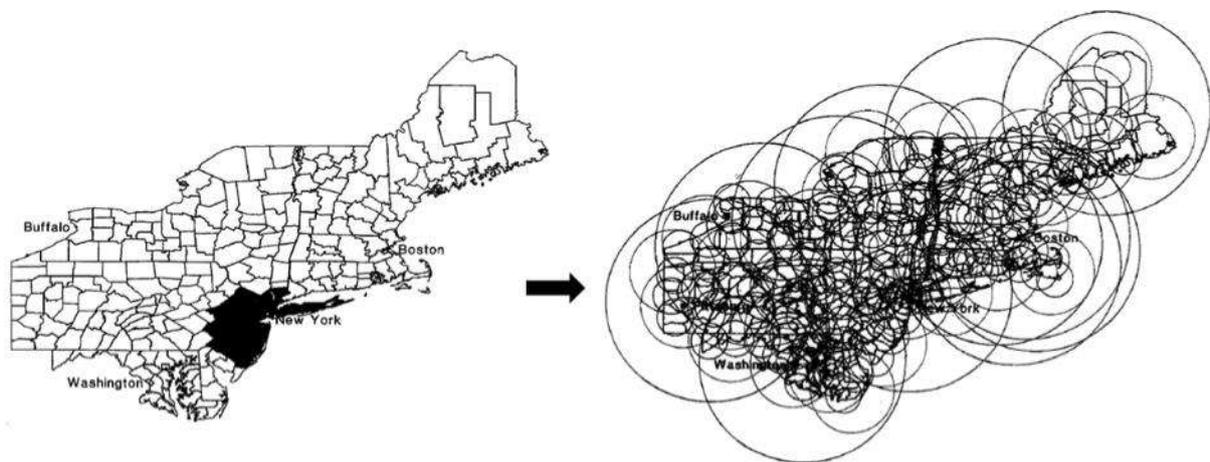


Figure 18 : Exemple de phase de détection de cluster dans l'Etat de New-York, Etats-Unis. Source : Martin Kulldorf, 1999

2.2.3 Phase d'inférence

Le cluster ayant maximisé ce rapport est donc celui qui a le plus de chance de se produire. Or, sa significativité n'a pas encore été testé, alors une *p-value* doit lui être associé. Néanmoins, la distribution de λ sous l'hypothèse nulle n'ayant pas une forme analytique (i.e. dont l'expression mathématique n'est connue), la recherche d'une valeur de significativité n'est

pas possible. Pour pallier ce problème, une technique de simulation dite « de Monte Carlo » est réalisée.

Ces tests de Monte-Carlo ont pour principe de simuler des données d'incidence sous l'hypothèse H_0 partout dans la zone d'étude et de refaire la phase de détection un grand nombre de fois (généralement 999 ou 9999 fois). Le résultat obtenu est une distribution de 999 statistiques de test. Pour vérifier la significativité du MLC, le rang de sa statistique de test va donc être observé. La *p-value* associée sera donc : $1 - \frac{\text{Rang de } \lambda_{\text{observés}}}{999+1}$. Par exemple, si le rang $\lambda_{\text{observée}}$ dans la distribution est 980^{ème} sur 999+1, sa *p-value* sera donc de 0,02, et ainsi le MLC sera significatif. L'hypothèse H_0 sera alors rejetée, ce qui validera H_1 , donc qu'il existe un cluster au sein de la zone d'étude. Si au contraire son rang est le 500^{ème} dans la distribution, alors sa *p-value* sera de 0,5 et donc H_0 ne sera pas rejeté, donc il n'existe pas de cluster au sein de la zone d'étude.

2.2.4 Choix du type de scan

Il existe différents modèles de statistiques de scan spatial implémentés au sein du logiciel SaTScan. En ce qui concerne l'étude MONICA, le choix d'un scan « classique » avait été tout d'abord considéré, mais un autre type de scan a été préféré au fil de l'étude : le scan « isotonique ».

Le terme « isotonique » se compose du mot « iso » qui veut dire « même » et « tonique », venant du grec « tonus » qui signifie « intensité ». Cela se traduit par le fait que, contrairement à un scan circulaire « classique » où une fenêtre homogène s'affiche lorsqu'un cluster est détecté, un scan isotonique renverra une fenêtre elle aussi circulaire, mais qui sera divisée en plusieurs niveaux d'intensités (ici, de niveaux égaux d'incidence). La figure 19 illustre la façon dont le cluster et ses niveaux sont calculés, ainsi que sa représentation graphique :

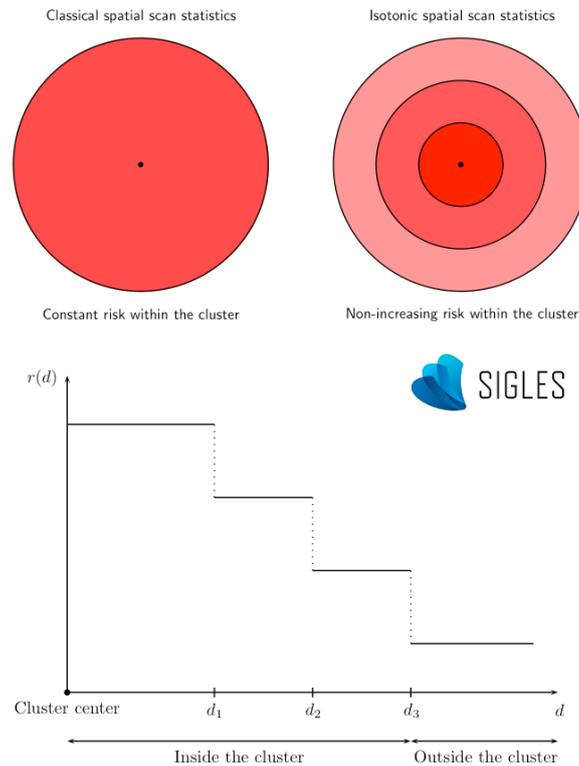


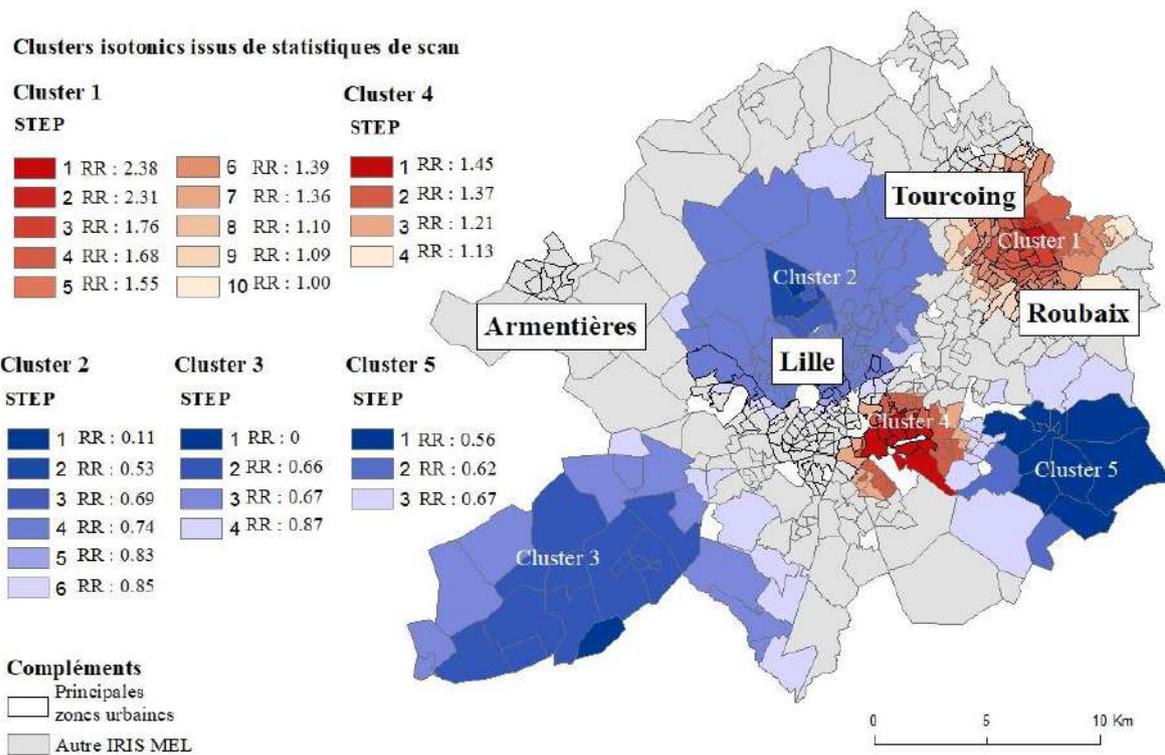
Figure 19 : Schéma de réalisation d'un scan isotonique. Source : Michaël Génin.

Cette forme de cluster montre bien que plusieurs cercles de différentes intensités se forment autour d'un épicycle, qui ici serait un SIR de sur-incidence plus élevé qu'ailleurs, ou de sous-incidence plus faible qu'ailleurs. Pour un cluster de sur-incidence, le risque est plus élevé dans le premier cercle autour de l'épicycle, et diminue au fur et à mesure des cercles de différents niveaux, et inversement pour les clusters de sous-incidence.

Ce type de cluster permet d'atténuer l'une des limites des clusters classiques. En effet, lors de la détection d'un cluster classique, l'hypothèse est que le risque est le même sur l'ensemble du cluster. Or les calculs de SIR prouvent que les risques varient d'un IRIS à un autre, et qu'il est peu probable d'avoir le même niveau de risque partout sur des zones pouvant contenir jusqu'à 50% de la population. L'avantage d'un cluster isotonique est donc d'afficher le niveau de risque à l'épicycle et l'évolution progressive autour de celui-ci, ce qui peut contredire l'hypothèse que le risque est le même sur l'ensemble du cluster.

2.3 Résultats des détections

La carte ci-dessous montre la détection de clusters isotoniques dans la métropole lilloise.



Antoine Longo, source carto/stats : IGN, INSEE, étude MONICA, 2018

Figure 20 : Clusters isotoniques de sur-incidence et de sous-incidence dans la MEL, ainsi que la valeur moyenne du risque relatif de chaque niveau

La détection de clusters dans la MEL a permis d'observer l'apparition de 5 zones d'agrégation de profils d'incidences similaires. Ainsi, sur la figure 20, apparaissent 3 clusters de sous-incidence et 2 de sur-incidence. Le cluster 1, qui est le *Most Likely Cluster*, donc celui qui a le plus de chance d'être observé, montre une plus forte sur-incidence que le reste de la zone ($p\text{-value} < 0,001$, Tableau 7). Son épicentre se trouve à cheval sur les communes de Tourcoing et Roubaix, ce qui confirme la première impression que laissait les SIR lissés, à savoir qu'il existait un risque plus élevé autour de ces communes que dans le reste de la MEL. Un autre cluster de sur-incidence est aussi détecté par le logiciel SaTScan, mais celui-ci, outre révéler une taille plus « petite » que le Cluster 1, s'avère moins « à risque » que le premier. En effet, son épicentre affiche un RR 1,45, soit plus faible que l'épicentre du premier cluster culminant à 2,38. Le risque est donc légèrement plus faible dans ce second cluster de sur-incidence. De plus, il est à ajouter que contrairement aux SIR lissés observés précédemment, ce cluster ne s'étend pas sur tous les IRIS limitrophes du Sud de Lille, mais se concentre plutôt à l'Est de la commune, autour de Villeneuve d'Ascq.

A l'inverse, 3 clusters de sous-incidence sont révélés par la détection, et les deux premiers (clusters 2 et 3) concordent avec les premières impressions que laissaient l'observation des SIR lissés. De plus, un dernier cluster est détecté alors qu'il était moins visible d'après un simple regard sur les SIR Lissés. Celui-ci se situe plus à l'Est que le Cluster 4, s'étendant jusqu'aux limites de la MEL. Il est aussi à constater que sa limite rentre en contact avec le Cluster 4 de sur-incidence. Cela peut donc laisser penser que ces zones de contact entre de la faible sur-incidence et de la faible sous-incidence sont des zones qui soit tendent vers un risque neutre (donc une absence de risque), soit des zones fortement liées à leur épicode.

De plus, en se référant à la carte de l'occupation du sol dans la MEL, il semblerait que les clusters de sur-incidence se retrouvent plutôt dans les zones urbaines, et les clusters de sous-incidence se situeraient dans des zones plus rurales.

Le détail de ces clusters est illustré dans le tableau 7 ci-dessous.

Tableau 7 : Description des clusters isotoniques par niveaux de cercles concentriques

Numéro du cluster	Niveaux Isotoniques	Radius	Nombre d'iris	Cas observés	Cas attendus	RR	p-value
<i>Cluster de sur-incidence</i>							
1	Global	3,90	96	848	648.67	1.41	
	1	0	1	14	6.38	2.38	
	2	0,71	4	51	23.94	2.31	
	3	1,15	6	65	39.91	1.76	
	4	1,98	15	154	99.25	1.68	
	5	2,11	5	48	33.58	1.55	<0.0001
	6	3,15	36	299	233.53	1.39	
	7	3,234	3	27	21.5	1.36	
	8	3,238	1	5	4.92	1.1	
	9	3,77	22	157	155.46	1.09	
	10	3,90	3	28	30.21	1	
4	Global	2.56	44	335	256	1.34	
	1	1.48	11	97	68.51	1.45	
	2	2.21	20	146	109.28	1.37	0.025
	3	2.54	12	87	73.66	1.21	
	4	2.56	1	5	4.56	1.13	
<i>Cluster de sous-incidence</i>							
2	Global	5,74	85	492	628.9	0.74	
	1	0	1	1	8.97	0.11	
	2	0,4	1	4	7.16	0.53	
	3	1,79	3	16	22.03	0.69	<0.0001
	4	5.23	65	388	497.17	0.74	
	5	5.3	2	12	13.82	0.83	
3	Global	8.37	30	163	228.65	0.7	
	1	0	1	0	4.57	0	
	2	5.77	11	57	84.39	0.66	0.0044
	3	7.99	11	60	88.22	0.67	
	4	8.37	7	46	51.47	0.87	
5	Global	4,86	20	89	141.96	0.62	
	1	2.9	8	36	62.75	0.56	0.045
	2	3.32	2	10	15.91	0.62	
	3	4.86	10	43	63.3	0.87	

Il existe donc des clusters spatiaux d'incidence remarquable dans la Métropole Lilloise : 2 de sur-incidence et 3 de sous-incidence. Ces clusters montrent donc que plusieurs IRIS dans la MEL connaissent des risques similaires, et donc pourraient avoir des profils similaires suivant plusieurs facteurs qui pourraient les caractériser. La caractérisation de ces clusters spatiaux par des variables environnementales est donc la prochaine étape qui permettrait de comprendre les variations spatiales de l'incidence des cas de CI dans la MEL.

Aussi, en plus des variables environnementales (défaveur sociale et pollution atmosphérique) incarnant les objectifs de l'étude, une caractérisation sur l'occupation du sol sera aussi réalisée dans le même temps, afin de vérifier si les clusters de sur-incidence sont significativement caractérisés par un environnement plus urbain et si les clusters de sous-incidence sont significativement plus représentés par un environnement plus rural. Ces tests pourraient servir à cerner les enjeux autour des sources des éventuels polluants qui pourraient être détectés comme significativement « à risque » par les tests statistiques.

Chapitre 3 : Caractérisation des clusters spatiaux et influence des facteurs de risque dans les variations spatiales d'incidence

1. Objectif

Les clusters qui ont été détectés montrent des zones d'agrégation de sur-incidence et de sous-incidence dans la MEL, ce qui signifie que la répartition du risque de cardiopathie ischémique dans la MEL n'est pas homogène. Cela implique que des facteurs influencent la distribution de ce risque. Ainsi, les facteurs de risque qui ont été soumis à l'étude sont la défaveur sociale, la pollution et l'occupation du sol, trois facteurs environnementaux.

Caractériser les clusters va donc permettre de savoir si l'un de ces facteurs, voire les deux, parvien(nen)t à expliquer la répartition du risque. En d'autres termes, c'est ici que sera déterminé si la défaveur sociale et la pollution ont un impact sur la variation spatiale de l'incidence des cas de CI dans la MEL. Pour caractériser ces clusters, une méthode de statistique de comparaison va être mise en place afin de connaître comment interagissent les groupes d'IRIS en fonction des facteurs de risque environnementaux. Pour réaliser cela, deux stratégies vont être mises en œuvre : une comparaison par groupe d'IRIS similaires (sur-incidence/sous-incidence/neutres) et une comparaison par clusters (5 clusters spatiaux + groupe d'IRIS neutres).

2. Les stratégies de comparaisons de groupes d'IRIS

Le principe de comparaison exprime ici une méthode de comparaison statistique. Chaque IRIS possède une valeur du Fdep, une valeur de concentration en polluants en $\mu\text{g}/\text{m}^3$, et une valeur de part d'occupation de son sol par de l'espace agricole et de l'espace urbain. Tous les IRIS vont donc être rassemblés par groupe suivant deux stratégies. Chaque groupe d'IRIS aura donc une distribution de valeur de Fdep, de valeur de concentration de polluant et d'occupation. L'objectif de cette approche est de comprendre, par ces tests de comparaison, si la répartition des valeurs du Fdep et des polluants par groupes d'IRIS sont particulièrement différents les uns des autres. Cela pourra permettre d'affirmer ou d'infirmer les relations entre les facteurs environnementaux et les variations d'incidence.

La première stratégie qui est mise en place consiste à rassembler les IRIS appartenant aux mêmes types de clusters afin de constituer 3 groupes : les IRIS de sur-incidence, ceux de sous-incidence et les IRIS qui n'appartiennent à aucun cluster. Ceux-là sont dits « neutres ». Les valeurs du Fdep et de concentrations de polluants de chaque groupe seront donc comparées aux valeurs des autres groupes. Trois tests seront alors effectués : Sur-incidence contre sous-incidence, sur-incidence contre neutres, sous-incidence contre neutres.

Seulement, il existe une limite à cette approche. Cette stratégie sous-tend une hypothèse forte selon laquelle les facteurs écologiques prédictifs de l'incidence sont identiques sur le territoire. En d'autres termes, La répartition de l'incidence n'étant pas homogène sur l'ensemble du territoire, et même au sein des clusters, il est donc difficile de partir du principe que les groupes de clusters possèdent des IRIS homogènes en termes de situation socioéconomique.

Pour pallier cette limite, une deuxième stratégie a été mise en place. Celle-ci consistera à établir les mêmes tests de comparaison statistique, mais cette fois-ci entre les 5 différents clusters, ainsi que le regroupement d'IRIS neutres.

3. Les méthodes statistiques de comparaison

Il existe plusieurs méthodes pour effectuer des tests de comparaison statistiques. Celles qui ont été choisies se décomposent en deux étapes. Tout d'abord, un test appelé « Test de Kruskal-Wallis » est réalisé, puis des tests dits « Post-hoc » viennent compléter ce premier test

3.1 Les tests de Kruskal-Wallis

Le test de Kruskal-Wallis est un test non-paramétrique permettant de comparer la distribution d'une variable quantitative entre plusieurs groupes. Il s'agit de l'équivalent non-paramétrique de l'analyse de variance (ANOVA). Cette dernière nécessite des conditions d'application (Normalité de la variable et homogénéité des variances entre les groupes). Dans le cas du Fdep et des variables mesurant la pollution, ces conditions n'étaient pas remplies, d'où l'utilisation de tests de Kruskal-wallis.

La valeur centrale qui sera comparée entre les différents groupes est la médiane.

Ce test peut se schématiser ainsi :

Tableau 8 : Schéma de réalisation d'un test de Kruskal-Wallis

	Neutres	Sous-incidence	Sur-incidence	<i>p-value</i>
Variable de caractérisation	Valeur centrale (M1)	Valeur centrale (M2)	Valeur centrale (M3)	

L'hypothèse H₀ de ce test est que tous les échantillons ont une valeur centrale identique, donc que M₁=M₂=M₃. Si la *p-value* est significative, H₀ est rejetée au profit de l'hypothèse alternative spécifiant qu'il existe au moins un groupe différent des autres.

Cependant, le résultat de ce test n'indique pas quels sont les groupes qui sont particulièrement différents les uns des autres, ce qui empêche de connaître quels sont les groupes les plus influencés par la défaveur sociale. Pour connaître ces groupes, des tests dits *Post-hoc* sont réalisés.

3.2 Les tests Post-Hoc et correction de Bonferroni

Les Tests Post-Hoc sont utilisés dans le cadre d'un test de comparaison de valeur centrales comme les analyses de variance (ANOVA) ou les tests de Kruskal-Wallis. Leur utilité est de connaître les groupes qui sont particulièrement différents les uns des autres après avoir rejeté l'hypothèse qu'il existait une égalité des valeurs centrales entre les groupes analysés.

En se servant des médianes calculées précédemment (cf. tableau 8), les groupes d'IRIS vont être analysés deux par deux, dans le but de constater quel est ou quels sont le.s groupe.s de clusters qui sont particulièrement différents des autres.

La « *p-value* » décrit si la différence est significative ou non. Si sa valeur est inférieure à 0,05, la différence est significative, et donc les clusters sont significativement différents les uns des autres.

Mais à l'abord de ces tests, un détail reste important à souligner. En effet, un biais statistique peut survenir lors d'une multiplicité des tests de comparaison : l'inflation du risque alpha.

Lors d'un test statistique, la valeur du risque d'échec est de 5%. Mais si les tests se multiplient, ce risque d'échec se multiplie aussi, ce qui laisse un risque d'obtenir des valeurs faussement positives ou faussement négatives. Pour pallier ce problème, et ce qui a été réalisé en amont de ces tests post-hoc, une correction dite « de Bonferroni » a été effectuée.

La correction de Bonferroni vient donc limiter le problème de multiplicité des risques. Sa formule prend alors en compte le nombre de tests qui vient diviser la limite du risque. Par ailleurs, les *p-value* des tests post-hoc sont corrigées (on peut les comparer à 5%). Pour corriger les *p-value*, le calcul est $p\text{-value}(\text{test Post-Hoc}) \times \text{nombre de comparaisons}$. Pour ces tests Post-Hoc, la formule sera alors $0,05 / 3$ pour fixer une nouvelle limite lors des tests de comparaisons de 3 groupes, et $0,05 / 15$ pour des tests de comparaisons de 6 groupes. C'est selon cette nouvelle limite que la significativité a donc été mise à l'épreuve.

4. Résultats

4.1 Première stratégie

La première stratégie d'analyse consiste donc à réaliser des comparaisons statistiques par groupes de clusters.

Le tableau 9 ci-dessous montre la réalisation des tests de Kruskal-Wallis (colonne « p ») et les tests Post-Hoc (3 dernières colonnes) par variables de caractérisation.

Tableau 9 : Tests de Kruskal-Wallis et post-hoc par groupe de clusters (première stratégie)

	Neutres N=198	Sous-incidence N=135	Sur-incidence N=140	p	p.0 vs 1	p.0 vs 2	p.1 vs 2
fdep09	-0.04 [-1.00;1.53]	-1.04 [-2.48;0.01]	1.62 [0.62;2.81]	<0.001	<0.001	<0.001	<0.001
Agricoles	9.83 [0.00;41.4]	15.0 [0.00;45.6]	0.00 [0.00;5.25]	<0.001	0,212	<0.001	<0.001
Urbain	62.9 [34.9;91.5]	58.2 [33.3;85.7]	89.5 [61.0;100]	<0.001	0,345	<0.001	<0.001
Al	1101 [942;1227]	1005 [938;1180]	924 [858;1007]	<0.001	0,059	<0.001	<0.001
As	1.13 [0.99;1.21]	1.11 [1.01;1.18]	1.02 [0.92;1.13]	<0.001	0,098	<0.001	0,001
Be	0.24 [0.23;0.25]	0.24 [0.22;0.25]	0.24 [0.23;0.24]	0,006	0,006	0,137	0,088
Cd	0.69 [0.48;0.84]	0.62 [0.44;0.71]	0.68 [0.47;1.00]	0,002	0,006	0,259	0,006
Co	0.72 [0.63;0.79]	0.64 [0.53;0.75]	0.62 [0.55;0.70]	<0.001	<0.001	<0.001	0,147
Cr	4.67 [3.87;5.16]	4.07 [3.56;4.70]	4.23 [3.65;4.85]	<0.001	<0.001	0,002	0,184
Cu	19.9 [15.3;29.7]	17.0 [13.2;19.7]	23.8 [17.2;32.0]	<0.001	<0.001	0,007	<0.001
Hg	0.14 [0.11;0.15]	0.13 [0.11;0.15]	0.11 [0.10;0.13]	<0.001	0,204	<0.001	<0.001
Mn	51.3 [43.5;55.1]	45.6 [42.4;48.7]	46.5 [41.5;50.9]	<0.001	<0.001	<0.001	0,8
Ni	2.95 [2.67;3.25]	2.53 [2.31;2.83]	2.99 [2.79;3.26]	<0.001	<0.001	0,261	<0.001
Pb	34.8 [19.2;47.2]	28.9 [16.6;42.8]	25.8 [21.3;35.7]	0,125	0,086	0,404	0,78
Pd	19.6 [17.4;22.6]	22.4 [17.8;28.9]	17.8 [15.8;24.6]	<0.001	<0.001	0,01	<0.001
Sb	2.17 [1.34;2.63]	1.77 [1.27;2.23]	1.96 [1.77;2.15]	<0.001	<0.001	0,035	0,001
Ti	15.9 [14.7;16.6]	13.5 [12.8;15.4]	16.3 [14.9;17.1]	<0.001	<0.001	0,077	<0.001
V	3.65 [3.36;3.89]	3.36 [3.00;3.94]	3.86 [3.41;4.03]	<0.001	0,001	0,026	<0.001
Zn	120 [96.0;137]	105 [88.9;123]	115 [100;132]	<0.001	0,001	0,461	0,001
RIM	1.87 [1.43;2.15]	1.62 [1.43;1.87]	1.69 [1.43;1.94]	0,002	0,001	0,105	0,105
Poussieres	1943 [1819;2029]	2054 [1938;2148]	1963 [1822;2069]	<0.001	<0.001	0,174	<0.001
Eutrophie	5.78 [5.64;5.85]	5.67 [5.61;5.80]	5.73 [5.66;5.82]	0,025	0,077	0,872	0,009
PM10	33.3 [31.6;36.0]	33.1 [30.8;35.5]	33.6 [32.6;34.5]	0,069	0,161	0,356	0,075
NO2	33.5 [30.0;38.5]	32.9 [28.6;36.8]	33.8 [31.7;36.1]	0,031	0,063	0,503	0,035
Score_Env	1150 [928;1546]	1168 [886;1481]	1144 [984;1396]	0,609	0,647	0,647	0,647

Le tableau 9 montre ici qu'à l'issue de ces tests, la *p-value* du test de Kruskal-Wallis sur le Fdep est inférieure à 0,05 ce qui signifie que H_0 est rejetée. Il existe bien une différence entre les groupes, ce qui signifie que le Fdep, et donc la défaveur sociale, est significativement différentes suivant les types de clusters. De plus, les tests Post-hoc montrent tous une « *p-value* » significative, ce qui prouve que tous les clusters ont bien une défaveur sociale différentes les uns des autres.

En ce qui concerne l'occupation du sol, celle-ci est significativement différentes entre les clusters de sur-incidence et le reste du territoire de la MEL, mais pas entre les clusters de sous-incidence et les neutres. Cela suppose que l'espace urbaine possède une grande influence sur la variation sur risque de CI dans la MEL.

A propos des polluants, plusieurs montrent des différences notables entre les groupes de clusters, à l'image des métaux lourds comme le cuivre (Cu) et le palladium (Pd) ou des gaz comme le dioxyde d'azote (NO₂).

Néanmoins, comme expliqué précédemment, cette première stratégie possède une limite : celle de partir du principe que l'incidence est la même partout sur l'ensemble des IRIS composant les regroupements de cluster. Donc pour affiner ces résultats, la deuxième stratégie impliquant de réaliser ces tests par cluster a été effectuée.

4.2 Deuxième stratégie

Le tableau situé en Annexe 3 montre la réalisation des tests de Kruskal-Wallis (colonne « p ») et les tests Post-Hoc (15 dernières colonnes) par variables de caractérisation pour la deuxième stratégie.

Au regard de ce tableau, plusieurs informations sont à souligner. Tout d'abord, la *p.overall* indique que tous les clusters sont significativement différents les uns des autres, ce qui signifie que pour l'ensemble des variables il existe des différences de distribution de valeur significative entre les clusters.

En ce qui concerne le Fdep, il est intéressant de constater que le cluster 1 est particulièrement différent de tous les autres clusters, y compris l'autre cluster de sur-incidence, le cluster 4. Cela indique que la défaveur sociale est différente entre ces deux clusters et que, comme le cluster 1 connaît une plus forte défaveur que le cluster 4, la défaveur y est significativement plus importante. Aussi, presque tous les clusters de sous-incidence semblent être similaires car leur *p-value* n'est pas significative, sauf la comparaison entre les clusters 3 et 5. En effet, celui-ci montre une *p-value* de 0,011, donc inférieure à 0,05. Cependant, étant donné que cette valeur est l'une des plus proche de la non-significativité, celle-ci pourrait révéler un manque de puissance de puissance statistique, notamment car le cluster 5 est composé de seulement 20 IRIS, ce qui fait fortement varier les résultats.

D'après ces tests, il semblerait que les clusters 1 et 4 soient relativement similaires en termes d'occupation agricole et urbaine du sol, tandis que les clusters 2 et 5 ainsi que 3 et 5 sont aussi très similaires. Néanmoins, il est possible d'observer que clusters 2 et 4 ne sont pas significativement différents, ce qui pose ici une limite à la fois sur l'interprétation des résultats, mais aussi sur la formation des variables, notamment la variable « urbaine » qui a agrégée plusieurs variables d'occupation du sol qui peuvent se retrouver à la fois en zone urbaine et en zone rurale (exemple : le logement. Il est plus fort en zone urbaine mais pas inexistant en zone rurale). Par conséquent, le reste des analyses concernant l'occupation du sol n'a pas été réalisée, puisque son interprétation reste trop fragile. Néanmoins, ces tests donnent une indication intéressante en termes notamment de pollution, puisque si la sur-incidence à tendance à se retrouver dans des zones urbaines, les sources de éventuels polluants significatifs à analyser seront donc ceux émis en zone urbaine, comme par exemple le chauffage résidentiel, le trafic routier, les industries...

A propos des polluants, il semblerait que la répartition des polluants ne soit pas forcément inégale en fonction des clusters. En effet, la répartition, par exemple, des Poussières (pour rappels, les « poussières » rassemblent l'ensemble des Matières Particulaires) indiquent qu'il n'existe pas de différence entre leur proportion au sein des IRIS neutres et des cluster 2 (sous-incidence) et 4 (sur-incidence), tout comme la répartition entre les clusters 1 (sur-incidence) et 3 (sous-incidence). *A contrario*, le Nickel (Ni) et le Dioxyde d'azote (NO₂) a tendance à être significativement différent entre les clusters, ce qui indique qu'il n'est pas également réparti sur le territoire.

Cependant, après avoir constaté ces résultats, un détail semble être important à souligner et peut constituer une limite à ces tests. Il s'agit des résultats sur les comparaisons liées au Fdep.

5. Limites des stratégies d'analyse de la caractérisation des clusters

Les résultats concernant le Fdep montrent que la défaveur sociale caractérise fortement les clusters. En effet, les clusters de sur-incidence montrent une relative défaveur sociale et les clusters de sous-incidence sont eux caractérisés par une population plus favorisée, ou en tout cas, moins défavorisée. De ce fait, tout porte à croire que la défaveur sociale est un facteur de risque des cardiopathies ischémique dans la MEL. Néanmoins une question se pose : comment être sûr que les clusters qui ont été détectés ne sont pas simplement des clusters de forte défaveur sociale et des clusters de populations favorisés ? La défaveur sociale serait-elle un facteur de confusion ?

Pour vérifier cela, un ajustement a été effectué sur la défaveur sociale, à l'instar des standardisations effectuées plus tôt sur l'âge et le sexe. L'objectif ici est de partir du principe que la défaveur sociale est la même partout. Après cet ajustement, des nouveaux SIR lissés et ajustés sur la défaveur seront donc établis. L'objectif est d'observer si, à la suite de ces ajustements, il existe toujours une variabilité de l'incidence qui ne serait pas expliquée par la défaveur sociale. Si le cas se présente, la question qui se posera alors sera : cette variabilité non-expliquée pourrait-elle être expliquée par la pollution ?

PARTIE III – La défaveur sociale : unique facteur de risque ?

SOMMAIRE

PARTIE III – La défaveur sociale : unique facteur de risque ?	75
Chapitre 1 : Ajustement des SIR lissés sur la défaveur sociale et distribution spatiale de l'incidence7!	
1. Objectif de l'ajustement des nouveaux SIR lissés	75
2. Ajustement et lissage des SIR ajustés sur la défaveur sociale par régression écologique.....	76
3. Résultats des SIR lissés ajustés sur la défaveur sociale	76
Chapitre 2 : Détection de clusters spatiaux ajustés sur la défaveur sociale	78
1. Objectif de la détection de nouveaux clusters spatiaux	78
2. Technique d'ajustement par régression de Poisson	79
Chapitre 3 : Régressions écologiques ajustées sur la défaveur sociale.....	80
1. Objectif de la régression écologique ajustée	80
2. Méthodologie de l'analyse bivariée	80
3. Résultat des régressions écologiques ajustées sur la défaveur sociale	82

Chapitre 1 : Ajustement des SIR lissés sur la défaveur sociale et distribution spatiale de l'incidence

1. Objectif de l'ajustement des nouveaux SIR lissés

Comme expliqué pour conclure la partie précédente, la défaveur sociale semble constituer un facteur de confusion des résultats tellement les liens entre situation socioéconomique et incidence démontrés sont significatifs.

Le principe de cette nouvelle approche est de calculer désormais des rapports d'incidence standardisés sur les facteurs de confusions connus (l'âge et le sexe), lissés, et ajustés cette fois-ci sur les valeurs du Fdep pour chaque IRIS. L'objectif de cet ajustement est d'observer si, en partant du principe que la défaveur sociale est la même sur l'ensemble de la MEL, des IRIS affichent toujours des SIR remarquables (sur-incidence ou sous-incidence).

Pour réaliser cet ajustement, une technique statistique appelée « régression écologique » a été réalisée. Dans un premier temps, il s'agira d'observer quelle serait la répartition des SIR si la défaveur sociale était la même sur l'ensemble du territoire, puis, dans un second temps et ce qui constitue l'objectif principal de cet ajustement, il s'agira d'obtenir des SIR résiduels. Ces

résidus statistiques détermineront la part de la variabilité spatiale qui n'est pas expliquée par le Fdep.

2. Ajustement et lissage des SIR ajustés sur la défaveur sociale par régression écologique

Une régression écologique est un modèle prédictif qui va permettre de déterminer comment une valeur va évoluer suivant l'évolution d'une des variables. L'adjectif « écologique » désigne ici le fait que les variables prises en compte dans le modèle sont calculées à l'échelle d'un groupe/d'une population et non à l'échelle individuelle (comme le Fdep). La formule est basée sur le même modèle que le lissage des SIR, dans lequel est introduit une variable écologique d'ajustement, ici le Fdep. Cette formule est la suivante :

$$\log(\theta_i) = \beta_0 + U_i + V_i - \log(e_i) + \beta_1 Fdep_i,$$

Où $Fdep_i$ = Valeur du Fdep pour l'unité spatiale i .

Le résultat de cette opération indiquera le SIR ajusté sur la défaveur sociale. Afin de connaître l'estimation du SIR non-imputable à la défaveur sociale, les SIR résiduels ont été calculés de la manière suivante :

$$\theta_i^{res} = e^{\beta_0 + U_i + V_i - \log(e_i)}$$

Cet indicateur représente la part du risque relatif de CI non-imputable à la défaveur sociale. En d'autres termes, et en ce qui concerne la régression écologique sur le Fdep, les résidus statistiques représentent donc le risque relatif restant après ajustement sur le Fdep, d'où le terme de « résidus ». Ici, cette méthode statistique sert donc à démontrer si la défaveur sociale peut à elle seule expliquer la variation de l'incidence des cas de CI dans la MEL.

3. Résultats des SIR lissés ajustés sur la défaveur sociale

Une fois la régression effectuée, la répartition de la défaveur sociale est considérée comme étant identique sur tout le territoire de la MEL. La régression écologique sur la défaveur sociale donne donc lieu à une nouvelle répartition spatiale des SIR. La carte 22 à gauche montre les SIR lissés après cet ajustement, tandis que la carte 23 à gauche affiche les résidus

statistiques de cet ajustement, afin de constater s'il existe des IRIS où la défaveur sociale ne peut expliquer les variations spatiales d'incidence.

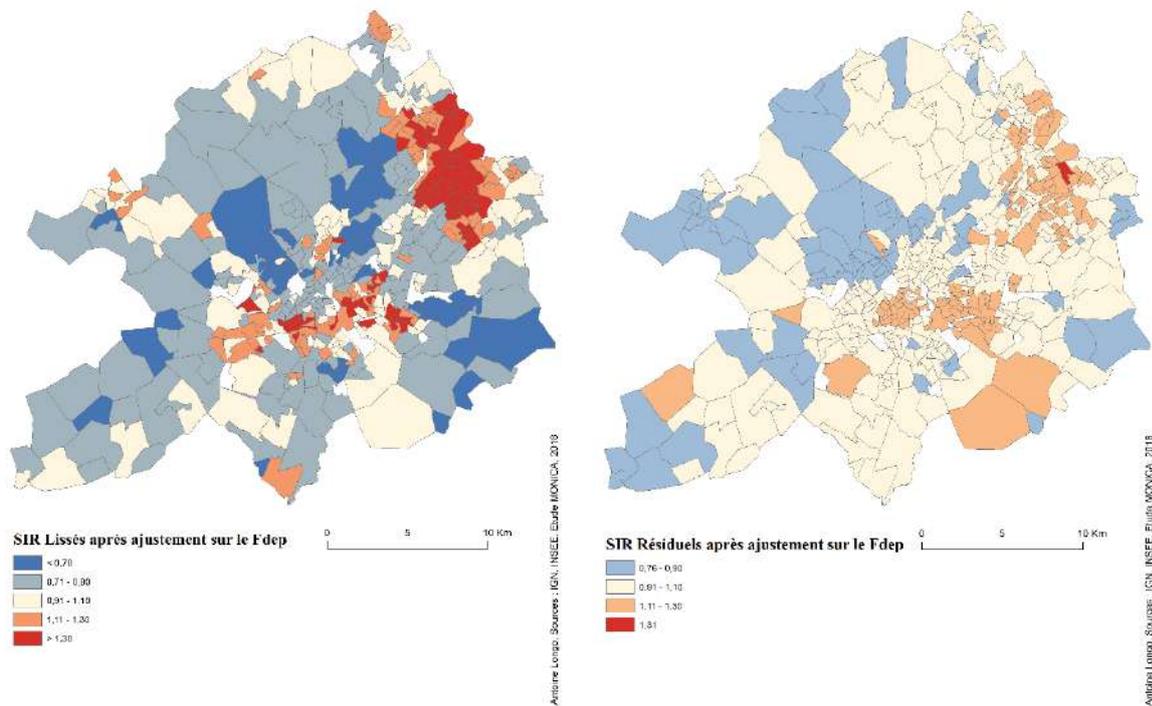


Figure 21 : Carte des SIR Lissés après ajustement sur la défaveur (gauche)

Figure 22 : Carte des résidus de SIR lissés après ajustement sur la défaveur sociale (droite)

D'après la carte 21, la répartition des SIR lissés ajustés diffère de la répartition des SIR lissés sans ajustement. Plusieurs zones recueillant des IRIS neutres se montrent désormais avec une faible sur-incidence. Mais plus précisément, si le niveau de défaveur sociale était le même sur toute la MEL, les zones de sur-incidence seraient toujours aussi remarquables, notamment les IRIS du Sud de la commune lilloise et les communes de Tourcoing et Roubaix.

Aussi, plusieurs zones semblent désormais voir leur SIR lissés être modifiées. Par exemple, le Sud de la MEL montre une répartition des SIR lissés ajustés plus commune, avec des profils identiques de sous-incidence, là où quelques IRIS montraient de la sur-incidence avant l'ajustement. Cela est aussi le cas pour quelques IRIS au Sud-Ouest de Lille ou au Sud-Est de la Métropole. L'interprétation qui peut en être faite est que la défaveur sociale y joue un rôle primordial dans les variations spatiales d'incidence, et qu'après un ajustement sur celle-ci, le risque de CI s'amointrie, jusqu'à parfois même devenir négatif et définir un territoire de sous-risque.

En outre, d'après la carte 22 qui représente la répartition des SIR résiduels après ajustement sur le Fdep, les IRIS qui ne sont pas neutres sont ceux qui indiquent que la seule défaveur sociale ne peut expliquer la variation de leur incidence. Sur cette carte, il est possible d'observer qu'après un ajustement des SIR lissés, il persiste une hétérogénéité spatiale de la répartition de l'incidence des CI dans la MEL. Quelques IRIS au Nord-Ouest de Lille et au Sud de la Métropole montrent un SIR résiduel plus faible que les autres. Ceux-là faisaient déjà parti des clusters de sous-incidence précédemment. Dans le même temps, une grande majorité des IRIS de la commune de Lille ainsi que ceux de l'Est de Lille et une grande partie de ceux des communes du Nord composés de Tourcoing-Roubaix-Wattrelos montrent un SIR résiduels supérieurs aux autres. L'un de ces IRIS, celui de Crétinier à Wattrelos, semble d'ailleurs toujours afficher un SIR résiduel plus fort que les autres. Celui-ci ainsi que l'ensemble des SIR résiduels de sur-incidence indique que la défaveur sociale ne peut pas expliquer à elle seule la variation de l'incidence des cas de CI dans la MEL.

A présent, la question est de savoir si des clusters peuvent encore être détectés après cet ajustement sur la défaveur sociale. Pour vérifier cela, une nouvelle détection de cluster spatiaux a été réalisée. Cette détection permettra de savoir si des clusters peuvent être détectés grâce à ces nouveaux SIR lissés ajustés. Ceux-là pourraient potentiellement être expliqué par la pollution atmosphérique.

Chapitre 2 : Détection de clusters spatiaux ajustés sur la défaveur sociale

1. Objectif de la détection de nouveaux clusters spatiaux

Après le calcul des SIR lissés et ajustés sur la défaveur sociale, une nouvelle détection de cluster est relancée. Cette nouvelle détection répond à la question : « en partant du principe que la défaveur sociale est la même sur l'ensemble de la MEL, y a-t-il toujours des zones d'incidence remarquable ? ». L'objectif est donc d'observer si des IRIS aux profils d'incidence similaires se regroupent toujours en clusters déterminant de zones « à risque » et des zones de moindre risque ou de non-risque. Cette détection vise donc à connaître l'impact de l'autre facteur environnemental mis à l'étude dans le projet MONICA, à savoir la pollution atmosphérique.

2. Technique d'ajustement par régression de Poisson

Au sein du logiciel SaTScan qui met en œuvre une grande majorité des modèles de statistiques de scan spatiaux, il existe deux méthodes afin de réaliser des ajustements sur des facteurs de confusion de l'incidence. La première consiste à utiliser un fichier de population stratifié sur les facteurs de confusion afin de réaliser une standardisation indirecte. Cette méthode, simple de prime abord, nécessite tout de même de disposer des données de population sous-jacente pour chacune des combinaisons de strates des facteurs de confusion (*e.g.* : la population d'hommes âgés de 35 à 44 ans d'un IRIS). Par ailleurs, le facteur de confusion se doit d'être qualitatif afin de réaliser une standardisation indirecte. La deuxième méthode d'ajustement consiste à modéliser le nombre de cas observés par unité spatiale par un modèle de régression de Poisson considérant le facteur de confusion comme variable explicative. Ainsi, les effectifs prédits par le modèle sont considérés comme ajustés sur le facteur de confusion et peuvent être introduits directement dans le modèle de statistique de scan spatial.

Dans le cas de l'indice de défaveur sociale, la première méthode d'ajustement n'est pas applicable car i) l'indice est de nature quantitative et ii) même après discrétisation de l'indice (*e.g.* : quartiles), les données de population par sexe, classe d'âge et classe de défaveur ne sont, à notre connaissance, pas disponibles. Aussi, la deuxième méthode basée sur une régression de Poisson a été utilisée. De surcroît, en prenant comme variable d'*offset* les effectifs attendus issus de la standardisation indirecte sur l'âge et le sexe, cette méthode a permis d'ajuster la détection de cluster à la fois sur l'âge, le sexe et la défaveur sociale.

3. Résultats

Les *p-value* de la seconde détection de cluster sont décrits dans le tableau 10 ci-dessous.

Tableau 10 : détection de clusters avec ajustement sur la défaveur sociale

Numéro du cluster	<i>p-value</i>
1	0.1
2	0.2
3	0.265
4	0.968
5	0.994

Le tableau 10 ci-dessus montre qu'après ajustement sur la défaveur sociale, 5 clusters sont détectés. Néanmoins, il ne s'avère qu'aucun de ces clusters n'est significatifs. Cela constitue

un résultat fondamental car cela signifie qu'après ajustement sur la défaveur sociale, la pollution atmosphérique n'a pas d'influence globale sur les variations spatiales d'incidence des CI dans la MEL.

Ces résultats expriment donc le fait que la défaveur sociale constitue un facteur capital dans l'étude des variations spatiales d'incidences des cas de cardiopathies ischémiques dans la MEL. La défaveur sociale constitue donc une problématique et un enjeu majeur en ce qui concerne la santé cardiaque au sein de la Métropole lilloise.

A présent, malgré cette absence de cluster liée à la pollution atmosphérique, la question est de savoir si, parmi les polluants analysés dans l'étude, certains pouvaient avoir une influence dans la variabilité des SIR. Pour vérifier cela, une régression écologique a été réalisée, mais cette fois-ci ajustée sur la défaveur sociale.

Chapitre 3 : Régressions écologiques ajustées sur la défaveur sociale

1. Objectif de la régression écologique ajustée

Le principe de cette nouvelle approche est de calculer désormais le risque relatif imputable à chaque polluant, ajustés sur la défaveur sociale. L'objectif de cette régression est de déterminer des facteurs prédictifs de variabilité spatiale d'incidence des CI. En d'autres termes, il s'agira d'observer si, en partant du principe que la défaveur sociale est la même sur l'ensemble de la MEL, certains polluant affichent un risque relatif significatif.

Cette régression a été réalisée en deux temps. Premièrement des analyses bivariées, ajustées sur la défaveur sociale ont été réalisées. Deuxièmement, un modèle multivarié a été réalisé.

2. Méthodologie de l'analyse bivariée

Une analyse bivariée est modèle de régression écologique qui va permettre ici d'identifier quels sont les polluants qui sont associées à la variabilité spatiale des CI. Il y a donc ici 20 analyses bivariées. Toutes les analyses sont ensuite agglomérées dans un graphique appelé « *forestplot* ».

Ces modèles se caractérisent par un graphique de type *forestplot*. Au sein de ces graphiques, chaque polluant est intégré et la distribution de leurs valeurs est divisée en 3 groupes, représentés par 3 points noirs. Chaque point représente la valeur du risque relatif d'un des trois quartiles les plus élevés de la distribution des valeurs du polluant (*i.e* les 75% d'IRIS¹⁷ les plus impactés par le polluant analysé) comparé au premier quartile (25% les moins touchés). En d'autres termes, lorsque que le premier point d'une variable affiche un Risque Relatif (RR) de 2, il s'agit du RR concernant le deuxième quartile de la variable représentée. Cela signifie que pour le deuxième quartile de cette variable, les IRIS qui en font partie multiplient leur risque de CI par 2 par rapport au fait d'être dans le premier quartile le risque de CI.

Une barre horizontale traverse chaque point : représente l'intervalle de crédibilité, qui indique la zone de valeur dans laquelle le risque relatif peut varier. Enfin, une ligne verticale à gauche du graphique représente la valeur 1 du RR.

Un quartile de polluant est significatif lorsque son intervalle de crédibilité ne contient pas la valeur 1. Si cet intervalle comprend la valeur 1, cela signifie que le risque relatif varie au-dessus et sous la valeur 1, et que donc il est impossible de déterminer si le polluant a une influence sur un sur-risque ou un sous-risque. *A contrario*, un intervalle de crédibilité étant strictement au-dessus de la valeur 1 révélera donc que le polluant en question peut être un facteur prédictif significatif de la variabilité des CI dans la MEL.

La formule pour la régression écologique ajustée sur les polluants revient à réaliser la même régression que celle ajustée sur le Fdep, en y introduisant une valeur de pollution.

$$\log(\theta_i) = \beta_0 + U_i + V_i - \log(e_i) + \beta_1 Fdep + \beta_2 Polluant_j$$

Où Polluant_j = Valeur d'un polluant pour l'unité spatiale observée.

Si des polluants affichent des résultats significatifs dans les analyses bivariées, un modèle multivarié sera alors mis en place. Ce modèle sert à vérifier comment évolue une variable en fonction de toutes les autres variables. De plus, la défaveur sociale sert de variable d'ajustement à ce modèle.

¹⁷ Les quartiles sont représentés par Q2, Q3 et Q4 et sont comparés à Q1. Q1 représente les 25% d'IRIS les moins touchés, Q2 représente les IRIS situés dans les 25% à 50% les moins touchés, Q3 les 50% à 75% et Q4 les 75% à 100%.

3. Résultat des régressions écologiques ajustées sur la défaveur sociale

3.1 *Analyse bivariée*

Les tableaux 11 et 12 ci-dessous montrent les résultats des 19 analyses bivariées (correspondant au nombre de variables de polluants).

Tableau 11 : Risque relatif (Intervalle de crédibilité à 95%) de cardiopathie ischémique par variable de pollution (Partie 1))

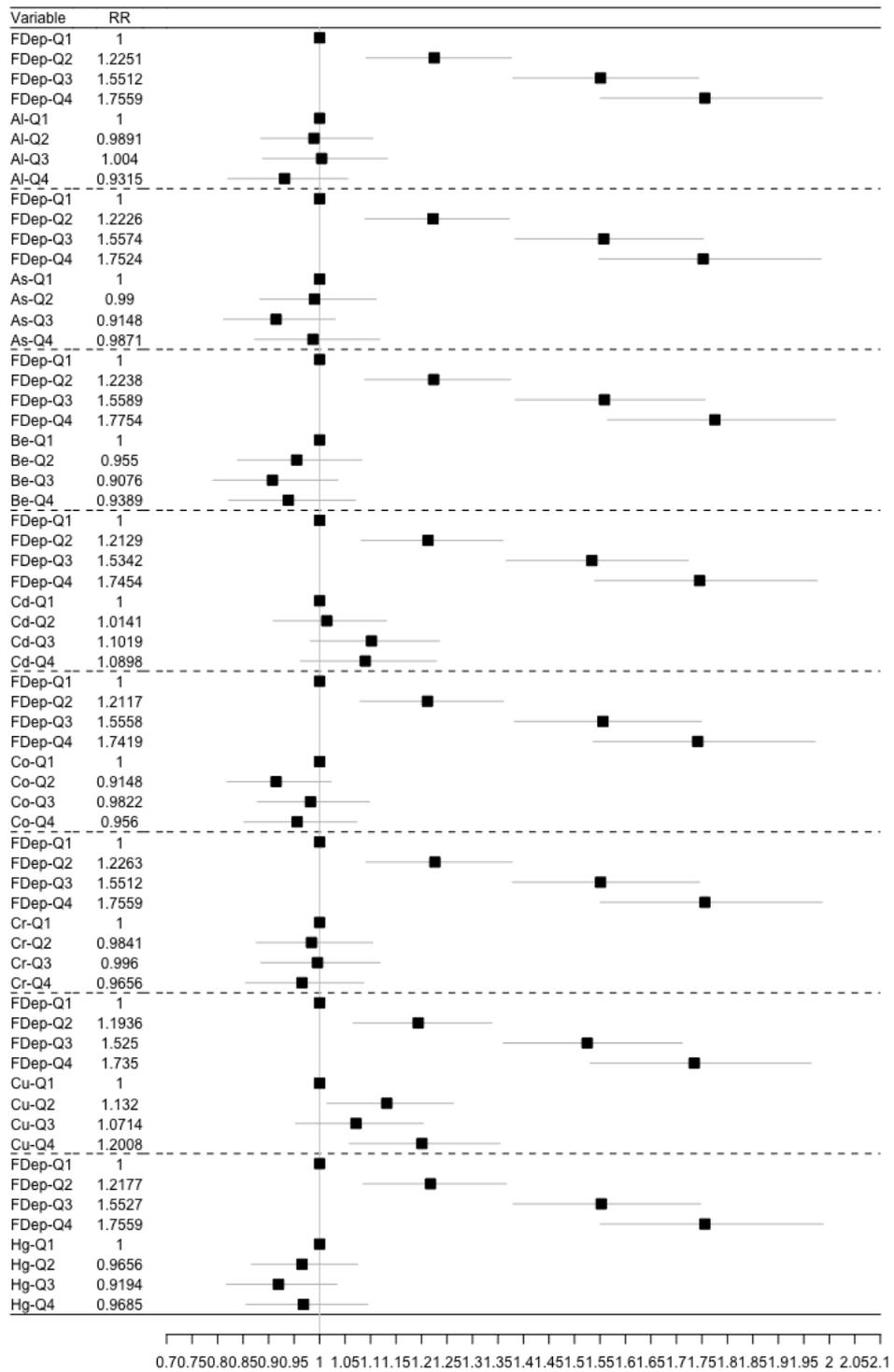
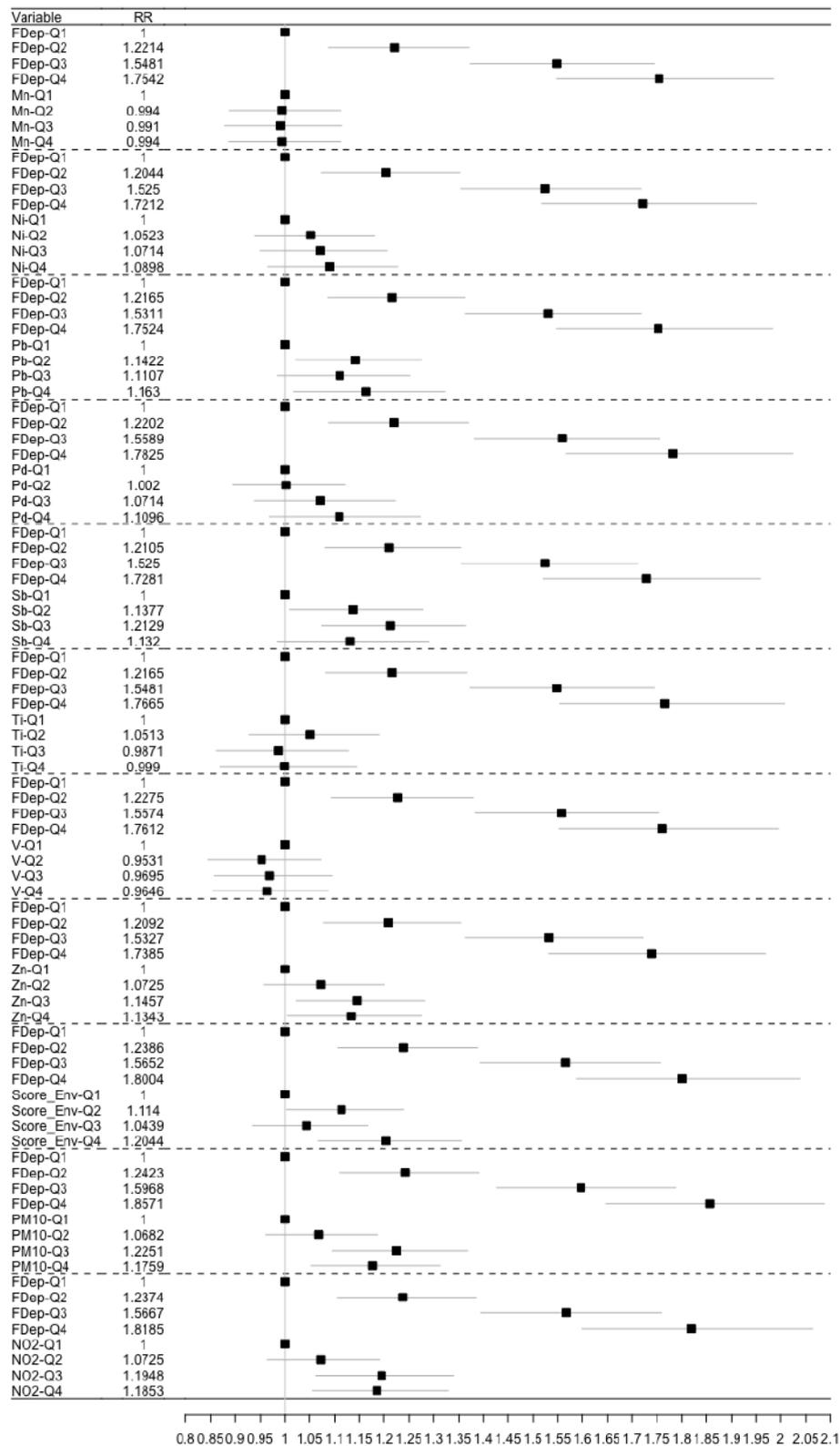


Tableau 12 : Risque relatif (Intervalle de crédibilité à 95%) de cardiopathie ischémique par variable de pollution (Partie 2)



D'après ces *forestplot*, plusieurs résultats sont à signaler. Tout d'abord, il est possible d'observer plusieurs différentes répartitions des polluants. En effet, on distingue tout d'abord que certains polluants n'ont pas de variabilité significative, ce qui indique que leur seule

présence ne suffit pas à faire varier l'incidence, comme par exemple le Manganèse (Mn, tableau 12), le plomb (Pb, tableau 12) ou encore l'Aluminium (Al, table 11).

D'autre part, certains polluants suivent ce qui s'apparente à un « effet dose ». Cela indique que le risque relatif augmente lorsque la dose de polluant augmente à son tour. Cet effet est d'ailleurs déjà illustré avec la défaveur sociale : dès que le niveau de défaveur sociale augmente, plus le risque relatif augmente. En ce qui concerne les polluants, c'est le cas notamment pour le Nickel (Ni, tableau 12) ou encore pour le Palladium (Pd, tableau 12). Néanmoins, les intervalles de crédibilités franchissent tous la valeur 1, ce qui indique que le risque relatif varie au-dessus et sous la valeur 1, et que donc il est impossible de déterminer si le Palladium a une influence sur un sur-risque ou un sous-risque. Il n'y a donc aucun polluant qui suit un « effet dose » significatif ».

Enfin, certains polluants ont une distribution qui suivent un « effet seuil ». Cela indique que, dépassé un certain seuil, leur impact devient significatif. Il est possible de distinguer cet effet notamment en regardant la pour le Dioxyde d'Azote (NO₂, tableau 12) ainsi que sur le Zinc (Zn, tableau 12). Pour ces polluant, ce seuil semble être représenté par leur médiane. Cela indique qu'à partir du moment où la concentration de polluant dépasse la médiane des valeurs répartie sur la MEL, le risque relatif lié à ces polluants devient plus fort. Cette répartition semble être aussi la même en ce qui concerne le Cadmium (Cd, tableau 11), mais les intervalles de crédibilités franchissent tous

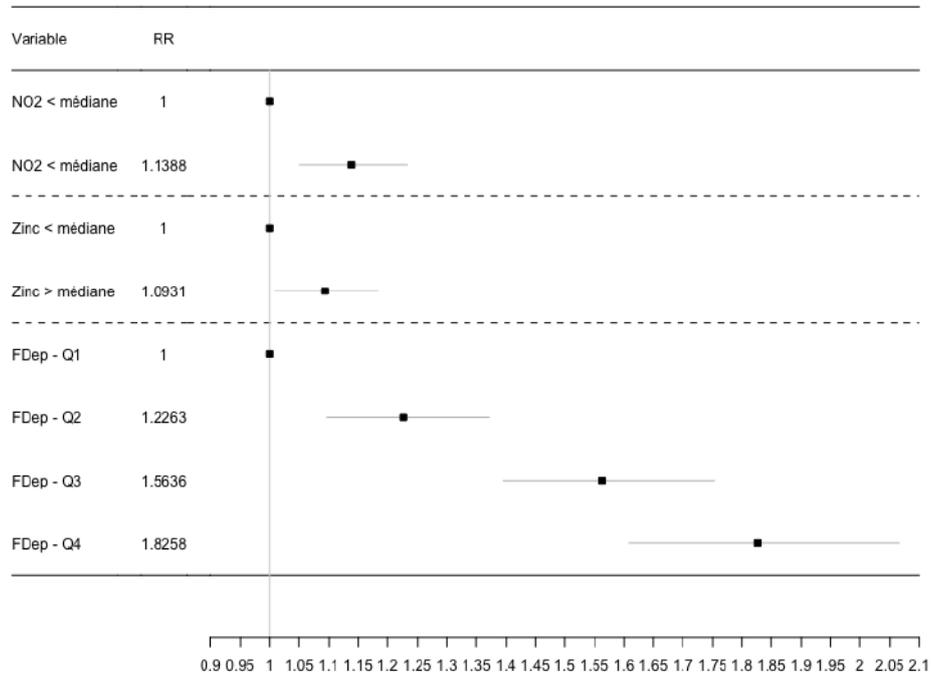
D'après ces résultats, il est donc possible de distinguer 2 polluants ayant un « effet seuil » significatifs (NO₂ et Zinc) qui peuvent avoir un impact sur le risque relatif d'incidence des cas de CI dans la MEL. Ces deux polluants sont donc intégrés dans un modèle d'analyse multivariée

3.2 Analyse multivariée

Le graphique X va donc indiquer comment varient le NO₂ et le Zinc par rapport à l'incidence ajustée sur le niveau du Fdep. Le point noir représente risque relatif lié aux polluants, et la ligne verticale représente la médiane du RR (1). Cependant, contrairement aux analyses bivariées, les polluants ont cette fois été divisés en deux classes. Les 50% des valeurs les plus élevées sont comparées aux 50% des valeurs les plus faibles. En effet, étant donné que les effets « seuils » de ces polluants se situent au niveau de leur médiane, ce sera donc celle-ci qui sera utilisée pour l'analyse. Elle est représentée par un point dans le tableau 13.

Encore une fois, si la risque relatif associé à chaque modalité de chaque polluant, ainsi que leurs intervalles de crédibilité, ne franchissent pas la valeur 1, le polluant sera alors considéré comme un facteur prédictif significatif de la variabilité des cardiopathies ischémiques.

Tableau 13 : Forestplot multivarié par polluant montrant un "effet seuil"



Pour les deux polluants, les risques relatifs ainsi que leur intervalle de crédibilité dépassent la valeur 1. Par conséquent cela signifie que, dans la MEL, le Zinc et le Dioxyde d'Azote ont un impact significatif sur les variations spatiales du risque d'incidence de cardiopathies ischémiques.

DISCUSSION

Cette étude concernant le registre MONICA de la Métropole Européenne de Lille a permis de mettre en lumière plusieurs points forts et inédits en ce qui concerne les variations spatiales d'incidence sur le territoire de la MEL à une échelle fine : l'IRIS.

Tout d'abord, des calculs d'incidence bruts, puis standardisés aux facteurs de confusion connus (l'âge et le sexe) ont été calculés afin de connaître le nombre de cas attendus pour chaque IRIS. Ce calcul de cas attendus a permis par la suite de calculer des rapports d'incidence standardisés (SIR) bruts qui ont servi à avoir une première approche sur les variations spatiales de l'incidence. Ces SIR ont ensuite été lissés par une méthode de lissage bayésien, qui prend notamment en compte le comportement des IRIS voisins. Grâce à ce lissage, des SIR lissés ont pu être calculés afin de constater qu'il existe une. Puis grâce à ces SIR lissés, une détection de cluster a été réalisée.

Trois clusters de sous-incidence et deux clusters de sur-incidence ont été détectés. Ces clusters ont ensuite été caractérisés tout d'abord par trois facteurs environnementaux : la défaveur sociale, la pollution atmosphérique et l'occupation du sol, bien que celle-ci n'ait pas donné lieu à des analyses plus poussées car sa composition n'était pas assez fiable. Cela est notamment dû à la source de données provenant de la Région Hauts-de-France. Le découpage des différents postes d'occupation effectuée par leurs équipes réunissant 52 postes dont 49 ne définissaient seulement que les espaces ruraux. D'autres sources ont été aussi discutées, notamment la base Corinne Land Cover qui est l'une des plus réputées, mais encore une fois les postes d'occupation du sol étaient plus détaillés pour décrire les espaces ruraux que pour décrire les espaces urbains. Le regroupement des postes effectués par la suite s'est inspiré des postes d'occupation du sol des bases de l'Institut d'Urbanisme de Paris. Néanmoins, comme l'espace urbain fut moins finement détaillé que l'espace rurale, cela a rendu les regroupements peu fiables et donc une caractérisation du territoire moins fine.

La caractérisation a ensuite été effectuée suivant deux stratégies : des tests de comparaisons par groupe de clusters, puis par clusters.

Après cette caractérisation, il a semblé que la défaveur sociale avait un impact conséquent sur l'influence des résultats. Pour affiner les recherches, la défaveur sociale a alors été considérée comme facteur de confusion, et un ajustement des SIR lissés a été réalisé afin de rendre compte de la variabilité de l'incidence des cardiopathies ischémiques dans la MEL en partant

du principe que la défaveur sociale était la même sur l'ensemble du territoire. Ainsi, après un ajustement sur la défaveur sociale, une hétérogénéité spatiale subsistait. En effet, plusieurs IRIS ont montré des SIR résiduels élevés, ce qui indiquaient que le risque ne pouvait pas seulement s'expliquer par la défaveur sociale.

De ce fait, une nouvelle détection de cluster ajustée sur la défaveur sociale a été réalisée. Cependant, les clusters spatiaux détectés ne se sont pas avérés statistiquement significatifs. Alors une régression écologique a été menée sur les polluants, afin de déterminer si, parmi les polluants de l'étude, certains pouvaient être considérés comme facteurs prédictifs tout en ajustant sur la défaveur sociale. Ce fut le cas pour le Zinc et pour le Dioxyde d'azote, d'après une analyse bivariée dans un premier temps, puis une analyse multivariée.

Les résultats de cette recherche indiquent donc qu'après avoir prouvé qu'il existait une hétérogénéité spatiale des variations d'incidence dans la MEL, des clusters d'incidence ont été révélés. Ceux-ci montrent que plus la défaveur sociale est importante, plus le risque de survenu d'un événement de CI est important. Puis après un ajustement sur la défaveur sociale, il a été prouvé que les pollutions atmosphériques constituées de Zinc et de Dioxyde d'Azote pouvaient expliquer la variabilité résiduelle, c'est-à-dire celle qui n'était pas expliquée par la défaveur sociale.

Parmi ces clusters d'IRIS, un IRIS semble toujours afficher un risque plus élevé que le reste : il s'agit de l'IRIS Crétinier, à Wattrelos. De manière plus globale, ce sont la majorité des IRIS à grande majorité urbaine ainsi que ceux qui composent les communes de Wattrelos, de Roubaix et de Tourcoing qui sont parmi les zones les plus « à risque » dans l'ensemble du territoire de la MEL. *A contrario*, des IRIS, plus ruraux, ont été détectés comme faisant partie de cluster de sous-incidence. Ils sont situés plutôt au Nord-Ouest de la commune de Lille, à l'Est et dans la pointe Sud de la MEL. Ceux-ci montrent une sous-incidence qui semble être liée à moins de défaveur sociale, mais aussi à une moindre présence de polluant, en général.

Une recherche récente (Backholer et *al.*, 2016) a été menée pour réaliser une méta-analyse¹⁸ sur 44 articles ayant établi un lien entre défaveur sociale et maladies cardiovasculaires (MCV). 22 millions d'individus répartis dans 116 cohortes (22 d'Asie, 75 d'Europe, 7

¹⁸ Une méta-analyse en santé désigne un regroupement de plusieurs recherches de tailles plus modestes afin de réaliser une étude statistique où le nombre de cas serait plus élevé. L'objectif est d'obtenir, par la suite, des conclusions plus générales.

d'Amérique du Nord, 12 d'Australasie¹⁹) ont été concernés par cette étude. Les individus étaient âgés de 20 à 107 ans. Toutes les études n'avaient pas les mêmes variables de caractérisation de la défaveur sociale, mais celle qui revenait majoritairement était le niveau d'éducation. L'analyse sur cette première variable a montré que le Risque Relatif de CI était plus élevé chez les hommes et les femmes ayant les niveaux d'éducatifs les plus faibles comparés à ceux dont les niveaux étaient plus élevés. Puis des analyses concernant des variables alternatives d'estimation de la défaveur sociale (le revenu, la catégorie socio-professionnelle et la défaveur de l'air urbaine) ont été réalisées et il s'avère que ces variables indiquent les mêmes résultats que ceux du niveau d'éducation. Une association significative a donc été trouvée par la suite entre l'augmentation de la défaveur sociale et le risque de CI.

D'autres études avaient déjà mis en avant les relations clusters de défaveur sociale et mortalité due aux maladies cardiovasculaires, notamment dans le comté de Harris, au Texas aux États-Unis entre 2008 et 2010 (Ford & Highfield, 2016), ou bien dans la commune d'Adélaïde, au sud de l'Australie, entre 2010 et 2012 (Bagheri *et al.*, 2015). Toutes ces études avaient une échelle d'étude comparable à l'IRIS française. L'étude MONICA basée sur le registre Lillois s'inscrit donc dans la continuité de ces recherches et figure parmi les études pertinentes concernant la recherche de liens entre défaveur sociale et risques autour des maladies cardiovasculaires de toute nature.

En ce qui concerne la pollution atmosphérique comme facteur explicatif de la variabilité spatiale de l'incidence, les études sont de plus en plus nombreuses à s'y intéresser. Par exemple, une étude concernant aussi le projet MONICA a été réalisée dans l'objectif d'observer des potentiels liens entre incidence des MCV et pollution de l'air, mais cette fois-ci dans la ville de Göteborg, en Suède (Stockfelt *et al.*, 2017). Celle-ci s'est effectuée sur 10 350 individus répartis dans 2 cohortes : la *Primary Prevention Cohort* (PPS) qui s'intéresse à un tiers des hommes Götensbourgeois nés entre 1915 et 1925, et une cohorte tirant des cas incidents au hasard du registre MONICA des années 1985, 1990 et 1995 âgés entre 25 et 64 ans. Ces deux cohortes ont été suivies entre 1990 et 2011. Toutes ces personnes ont renseigné leur adresse et ont été géocodées au quartier de résidence, puis un redécoupage en carroyage a été effectué afin de représenter plus facilement les variations des polluants. Les polluants pris en compte étaient les Matières Particulaires (PM 2,5 et 10) et le Noir de Carbon (BC). Les résultats de cette étude ont montré qu'il existait une relation significative entre les PM 2,5 et 10 et les cas

¹⁹ Australasie : région d'Océanie ayant des délimitations floues, composée principalement de l'Australie, de la Nouvelle-Zélande et des îles du Pacifique alentour.

de cardiopathies ischémiques issus de la cohorte PPS, notamment lors des cinq dernières années de l'étude. Mais surtout, il a été montré que ces polluants avaient aussi une relation avec des cas féminins du registre MONICA. Ce résultat indique l'une des perspectives des études concernant le projet MONICA, à savoir comment affiner la prévention des maladies cardiovasculaire et des CI par sexe. Quelques avancées ont déjà été soulignée, principalement en ce qui concerne l'études des différents symptômes d'un cas de CI (Berg et *al.*, 2009 ; Goldberg et *al.*, 1998).

Les polluants qui ont été pris en compte pour l'étude MONICA lilloise n'ont, eux, pas tous montré une relation avec une augmentation du risque d'infarctus. Néanmoins, ceux qui ont été détectés comme étant plus dangereux ont tous un lien avec le trafic autoroutier (diesel ou essence) et les activités industrielles. Bien que leur niveau ait baissé ces dernières années (Atmo Hauts-de-France, 2018) il est possible que les IRIS les plus touchés soient influencés par la traversée d'une grande voie (autoroute au sud de la ville) ou bien connaissent un fort passif industriel. L'exemple le plus marquant est celui concernant l'IRIS Crétinier, dans la commune de Wattrelos. Si aucune raison statistique ne vient expliquer pourquoi il est le seul des 473 peuplés de la MEL à connaître un SIR lissé et ajusté dépassant 1,30, l'une des théories seraient ce SIR pourraient être influencé à la fois par la chaudronnerie au Nord de l'IRIS et par l'entreprise de production de plastique à l'Est, tous deux pouvant rejeter des métaux lourds.

Néanmoins, la prise en compte d'un grand nombre de polluants constitue l'une des forces de cette étude, qui a su combiner des connaissances biologiques et environnementales pour récolter et analyser les données sur les polluants, mais aussi des connaissances géomatiques pour transposer et les affiner à l'échelle des IRIS sur le territoire de la MEL, et des connaissances statistiques pour connaître leurs variations spatiales et le degré de significativité lié à l'incidence des cas de CI. Cette étude est d'autant plus importante puisqu'elle s'inscrit dans une démarche mondiale d'analyse de facteur environnementaux, comme il est possible de le constater avec l'étude quasi-similaire à Göteborg. Enfin, l'étude du registre MONICA lillois s'est servie de techniques statistiques innovantes comme la détection de cluster par statistique de scan spatial, les calculs de SIR lissés, ou l'utilisation de régressions écologiques. Ces méthodes ont permis à la fois d'identifier une hétérogénéité spatiale sur l'ensemble de la MEL, les zones de sur-incidence et de sous-incidences les plus significatives, à une échelle fine qui sont les IRIS, et enfin de trouver les facteurs prédictifs

significatifs qui conditionnent cette hétérogénéité, à savoir la défaveur sociale et des polluants atmosphériques tels que le Zinc et le Dioxyde d'Azote.

Cependant, cette étude connaît quelques biais qui pourraient potentiellement influencer quelques chiffres dans les résultats. Premièrement, les tranches d'âges ont été découpées en 3 catégories, dont les deux dernières regroupent des cas ayant 15 années d'écart. Or, après réunion avec l'un des professeurs en épidémiologie de l'équipe de recherche de MONICA à Lille, il existe une forte variabilité intra-classe d'âge de l'incidence en général. En d'autres termes, l'incidence ne serait pas la même à 45 ans qu'à 59, et encore moins entre 60 et 74 ans. Un ajustement effectué en considérant des classes d'âge plus fines pourraient améliorer la qualité de l'ajustement sur l'âge et ainsi avoir un effet sur les résultats et les variations spatiales, notamment sur les IRIS connaissant une population majoritairement âgée ou majoritairement jeune.

De plus, l'étude n'a pas pu prendre en compte les effets de pics de pollution au sein des IRIS et durant la période d'étude, car les méthodes de biosurveillance ne permettent pas de prendre en compte cet effet. Les variations temporelles d'intensité de pollution ont donc pu être estompées par ces méthodes de captation. Mais dans ce cas il aurait aussi fallu établir une recherche sur les variations temporelles d'incidence, ce qui n'était pas l'objectif premier de cette étude mais ce qui constitue l'une des perspectives à venir.

D'un côté plus géographique, il faut rappeler que la MEL n'est pas une île mais qu'elle est un territoire continu connecté à différents espaces à la fois régionaux (Région Haut-de-France, dans le département du Nord et frontalière au sud au département du Pas-de-Calais) et nationaux (frontière nord avec la Belgique). Il est alors possible de s'intéresser à ce qui est appelé les « effets de bords », c'est-à-dire l'influence des territoires frontaliers dans l'étude. Et cela peut constituer un biais en plusieurs points. En effet, pour rappel, les cas du registre MONICA sont ceux qui ont été pris en charge dans l'un des hôpitaux situés dans la MEL et qui habitent dans la MEL. Or, une personne ayant subi un cas de CI mais qui se serait fait soigner hors de la MEL ne serait donc pas pris en compte dans le registre. Cette théorie peut notamment influencer les IRIS de la MEL proches d'un centre de soin situé hors de la MEL. Mais aussi, ces effets de bords peuvent avoir une influence sur la pollution de l'air. Effectivement, les industries à la frontière belge comme celles situées dans les alentours français de la MEL peut avoir un impact notoire sur les variations spatiales d'incidence. La figure 24 ci-dessous montre, d'après les données de l'aéroport de Lille-Lesquin, le sens des

vents dominants dans la MEL. Ces vents participent au transport de particules et peuvent jouer un rôle majeur dans ces variations et dans ce biais d'études

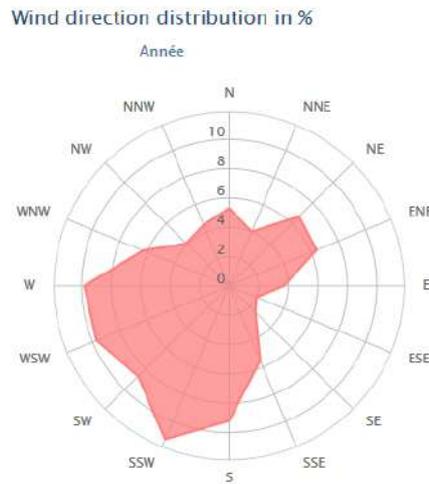


Figure 23 : Direction du vent d'après les informations de l'aéroport de Lille-Lesquin.

Ces biais constituent plusieurs perspectives des études à mener à la suite de celle-ci. En effet, après avoir analysé les variations spatiales de l'incidence des cardiopathies ischémiques dans la MEL, une analyse temporelle voire spatio-temporelle pourrait être utile, notamment pour mettre en lumière des potentiels liens entre incidence et pics de pollution, ou bien pour rechercher d'éventuelles variations saisonnières. Aussi, une prise en compte des effets de bords pourrait être à développer pour la suite car ces effets peuvent être amplifiés par plusieurs facteurs comme les différences de la qualité de la prise en charge entre les hôpitaux de la MEL et les autres ou bien distance à un cardiologue et les différences de prix pour un rendez-vous préventif.

Parmi les autres perspectives d'études en marge de celle-ci, il est à noter que des études concernant l'influence du bruit comme facteur de risque sont menés, mais pour l'instant difficile à mettre en place, la captation des mesures de bruits étant très fortement variable d'un côté d'une rue ou d'un bâtiment à un autre. Il serait donc intéressant de se pencher sur une méthode d'agrégation spatiale de données de mesure du bruit afin d'y travailler si possible à l'IRIS. Aussi, alors que cette étude s'intéresse aux cas incidence de CI dans la MEL, un travail complémentaire sur les premières récurrences seraient tout aussi important, car il s'agirait de déterminer les facteurs environnementaux qui influenceraient les récurrences et, par conséquent, les comparer avec les facteurs environnementaux d'incidence pour voir l'influence du territoire et de l'environnement dans la santé cardiaque à long-terme.

Enfin, l'une des perspectives les plus primordiales seraient de se servir de cette étude pour inciter à mettre en place une campagne de prévention à l'échelle de la MEL. Pour rappel de l'introduction, les MCV sont les premières causes de décès en France et dans le Monde, les infarctus du myocarde et les accidents vasculaires cérébraux en tête. Si un dépistage des maladies cardiovasculaires n'est pas réellement à l'ordre du jour car coûteux à réaliser (47,73€ en secteur 1, jusqu'à 80€ en secteur 2) et que les accidents qui en résultent ne peuvent pas être détectés à l'avance, la prévention sur les facteurs individuels restent néanmoins d'actualités, même de façon détournée (manger 5 fruits et légumes par jour, ne pas manger trop gras, salé, sucré, faire une activité sportive sont des conseils pour la bonne santé en générale et non uniquement pour la santé cardiaque). Une prévention sur les facteurs environnementaux seraient donc une façon d'alerter les populations sur les risques potentiels liés à une mauvaise hygiène de vie issue d'une éventuelle défaveur sociale, mais aussi sur l'impact de leur consommation d'énergie qui seraient en lien avec la pollution et *a fortiori*, avec leur santé cardiaque (industrie chimique et métallurgique, utilisation de véhicules polluants, chauffage au gaz ou au bois...).

CONCLUSION

L'étude MONICA est un des projets de l'OMS les plus importants. Les enjeux de santé publique auxquels elle s'intéresse concerne l'une des maladies les plus répandues au monde : les maladies cardiovasculaires. Si les facteurs individuels de l'étude sont connus et étudiés depuis les années 1980, les facteurs environnementaux restent, eux, moins étudiés à l'heure actuelle. Néanmoins, de plus en plus de recherches s'intéressent à ces facteurs. La défaveur sociale peut servir d'introduction aux études sur les facteurs individuels, tandis que la pollution atmosphérique est un facteur aux enjeux grandissant et est l'un des plus grands défis du XXI^e siècle sur le plan sanitaire.

Le projet MONICA de Lille a permis de faire avancer cette étude globale. Grâce à cette recherche, il a été prouvé que dans la Métropole Européenne de Lille il existe une hétérogénéité spatiale de la variation d'incidence des cas de cardiopathies ischémiques, qui désignent un type de crise aiguë déclenchée par une maladie cardiovasculaire. Au sein de cette Métropole, il existe des clusters de profils d'incidences similaires qui appuient cette découverte de variation spatiale d'incidence : 2 clusters de sur-incidence et 3 clusters de sous-incidence. Ces clusters ont été caractérisés par deux facteurs environnementaux : la défaveur sociale et la pollution atmosphérique. Grâce à cette caractérisation, il a été prouvé que la défaveur sociale avait un impact conséquent sur l'influence des résultats. Ensuite, il a été vérifié que certains polluants pouvaient être considérés comme facteurs prédictifs du risque de cardiopathies ischémique, notamment le zinc et le dioxyde d'azote. Les résultats de cette recherche indiquent donc que la sur-incidence des cas de cardiopathies ischémique dans la MEL est liée à la défaveur sociale et à la pollution atmosphérique. Plus la population d'un IRIS connaît une défaveur sociale et plus elle est confrontée au zinc et au dioxyde d'Azote, plus le risque relatif à l'incidence des cardiopathies ischémique dans la MEL augmente.

L'étude MONICA et ce mémoire constituent donc une recherche originale, à la fois par leur méthodologie que par leurs résultats. C'est pourquoi, dans le cadre de ce projet global, une étude similaire menée à l'échelle d'une autre ville, d'autres quartiers présentant des caractéristiques sociales et environnementales différentes, avec un échantillon plus important et/ou explorant d'autres critères environnementaux, serait intéressante et surtout complémentaire à ces premiers travaux de recherche. Les facteurs environnementaux liés au risque pour la santé cardiaque représentent, aujourd'hui, l'un des plus grands enjeux de santé publique mondiale.

BIBLIOGRAPHIE

- Anglade, C., Créatin C. « Les actions et programmes de prévention ». *Actualités et Dossiers en Santé Publique*, n°47, juin 2004.
- B.A, Franklin, R, Brook, C, Arden Pope III. « Air Pollution and Cardiovascular Disease ». *Current Problems in Cardiology* (2015) ; 40:207–238
- Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics, *Ann. Inst. Statist. Math.*, 43, 1-59.
- Cai, X., Li, Z., Scott, E.M. et al. *Environ Sci Pollut Res* (2016) 23: 6139.
- Clément, G. « Prévalence des principaux facteurs de risque cardiovasculaire dans les agglomérations de Lille et Dunkerque entre 2011 et 2013, et évolution à Lille entre 1985 et 2013 », Thèse de Doctorat en Médecine, Université de Lille, 2015.
- Criqui, M. H., et B. L. Ringel. « Does Diet or Alcohol Explain the French Paradox ? » *The Lancet* 344, n° 8939 (24 décembre 1994) : 1719-23.
- De Perreti, C., Chin, F., Tuppin, P., Danchin, N. « Personnes hospitalisées pour infarctus du myocarde en France : tendances 2002-2008 ». *Bulletin Épidémiologique Hebdomadaire*, n°41 (2012).
- Ducimetière, P, et Girard, X. « La fréquence des maladies cardio-vasculaires en France », *Actualités et Dossiers en Santé Publique*, n°47, juin 2004.
- Ducimetière, P, Haas, B., Ruidavets, J.B., Montaye, M., Wagner, A., Ferrières, J., Dallongeville, J., Bingham, A. « Fréquence et mortalité à 28 jours des divers épisodes d'insuffisance coronaire aiguë dans trois régions françaises en 2006 », *Bulletin Épidémiologique Hebdomadaire*, n°40-41 (2012)
- Dujardin, J.-J., et J.-P. Cambou. « Épidémiologie de l'infarctus du myocarde ». *EMC - Cardiologie-Angéiologie* 2, n° 4 (novembre 2005): 375-87. <https://doi.org/10.1016/j.emcaa.2005.07.010>.
- Evans, Alun. « The French Paradox and Other Ecological Fallacies ». *International Journal of Epidemiology* 40, n° 6 (1 décembre 2011): 1486-89.
- Ezzati, Majid. « How Can Cross-Country Research on Health Risks Strengthen Interventions? Lessons from INTERHEART ». *The Lancet* 364, n° 9438 (septembre 2004): 912-14.
- Garfield, Richard. « Health Care in Cuba and the Manipulation of Humanitarian Imperatives ». *The Lancet* 364, n° 9438 (septembre 2004) : 1007.
- Gibbons, Jean D., et John W. Pratt. « P-Values: Interpretation and Methodology ». *The American Statistician* 29, n° 1 (1975): 20-25. <https://doi.org/10.2307/2683674>.
- Global Status Report on Noncommunicable Diseases 2010*.
- Hulka, Barbara S., et Patricia G. Moorman. « Breast Cancer: Hormones and Other Risk Factors ». *Maturitas* 38, n° 1 (28 février 2001): 103-13.
- Kulldorf, Martin. « A Spatial Scan Statistic ». *Communications in Statistics - Theory and Methods.*, 26(6), 1481-1496 (1997)
- Kulldorf, Martin. « Spatial Disease Clusters: Detection and Inference». *Statistics in Medicine.*, Vol. 14, 799-810 (1995)
- M Kulldorff, W F Athas, E J Feurer, B A Miller, and C R KeyDivision of Cancer Prevention, National Cancer Institute, Bethesda, Md 20892-7368, USA. "Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico.", *American Journal of Public Health* 88, no. 9 (September 1, 1998): pp. 1377-1380.
- Kulldorff M. 1999. An isotonic spatial scan statistic for geographical disease surveillance. *Journal of the National Institute of Public Health*, 48: 94–101.
- Lang, Thierry, Chantal Cases, Pierre Chauvin, Eric Jouglu, Florence Jusot, Anne Laporte, Pierre Lombrail, Gwenn Menvielle. « Indicateurs de suivi de l'évolution des inégalités

- sociales de santé dans les systèmes d'information en santé », *Haut Conseil en Santé Publique*, Rapport du mois de juin, 2013.
- Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés (s. d.).
- LOI n° 2014-58 du 27 janvier 2014 de modernisation de l'action publique territoriale et d'affirmation des métropoles, 2014-58 § (2014).
- LOI n° 2015-991 du 7 août 2015 portant nouvelle organisation territoriale de la République, 2015-991 § (2015).
- Montaye, M., Bingham, A., Arveiler, D., Ruidavets, J.B., Ducimetière, P., Haas, B., Dallongeville, J. « Caractéristiques cliniques, biologiques et électriques des épisodes inauguraux d'insuffisance coronaire dans trois régions françaises en 2006 : relation avec la survie à 28 jours ». *Bulletin Epidémiologique Hebdomadaire*, n°40-41 (2012)
- Perel, C., Chin, F., Tuppin, P., Danchin, N., Alla, F., Juillière, Y., de Peretti, C. «Taux de patients hospitalisés pour insuffisance cardiaque en 2008 et évolution en 2002-2008, France » *Bulletin Epidémiologique Hebdomadaire*, n°41 (2012).
- Pornet, Carole, Cyrille Delpierre, Olivier Dejardin, Pascale Grosclaude, Ludivine Launay, Lydia Guittet, Thierry Lang, et Guy Launoy. «Construction of an Adaptable European Transnational Ecological Deprivation Index: The French Version ». *Journal of Epidemiology And Community Health* 66, N° 11 (Novembre 2012): 982-89.
- PORNET, Carole, Cyrille DELPIERRE, Olivier DEJARDIN, Ludivine Launay, Lydia Guittet, Thierry Lang, Et Guy Launoy. « Construction D'un Indice Ecologique De Défavorisation A Vocation Européenne : The French EDI » *Journée de Méthodologie Statistiques*, INSEE, 2012
- Renaud, S., et M. de Lorgeril. « Wine, alcohol, platelets, and the French paradox for coronary heart disease ». *The Lancet*, Originally published as Volume 1, Issue 8808, 339, n° 8808 (20 juin 1992): 1523-26.
- Renaud, S., et M. de Lorgeril. « Wine, Alcohol, Platelets, and the French Paradox for Coronary Heart Disease ». *The Lancet* 339, n° 8808 (20 juin 1992): 1523-26.
- Richard, J. L. « [Coronary risk factors. The French paradox] ». *Archives des maladies du coeur et des vaisseaux* 80 Spec No (avril 1987): 17-21.
- RICHARD, J.L. « Le Projet MONICA : un projet OMS de recherche carcio-vasculaire ». *Les registres de morbidité* 36, n° 4-5 (1988): 325-34.
- Rosengren, Annika, Steven Hawken, Stephanie Ôunpuu, Karen Sliwa, Mohammad Zubaid, Wael A Almahmeed, Kathleen Ngu Blackett, Chitr Sitthi-amorn, Hiroshi Sato, et Salim Yusuf. « Association of Psychosocial Risk Factors with Risk of Acute Myocardial Infarction in 11 119 Cases and 13 648 Controls from 52 Countries (the INTERHEART Study): Case-Control Study ». *The Lancet* 364, n° 9438 (septembre 2004): 953-62.
- Stockfelt, L., Anderson, E.M., Molnar, Peter., Gidhagen, L., Segersson, D., Rosengren, A., Barregard, L., Sallsten, G. « Long-term effects of total and source-specific particulate air pollution on incident cardiovascular disease in Gothenburg, Sweden ». *Environmental Research* 158 (2017) 61-71.
- Temam, Sofia, Raphaëlle Varraso, Carole Pornet, Margaux Sanchez, Aurélie Affret, Bénédicte Jacquemin, Françoise Clavel-Chapelon, Grégoire Rey, Stéphane Rican, et Nicole Le Moual. « Ability of ecological deprivation indices to measure social inequalities in a French cohort ». *BMC Public Health* 17, n° 1 (15 décembre 2017): 956.
- The Lancet. « China Must Prioritise Health Opportunities for All ». *The Lancet* 364, n° 9438 (septembre 2004): 909-10.
- The Lancet. « Choose advice wisely for health », *The Lancet* 364, n° 9438 (septembre 2004) : 909-10.

- Vienneau, D., Schindler, C., Perez, L., Probst-Hensch, N. « The relationship between transportation noise exposure and ischemic heart disease : A meta-analysis », *Environmental Research* 138 (2015) 372-380
- Wagner, A., Ruidavets, J.B., Montaye, Michèle., Ferrières, J., Amoyel, P. « Evolution de la maladie coronaire en France de 2000 à 2007 », *Bulletin Epidémiologique Hebdomadaire*, n°40-41 (2012)
- WHO MONICA Project principal investigators. « The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): A major international collaboration ». *J Clin Emiol* Vol. 41, n° No. 2 (1988): 105-114.
- Yusuf, Salim, Steven Hawken, Stephanie Ôunpuu, Tony Dans, Alvaro Avezum, Fernando Lanas, Matthew McQueen, et al. « Effect of Potentially Modifiable Risk Factors Associated with Myocardial Infarction in 52 Countries (the INTERHEART Study): Case-Control Study » 364 (2004): 16.

SITOGRAPHIE

- « Cardiovascular diseases (CVDs) ». World Health Organization.
<http://www.who.int/fr/>
- « Douleur aiguë, chronique - Douleur | Institut National Du Cancer ». <http://www.e-cancer.fr>
- Mandard, Stéphane. « La pollution de l'air tue 7 millions de personnes par an dans le monde, alerte l'OMS ». *Le Monde.fr*, 2 mai 2018, sect. Planète.
<https://www.lemonde.fr>
- « Neuf personnes sur 10 respirent un air pollué dans le monde ». World Health Organization.
<http://www.who.int/fr>
- « Plan d'action 2013-2020 » OMS (2017).
<http://www.who.int/fr>
- « Rapport sur la santé dans le monde » OMS (2010)
<http://www.who.int>

TABLE DES ILLUSTRATIONS

Figure 1 : Carte de situation de la MEL. Sources : IGN, Géoportail.....	15
Figure 2 : Aéroport Lille-Lesquin. Source : France 3 Région Hauts-de-France (à gauche)	16
Figure 3 : Centre Commerciale et d'activités Euralille. Source : Site internet Euralille (à droite).....	16
Figure 4 : Tracé des lignes du Transpole de la MEL en 2018. Source : Transpole.....	17
Figure 5 : Occupation du sol dans la MEL en 2009. Sources : Base de données SIGALE, Région Hauts-de-France.....	18
Figure 6 : Ancienne industrie métallurgique fermée, à Roubaix. Source : Antoine Longo (à gauche) .	19
Figure 7 : Anciens logements ouvriers en briques rouges à Roubaix. Source : Antoine Longo (à droite)	19
Figure 8 : Organigramme du processus d'inclusion des cas et des IRIS dans l'étude MONICA	29
Figure 9 : Cartographie de la défaveur sociale suivant les valeurs du French EDI	38
Figure 10 : Cartographie de la défaveur sociale suivant les valeurs du Fdep (droite).	38
Figure 11 : Capteurs de surveillance physico-chimique. Photos : Florent Occeci.....	41
Figure 12 : Schéma d'une interpolation spatiale.....	41
Figure 13 : Schéma des espèces végétales et fongiques utilisés pour de la biosurveillance. Schéma : Association pour la Prévention de la Pollution Atmosphérique	42
Figure 14 : Taux d'incidence brutes.....	49
Figure 15 : Graphique d'incidence standardisée par âge	52
Figure 16: Répartition spatiales des SIR Bruts dans la MEL	55
Figure 17 : Répartition des SIR Lissés à l'échelle des IRIS de la MEL.....	58
Figure 18 : Exemple de phase de détection de cluster dans l'Etat de New-York, Etats-Unis. Source : Martin Kulldorf, 1999	61
Figure 19 : Schéma de réalisation d'un scan isotonique. Source : Michaël Génin.....	63
Figure 20 : Cluster isotoniques de sur-incidence et de sous-incidence dans la MEL, ainsi que la valeur moyenne du risque relatif de chaque niveau.....	64
Figure 21 : : Carte des SIR Lissés après ajustement sur la défaveur	77
Figure 22 : Carte des résidus de SIR lissés après ajustement sur la défaveur sociale.....	77
Figure 23 : Direction du vent d'après les informations de l'aéroport de Lille-Lesquin.....	92
Tableau 1 : Résumé de la revue de la littérature.....	21
Tableau 2 : Résumé des indices de défaveur sociale en France	33
Tableau 3 : Les principales sources des polluants étudiés	43
Tableau 4 : Description du registre MONICA	47
Tableau 5 : Incidences brutes par années et classes de populations.....	49
Tableau 6 : Taux d'incidence standardisé par âge, pour 100 000 habitants.....	51
Tableau 7 : Description des clusters isotoniques par niveaux de cercles concentriques	66
Tableau 8 : Schéma de réalisation d'un test de Kruskal-Wallis.....	69
Tableau 9 : Tests de Kruskal-Wallis et post-hoc par groupe de clusters (première stratégie)	71
Tableau 10 : détection de clusters avec ajustement sur la défaveur sociale.....	79
Tableau 11 : Risque relatif (Intervalle de crédibilité à 95%) de cardiopathie ischémique par variable de pollution (Partie 1)	83
Tableau 12 : Risque relatif (Intervalle de crédibilité à 95%) de cardiopathie ischémique par variable de pollution (Partie 2)	84
Tableau 13 : Forestplot multivarié par polluant montrant un "effet seuil"	86

ANNEXES

ANNEXE 1 : GEOCODAGE D'ADRESSE A L'IRIS

Le géocodage des adresses à l'IRIS intervient au moment où de nouveaux cas sont introduits dans l'étude : ceux des années 2010 et 2011. Au moment de leur survenue de cardiopathie ischémique, les patients sont pris en charge par les urgences et sont admis à l'hôpital. Lors de leur admission, une fiche de renseignement est remplie pour identifier le patient. Son adresse est aussi inscrite. Néanmoins, les adresses ne pouvant pas être révélées publiquement et l'étude MONICA s'axant sur une échelle d'IRIS, il faut que chaque adresse se voie associer son IRIS d'appartenance. L'exercice de trouver les coordonnées spatiales d'une adresse afin de trouver son IRIS d'appartenance s'appelle un géocodage. Pour cela, les informations qu'il faut récolter sont les coordonnées spatiales de l'adresse recherchée.

Le géocodage réalisé lors de l'étude ne s'effectue donc pas via des logiciels internet de géocodage étant donné qu'il existe un risque de fuite d'informations. Pour éviter ce risque, un géocodage sans internet est effectué. Celui-ci se base sur des registres de données résultant des recensements d'adresses existantes : la BD Adresses de l'IGN et la Base Adresses Nationale (gouvernement). Dans ces bases, les coordonnées spatiales des adresses sont indiquées.

L'objectif est donc ici d'assimiler des coordonnées spatiales aux adresses des patients de l'étude MONICA grâce aux bases nationales.

L'annexe suivante ne s'appuiera pas sur le fichier source étant donné que ses informations ne doivent pas être rendues publiques. Il s'appuiera donc sur un exemple.

Ce géocodage s'effectue en plusieurs étapes.

ETAPE 1 : ACCORDER LES BASES

Le fichier à géocoder doit **impérativement** posséder les informations suivantes pour pouvoir réaliser une correspondance avec les bases nationales :

- Le numéro de la voie (s'il y a un « bis » ou « ter », les ajouter aussi)
- Le type de voie (rue, allée, boulevard, route...)
- Le nom de la rue (uniquement).

- La commune
- Le code postal

Le tout doit être écrit **en majuscule**, afin d'éviter les confusions à cause des accents. Il ne doit pas non plus y avoir de tirets entre les mots.

Au sein de toutes les bases, les adresses doivent être notées comme ceci :

	A	B
1	RUE	CP
2	9 RUE LAVOISIER LILLE	59350
3	100 RUE DU LUXEMBOURG ROUBAIX	59100
4	53 RUE DE FIVES VILLENEUVE D ASCQ	59650
5	36 RUE JEAN BAPTISTE DUCROCQ MARQ EN BAROEUL	59700
6	15 PLACE DU THEATRE TOURCOING	59200
7	1 RUE COUTEAU WATTRELOS	59150
8		
9		
10		
11		
12		

Il est possible que les adresses ne soient pas classées directement comme cela. Néanmoins, il existe deux outils dans le logiciel Excel afin de désagréger et d'agréger des adresses : les outils « Concaténer » et « Convertir ».

ETAPE 1 bis : LES OUTILS « CONCATENER » ET « CONVERTIR »

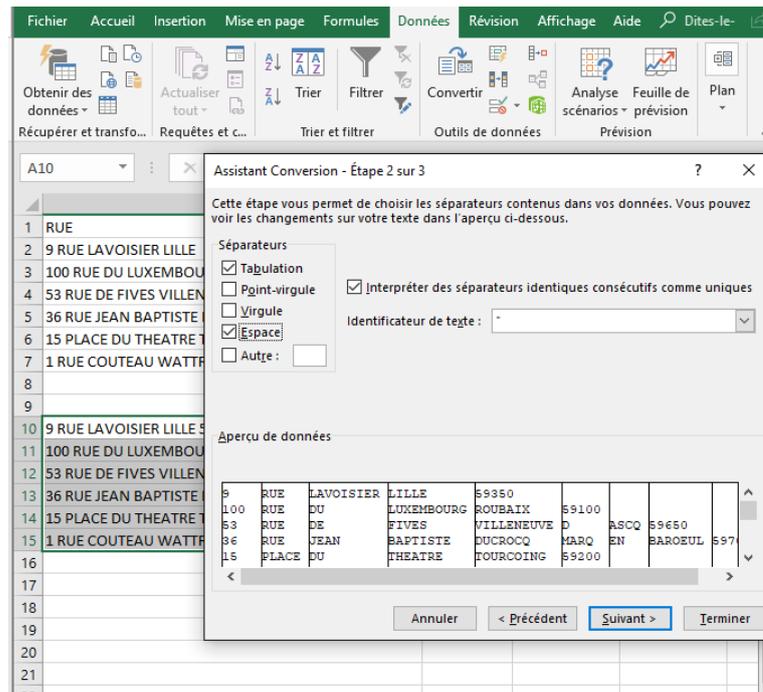
Si parmi les adresses de départs il y a des caractères qui sont rangées dans la même case mais qui ne devrait pas s'y trouver, il faut alors utiliser l'outil « convertir ». Celui-ci permet de **diviser** le contenu d'une cellule dans plusieurs autres cellules.

Cette méthode s'applique si les adresses sont inscrites comme ceci :

9 RUE LAVOISIER LILLE 59350	
100 RUE DU LUXEMBOURG ROUBAIX 59100	
53 RUE DE FIVES VILLENEUVE D ASCQ 59650	
36 RUE JEAN BAPTISTE DUCROCQ MARQ EN BAROEUL 59700	
15 PLACE DU THEATRE TOURCOING 59200	
1 RUE COUTEAU WATTRELOS 59150	

La démarche à suivre est la suivante :

- Cliquer sur « Données » dans la barre d'outils
- Sélectionner « Convertir »
- Sélectionner « Délimité » dans la boîte de dialogue puis « suivant »
- Cocher « tabulation » et « espace » afin de diviser les caractères à l'intérieur des cellules par mots.



- Sélectionner « suivant » jusqu'à « terminer ».

Le résultat doit être similaire à ceci :

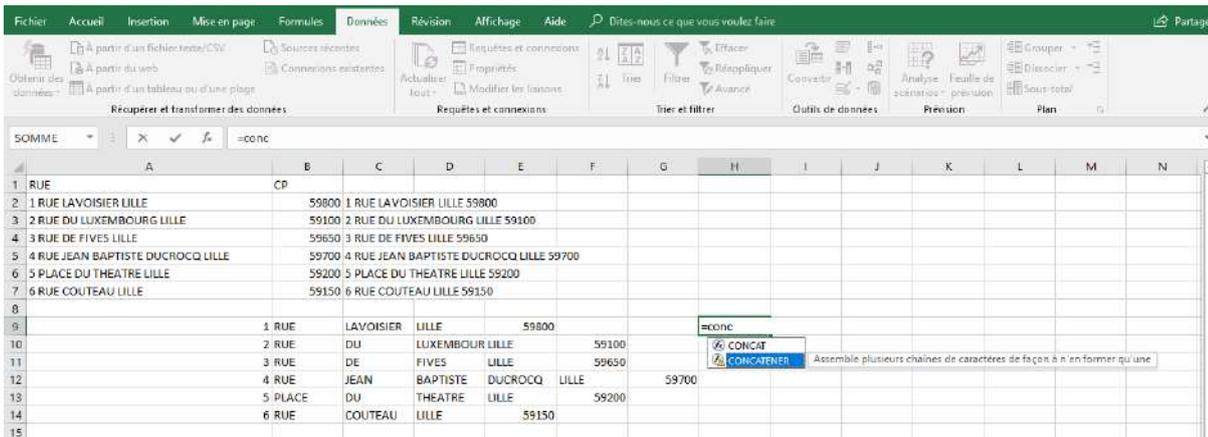
8						
9		1 RUE	LAVOISIER	LILLE	59800	
10		2 RUE	DU	LUXEMBOURG	LILLE	59100
11		3 RUE	DE	FIVES	LILLE	59650
12		4 RUE	JEAN	BAPTISTE	DUCROCQ	LILLE 59700
13		5 PLACE	DU	THEATRE	LILLE	59200
14		6 RUE	COUTEAU	LILLE	59150	
15						

Grâce à cette division, toutes les informations se retrouvent divisées. Pour les réagréger, il faut utiliser l'outil « concaténer ». Celui-ci permet de **rassembler** plusieurs cellules en une seule.

La démarche à suivre est la suivante :

- Sélectionner la cellule dans laquelle l'agrégation d'informations est souhaitée.

- Ecrire « = » suivi de « concaténer » puis double-cliquer sur celui-ci.



- Ouvrir la parenthèse et sélectionner la première cellule à agréger.
- Ajouter « ; » pour passer à une autre cellule. **Attention**, il faut veiller à rajouter des guillemets et un espace entre chaque guillemet afin de rajouter un espace entre les mots lors de la concaténation. Exemple : =CONCATENER(A1; ' ' ;B1)

8															
9															
10			9 RUE	LAVOISIER	LILLE		59350	=CONCATENER(A10;" ";B10;" ";C10;" ";D10)							
11			100 RUE	DU	LUXEMBOURG	ROUBAIX		59100							
12			53 RUE	DE	FIVES	VILLENEUVE D	ASCQ	59650							
13			36 RUE	JEAN	BAPTISTE	DUROCQ	MARQ	EN	59700						
14			15 PLACE	DU	THEATRE	TOURCOING		59200							
15			1 RUE	COUTEAU	WATTRELOS		59150								
16															
17															

- Lorsque le résultat s'affiche, effectuer un « copier » puis « collage spécial par valeurs » afin de n'avoir uniquement que le texte et non le résultat de l'opération.
- Réorganiser les lignes d'adresses.

Le résultat doit être similaire à ceci :

9 RUE	LAVOISIER	LILLE	59350					9 RUE LAVOISIER LILLE		
100 RUE	DU	LUXEMBOURG	ROUBAIX	59100				100 RUE DU LUXEMBOURG ROUBAIX		
53 RUE	DE	FIVES	VILLENEUVE D	ASCQ	59650			53 RUE DE FIVES VILLENEUVE D ASCQ		
36 RUE	JEAN	BAPTISTE	DUROCQ	MARQ	EN	59700		36 RUE JEAN BAPTISTE DUCROCQ MARQ EN BAROEUL		
15 PLACE	DU	THEATRE	TOURCOING	59200				15 PLACE DU THEATRE TOURCOING		
1 RUE	COUTEAU	WATTRELOS	59150					1 RUE COUTEAU WATTRELOS		

Il est impératif d'effectuer aussi ces changements sur les bases nationales. Celles-ci ne sont pas forcément classifiés dans les mêmes ordres ni de manière identique (majuscules,

minuscules, accents...). Il faut donc réaliser ces mêmes modifications afin d'avoir des bases similaires.

ETAPE 2 : JOINDRE LES COORDONNEES SPATIALES D'UNE BASE A UNE AUTRE

Pour joindre les coordonnées spatiales des adresses des bases nationales aux adresses du fichier de départ, l'outil utilisé sera « RECHERCHEV » sur le logiciel Excel. Cet outil peut chercher une valeur dans la colonne d'un tableau et renvoyer la valeur inscrite dans la même ligne à partir de la colonne spécifiée. En d'autres termes, si deux adresses sont identiques dans deux bases différentes mais que les coordonnées spatiales sont inscrites dans seulement l'une des deux bases, il est possible de restituer les coordonnées spatiales dans la base qui en a besoin.

Les coordonnées spatiales peuvent différer suivant les bases. Il peut y avoir des données X et Y correspondant à un système de projection type Lambert 93, ou bien des données Longitude et Latitude correspondant plutôt à un système WGS 1984.

Ici, il s'agit de points de Longitude et Latitude. La démarche à suivre est la suivante :

- Sélectionner la cellule dans laquelle la donnée recherchée va apparaître (longitude ou latitude)
- Ecrire « = » suivi de « RECHERCHEV ». L'outil a besoin de quatre informations.
 - En premier lieu, cliquer sur la cellule qui contient l'adresse dont les coordonnées sont demandées.
 - Deuxièmement, sélectionner la plage de données dans laquelle sont inscrites les adresses recherchées ainsi que les coordonnées spatiales.
 - Troisièmement, écrire le numéro de la colonne **de la plage de données sélectionnée** dans laquelle se trouve la coordonnée spatiale recherchée. Exemple, si la plage de données sélectionnée contient, dans l'ordre, l'adresse, la longitude et la latitude, et que c'est la longitude qui est recherchée, alors le numéro de la colonne de la longitude est 2.
 - Enfin, cliquer sur « FAUX » afin de n'avoir une donnée que **si et seulement si** les deux adresses inscrites dans les deux bases correspondent bien l'une et l'autre.

	H	I	J	K
1	NOM_VOIE	CP	Lat	Long
2	9 RUE LAVOISIER LILLE	59800	50,6297165	3,047845
3	100 RUE DU LUXEMBOURG ROUBAIX	59100	50,6911696	3,156384
4	53 RUE DE FIVES VILLENEUVE D ASCQ	59650	50,6320159	3,124142
5	36 RUE JEAN BAPTISTE DUCROCQ MARQ EN BAROEUL	59700	50,6657161	3,09935
6	15 PLACE DU THEATRE TOURCOING	59200	50,7201995	3,152511
7	1 RUE COUTEAU WATTRELOS	59150	50,6965797	3,202417

	A	B	C	D
1	RUE	CP	Latitude	Longitude
2	9 RUE LAVOISIER LILLE	59350	=RECHERCHEV(A2;	
3	100 RUE DU LUXEMBOURG ROUBAIX	59100	geoloc_bdadresse_mel.csv!	
4	53 RUE DE FIVES VILLENEUVE D ASCQ	59650	\$K\$7;3;FAUX)	
5	36 RUE JEAN BAPTISTE DUCROCQ MARQ EN BAROEUL	59700		
6	15 PLACE DU THEATRE TOURCOING	59200		
7	1 RUE COUTEAU WATTRELOS	59150		

Le résultat doit être similaire à ceci :

	H	I	J	K
1	NOM_VOIE	CP	Lat	Long
2	9 RUE LAVOISIER LILLE	59800	50,6297165	3,04784519
3	100 RUE DU LUXEMBOURG ROUBAIX	59100	50,6911696	3,15638462
4	53 RUE DE FIVES VILLENEUVE D ASCQ	59650	50,6320159	3,12414219
5	36 RUE JEAN BAPTISTE DUCROCQ MARQ EN BAROEUL	59700	50,6657161	3,0993553
6	15 PLACE DU THEATRE TOURCOING	59200	50,7201995	3,15251174
7	1 RUE COUTEAU WATTRELOS	59150	50,6965797	3,20241739

	A	B	C	D
1	RUE	CP	Latitude	Longitude
2	9 RUE LAVOISIER LILLE	59350	50,62971653	3,047845189
3	100 RUE DU LUXEMBOURG ROUBAIX	59100	50,69116957	3,156384617
4	53 RUE DE FIVES VILLENEUVE D ASCQ	59650	50,63201585	3,124142194
5	36 RUE JEAN BAPTISTE DUCROCQ MARQ EN BAROEUL	59700	50,66571607	3,099355296
6	15 PLACE DU THEATRE TOURCOING	59200	50,7201995	3,15251174
7	1 RUE COUTEAU WATTRELOS	59150	50,69657966	3,202417388

ETAPE 3 : LE GEOCODAGE MANUEL

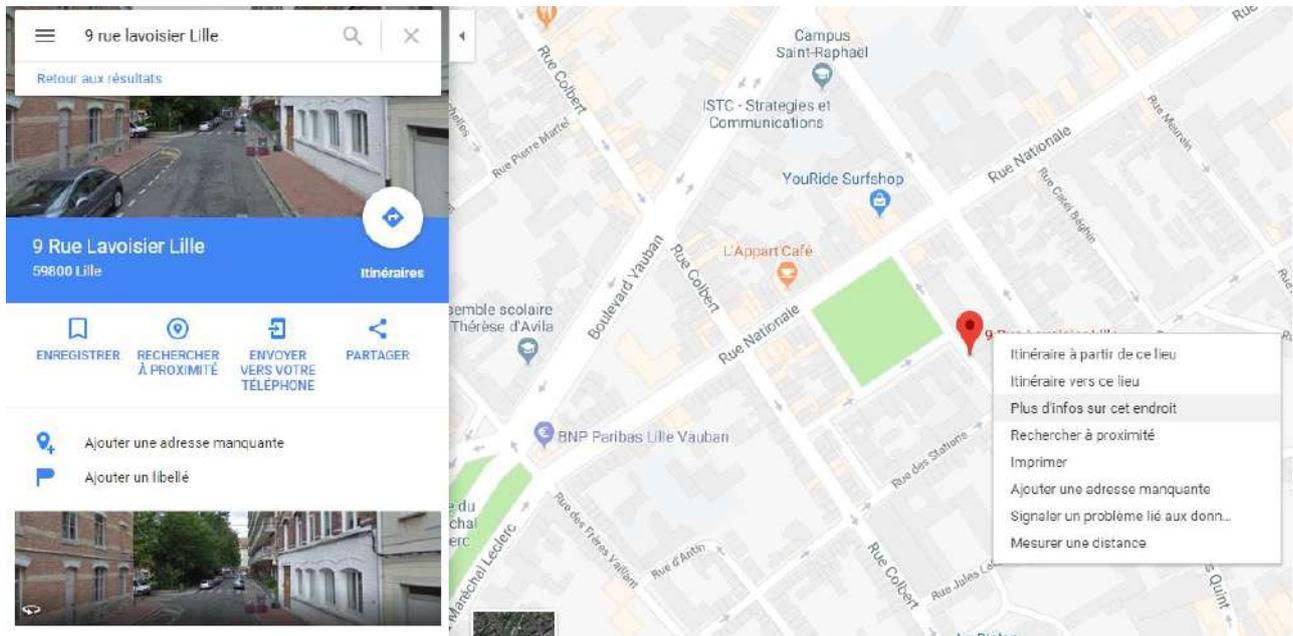
L'étape précédente est réalisée dans le but de connaître les coordonnées spatiales de l'ensemble des adresses du fichier de base. Néanmoins, il est possible que certaines adresses recherchées ne soient ni dans la BAN, ni dans la BD Adresses. Si cela se produit et qu'il manque des adresses, la dernière solution est ce qu'on appelle un « géocodage manuel ». Cela implique de rechercher les coordonnées spatiales grâce à l'outil Google Maps. Cette étape peut s'avérer chronophage.

Ici, le problème de « fuite » concernant les adresses protégées par la CNIL n'a plus lieu d'être puisque ce n'est pas un fichier qui est diffusé publiquement mais une recherche ponctuelle d'un lieu, sans aucune information quelconque ajoutée.

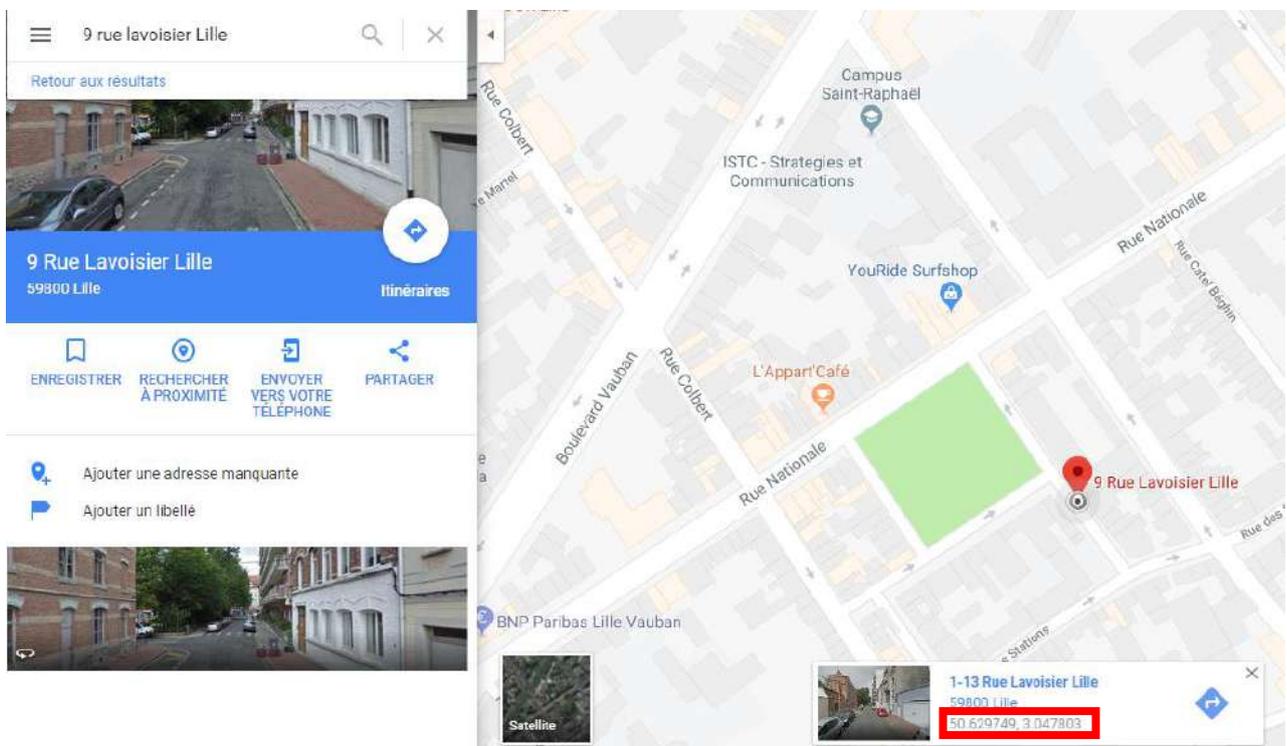
Pour connaître les points des adresses manquantes, la démarche est la suivante :

- Taper l'adresse dans Google Maps

- Effectuer un clic-droit sur le point de l'adresse et sélectionner « plus d'infos sur ce point »



- Copier/coller les coordonnées spatiales indiquées en bas de la page.



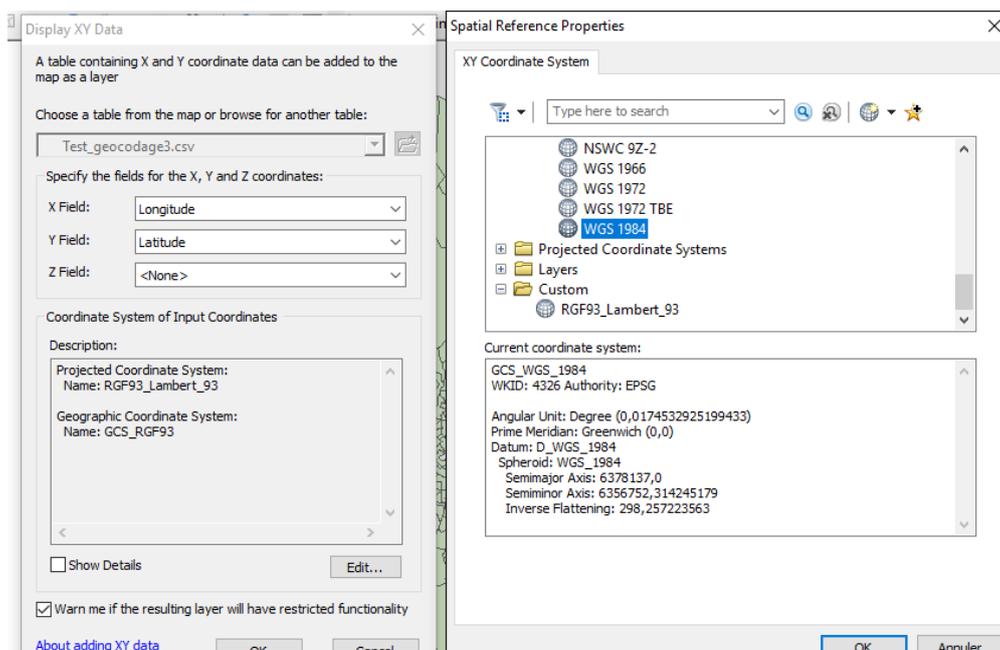
Le fichier d'adresses peut alors ainsi être complété.

ETAPE 4 : AFFICHER LES ADRESSES DANS UN SIG

Une fois que toutes les adresses du fichier de base ont des coordonnées spatiales, l'objectif est de les intégrer dans un SIG avant de leur attribuer un IRIS d'appartenance.

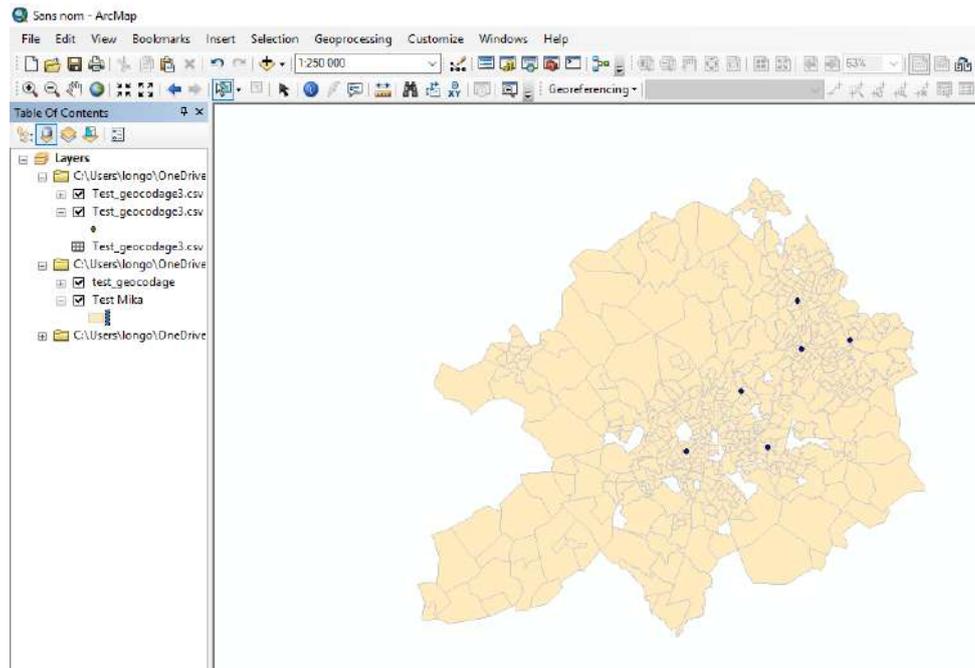
La démarche à suivre est la suivante :

- Dans un SIG, intégrer un fichier .shp du territoire d'étude ainsi que le fichier d'adresse de base.
- Effectuer un clic-droit sur le fichier d'adresse et cliquer sur *Display X & Y* afin d'afficher les points de latitude et longitude.
- Rentre la colonne « Latitude » à Y et « Longitude » à X.
- Avant de valider, trouver « Edit » en bas de la page afin de bien afficher le système de projection. Celui-ci doit être, pour le cas présent, en WGS 1984.
- Pour trouver ce système de projection, aller dans *Geographic Coordinate Systems > World > WGS 1984*.



- Cliquer sur *OK* sur la première fenêtre puis *OK* sur la deuxième.
- Un nouveau fichier d'adresses géocodées apparaît ainsi que les points sur la carte.

Le résultat doit être similaire à ceci :

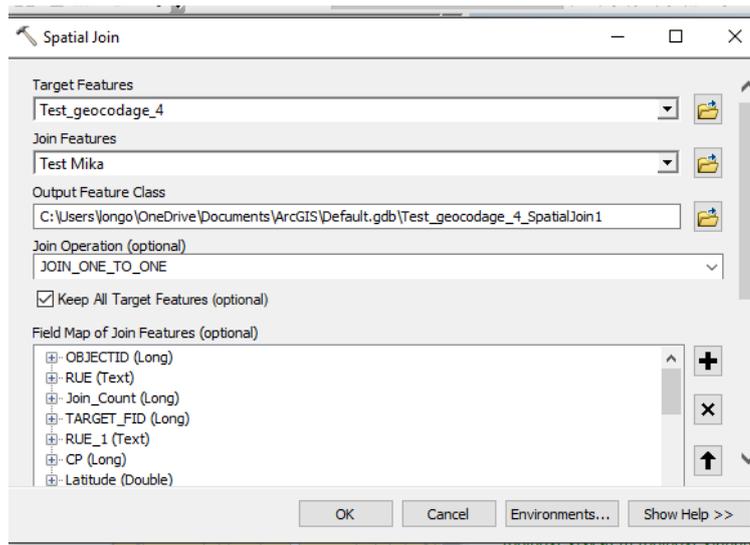


ETAPE FINALE : TROUVER LES IRIS D'APPARTENANCE DE CHAQUE ADRESSE

Maintenant que toutes les adresses sont affichées sur le territoire, il s'agit désormais de connaître l'IRIS dans lesquelles elles se trouvent. Pour cela, l'outil qui sera nécessaire sera une « jointure spatiale ».

La démarche à suivre est la suivante :

- Dans la *ArcToolbox*, aller dans *Analysis tool > Overlay > Spatial Join*
- Dans la fenêtre de jointure spatiale, rentrer :
 - Le fichier d'adresses géocodées comme cible dans *Target Features*
 - Le fichier .shp du territoire d'étude comme fichier à joindre dans *Join Features*
 - Le chemin ainsi que le nouveau nom du fichier en sortie dans *Output Feature Class*
 - Cocher la case *Keep All Target Feature* afin de conserver l'ensemble des adresses pour constater si certaines n'ont pas pu être jointes.



- Cliquer sur *OK*.
- Un nouveau fichier d'adresses géocodées s'ouvre. Celui-ci affiche les informations spatiales de chaque adresse, dont les IRIS d'appartenance.

Le résultat doit être similaire à ceci :

Shape*	OBJECTID	RUE	Join_Count	TARGET_FID	CP	Latitude	Longitude	LMCU Iris1 1	LMCU Iris1 12	LMCU Iri 12	LMCU Iri 23	LMCU Iri 34	LMCU Iri 45
Point	1	RUE LAVOISIER LILLE	1	0	5935	50.629717	3.047845	59350	LILLE	407	593500407	Stations	
Point	2	00 RUE DU LUXEMBOURG LILLE	1	1	5951	50.69117	3.156385	59512	ROUBAIX	101	595120101	Macellerie	
Point	3	3 RUE DE FIVES LILLE	1	2	5900	50.632016	3.124142	59009	VILLENEUVE-D'ASCQ	204	590090204	Les Facultés	
Point	4	8 RUE JEAN BAPTISTE DUCROCQ LILLE	1	3	5937	50.685716	3.099355	59378	MARCO-EN-BAROEUL	602	593780602	Hotel de ville	
Point	5	5 PLACE DU THEATRE LILLE	1	4	5959	50.720199	3.152512	59599	FOURCOING	901	595990901	Nationale	
Point	6	RUE COUTEAU LILLE	1	5	5965	50.69658	3.202417	59650	WATTRELOS	302	596500302	Laboureur	

20

Cette table peut s'exporter par la suite dans le fichier d'adresse de base. Les adresses peuvent donc ensuite être supprimées afin de ne garder que le nom des communes et des IRIS. Le fichier de base peut donc être anonymisé.

²⁰ Une erreur s'est glissée dans la colonne d'adresse : toutes les communes affichent « Lille » au lieu de leur commune d'appartenance. Cela est simplement dû à une erreur dans le fichier importé, mais les coordonnées spatiales sont tout de même en raccord avec le fichier Excel du début. Cela n'altère en rien le résultat final. Les communes sont d'ailleurs retrouvées après la jointure spatiale.

ANNEXE 2 : DETECTION DE CLUSTERS ISOTONIQUES PAR LE LOGICIEL SATSCAN

Le logiciel SaTScan a été employé pour détecter des clusters isotoniques. Son utilisation s'effectue en plusieurs étapes.

ETAPE 1 : VERIFIER LES FICHERS SOURCES

Avant d'utiliser le logiciel SaTScan, il est nécessaire de posséder **3 fichiers** ou informations différentes :

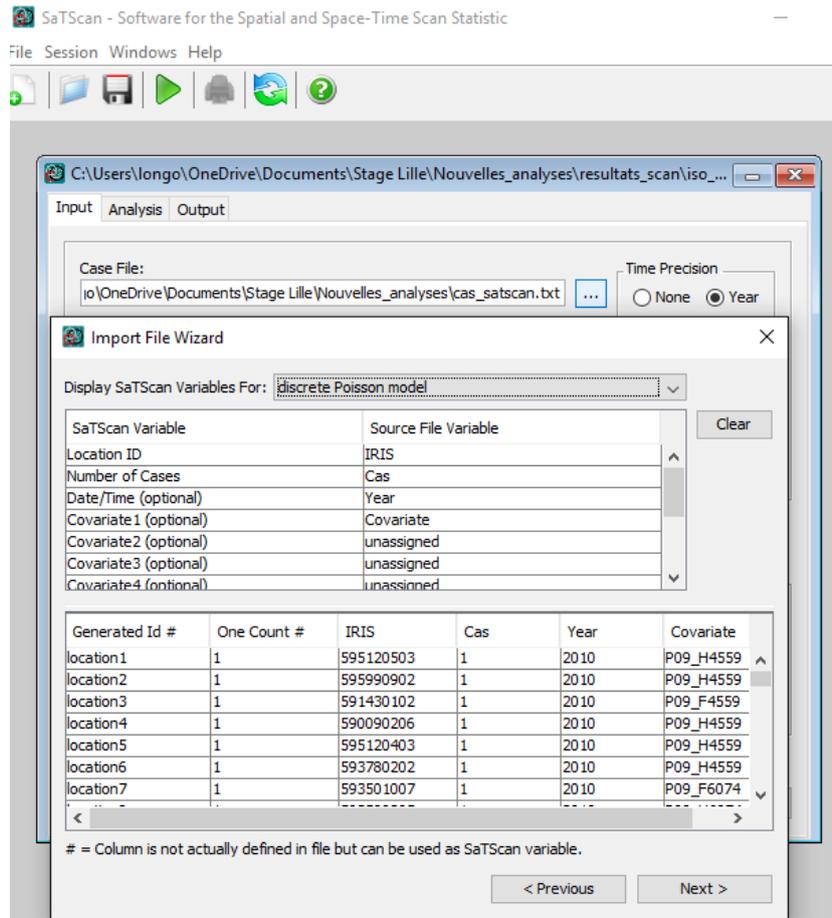
- Un fichier recensant les cas par IRIS
- Un fichier recensant la population à risque par IRIS
- Un fichier de coordonnées spatiales par IRIS.

ETAPE 2 : INTEGRER LE FICHER DE CAS

En ouvrant le logiciel SaTScan, une fenêtre s'ouvre avec plusieurs potentiels emplacements de fichier. Pour l'étude MONICA, seuls les emplacements qui sont utiles pour des données de **modèles de Poisson** seront renseignés.

Le premier emplacement est le *Case File*. Il faut intégrer le fichier recensant les cas par IRIS. La démarche est la suivant :

- Cliquer sur la case montrant trois points de suspensions à droite de l'emplacement du fichier de cas.
- Dans la nouvelle fenêtre, indiquer si la première ligne du fichier de cas est un titre, puis indiquer comment sont séparées les colonnes (espaces, virgule, point-virgule...)
- Dans une autre nouvelle fenêtre, ajouter :
 - Les IRIS comme *Location ID*
 - Les cas comme *Number of case*
 - La colonne « Covariate » comme *Covariate 1*. Celle-ci indique les classes d'âge.
 - La colonne *Year* est tout de même remplie, bien que les analyses ne soient pas temporelles mais spatiales.



- Cliquer sur *Next*.

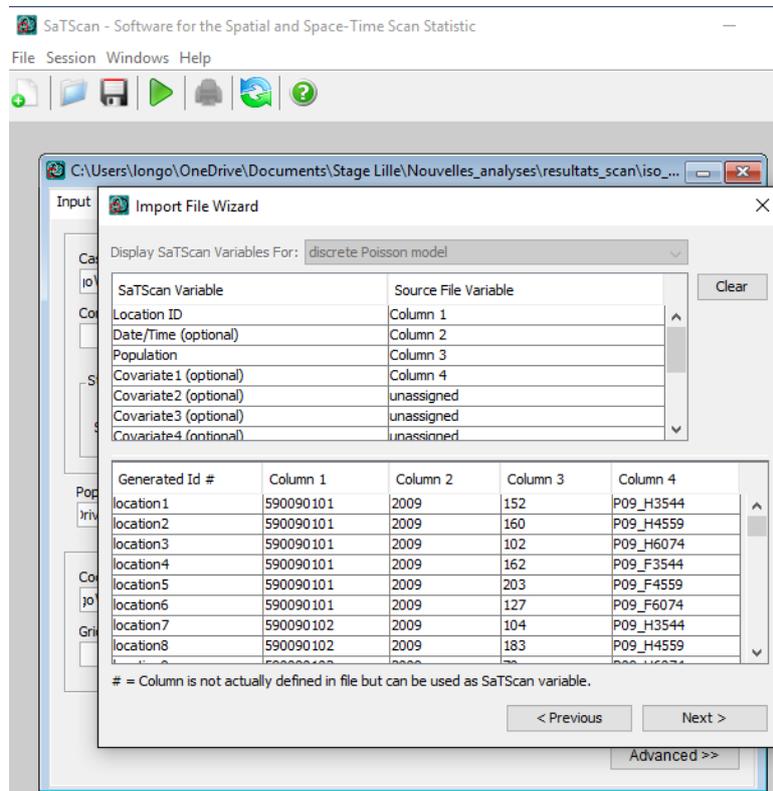
ETAPE 3 : INTEGRER UN FICHER DE POPULATION

Après le fichier de cas, il faut directement se rendre à l'emplacement *Population File*.

Pour le fichier de population, la démarche reste similaire. La seule différence réside dans la colonne « Column 3 » où le nombre d'habitant à risque par IRIS est inscrit à la place des cas dans le fichier de cas.

- Cliquer sur la case montrant trois points de suspensions à droite de l'emplacement du fichier de population.
- Dans la nouvelle fenêtre, indiquer si la première ligne du fichier de cas est un titre, puis indiquer comment sont séparées les colonnes (espaces, virgule, point-virgule...)
- Dans une autre nouvelle fenêtre, ajouter :
 - o La « Column 1 », qui désigne les IRIS, comme *Location ID*
 - o La « Column 3 », qui désigne la population, comme *Population*

- La colonne « Column 3 » comme *Covariate 1*. Celle-ci indique les classes d'âge.
- La « Column 2 » comme *Year*.



- Cliquer sur *Next*.

ETAPE 3 : INTEGRER UN FICHER DE COORDONNEES SPATIALES

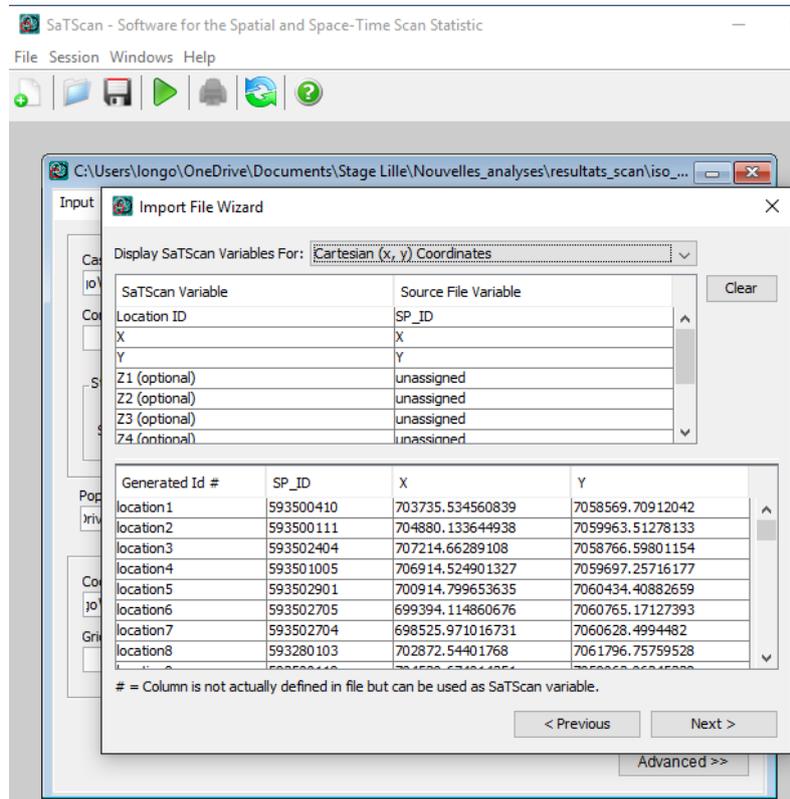
Enfin, il faut se rendre à l'emplacement *Coordinates File*.

Avant d'intégrer ce fichier, il faut choisir dans l'encart à droite de l'emplacement, le type de coordonnées spatiales. Les données de l'étude MONICA ayant une projection de type Lambert 93, il faut cocher *Cartesian*.

La démarche pour intégrer les données spatiales restent similaires aux précédents fichiers.

- Cliquer sur la case montrant trois points de suspensions à droite de l'emplacement de coordonnées spatiales.
- Dans la nouvelle fenêtre, indiquer si la première ligne du fichier de cas est un titre, puis indiquer comment sont séparées les colonnes (espaces, virgule, point-virgule...)
- Dans une autre nouvelle fenêtre, ajouter :

- « SP_ID », qui désigne les IRIS, comme *Location ID*
- « X », qui désigne la longitude, comme *X*
- « Y », qui désigne la latitude, comme *Y*.

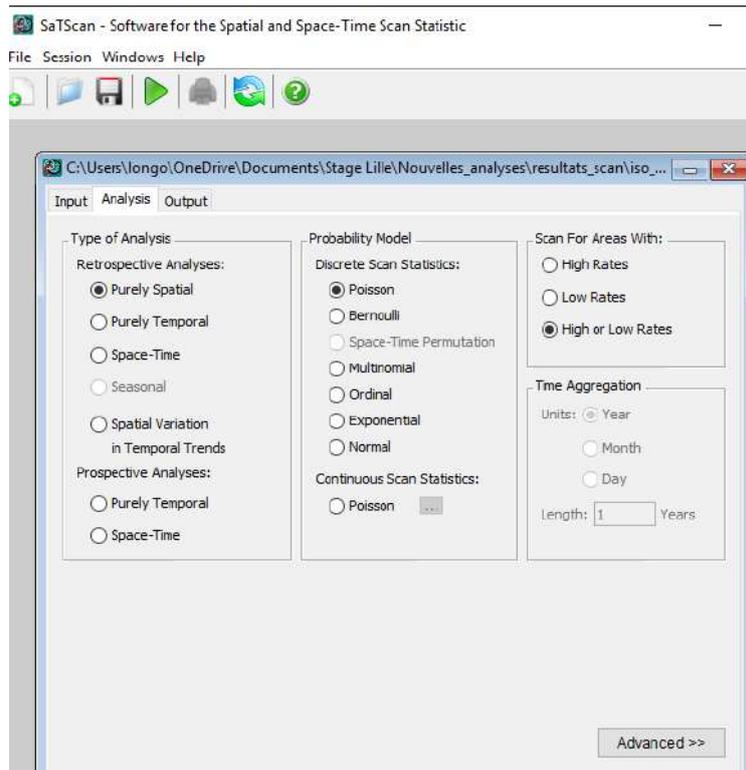


- Cliquer sur *Next*.

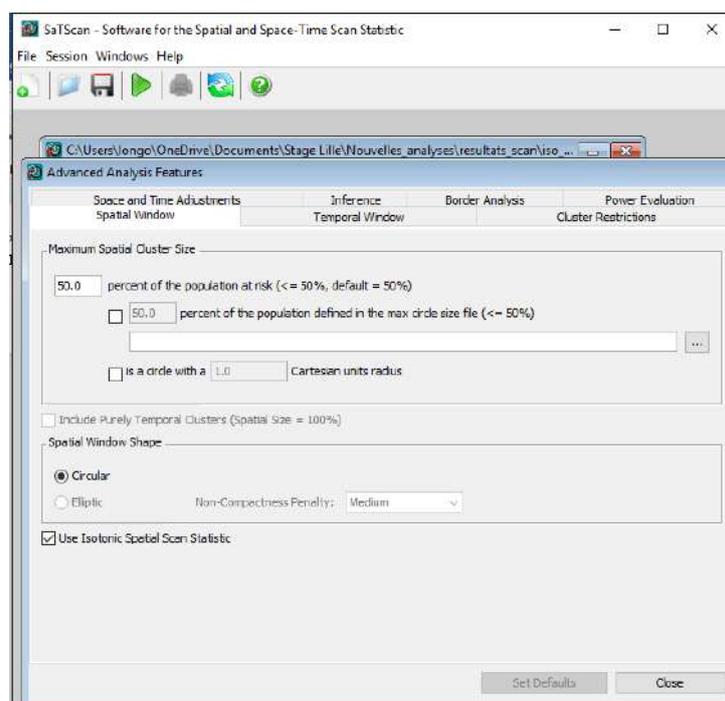
ETAPE 4 : SELECTIONNER LES TYPES D'ANALYSE

L'étude MONICA est une étude de variations spatiales d'incidence, et les clusters de sur-incidence et sous-incidence ont été nécessaire. Pour les réaliser, la démarche est la suivante :

- Sélectionner « *Analysis* » en haut de la fenêtre de départ
- Dans le premier encart, cocher « *Purely Spatial* » comme type d'analyse.
- Dans le second encart, cocher « *Poisson* » comme modèle de probabilité.
- Dans le troisième encart, cocher « *High and Low Rates* » afin d'avoir des clusters de sur-incidence et de sous-incidence.



- Cliquer ensuite sur « *Advanced >>* » afin de voir les options avancées.
- En bas de la fenêtre, cocher la case « *Use Isotonic Spatial Scan Statistics* » afin d'obtenir des clusters isotoniques.

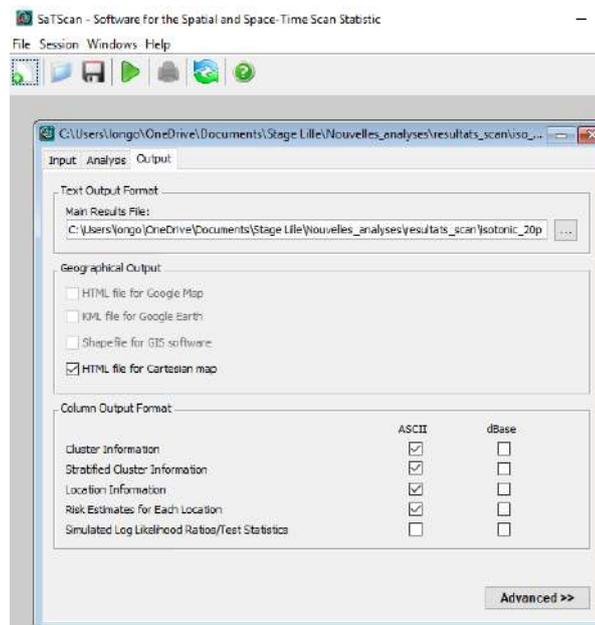


- Laisser le reste des informations par défaut et cliquer sur « *Close* ».

ETAPE FINALE : CALCULER LES CLUSTERS

Une fois tous les fichiers intégrés au logiciel, celui-ci peut désormais calculer les clusters et déterminer les IRIS qui les intégreront. Mais avant cela, il faut indiquer au logiciel le nom du fichier en sortie, ainsi que les différents types de données nécessaires. Pour cela, la démarche est la suivante :

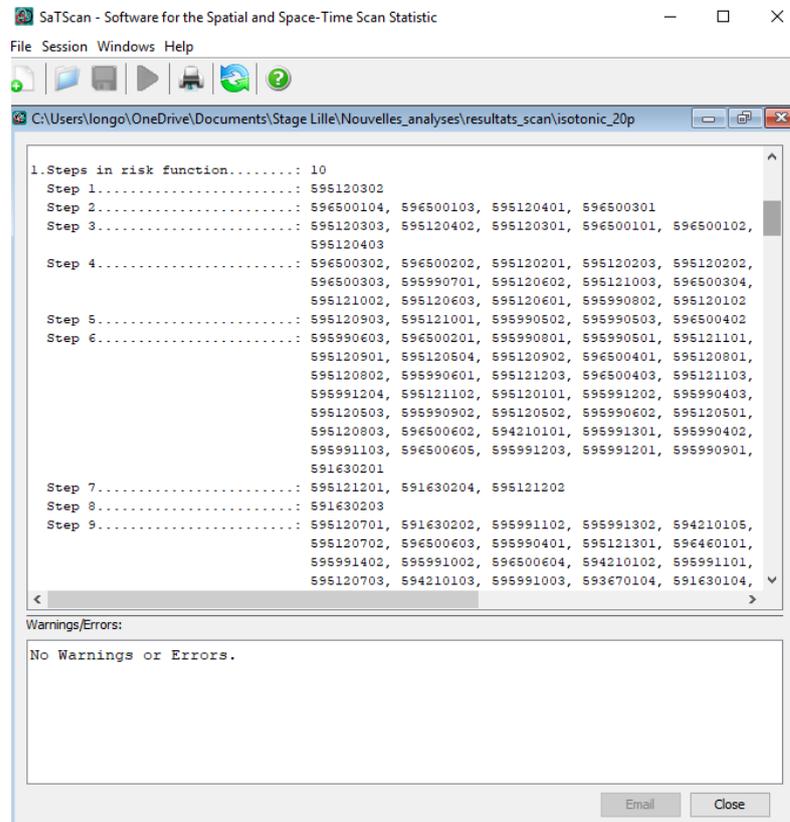
- Dans l'encart *Text Output Format*, rédiger le nom du fichier en sortie.
- Dans l'encart *Geographical Output*, sélectionner *HTML for Cartesian map*. Celui-ci donnera un fichier HTML recensant les informations de chaque cluster. Ces informations sont choisies dans l'encart suivant.
- Dans l'encart *Column Output Format*, cocher les informations qui seront utiles à l'étude. Ici, cocher :
 - *Cluster Information* pour connaître la significativité des clusters
 - *Stratified Cluster Information* pour avoir une division des informations par cercles concentriques
 - *Location Information* pour connaître les IRIS des clusters, par cercles concentriques
 - *Risk Estimates for Each Location* pour connaître le risque relatif de chaque cercle concentrique



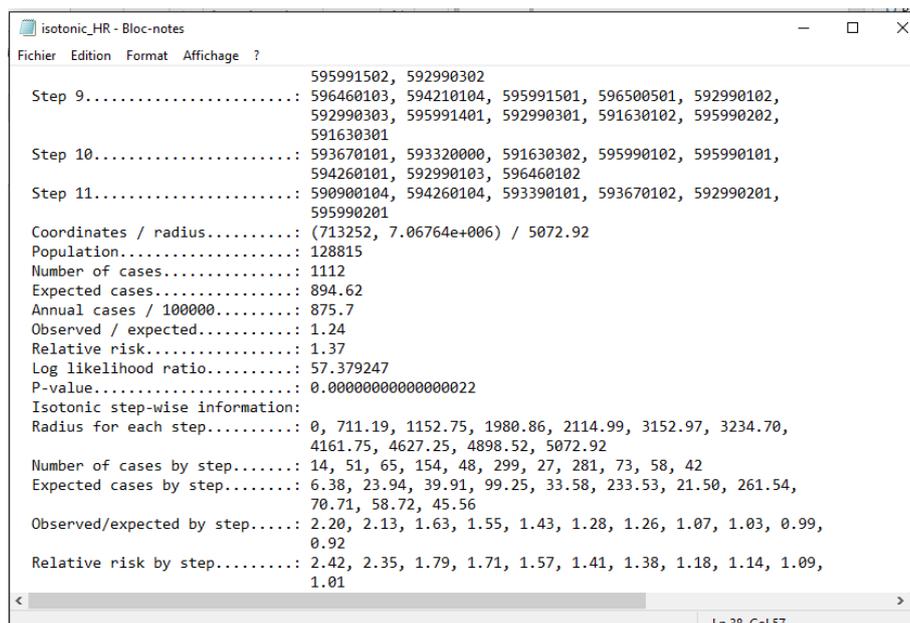
Pour démarrer les calculs, sélectionner la flèche verte en haut de la fenêtre, à droite de la disquette.

Le résultat doit être similaire à ceci :

- Sur le logiciel SaTScan :



- Sur le fichier HTML en sortie :



ANNEXE 3 : LES TESTS POST HOC DES TESTS KW SUR FDEP, POLLUTION ET MOS

	Neutre	Clusters de sous-incidence			Clusters de sur-incidence		p.overall	p.0 vs 1	p.0 vs 2	p.0 vs 3	p.0 vs 4	p.0 vs 5	p.1 vs 2	p.1 vs 3	p.1 vs 4	p.1 vs 5	p.2 vs 3	p.2 vs 4	p.2 vs 5	p.3 vs 4	p.3 vs 5	p.4 vs 5	
	N=198	Cluster 2 N=85	Cluster 3 N=30	Cluster 5 N=20	Cluster 1 N=96	Cluster 4 N=44																	
fdep09	-0.04 [-1.00;1.53]	-1.16 [-2.88;0.15]	-0.64 [-1.41;0.14]	-1.46 [-2.32;-0.93]	2.11 [1.04;3.54]	0.88 [0.20;1.36]	<0.001	<0.001	<0.001	0,014	0,071	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0,081	<0.001	0,458	<0.001	0,011	<0.001
Agricoles	9.83 [0.00;41.4]	1.45 [0.00;27.0]	40.0 [29.0;59.2]	41.4 [25.8;65.7]	0.00 [0.00;4.15]	0.00 [0.00;13.0]	<0.001	<0.001	0,041	<0.001	0,003	0,002	0,004	<0.001	0,22	<0.001	<0.001	0,237	<0.001	<0.001	0,797	<0.001	
Urbain	62.9 [34.9;91.5]	65.8 [38.2;90.2]	46.6 [33.5;62.3]	42.5 [25.0;68.5]	91.6 [71.8;100]	82.3 [51.6;100]	<0.001	<0.001	0,474	0,026	0,028	0,109	0,002	<0.001	0,348	<0.001	0,01	0,115	0,045	<0.001	0,961	0,011	
Al	1101 [942;1227]	1004 [957;1100]	1348 [1058;1521]	785 [662;838]	899 [841;964]	1046 [989;1197]	<0.001	<0.001	0,033	<0.001	0,804	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0,089	<0.001	<0.001	<0.001	<0.001	
As	1.13 [0.99;1.21]	1.12 [1.03;1.18]	1.17 [1.10;1.22]	0.72 [0.69;0.85]	1.09 [1.01;1.16]	0.89 [0.81;0.94]	<0.001	0,106	0,501	0,14	<0.001	<0.001	0,229	0,004	<0.001	<0.001	0,029	<0.001	<0.001	<0.001	<0.001	0,004	
Be	0.24 [0.23;0.25]	0.22 [0.21;0.24]	0.24 [0.24;0.25]	0.25 [0.24;0.25]	0.24 [0.23;0.25]	0.23 [0.22;0.23]	<0.001	0,372	<0.001	0,335	<0.001	0,004	<0.001	0,567	<0.001	0,013	<0.001	0,031	<0.001	<0.001	0,041	<0.001	
Cd	0.69 [0.48;0.84]	0.65 [0.57;0.72]	0.52 [0.37;0.66]	0.44 [0.28;0.84]	0.55 [0.42;0.69]	1.08 [0.88;1.19]	<0.001	0,003	0,224	0,001	<0.001	0,139	0,008	0,207	<0.001	0,62	0,001	<0.001	0,351	<0.001	0,692	<0.001	
Co	0.72 [0.63;0.79]	0.64 [0.56;0.75]	0.74 [0.67;0.79]	0.38 [0.33;0.52]	0.66 [0.58;0.74]	0.55 [0.49;0.60]	<0.001	0,002	0,005	0,575	<0.001	<0.001	0,701	0,007	<0.001	<0.001	0,019	<0.001	<0.001	<0.001	<0.001	0,002	
Cr	4.67 [3.87;5.16]	4.01 [3.61;4.47]	4.74 [4.34;5.22]	1.93 [1.71;2.73]	4.21 [3.68;4.70]	4.36 [3.57;5.00]	<0.001	0,002	<0.001	0,46	0,207	<0.001	0,477	<0.001	0,46	<0.001	<0.001	0,252	<0.001	0,066	<0.001	<0.001	
Cu	19.9 [15.3;29.7]	18.7 [16.4;22.5]	13.8 [12.6;17.2]	9.91 [8.70;11.1]	21.8 [17.0;26.0]	45.7 [20.5;70.4]	<0.001	0,599	0,375	<0.001	<0.001	<0.001	0,075	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	
Hg	0.14 [0.11;0.15]	0.14 [0.13;0.15]	0.12 [0.10;0.13]	0.09 [0.08;0.10]	0.11 [0.09;0.13]	0.12 [0.10;0.13]	<0.001	<0.001	0,043	0,011	<0.001	<0.001	<0.001	0,152	0,156	0,001	<0.001	<0.001	<0.001	0,545	<0.001	<0.001	
Mn	51.3 [43.5;55.1]	45.6 [43.4;48.3]	55.4 [44.6;61.0]	38.8 [35.0;45.3]	45.4 [40.4;53.3]	46.8 [43.4;47.8]	<0.001	0,006	<0.001	0,316	<0.001	<0.001	0,651	0,009	0,843	0,009	0,003	0,686	0,006	0,009	<0.001	0,003	
Ni	2.95 [2.67;3.25]	2.53 [2.37;2.84]	2.71 [2.53;2.92]	1.61 [1.28;1.78]	3.10 [2.93;3.42]	2.79 [2.48;2.89]	<0.001	<0.001	<0.001	0,01	0,001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0,082	0,216	<0.001	0,613	<0.001	<0.001
Pb	34.8 [19.2;47.2]	38.2 [25.4;48.0]	24.7 [13.3;29.0]	14.0 [9.36;23.1]	23.2 [20.8;26.4]	52.6 [39.1;77.8]	<0.001	<0.001	0,386	<0.001	<0.001	<0.001	<0.001	0,945	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0,021	<0.001
Pd	19.6 [17.4;22.6]	25.7 [21.8;31.5]	17.0 [16.2;17.8]	17.6 [15.3;22.7]	16.2 [15.5;18.0]	26.1 [24.8;27.1]	<0.001	<0.001	<0.001	<0.001	<0.001	0,15	<0.001	0,024	<0.001	0,223	<0.001	0,945	<0.001	<0.001	<0.001	0,936	<0.001
Sb	2.17 [1.34;2.63]	2.11 [1.59;2.34]	1.25 [1.09;1.77]	0.97 [0.75;1.59]	1.89 [1.71;2.04]	2.23 [2.06;2.47]	<0.001	0,001	0,086	<0.001	0,358	<0.001	0,007	<0.001	<0.001	<0.001	<0.001	0,003	<0.001	<0.001	0,05	<0.001	
Ti	15.9 [14.7;16.6]	13.1 [12.0;13.8]	17.1 [15.8;18.1]	13.5 [13.2;14.3]	16.8 [16.3;17.4]	14.2 [13.9;14.8]	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0,429	<0.001	<0.001	<0.001	<0.001	0,016	<0.001	<0.001	0,003	
V	3.65 [3.36;3.89]	3.37 [3.19;3.85]	3.84 [3.60;4.16]	2.32 [2.11;2.51]	3.97 [3.85;4.08]	3.29 [3.05;3.47]	<0.001	<0.001	0,007	0,037	<0.001	<0.001	<0.001	0,218	<0.001	<0.001	0,004	0,032	<0.001	<0.001	<0.001	<0.001	

	Neutre	Clusters de sous-incidence	Clusters de sur-incidence																				
		Cluster 2	Cluster 3	Cluster 5	Cluster 1	Cluster 4	p.overall	p.0 vs 1	p.0 vs 2	p.0 vs 3	p.0 vs 4	p.0 vs 5	p.1 vs 2	p.1 vs 3	p.1 vs 4	p.1 vs 5	p.2 vs 3	p.2 vs 4	p.2 vs 5	p.3 vs 4	p.3 vs 5	p.4 vs 5	
Zn	120 [96.0;137]	108 [97.0;122]	100 [82.1;125]	77.1 [61.3;108]	106 [95.3;118]	134 [124;141]	<0.001	0,002	0,031	0,036	<0.001	0,001	0,343	0,618	<0.001	0,006	0,343	<0.001	0,004	<0.001	0,031	<0.001	
RIM	1.87 [1.43;2.15]	1.72 [1.50;1.93]	1.59 [1.41;1.84]	1.09 [0.87;1.35]	1.57 [1.41;1.77]	2.07 [1.86;2.28]	<0.001	<0.001	0,158	0,01	0,001	<0.001	0,001	0,567	<0.001	<0.001	0,051	<0.001	<0.001	<0.001	<0.001	<0.001	
Poussieres	1943 [1819;2029]	2110 [2046;2177]	1878 [1822;1966]	1937 [1790;2014]	1906 [1798;2015]	2067 [1962;2149]	<0.001	0,298	<0.001	0,239	<0.001	0,835	<0.001	0,844	<0.001	0,907	<0.001	0,019	<0.001	<0.001	0,741	0,004	
Eutrophie	5.78 [5.64;5.85]	5.63 [5.59;5.70]	5.92 [5.89;5.94]	5.66 [5.64;5.69]	5.76 [5.69;5.90]	5.67 [5.61;5.74]	<0.001	0,08	<0.001	<0.001	0,001	0,004	<0.001	<0.001	<0.001	<0.001	<0.001	0,119	0,41	<0.001	<0.001	0,562	
PM10	33.3 [31.6;36.0]	34.1 [32.5;37.2]	30.9 [29.8;33.3]	30.7 [29.7;32.7]	33.4 [32.3;34.3]	33.9 [33.1;37.1]	<0.001	0,765	0,154	<0.001	0,021	0,001	0,024	<0.001	0,008	<0.001	<0.001	0,345	<0.001	<0.001	0,765	<0.001	
NO2	1150 [30.0;38.5]	1355 [31.3;39.1]	958 [25.5;33.5]	547 [25.4;33.0]	1092 [31.1;35.4]	1406 [33.4;40.0]	<0.001	0,452	0,452	<0.001	0,01	0,003	0,065	<0.001	0,001	0,003	<0.001	0,081	0,001	<0.001	0,905	<0.001	
Score_Env	928 [928;1546]	1543 [1106;1543]	1212 [863;1212]	952 [354;952]	1255 [945;1255]	1718 [1212;1718]	<0.001	0,096	0,048	0,021	<0.001	<0.001	<0.001	0,045	<0.001	<0.001	<0.001	0,104	<0.001	<0.001	0,001	<0.001	