



**HAL**  
open science

# Disparités spatiales de mortalité par cause en population générale

Walid Ghosn

► **To cite this version:**

Walid Ghosn. Disparités spatiales de mortalité par cause en population générale. Santé publique et épidémiologie. Université Paris Saclay (COmUE), 2017. Français. NNT: 2017SACLS066 . tel-01542406

**HAL Id: tel-01542406**

**<https://theses.hal.science/tel-01542406>**

Submitted on 19 Jun 2017

**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.

NNT : 2017SACLS066

THESE DE DOCTORAT  
DE  
L'UNIVERSITE PARIS-SACLAY  
PREPAREE A  
"L'UNIVERSITE PARIS-SUD"

ECOLE DOCTORALE N°570  
De Santé Publique

Spécialité de doctorat : Epidémiologie

Par

**Walid Ghosn**

**Disparités spatiales de mortalité par cause en population générale**

**Thèse présentée et soutenue au Kremlin-Bicêtre, le 21 Mars 2017 :**

**Composition du Jury :**

Pr, Bard, Denis	Professeur à l'EHESP	Président
Dr, Khat, Myriam	Directeur de Recherche Ined	Rapporteur
Dr, Zeitlin, Jennifer	Directeur de Recherche Inserm	Rapporteur
Dr, Eilstein, Daniel	Directeur de Recherche SP France	Examineur
Dr, Rey, Grégoire	Directeur du CépiDc-Inserm	Directeur de thèse
Dr, Rican, Stéphane	Maitre de Conférence Univ Paris X	Co-directeur de thèse



## **Remerciements**

Je dédie cette thèse à ma famille, aux personnes qui m'ont accompagnées dans ce travail et durant cette période importante de ma vie.

Je remercie tout particulièrement Grégoire Rey, Stéphane Rican et Eric Jouglà de m'avoir généreusement transmis leur savoir et leur passion sans compter.



## **Titre : Disparités spatiales de mortalité par cause en population générale**

**Mots clefs :** disparités géographiques, mortalité par cause, inégalités sociales, multiniveaux

**Résumé :** L'analyse des disparités géographiques de mortalité en population générale est essentielle pour la planification sanitaire. Elle a pour objectif d'identifier les facteurs impliqués dans la genèse de ces différences. Ceci contribue à orienter les décisions portant sur les populations à cibler et l'allocation de ressources. L'objectif de ce travail est de participer à expliquer les variations géographiques de la mortalité par cause de décès en population générale. En première partie, une estimation des disparités géographiques de mortalité par cause calculée depuis la base nationale exhaustive des causes médicales de décès a été réalisée à l'aide d'une mesure synthétique. Sur la base de ces constatations, une analyse écologique de l'évolution de la distribution géographique de la mortalité a été proposée. Faisant l'hypothèse que les changements démographiques sont un marqueur des changements socio-sanitaires, une typologie des changements démographiques, ainsi qu'un modèle de Poisson ont aidé à investiguer cette association avec l'évolution de la mortalité. De manière générale les disparités géographiques de mortalité ont augmenté sensiblement à partir de 1990. La contribution de l'échelle régionale aux disparités spatiales qui était très élevée en 1975, va évoluer de manière différenciée selon le sexe et la cause de décès. L'évolution de la mortalité était par ailleurs associée négativement à l'accroissement de la population. Modulé par le désavantage social et le degré d'urbanisation, ce lien était plus fort pour les maladies alcool-tabagiques et les morts violentes. La deuxième partie a été consacrée à l'investigation des facteurs de risques individuels et contextuels à l'aide d'un modèle multiniveaux. Dans un premier temps, la problématique d'une telle analyse a été élargie à événements rares de santé. Evaluée à l'aide de simulations, les recommandations sur la faisabilité d'une analyse en population générale

ont été appliquées dans un second temps sur l'Echantillon Démographique Permanent (EDP). Nous avons ainsi tenté d'expliquer les disparités géographiques de mortalité par la composition sociodémographique, l'environnement physique étant caractérisé par l'accessibilité géographique aux soins, et l'environnement social étant mesuré par un score de désavantage social. L'étude de faisabilité de l'analyse multiniveaux a mis en évidence la très forte sensibilité de la fiabilité des estimations au choix de l'échelle géographique et à l'incidence de l'évènement de santé étudié. Il existe une association contextuelle en population générale avec le désavantage social du lieu de résidence. Cette association est systématique et régulière pour toutes les causes de décès investiguées. Seules les disparités géographiques de mortalité par maladies circulatoires ont été en grande partie expliquées par les facteurs individuels et contextuels. L'accessibilité géographique aux soins n'était pas retrouvée associée à la mortalité.

Ce travail met l'accent sur la multitude des facteurs en jeu et la nécessaire complémentarité des approches utilisées afin d'appréhender les disparités géographiques de mortalité. La rareté de l'évènement étudié dans l'EDP et l'absence d'information socio-économique individuelle fiable dans la base exhaustive des causes de décès sont deux contraintes majeures de ce type d'analyse qui nécessitent le déploiement de méthodologies appropriées. Néanmoins, les perspectives d'enrichissement des données par des bases administratives et médico-administratives laissent présager des retombées importantes. Malgré l'incertitude résiduelle non négligeable sur certains résultats, l'étude met en évidence avec des éléments plus précis l'importance simultanée de facteurs individuels et contextuels en population générale.

**Title : Spatial mortality disparities by cause in general population****Keywords :** geographical disparities, cause-specific mortality, social inequalities, multilevel

**Abstract :** The analysis of geographical disparities in mortality in the general population is essential for health planning. The aim is to identify the factors involved in the genesis of these differences. This helps to guide decisions about the populations to be targeted and the allocation of resources. The objective of this work was to help explain geographic variations in mortality by cause of death in the general population.

In the first part, an estimate is made of the geographical disparities in mortality by cause calculated from the comprehensive national database of medical causes of death using a synthetic measure. On the basis of these findings, an ecological analysis of the evolution of the geographical distribution of mortality is proposed. Assuming that demographic changes are a marker of health changes in society, a typology of demographic changes and a Poisson model were used to investigate this association with mortality trends.

In general, geographical disparities in mortality have increased markedly since 1990. The contribution of the regional scale to spatial disparities, which was very high in 1975, has evolved differently according to sex and cause of death. The evolution of mortality is also negatively associated with the increase in population. Depending on the degree of social disadvantage and degree of urbanization, this link is stronger for alcohol-tobacco and violent deaths. The second part is devoted to investigating individual and contextual risk factors using a multilevel model. Initially, the problem of such an analysis was extended to the reliability of multilevel analysis of rare health events. Evaluated by using simulations, the recommendations on the feasibility of a general population analysis were then applied to the Permanent Demographic Sample (PDS).

We thus attempted to explain geographical disparities in mortality by socio-demographic composition, the physical environment being characterized by geographical accessibility to care and the social environment being measured by a social disadvantage score.

The multilevel feasibility analysis revealed the very high sensitivity of the reliability of the estimates to the geographical scale and the impact of health events studied. In the general population, there was a contextual association with the social disadvantage of the place of residence. This association was systematic and regular for all causes of death investigated. Only geographical disparities in mortality from circulatory diseases were largely explained by individual and contextual factors. Geographic access to care was not found to be associated with mortality.

This work emphasizes the multitude of factors involved and the need to combine approaches when dealing with geographical disparities in mortality. The scarcity of events studied in the PDS and the lack of reliable individual socio-economic information in the exhaustive database of causes of death are two major constraints of this type of analysis, in which the appropriate methodologies should be used. Nevertheless, the prospect of obtaining richer data thanks to the use of administrative and medico-administrative bases suggests that significant benefits may be obtained from this approach. Despite the considerable residual uncertainty on certain results, the study clearly highlights the simultaneous importance of individual and contextual factors in the general population.

## PUBLICATIONS

### Communications orales

Evolutions Sociodémographiques, Changements Territoriaux et Inégalités de Mortalité face aux Cancers en France. Séminaire de restitution de l'appel à projets libres « recherche en sciences humaines et sociales, épidémiologie et santé publique », INCa Institut National du Cancer, Boulogne-Billancourt, 15 Janvier 2015

ILEM Inégalités sociales, lieux de résidence et mortalité par causes : Analyses multiniveaux de données longitudinales en population générale. Séminaire de clôture de l'IReSP, " Soutien à la recherche sur les inégalités d'accès aux soins en cancérologie", Paris, 26 Novembre 2015.

Analyse multiniveaux d'un évènement rare de santé, Etude de faisabilité, application à la mortalité par cause en population générale. Réunion du groupe utilisateurs Echantillon Démographique Permanent, Insee – Division Enquêtes et Etudes démographiques, 19 Mai 2016, Malakoff

Analyse multiniveaux d'un évènement rare de santé, Etude de faisabilité, application à la mortalité par cause en population générale. Séminaire "Les rencontres de statistique appliquée" Analyses multiniveaux, Ined, 23 Mai 2016, Paris

### Articles

#### Publié dans l'European Journal of Public Health

Trends in geographic mortality inequalities and their association with population changes in France, 1975–2006. Eur J Public Health 2012. Ghosn W, Kassié D, Jougla E, Salem G, Rey G, Rican S

#### Soumis à Population Health Metrics

In which conditions could a multilevel analysis of rare health events be conducted? A simulation study for multilevel Poisson regression. Population Health Metrics. Ghosn W, Latouche A, Rican S, Rey G

#### En relecture au BMC Public Health

Associations of cause-specific mortality with area level deprivation and travel time to health care in France from 1990 to 2007, a multilevel analysis. BMC public Health. Ghosn W, Menvielle G, Rican S, Rey G





## Tables des matières

Liste des Figures .....	5
Liste des Tableaux .....	6
Acronymes .....	7
<b>INTRODUCTION GENERALE .....</b>	<b>9</b>
Historique des disparités géographiques de mortalité .....	9
Inégalités sociales de mortalité .....	9
L'accès aux soins .....	10
Chute de la mortalité et inégalités .....	10
L'analyse écologique comme démarche scientifique .....	11
Changements de paradigmes et modélisation .....	11
Le modèle multiniveaux .....	12
L'analyse multiniveaux des disparités géographiques de santé .....	13
Position du travail et plan .....	13
<b>METHODES ET SOURCES GENERALES .....</b>	<b>14</b>
Données CépiDc .....	14
Causes de décès et mortalité évitable .....	15
Données socio-démographiques Insee .....	15
Données individuelles : Echantillon Démographique Permanent .....	16
Echelles Géographiques .....	16
Indice de Désavantage Social .....	17
Ratio standardisé de mortalité .....	17
Modèle de Poisson .....	17
<b>I. ANALYSE ECOLOGIQUE DES DISPARITES GEOGRAPHIQUES DE MORTALITE .....</b>	<b>19</b>
Introduction .....	19
Méthodes et sources .....	21
Cartographie .....	21
Mesure synthétique des disparités géographiques .....	21
Contribution des échelles aux disparités géographiques .....	21
Corrélations temporelles des disparités .....	22
Taux d'accroissement annuel moyen de la population .....	22
Typologie des évolutions démographiques .....	23

Données de mortalité et causes de décès.....	23
Modèle statistique .....	24
Résultats.....	25
Disparités géographiques de mortalité .....	25
Association écologique entre changements démographiques et évolution de la mortalité .....	32
Discussion.....	39
<b>II. ANALYSE MULTINIVEAUX DE LA MORTALITE PAR CAUSE - ETUDE DE FAISABILITE.....</b>	<b>42</b>
Introduction.....	42
Epidémiologie et effets du contexte .....	42
Fiabilité et analyse multiniveaux des évènements rares.....	42
Les effets du contexte sur les comportements, la santé et la mortalité .....	43
Définition du contexte, problématique de l'échelle géographique.....	43
Source de données pour l'analyse multiniveaux de la mortalité en population générale .....	44
Mesure de concordance des données de mortalité .....	44
Objectifs .....	45
Méthodes et sources.....	45
Concordance des données de mortalité entre 2 sources - Analyse graphique.....	45
Mesure de concordance des RSM à plusieurs échelles .....	47
Fiabilité du Modèle de Poisson Multiniveaux - Définition du modèle.....	49
Protocole de simulation.....	49
Résultats.....	52
Design équilibré .....	52
Design déséquilibré - Population EDP.....	53
Discussion.....	64
Synthèse des résultats.....	64
Implications.....	64
Comparaison avec les études précédentes.....	64
Problème des échelles d'analyses .....	65
Problèmes de modélisation.....	65
Problèmes des données.....	66
Problèmes de méthodes statistiques .....	66

Conclusion.....	66
<b>III. ANALYSE MULTINIVEAUX DE LA MORTALITE PAR CAUSE – APPLICATION .....</b>	<b>68</b>
Introduction.....	68
Méthodes et sources .....	69
Population.....	69
Causes de décès.....	70
Echelle géographique .....	70
Analyse statistique.....	70
Données d’accessibilité géographique aux structures de soins.....	70
Distribution des Temps de Trajet aux Spécialistes et Spécialités Hospitalières .....	71
Corrélations spatiales des temps de trajet .....	73
Analyse en Composante Principale des temps de trajet aux structures de soins.....	75
Corrélations spatiales des caractéristiques contextuelles .....	78
Score d'accessibilité géographique aux structures de soins .....	78
Résultats.....	80
Discussion .....	87
Résumé des résultats .....	87
Comparaison avec la littérature.....	87
Interprétation .....	88
Limites.....	89
Conclusion.....	90
<b>DISCUSSION GENERALE .....</b>	<b>92</b>
Mesure synthétique des disparités géographiques.....	92
Analyse écologique des tendances de la distribution géographique .....	92
Fiabilité de l'analyse multiniveaux des effets du contexte .....	93
Sources de données pour l'analyse multiniveaux en population générale.....	93
Définition du contexte - problématique méthodologique et statistiques.....	93
Les associations contextuelles en population générale .....	94
Conclusion.....	96
<b>BIBLIOGRAPHIE .....</b>	<b>97</b>
<b>ANNEXE .....</b>	<b>108</b>
<b>ARTICLES .....</b>	<b>114</b>



## Liste des Figures

Figure I.1: Disparités géographiques de mortalité toutes causes des femmes: cartographie, mesure synthétique et décomposition par échelle.....	25
Figure I.2: Disparités géographiques de mortalité toutes causes des hommes: cartographie, mesure synthétique et décomposition par échelle.....	26
Figure I.3: Disparités géographiques de mortalité prématurée des hommes par cardiopathies ischémiques: cartographie, mesure synthétique et décomposition par échelle.....	27
Figure I.4: Disparités géographiques de mortalité prématurée des femmes par maladies cérébrovasculaires: cartographie, mesure synthétique et décomposition par échelle.....	28
Figure I.5: Disparités Géographiques de Mortalité Evitable Prématurée des Femmes imputable aux systèmes de soins : Cartographie, Mesure Synthétique et Décomposition par Echelle.....	30
Figure I.6: Disparités Géographiques de Mortalité Evitable Prématurée des Hommes imputable à l'absence de prévention primaire: Cartographie, Mesure Synthétique et Décomposition par Echelle.....	31
Figure I.7: Cartographie de la typologie des accroissements démographiques de la population entre 1962 et 2006.....	32
Figure I.8: Carte synthétique de la typologie des accroissements démographiques dans les quintiles extrêmes de désavantage social (FDep90).....	33
Figure I.9: Evolutions des RSM tous âges et prématurés (moins de 65 ans) entre 1975 et 2006 dans les catégories de la typologie des évolutions démographiques des communes.....	34
Figure I.10: Evolutions des RSM prématurés entre 1975 et 2006 dans les catégories de la typologie des évolutions démographiques des communes les moins et les plus socialement désavantagées.....	35
Figure I.11: Evolutions des RSM prématurés entre 1975 et 2006 dans les catégories de la typologie des évolutions démographiques par degré d'urbanisation des communes.....	36
Figure II.1: Comparaison des RSM régionaux entre EDP et CépiDc.....	46
Figure II.2: Estimation de l'effet de contexte en fonction du nombre de clusters et d'individus - Design équilibré et variance fixée à 0.05.....	57
Figure II.3: Estimation de la variance spatiale en fonction du nombre de clusters et d'individus - Design équilibré et variance fixée à 0.05.....	58
Figure II.4: Estimation de la variance spatiale en fonction de la prévalence de l'évènement - Design déséquilibré EDP et variance fixée à 0.01.....	62
Figure II.5: Biais (%) d'estimation de la variance spatiale en fonction de son amplitude - Design déséquilibré EDP et prévalence fixée à 0.01.....	63
Figure III.1: Corrélations spatiales entre les temps de trajet intercommunaux aux différentes structures de soins en 2007.....	74
Figure III.2: Analyse en composantes principales des temps de trajets aux spécialistes.....	75
Figure III.3: Analyse en composantes principales des temps de trajets aux spécialités hospitalières.....	76
Figure III.4: Analyse en composantes principales des temps de trajets aux structures de soins.....	77
Figure III.5: Variabilité spatiale de la mortalité : avant/après ajustements.....	85

## **Liste des Tableaux**

Tableau I.1: Association entre évolution de la mortalité prématurée et taux d'accroissement annuel moyen par période et cause de décès .....	38
Tableau II.1: Indice de concordance des distributions spatiales de la mortalité provenant des 2 sources: CépiDc et EDP .....	48
Tableau II.2: Biais (%) en fonction du nombre de clusters et d'individus par clusters - Design équilibré et variance fixée à 0.05 .....	54
Tableau II.3: EQM ( $10^{-2}$ ) en fonction du nombre de clusters et d'individus par clusters - Design équilibré et variance fixée à 0.05 .....	55
Tableau II.4: Probabilité de couverture (%) en fonction du nombre de clusters et d'individus par clusters - Design équilibré et variance à 0.05 .....	56
Tableau II.5: Biais (%) en fonction de la prévalence de l'évènement - Design déséquilibré EDP et variance fixée à 0.01 .....	59
Tableau II.6: EQM ( $10^{-3}$ ) en fonction de la prévalence de l'évènement - Design déséquilibré EDP et variance fixée à 0.01 .....	60
Tableau II.7: Probabilité de couverture (%) en fonction de la prévalence de l'évènement - Design déséquilibré EDP et variance fixée à 0.01 .....	61
Tableau III.1: Distribution des temps de trajet aux spécialistes .....	71
Tableau III.2: Distribution des temps de trajet aux spécialités hospitalières .....	72
Tableau III.3: Corrélations spatiales entre les caractéristiques contextuelles des communes .	78
Tableau III.4: Corrélations spatiales entre les caractéristiques contextuelles des Zones d'Emploi .....	78
Tableau III.5: Distribution de la population EDP des plus de 30 ans le 1er Janvier 1990, vivant au 31 Décembre 1991 et ayant rempli leur bulletin de recensement RP90 .....	80
Tableau III.6: Association entre mortalité toutes causes, sexe, niveau d'études individuel, désavantage social et temps de trajets aux soins .....	81
Tableau III.7: Association entre mortalité par cause, sexe, niveau d'études individuel, désavantage social et temps de trajets aux soins .....	82
Tableau III.8: Effets du contexte sur la mortalité par cause (ajusté sur le niveau d'études individuel) .....	84
Tableau III.9: Association entre mortalité par cause, sexe, niveau d'études individuel, désavantage social du contexte, temps de trajets aux soins et densité de population .....	86

## Acronymes

ACS : Aide au paiement d'une Complémentaire Santé

CIM : Classification Internationale des Maladies

CMU-C : Couverture Maladie Universelle complémentaire

EAR : Enquêtes Annuelles de Recensements

EDP : Echantillon Démographique Permanent

EQM : Ecart Quadratique Moyen

GEE : Equations d'Estimations Généralisées

INSEE : Institut National de la Statistique et des Etudes Economiques

ipc : individus par clusters

IRDES : Institut de Recherche et Documentation en Economie de la Santé

ISCED : International Standard Classification of Education (classification internationale des niveaux d'études)

MAUP : Modifiable Areal Unit Problem (Problème des unités géographiques modifiables)

PSDC : Population Sans Double Compte

RNIPP : Répertoire National d'Identification des Personnes Physiques

RP : Recensement de la Population

RSM : Ratio Standardisé de Mortalité

SNDS : Système National des Données de Santé

TAAM : Taux d'Accroissement Annuel Moyen

UGCoP : Uncertain Geographic Context Problem (Problème d'incertitude du contexte géographique)

VADS : Voies Aéro-Digestives Supérieures

VPC : Variance Partition Coefficient (coefficient de partition de la variance)

ZE94 : Zone d'Emploi 1994





## INTRODUCTION GENERALE

### **Historique des disparités géographiques de mortalité**

Les premières données reconstituées sur les disparités géographiques de mortalité en population générale en France métropolitaine datent du début du XIX<sup>ème</sup> siècle [1]. De manière générale, le XX<sup>ème</sup> siècle est marqué par une réduction considérable de la mortalité et de ses disparités départementales [2]. L'observation de cartes de mortalité, disponibles à une échelle plus fine depuis 1968, fait apparaître des différentiels spatiaux très marqués par le fait régional, mais faisant de plus en plus de place à des disparités intra-régionales [3]. A partir des années 1990, les différentiels entre les grandes agglomérations et les villes situées en marge du réseau urbain principal d'une part, et à l'intérieur même des espaces urbains entre centres urbains, première et deuxième couronnes périurbaines d'autre part, vont continuer de s'accroître [4]. Exception faite des anciennes grandes régions industrielles de l'arc Nord-Est qui restent en forte surmortalité, les disparités spatiales entre régions urbanisées et régions traditionnellement rurales, de même que les écarts de mortalité entre communes urbaines socialement avantagées et désavantagées vont se creuser à partir de 1990 [4,5]. Globalement, ces structurations spatiales de la mortalité générale sont fortement dépendantes des structurations associées aux causes de décès liées aux comportements à risques (consommation d'alcool, de tabac, accident de la route, suicide) [6].

### **Inégalités sociales de mortalité**

En dépit d'un système de santé présenté comme l'un des plus performants au monde (selon l'OMS au début des années 2000 [7]), caractérisé par une égalité d'accès à de nombreux soins, et se traduisant par une espérance de vie à la naissance parmi les plus longues (78,9 ans pour les hommes et 85,1 ans pour les femmes en 2013), les inégalités sociales de mortalité restent fortes comparativement à celles d'autres pays industrialisés [8]. L'espérance de vie à 35 ans d'un cadre est de 6,3 années plus longue que celle d'un ouvrier. Ce même écart est de 3 années pour les femmes sur la période 2000-2008. Parmi les catégories sociales les plus vulnérables, les hommes inactifs non retraités ont une espérance de vie à 35 ans qui est de 10,2 années plus courte que pour le reste de la population [9]. Au cours de la période 1990-1996, le risque de mortalité prématurée était 2,3 fois plus élevé pour un homme et 1,9 pour les femmes n'ayant pas achevé leur cycle d'éducation primaire par rapport aux bacheliers et niveau d'études supérieures [10]. La contribution des causes de décès à la surmortalité prématurée des ouvriers par rapport aux cadres est très marquée par les causes liées à la surconsommation d'alcool et de tabac, aux accidents et aux suicides [11]. Au-delà de ces oppositions entre catégories sociales extrêmes, un gradient social de mortalité est constaté dans la plupart des travaux portant sur les inégalités sociales de mortalité et ce quelle que soit la mesure des catégories sociales retenue : classe sociale, catégories socio-professionnelle, statut socio-économique, niveau d'études, niveaux de revenu, etc. [11-14]. Pourtant, aucune réelle stratégie de réduction de ces inégalités n'a jusqu'à présent été mise en œuvre au niveau national [15].

## **L'accès aux soins**

L'accès aux soins est défini, entre autres, par des facteurs influençant l'entrée dans le système de soins, le recours ou non-recours au système de santé et module de manière hétérogène la santé des individus [16]. Suivant le modèle d'Andersen [17], les caractéristiques du système de santé favorisant la consommation de soins s'articulent autour de trois notions: la disponibilité, l'accessibilité, et le coût des soins. Le système de santé français combine à la fois une couverture maladie pour tous mais qui s'appuie majoritairement sur le travail et la capacité à s'ouvrir des droits complétée par l'accès à une complémentaire santé couvrant le reste à charge (mutuelle ou assurance privée). Ces dernières années, la mise en place de la Couverture Maladie Universelle complémentaire pour les plus socialement vulnérables (CMU-C) et de l'Aide au paiement d'une Complémentaire Santé pour les situations intermédiaires (ACS), a permis d'atténuer le fardeau que peut représenter ce reste à charge. Malgré ces efforts pour améliorer l'accès aux soins, il semble que des barrières culturelles et financières de recours aux soins persistent, et sont socialement distribuées [18]. Ceci se traduit par des inégalités sociales de recours aux soins. Il s'avère en effet que les individus ayant les besoins les plus importants consomment moins de soins [19]. Le système de soins repose sur un mélange d'acteurs publics et privés fournissant des soins hospitaliers et ambulatoire, une liberté d'installation pour les donneurs de soins et un paiement à l'acte pour les usagers. Afin de lutter contre ces inégalités individuelles d'accès aux soins, la dernière réforme initiée en 2014 et votée en 2016 généralise le tiers-payant et rend obligatoire l'accès à une assurance complémentaire. Il existe en plus des déterminants sociaux individuels classiquement décrits dans le recours et l'accès aux soins, d'importantes disparités territoriales d'accès aux soins [20]. Partant de l'après-guerre, plusieurs réformes nationales du système de santé ont été entreprises afin de contrôler les dépenses publiques et de rééquilibrer l'inégale distribution géographique des ressources et structures de soins [21]. Longtemps centralisées, ces compétences (ou responsabilités) ont progressivement été transférées vers une gestion régionale, en particulier par la mise en place des Agences Régionales de la Santé en 2009. Cette réforme du système de santé introduit pour la première fois la notion d'accessibilité géographique dans la planification sanitaire [22]. Un des autres objectifs de la loi de santé appliquée en 2016 vise également à développer les soins primaires et à améliorer la coordination des différents acteurs locaux de la santé et ce, au niveau régional [23].

## **Chute de la mortalité et inégalités**

Depuis 1990 en France, de fortes réductions des taux standardisés de mortalité ont été constatées pour les maladies circulatoires, les maladies respiratoires, les maladies digestives et les morts violentes. Dans une moindre mesure, on observe également une diminution de la mortalité par cancer (différenciée selon la localisation de la tumeur et selon le sexe). Les cancers constituent depuis 2004 la première cause de mortalité en France [24]. Cette rapide diminution de la mortalité par maladies circulatoires, cancers et maladies alcoolo-tabagiques aurait eu lieu de manière différenciées selon les groupes sociaux. Ainsi, l'amélioration globale de la santé des populations s'accompagnerait d'une augmentation des inégalités sociales relatives de mortalité comme cela a été constaté dans plusieurs pays [8,25,26]. On constate également une augmentation des différentiels spatiaux de mortalité à l'échelle des communes, notamment associés au niveau socio-économique moyen de la population résidente [27,28].

Aussi, la contribution des processus spatiaux liés au développement des territoires et aux changements sociaux et sanitaires a jusqu'ici été peu investiguée. Ces évolutions difficilement quantifiables car multidimensionnelles peuvent cependant être explorées et surveillées à l'aide de marqueurs. Parmi les marqueurs de processus spatiaux, le taux d'accroissement de la population reflète une dimension sociodémographique du territoire sommant accroissement naturel de la population et attractivité. Certaines études au Royaume-Uni ont mis en évidence que les zones défavorisées perdaient leur population en meilleure santé au profit des zones plus favorisées. Ce type d'échanges déséquilibrés aboutiraient au regroupement géographique des individus en meilleure santé laissant les individus les plus vulnérables dans les zones en perte de population [29–33].

### **L'analyse écologique comme démarche scientifique**

La démarche quantitative couramment utilisée pour étudier les disparités géographiques de santé est l'analyse écologique. Ce type d'étude met en relation des indicateurs de santé agrégés au niveau géographique (ex : le taux de mortalité à l'échelle départementale) avec des caractéristiques des unités géographiques (ex : le taux de chômage). La disponibilité et l'accessibilité des données constituent un atout majeur de cette démarche. Une mesure d'exposition à un facteur environnemental est toujours plus simple et moins coûteuse à mettre en place quand elle est estimée à partir d'une moyenne sur un espace délimité géographiquement. Certaines expositions environnementales comme la qualité de l'eau n'existent qu'à l'échelle du même réseau de distribution. Aussi, de nombreuses caractéristiques sociodémographiques, qui sont entre autres issues des recensements, sont facilement accessibles aux échelles administratives agrégées et ne posent pas, à l'inverse des données individuelles, de problèmes d'ordre réglementaire. L'inconvénient majeur de ce type d'analyse est l'impossibilité d'interpréter une association écologique au niveau individuel sans risquer de commettre une erreur écologique. Les deux exemples fondateurs de l'erreur écologique sont ceux de Durkheim et Robison. D'après les travaux de Durkheim, le ratio des taux de suicide entre régions de Prusse ayant un faible pourcentage de protestants contre celles qui en avaient le plus allait de 1 à 7.6, alors que ce ratio tombait à 2 à l'échelle individuelle [34]. L'exemple de Robison dans lequel une proportion élevée d'immigrés dans les états Américains était associée à un taux d'illettrisme plus faible illustre bien les limites de ce type d'analyse. Quand ce lien était exploré au niveau individuel, il s'avérait inversé conformément à l'hypothèse initiale de l'auteur. Ce lien écologique inattendu était dû à la forte attractivité pour les nouveaux arrivants des états industriels ayant en moyenne un niveau d'instruction plus élevé que les états ruraux où l'illettrisme était monnaie courante [35]. Une autre source d'erreur peu évoquée dans la littérature vient du possible temps de latence entre l'exposition à des facteurs de risque (sociaux ou environnementaux) et leurs effets sur la mortalité. Finalement, de par leurs avantages et limites les études écologiques sont principalement utiles pour la veille, explorer et générer de nouvelles hypothèses sur l'étiologie des maladies en vue d'études plus poussées incluant des données individuelles [36,37].

### **Changements de paradigmes et modélisation**

L'historique des méthodes utilisées dans les études épidémiologiques reflète les changements de paradigmes opérés depuis le XIX<sup>ème</sup> siècle. Sur fond de baisse générale de mortalité,

l'avènement des maladies chroniques comme principales causes de mortalité vers le milieu du XX<sup>ème</sup> siècle pousse les épidémiologistes à revoir progressivement le modèle classique hôte-agent-environnement. Après la seconde guerre mondiale, l'épidémiologie des maladies infectieuses fait place au paradigme des facteurs de risques [38]. L'étiologie des maladies chroniques étant largement multifactorielles, on utilise le terme de "réseau de causalité" ("web of causation") pour exprimer l'idée que ces causes sont liées entre elles. L'apport de l'épidémiologie pour la santé publique repose sur l'idée qu'identifier, isoler et mesurer les effets des facteurs de risques dans les chaînes causales permet de fournir des informations sur les améliorations possibles de la santé des populations [39]. Dans ce champ, les études individuelles traduisent une vision individualisée du risque destinée à améliorer l'action réparatrice. Ce paradigme a cependant été critiqué pour son apport partiel en matière de santé publique. En effet, adopter l'individualisme méthodologique (ou biomédical) pour s'attaquer aux inégalités sociales de santé reviendrait à expliquer la santé des populations par des différences physiopathologiques individuelles, ignorant l'influence préalable de la position socio-économique et de l'environnement dans lequel les populations évoluent. La surmortalité des couches inférieures du spectre social et de certaines zones géographiques abritant des populations vulnérables est très marquée par des causes liées aux pratiques de santé délétères. D'un point de vue populationnel, ces comportements qui sont socialement distribués ne peuvent être imputables aux seuls individus et à leur choix propre. Dans cette optique, l'épidémiologie sociale qui cherche à comprendre les mécanismes par lesquels la position sociale influe sur notre santé a également été amenée à s'intéresser aux effets du contexte de résidence [40–42]. Plusieurs études ont mis en avant l'importance du rôle de l'environnement physique et social de résidence, vécu ou perçu comme un espace d'opportunité, sur les pratiques et comportements de santé [43–45]. La théorie relative aux facteurs de risques intervenant simultanément au niveau individuel et contextuel considère que la santé est aussi le produit de l'environnement social et matériel [46,47]. L'identification des effets dits « multiniveaux » et plus particulièrement ceux du contexte de résidence relève de l'action préventive en santé publique [48].

### **Le modèle multiniveaux**

Les outils d'analyse de l'effet du contexte sur la santé sont les modèles multiniveaux. Ils se seraient développés avec la convergence de deux courants dans la statistique et les sciences sociales dans les années 1980 [49–51]. D'une part, les modèles mixtes qui sont capables d'inclure simultanément des effets fixes et des effets aléatoires font l'objet d'un développement considérable. D'autre part, l'idée que les comportements individuels peuvent être influencés par le contexte fait son chemin dans les sciences sociales [52]. Plus particulièrement, ce sont les sciences de l'éducation qui ont le plus fortement contribué au développement et à l'utilisation de tels modèles. Les élèves groupés dans des classes au sein d'écoles qui elles-mêmes sont situées dans des environnements sociaux disparates confèrent à ces données une structure strictement hiérarchique [51]. Plus précisément, les modèles multiniveaux permettent de mesurer la dépendance inhérente aux observations, par exemple des individus partageant le même espace de vie, et de la considérer comme une information pertinente et exploitable. Les modèles à Equations d'Estimation Généralisées (GEE) considéraient jusqu'ici cette corrélation simplement comme un paramètre à prendre en compte

[49,53]. Cette méthode permettait seulement de tester si les différences entre groupes étaient significatives en prenant en compte la composition de ces groupes. Le modèle multiniveaux, permet lui, non seulement de prendre en compte les différences intervenants sur plusieurs niveaux, individu et contexte, mais peut s'atteler également à expliquer cette dispersion quand elle existe [54,55].

### **L'analyse multiniveaux des disparités géographiques de santé**

Utilisé dans le cadre de l'analyse des disparités géographiques de mortalité, l'analyse multiniveaux permet de répondre à plusieurs questions clés. Quelle est l'amplitude des disparités géographiques de mortalité ? A quelles échelles géographiques interviennent-elles ? Adopter une posture réductionniste de la santé individuelle qui serait influencée d'une part par les caractéristiques sociodémographiques et d'autre part par les caractéristiques contextuelles du lieu de résidence permet également de poser une question essentielle: les disparités géographiques de mortalité sont-elles le reflet de la composition sociale des territoires ? Après introduction des caractéristiques individuelles, reste-il une part de variabilité spatiale non expliquée ? Est-ce que cette part résiduelle peut être expliquée par un contexte particulier ? Pour analyser les disparités spatiales de santé, le modèle de Poisson multiniveaux est une des clés de la quantification du rôle du contexte de résidence sur l'incidence d'évènements rares de santé tels que la mortalité indépendamment des caractéristiques individuelles. Certaines études ont mis l'accent sur le fait que ce type d'analyse permettrait une meilleure identification des populations à risques, étape préalable à une éventuelle répartition des ressources plus efficace [40,45]. Une investigation des facteurs contextuels de la mortalité par cause spécifique en population générale nécessite une bonne précision des estimations. Cependant, les performances du modèle multiniveaux de Poisson, qui doit souvent être utilisé pour exploiter ce type de données, n'ont reçu jusqu'à présent aucune attention, alors même que des erreurs d'estimation et de diagnostics seraient préjudiciables à la qualité de la prise de décision en santé publique.

### **Position du travail et plan**

L'épidémiologie, pour être efficace, doit s'appuyer sur les apports et acquis venus d'autres disciplines, ce qui lui vaut même d'être qualifiée de discipline "carrefour" [56]. Pour analyser les disparités spatiales de santé, la diversité des approches pour les appréhender renseigne sur la multiplicité des points de vue liant l'individu, sa santé et son environnement. L'approche positiviste interroge l'arrangement spatial des taux à l'aide de méthodes quantitatives. Cette approche fait écho à l'épidémiologie moderne qui cherche un ordre dans les disparités pour étudier l'étiologie des maladies. Deux courants non imperméables s'y distinguent : le positivisme social qui cherche ses facteurs discriminants dans les caractéristiques sociodémographiques des lieux, et le positivisme biomédical cherchant une relation causale avec les différences physiologiques individuelles. L'approche interactionniste (constructiviste), qui est issue principalement d'études qualitatives, s'intéresse aux représentations de la maladie, à leur construction sociale dont les fondements se situent dans les interactions sociales entre individus [57]. Dans la mesure où le positivisme s'appuie sur ce qui est observable, quantifiable et généralisable, nous envisageons nos travaux comme s'inscrivant dans ce courant. Reconnaisant pleinement l'apport de la géographie de la santé,

ainsi que la complémentarité des approches constructivistes qui lui sont associées, aussi bien pour la formulation d'hypothèses que pour l'interprétation de résultats issues de méthodes quantitatives, nous tenterons de lui soustraire ses apports pour enrichir nos connaissances. L'ambition de ce travail de thèse est de participer à expliquer la distribution géographique de la mortalité par cause en population générale.

La première partie consiste en une analyse écologique des disparités géographiques de mortalité par cause. En premier lieu, nous proposerons de quantifier les disparités géographiques de mortalité et leurs évolutions par échelle et combinaison d'échelles depuis 1975. Cette analyse a pour vocation d'explorer la distribution des disparités géographiques et ses tendances en utilisant la base exhaustive des causes médicales de décès.

Dans un deuxième temps, nous tenterons d'expliquer l'évolution de la distribution géographique de la mortalité par cause par les changements démographiques. Nous ferons l'hypothèse que le taux d'accroissement de la population est un marqueur des changements socio-sanitaires. L'évolution de la mortalité sera observée à l'aide d'une typologie des évolutions démographiques. Elle a pour objectif de capturer les processus spatiaux ayant traversé cette période tels que la métropolisation, la périurbanisation, la désindustrialisation, etc.

La deuxième partie de ce travail est une étude de faisabilité d'une analyse multiniveaux des disparités géographiques de mortalité par cause en population générale qui prend en compte le niveau individuel. En généralisant cette problématique à la fiabilité d'une analyse multiniveaux d'un évènement rare de santé, nous tenterons d'établir des recommandations sur le nombre d'unités géographiques, le nombre d'individus par unités, l'incidence de l'évènement et l'échelle géographique à considérer dans le cas français. Ce travail a été effectué à l'aide de simulations et compare également la fiabilité de plusieurs méthodes statistiques d'estimations bayésiennes et fréquentistes.

La troisième partie est l'application d'une analyse multiniveaux de la mortalité par cause effectuée sur un échantillon représentatif de la population française. En suivant les recommandations de la partie précédente, nous tenterons d'établir s'il existe une association contextuelle de la mortalité par cause avec l'environnement social et la structuration des services de soins du lieu de résidence. L'environnement social sera caractérisé par un score de désavantage social et la structuration des services de soins sera mesurée par un score synthétique d'accessibilité géographique aux soins.

## METHODES ET SOURCES GENERALES

### **Données CépiDc**

Les données de mortalité par cause de décès survenus entre 1968 et 2013 ont été utilisées. Ces données individuelles sont produites chaque année par l'Inserm-CépiDc et font l'objet d'un appariement avec les données d'état civil et du Répertoire National d'Identification des Personnes Physiques (RNIPP) traitées par l'Insee. Elles sont de ce fait considérées comme

exhaustives. Pour chaque décès, les variables utilisées dans la base du CépiDc sont:

- l'âge,
- le sexe,
- la cause initiale du décès,
- la date de naissance
- la date de décès,
- la commune de domicile au moment du décès

Les causes de décès sont mentionnées par le médecin certificateur sur le certificat de décès. Elles sont codées au CépiDc selon la Classification Internationale des Maladies (CIM) qui a évolué au cours de la période étudiée : codées entre 1968 et 1978 selon la 8ème révision (CIM8), selon la 9ème révision (CIM9) entre 1979 et 1999 et selon la 10ème révision (CIM10) de 2000 à 2013. Sur chaque certificat de décès le médecin déclare dans l'ordre rétrospectif les causes qui ont mené au décès. Le processus morbide principal ainsi détaillé permet au CépiDc de sélectionner par codage automatique et/ou manuel la cause initiale du décès selon les règles internationales décrites dans la CIM (Volume 2).

### **Causes de décès et mortalité évitable**

Durant tout le long de ce travail, nous considérerons le concept de mortalité évitable [58] qui regroupe deux catégories de causes de décès :

- Les décès imputables à la prévention, c'est-à-dire qui auraient pu être prévenus par des interventions en santé publique ciblant les déterminants sociaux, les comportements, les modes de vie et les facteurs environnementaux (*preventable*).
- Les décès imputables au système de soins, où l'ensemble des décès qui aurait pu être évités par une prise en charge efficace et à temps par le système de soins (*amenable*).

Ces regroupements nous permettront de sonder ces deux mécanismes potentiels intervenant dans la constitution de différentiels de mortalité. Les causes ainsi que les codes CIM associés à ces deux catégories sont détaillés en Annexe (Tableau S2). Les codes CIM des regroupements plus fins de causes de décès investigués dans ce travail sont disponibles dans chacune des parties ainsi qu'en Annexe (Tableau S1).

### **Données sociodémographiques Insee**

Les données sociodémographiques sont issues des Recensements exhaustifs de la Population (RP) en 1968, 1975, 1982, 1990, 1999 et des Enquêtes Annuelles de Recensements (EAR) qui permettent d'obtenir des statistiques robustes à partir de 2006. Mises à part les informations pouvant être recoupées avec l'état civil, ces données agrégées sont donc principalement construites sur une base déclarative.

Auparavant effectués sur la population entière à intervalles plus ou moins réguliers allant de 7 à 9 ans, les recensements sont réalisés grâce à l'EAR sur des échantillons de population à partir de 2004. Pour les communes de moins de 10 000 habitants, l'enquête de recensement porte sur toute la population, à raison d'une commune sur cinq chaque année. Les communes de 10 000 habitants ou plus, réalisent tous les ans une enquête par sondage auprès d'un échantillon d'adresses représentant 8 % de leurs logements. En cumulant cinq enquêtes,



l'ensemble des habitants des communes de moins de 10 000 habitants et 40 % environ de la population des communes de 10 000 habitants ou plus sont pris en compte. Les résultats complets du recensement de l'année centrale produits à partir des cinq enquêtes annuelles les plus récentes sont disponibles pour les années allant de 2006 à 2013.

### **Données individuelles : Echantillon Démographique Permanent**

L'Echantillon Démographique Permanent (EDP) est une base de données longitudinale représentative de la population française. Cet échantillonnage à environ 1% de la population chaîne les données d'état civil avec les données déclarées sur les formulaires de recensements par les individus nés durant 4 dates calendaires de chaque année. Elle contient de nombreuses informations sociodémographiques : commune de résidence, niveau d'études atteint, catégorie socio-professionnelle, activité, statut d'occupation du logement, statut marital, etc. [59]. Dans ce travail de thèse, les données individuelles issues des recensements de 1968, 1975, 1982, 1990 et 1999 ont été utilisées.

Cette base de données a été appariée avec la base des causes médicales de décès à l'aide d'une méthode d'appariement indirecte. Au total, 99.18% des décès de l'EDP ont été appariés avec la base du CépiDc (voir Annexe - Méthodes et Résultats de l'appariement de la base du CépiDc avec l'EDP).

### **Echelles Géographiques**

Les frontières administratives communales ont évolué au cours du temps. Certaines communes ont fusionné pour ne former qu'une seule entité, d'autres se sont scindées pour devenir deux communes distinctes. Il existe aussi d'autres redéfinitions de ces frontières comme le fait qu'une commune donne, reçoive ou échange des parcelles pouvant abriter ou non des populations. L'analyse de l'évolution des disparités spatiales de mortalité nécessite une échelle unique afin de permettre les comparaisons temporelles.

Pour ce faire, une échelle géographique communale invariante au cours du temps a été construite d'après la règle suivante : si une interaction quelconque (fusion, scission, échange de parcelles) a eu lieu entre 1968 et 2013, les deux communes concernées sont réunies sous le code commune de la plus peuplée des deux au recensement de 1999.

Dans ce travail, plusieurs échelles géographiques ont été considérées à partir des correspondances (datant de 1999, table nommée basegeo99) entre communes harmonisées et les échelles suivantes : canton, Zone d'Emploi 94 (ZE94), département, région. Dans la suite de ce travail, l'échelle Zone d'Emploi 94 sera nommée indistinctement échelle Zone d'Emploi ou bien échelle ZE94. Ci-dessous un descriptif de la distribution de la population recensée en 1990 dans chacune de ces échelles.

Echelle	nombre unités	moyenne habitant	1 <sup>er</sup> quartile	médiane	3 <sup>ème</sup> quartile
Région	22	2575076	1444893	1964967	2798016
Département	96	590122	286687	483047	758793
Zone d'Emploi	348	162792	58618	91941	172562
Canton	3703	15299	4930	9133	17716
Commune*	36310	1560	172	368	892

\* l'échelle Commune correspond aux communes harmonisées selon la méthodologie précédemment décrite.

### Indice de Désavantage Social

L'indice de désavantage social FDep99 pour *French Deprivation index 1999* a été utilisé afin de caractériser le niveau de désavantage des communes [60]. Cet indice est un score construit à partir de la 1<sup>ère</sup> composante principale (68% de variance expliquée) issue d'une analyse en composante principale pondérée par la population de quatre variables socio-économiques : le taux de chômage dans la population active, le taux d'ouvriers dans la population active, le taux de bacheliers dans la population des plus de 15 ans, ainsi que le revenu médian des ménages par unité de consommation. Cette dernière information n'étant pas disponible pour les communes de moins de 50 ménages, environ 6 000 sur 36 600 communes au total furent exclues.

L'indice de désavantage social aux échelles plus agrégées que l'échelle communale est calculé à partir de la moyenne pondérée des scores communaux appartenant à l'unité géographique supérieure (canton, zone d'emploi, département, etc.).

### Ratio standardisé de mortalité

La Ratio Standardisé de Mortalité (RSM) est un ratio entre les décès observés sur une période donnée au sein d'une unité géographique  $i$  soit  $O_i$ , sur le nombre de décès attendus  $E_i$  si la mortalité par classe d'âge et sexe était identique à la moyenne française.

Ainsi, le ratio standardisé de mortalité de la commune  $i$  s'écrit :  $RSM_i = O_i/E_i$ ,

où  $E_i = \sum_{\substack{s \in \text{sexe} \\ a \in \text{age}}} Tx_{s,a} \cdot Pop_{a,s}^i$ ,

$Tx_{s,a}$  étant le taux de mortalité moyen pour la classe d'âge quinquennale  $a$  et le sexe  $s$ ,  $Pop_{a,s}^i$  l'effectif de population de la classe âge  $a$  pour le sexe  $s$  dans l'unité géographique  $i$ .

### Modèle de Poisson

Une analyse écologique classique de la mortalité fait usuellement appel à un modèle de Poisson dont la définition générique est la suivante :

$$\begin{cases} O_i \sim \text{Poisson}(\lambda_i) \\ \log \lambda_i = \alpha + \beta \cdot X_i + \log E_i \end{cases}$$

où  $X$  est la matrice des variables caractérisant l'unité géographique  $i$  et  $\beta$  les associations avec ces mêmes caractéristiques.



# I. ANALYSE ECOLOGIQUE DES DISPARITES GEOGRAPHIQUES DE MORTALITE

## Introduction

L'analyse des disparités géographiques de santé commence avec la représentation cartographique qui permet d'appréhender leurs distributions spatiales dans un but exploratoire et descriptif. Une carte des taux de mortalité rend compte de la structuration spatiale sous-jacente. L'articulation des échelles géographiques en jeu dans ces différentiels permet de repérer les ressorts de structurations spatiales. Les cartes de la mortalité sont de ce fait de puissants outils pour l'épidémiologie, aidant à la formulation d'hypothèses sur les possibles liens entre santé et environnement. Ainsi, plusieurs études ont mis en lumière une surmortalité par mésothéliome dans les zones côtières abritant des chantiers navals qui sont des zones d'exposition intense à l'amiante [3,36,61]. Ces cartes de la mortalité par cause offrent également un bon support en termes de planification sanitaire permettant d'orienter les débats portant sur l'allocation de ressources [62]. La puissance descriptive d'une carte repose dans les continuités et les ruptures qu'elle laisse apparaître, ouvrant de nombreuses pistes à l'explication des facteurs sous-jacents. Son apport à la connaissance commence donc à partir d'une observation à l'œil nu. Cependant, une mesure synthétique des disparités géographiques de mortalité renseigne également sur leur ampleur, leurs évolutions ainsi que sur les échelles géographiques impliquées dans ces différentiels. Le développement des méthodes statistiques d'analyse de la mortalité permet aujourd'hui de proposer une mesure synthétique et une description de sa distribution spatiale. Compte tenu des changements socio-territoriaux, une telle quantité pourrait permettre d'évaluer les changements intervenus dans la géographie de la mortalité, de mettre évidence ses structures spatiales et d'interroger les échelles qui révèlent ces disparités.

Par ailleurs, l'identification des facteurs explicatifs des disparités géographiques de santé est le plus souvent faite à partir d'études dites écologiques. Ce type d'analyse considère l'unité géographique comme un regroupement pertinent pour questionner une relation entre un risque pour la santé, généralement des taux calculés à une échelle géographique donnée, et des niveaux moyens d'expositions des populations résidant dans ces mêmes unités géographiques [36]. Jusqu'ici ces analyses se sont intéressées aux liens statistiques entre santé et exposition aux différents polluants [63], à la proximité avec des installations de l'industrie minière ou pétrochimique [64,65], à l'exposition au soleil [66], à la consommation d'alcool et tabac [67] ou encore à la densité des enseignes qui en fournissent etc. [68]. Si la présence d'une association écologique ne permet pas d'établir de liens causaux, ni de séparer les effets individuels des effets du contexte [37,69], de telles analyses interrogent cependant l'hypothèse de ces liens sur de larges populations à faible coût [70]. Ainsi, une association écologique entre niveau socio-économique de la zone de résidence et santé a été mise au jour de manière quasi-systématique dans plusieurs pays [27,71–75]. En particulier, en France, il existe une association positive entre la mortalité et un indice de désavantage social [27]. Conçu comme un score de cumul de désavantages, il explique une large partie des disparités spatiales de la

mortalité tous âges, de la mortalité avant 65 ans, et de la mortalité pour les causes dites "évitables liées aux comportements à risque et à la prévention primaire" en 1999. Cependant, ce type d'indicateurs peinent à capter les processus spatiaux et les recompositions à l'œuvre qui sont susceptibles d'influer sur les niveaux de santé et leurs évolutions.

En effet, depuis les années 70, la France est traversée par des changements sociaux et territoriaux majeurs. Le passage à un marché de l'emploi largement dominé par le secteur tertiaire et son pendant, la désindustrialisation, modifie en profondeur les territoires entraînant des flux migratoires toujours plus orientés vers les zones urbaines et les grandes métropoles [4]. La concentration urbaine amorcée depuis la révolution industrielle et la création de grandes régions industrielles comme l'arc Nord-Est fait place à la métropolisation, à l'étalement urbain, ou à la périurbanisation, avec la constitution de nouvelles zones de résidence éloignées des centres villes devenus les principales zones d'emploi. Ces changements ont induit une redirection des flux financiers et des services de l'état (écoles, postes, hôpitaux) avec leurs lots de conséquences sur les économies locales [76]. La fermeture d'une usine entraînant avec elle la fermeture de certains commerces est un exemple assez courant de ces mécanismes délétères pour les communautés. S'ensuivent les départs des actifs et jeunes arrivant sur le marché du travail vers les zones d'emploi entraînant ces territoires en déclin dans une spirale négative [77]. A l'inverse, l'attractivité d'une zone enregistrant toujours plus d'arrivées reflète une dynamique plutôt positive. Il est possible d'envisager que ces zones recèlent plus d'opportunités pour des populations (jeunes adultes, actifs, retraités) ayant des attentes diverses [78].

Parmi les marqueurs explicatifs des évolutions de la distribution spatiale de la mortalité, les changements démographiques ont souvent été mis en avant mais relativement peu étudiés. L'accroissement ou le déclin de la population au sein d'une unité géographique traduit pourtant des dynamiques sociales, économiques, culturelles et environnementales qui ne sont pas forcément prises en compte dans les indicateurs usuels de désavantage. L'évolution des effectifs de population avec de manière sous-jacente les mobilités résidentielles socialement différenciées [79] et l'instabilité résidentielle des ménages [80], reflète indirectement les dynamiques locales en terme d'emploi, de service et d'habitat qui peuvent influencer des changements dans la mortalité. Jusqu'à présent l'évolution des disparités géographiques de mortalité n'a pas fait l'objet d'étude poussée. Les marqueurs explicatifs de ces évolutions différenciées restent relativement inexplorés en France. Dans la littérature, l'association entre changements démographiques et mortalité a été étudiée au Royaume-Uni depuis la seconde moitié du XIX<sup>ème</sup> siècle. A cette époque, le flux migratoire de milliers de travailleurs venant des campagnes et se dirigeant vers les villes et bassins industriels était associé à une détérioration des conditions d'hygiène et de l'indicateur de mortalité. Depuis ce temps, l'association s'est inversée. Un accroissement de la population est généralement associé à une diminution de la mortalité. Plus spécifiquement, un taux d'accroissement démographique positif est associé à une réduction nette de la mortalité, alors qu'une faible diminution de celle-ci est observable pour les zones géographiques en déclin du Royaume-Uni et de l'Irlande [81]. L'association entre taux d'accroissement et mortalité est plus forte pour les hommes que les femmes en Grande-Bretagne [82,83], dans le comté de Värmland en Suède [84] et dans la

région de Madrid [85], et plus faible pour certaines causes spécifiques de décès comme les cancers, les maladies liées à la consommation d'alcool, les troubles mentaux, etc. [84,86]. Bien que l'on observe des tendances distinctes de la mortalité entre les zones riches et pauvres [31,87], l'association avec l'accroissement de la population reste significative même après ajustement sur le désavantage social [82,84,88]. L'amplitude de cette association dépend finalement de la période et du lieu étudié [86], mais aussi de l'échelle d'analyse et de la définition du désavantage social choisie [82,89,90].

La cartographie et les études écologiques constituent deux piliers de l'analyse des disparités géographiques de santé. La première partie de ce travail est consacrée à l'analyse descriptive des différentiels de mortalité par cause. A l'aide d'un modèle mixte de Poisson, ces disparités de mortalité par cause sont mesurées à plusieurs échelles et pour plusieurs périodes pour en observer l'évolution. La deuxième partie de ce chapitre interroge l'association écologique entre évolution de la mortalité par cause et taux d'accroissement de la population.

## **Méthodes et sources**

### **Cartographie**

Les RSM cartographiés ont été lissés à l'aide d'un modèle BYM [91] afin de faire ressortir les structurations spatiales de la mortalité. Ce modèle considère une dispersion de Poisson ainsi qu'une surdispersion se décomposant en une composante aléatoire gaussienne spatialement structurée et une composante aléatoire gaussienne non structurée. Les taux représentés sont calculés en sommant les deux composantes gaussiennes.

### **Mesure synthétique des disparités géographiques**

Le modèle de Poisson mixte suivant mesure les effets "unités géographiques" à l'aide des effets aléatoires gaussiens  $u_i$  :

$$\left\{ \begin{array}{l} O_i \sim \text{Poisson}(\lambda_i) \\ \log \lambda_i = \alpha + \log E_i + u_i \\ u_i \sim N(0, \sigma_{spat}^2) \end{array} \right.$$

où  $i \in I$ , l'ensemble des unités géographiques dont on souhaite mesurer la dispersion.

Ainsi, la variance  $\sigma_{spat}^2$  de la loi des effets aléatoires  $u_i$  constitue une mesure synthétique de la variabilité spatiale de la mortalité diminuée de sa composante poissonnienne.

### **Contribution des échelles aux disparités géographiques**

Le modèle de Poisson mixte suivant permet de mesurer simultanément les effets aléatoires à deux échelles emboîtées ou non. Nous considérons ici deux échelles emboîtées, soit le modèle suivant :

$$\begin{cases} O_i \sim \text{Poisson}(\lambda_i) \\ \log \lambda_i = \alpha + \log E_i + u_i + v_j \\ u_i \sim N(0, \sigma_{éch.1}^2); v_j \sim N(0, \sigma_{éch.2}^2) \end{cases}$$

où  $i \in I$ , l'ensemble des unités géographiques de l'échelle la plus fine (ex: commune) et  $j \in J$ , l'ensemble des unités géographiques de l'échelle supérieure (ex: région).

Ainsi, prenant pour exemple l'emboîtement d'échelles "région-commune", la contribution d'une échelle géographique aux disparités spatiales totales peut se mesurer en partitionnant la variance totale. La formule du VPC pour *Variance Partition Coefficient* est la suivante :

$$VPC_{commune} = \frac{\sigma_{commune}^2}{\sigma_{totale}^2}$$

$$VPC_{région} = \frac{\sigma_{région}^2}{\sigma_{totale}^2}$$

$$\text{où } \sigma_{totale}^2 = \sigma_{commune}^2 + \sigma_{région}^2$$

Les disparités spatiales expliquées par une échelle géographique sont exprimées en pourcentage de variance totale.

### Corrélations temporelles des disparités

Les changements survenus entre deux périodes dans la distribution spatiale de la mortalité sont mesurés à l'aide d'un modèle de Poisson mixte à effets aléatoires corrélés définie comme suit:

$$\begin{cases} O_{it} \sim \text{Poisson}(\lambda_{it}) \\ \log \lambda_{it} = \alpha + \log E_{it} + U_{it} \\ U_{it} \sim N(0, \Sigma) \text{ avec } \Sigma = \begin{pmatrix} \sigma_{spat\_t1}^2 & \rho \cdot \sigma_{spat\_t1} \cdot \sigma_{spat\_t2} \\ \rho \cdot \sigma_{spat\_t1} \cdot \sigma_{spat\_t2} & \sigma_{spat\_t2}^2 \end{pmatrix} \end{cases}$$

où  $t \in \{t1, t2\}$  les deux périodes à comparer,  $i \in I$ , l'ensemble des unités géographiques dont on souhaite mesurer les dispersions et  $\rho$  la corrélation entre les distributions spatiales des deux périodes.

### Taux d'accroissement annuel moyen de la population

Les effectifs de population aux années de recensement 1968, 1982, 1990, 1999 et 2006 sont fournis par l'INSEE. La variable utilisée dans ce travail jusqu'en 1999 inclus est la Population Sans Double Compte (PSDC).

En 2006, l'effectif de population est calculé d'après les EAR des années allant de 2004 à 2008.

A partir de ces données de population, les taux d'accroissements annuels moyens entre chaque période censitaire consécutive  $p_t$  et  $p_{t+1}$  ont été calculés (TAAM pour *Taux d'Accroissement Annuel Moyen*) aux échelles communales et cantonales selon la définition suivante:

$$TAAM^{(p_t, p_{t+1})} = \sqrt[a]{PSDC^{(p_{t+1})} / PSDC^{(p_t)}} - 1$$

où  $a$  est le nombre d'années écoulées entre deux périodes censitaires.

### Typologie des évolutions démographiques

Une catégorisation des profils d'évolutions démographiques des communes a été mise en œuvre afin de capter les différents processus migratoires qu'a connus la France d'après-guerre. La phase d'urbanisation et d'exode rural massif a été initiée avant les années 60 et a connu une intensification croissante jusqu'en 1990 où la déconcentration urbaine se généralise (périurbanisation). Afin de mieux appréhender ces changements, la typologie inclut également les évolutions démographiques intervenant en 1962 et 1975.

Les communes ont été classées dans chacune des catégories suivantes :

- DEC/INC : Population qui décline (TAAM < 0) de 1962 à 1990 puis croît (TAAM > 0) de 1990 à 2006.
- INC/DEC : Population qui croît entre 1962 et 1990, et décline de 1990 à 2006.
- DEC : Population déclinante (TAAM < 0) de 1962 à 1990 et également de 1990 à 2006.
- INC1 : Croissance inférieure à la moyenne nationale entre 1962 et 2006 (avec TAAM > 0 sur les deux sous-périodes 1962-1990 et 1990-2006).
- INC2 : Croissance supérieure à la moyenne nationale entre 1962 et 2006 (avec TAAM > 0 sur les deux sous-périodes 1962-1990 et 1990-2006).

### Données de mortalité et causes de décès

Les données de mortalité sont extraites de la base nationale de l'Inserm-CépiDc. Les décès ont été groupés par périodes quinquennales centrées sur les années de recensement (1975, 1982, 1990, 1999 and 2006) pour éviter les biais numérateur-dénominateur. Les RSM toutes causes ont été calculés séparément tous sexes, hommes, femmes et pour les moins de 65 ans (mortalité prématurée). Pour analyser les changements dans les niveaux de mortalité indépendamment du déclin général de la mortalité, le taux de mortalité national de référence utilisé pour calculer le RSM était spécifique à chaque période. Le nombre attendu de décès dans le canton  $i$  pour l'année  $p_t$  était défini par  $E_i^{p_t}$ . L'intervalle de confiance à 95% a été estimé à l'aide de la méthode exacte de Poisson.

Les causes de décès qui ont été investiguées dans les associations écologiques sont les cardiopathies ischémiques, les maladies cérébrovasculaires, les tumeurs, les tumeurs des voies aéro-digestives supérieures (VADS), du poumon, colorectale, de la vessie, du sein et de la prostate, les maladies respiratoires, les morts violentes et les maladies liées à la consommation d'alcool (foie, psychoses). Les codes CIM de ces causes sont disponibles en annexe (Tableau S2).



### Modèle statistique

Cette association écologique est mesurée à l'aide d'un modèle de Poisson surdispersé de type GEE [92] afin de prendre en compte la corrélation temporelle entre deux mesures faites sur une même unité géographique à deux périodes distinctes.

Le modèle statistique est défini comme suit :

$$\begin{cases} O_i \sim \text{Poisson}(\lambda_i) \\ \log \lambda_i^{(u)} = \alpha_0 + \log E_i^{(u)} + \beta_0 \cdot TAAM^{(p_t, p_{t+1})} \\ \quad + \beta_1 \cdot I_{p_{t+1}}(u) \cdot TAAM^{(p_t, p_{t+1})} \end{cases}$$

où  $u \in \{p_t, p_{t+1}\}$  les deux périodes censitaires entre lesquelles le changement est mesuré et  $I_{p_{t+1}}(u)$  la fonction indicatrice de la seconde période  $p_{t+1}$ .

Les paramètres  $\alpha_0$  et  $\beta_0$  ajustent la mortalité moyenne des deux périodes sur le taux d'accroissement annuel moyen, tandis que le paramètre  $\beta_1$  capte la variabilité du différentiel de mortalité entre les deux périodes expliquée par ce taux d'accroissement.

La mesure d'association  $\beta_1$  s'interprète comme un ratio de taux entre les deux périodes:

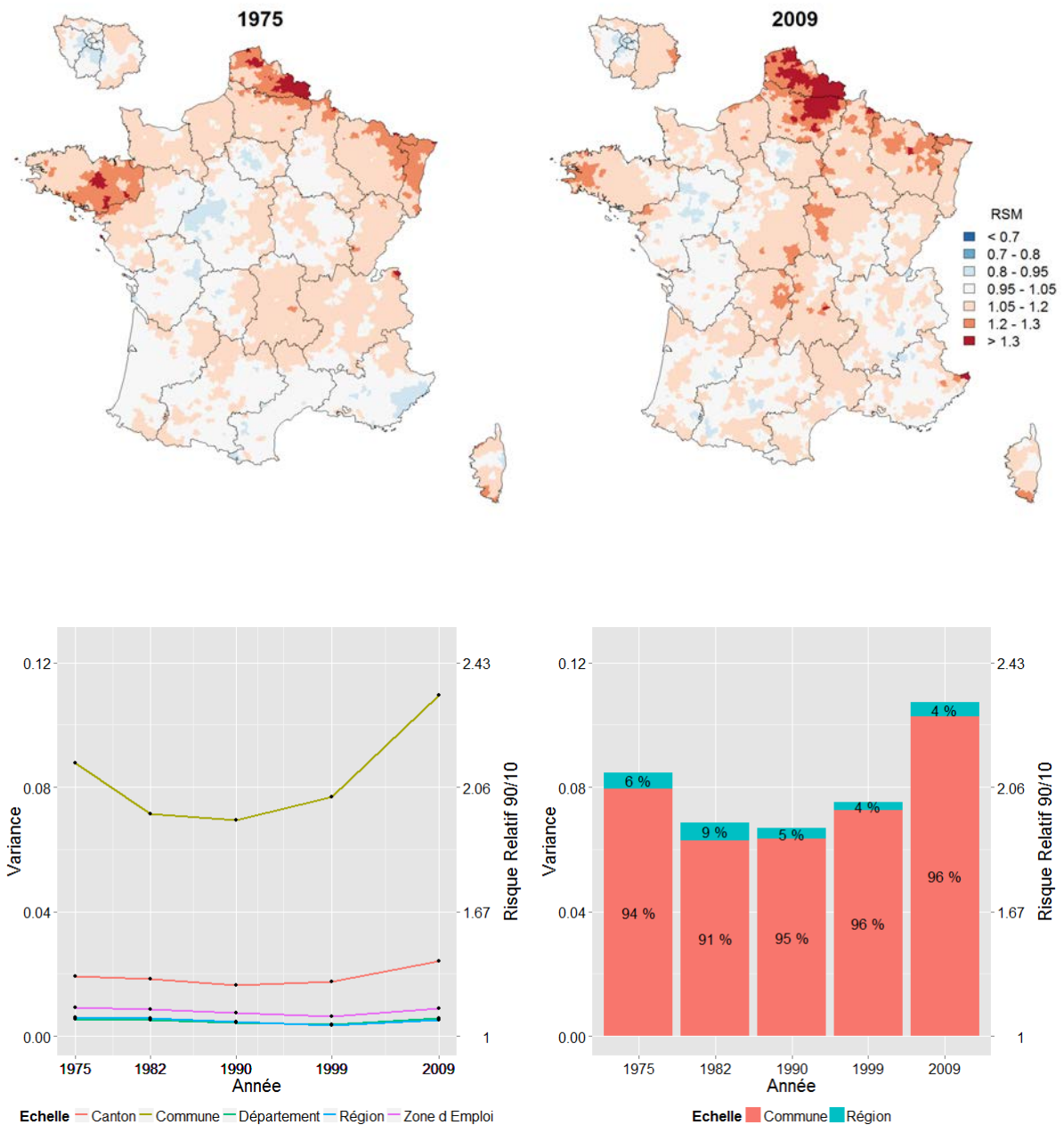
$$\frac{\frac{\lambda_i^{(p_t)}}{E_i^{(p_t)}}}{\frac{\lambda_i^{(p_{t+1})}}{E_i^{(p_{t+1})}}} = \exp(\beta_1 \cdot CAPGR^{(p_t, p_{t+1})}) \simeq 1 + \beta_1 \cdot CAPGR^{(p_t, p_{t+1})}$$

Ainsi, pour un taux d'accroissement annuel moyen positif de 1% de la population entre les périodes  $p_t$  et  $p_{t+1}$ , le RSM de l'unité géographique  $i$  varie de  $\beta_1\%$ . Ce dernier résultat est déduit de l'approximation linéaire de la fonction exponentielle au voisinage de 0.

## Résultats

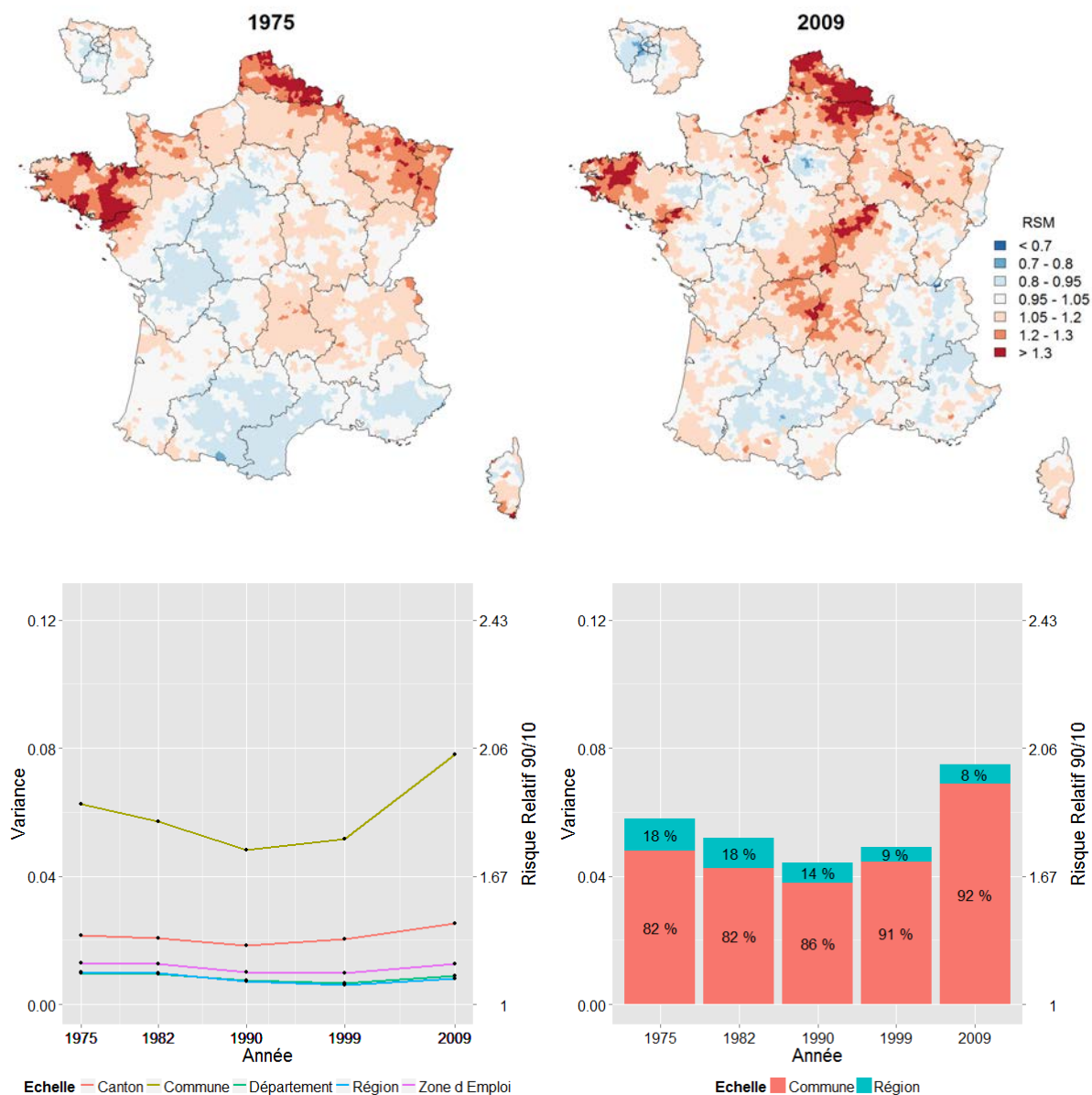
### Disparités géographiques de mortalité

Figure I.1: Disparités géographiques de mortalité toutes causes des femmes: cartographie, mesure synthétique et décomposition par échelle.



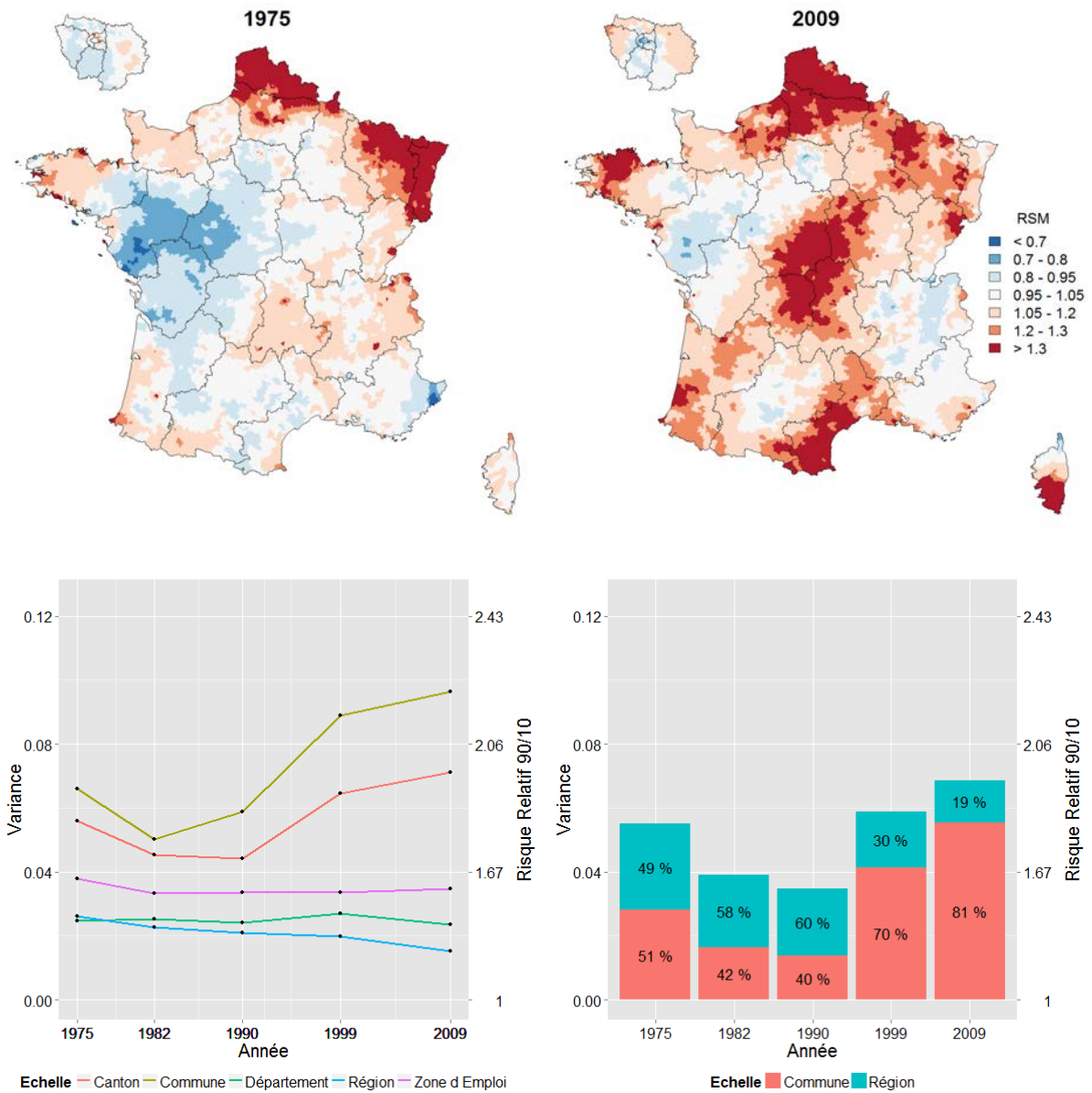
Les RSM représentés sur la carte à l'échelle cantonale ont été lissés (haut). Le graphique en bas à gauche représente la variance des effets aléatoires estimé à l'aide d'un modèle mixte de Poisson. Sur le graphique en bas à droite est représentée la somme des variances aux échelles communale et régionale d'un modèle mixte de Poisson à deux niveaux. La proportion affichée mesure la contribution de chaque échelle à la variance totale (VPC aux échelles communale et régionale).

Figure I.2: Disparités géographiques de mortalité toutes causes des hommes: cartographie, mesure synthétique et décomposition par échelle



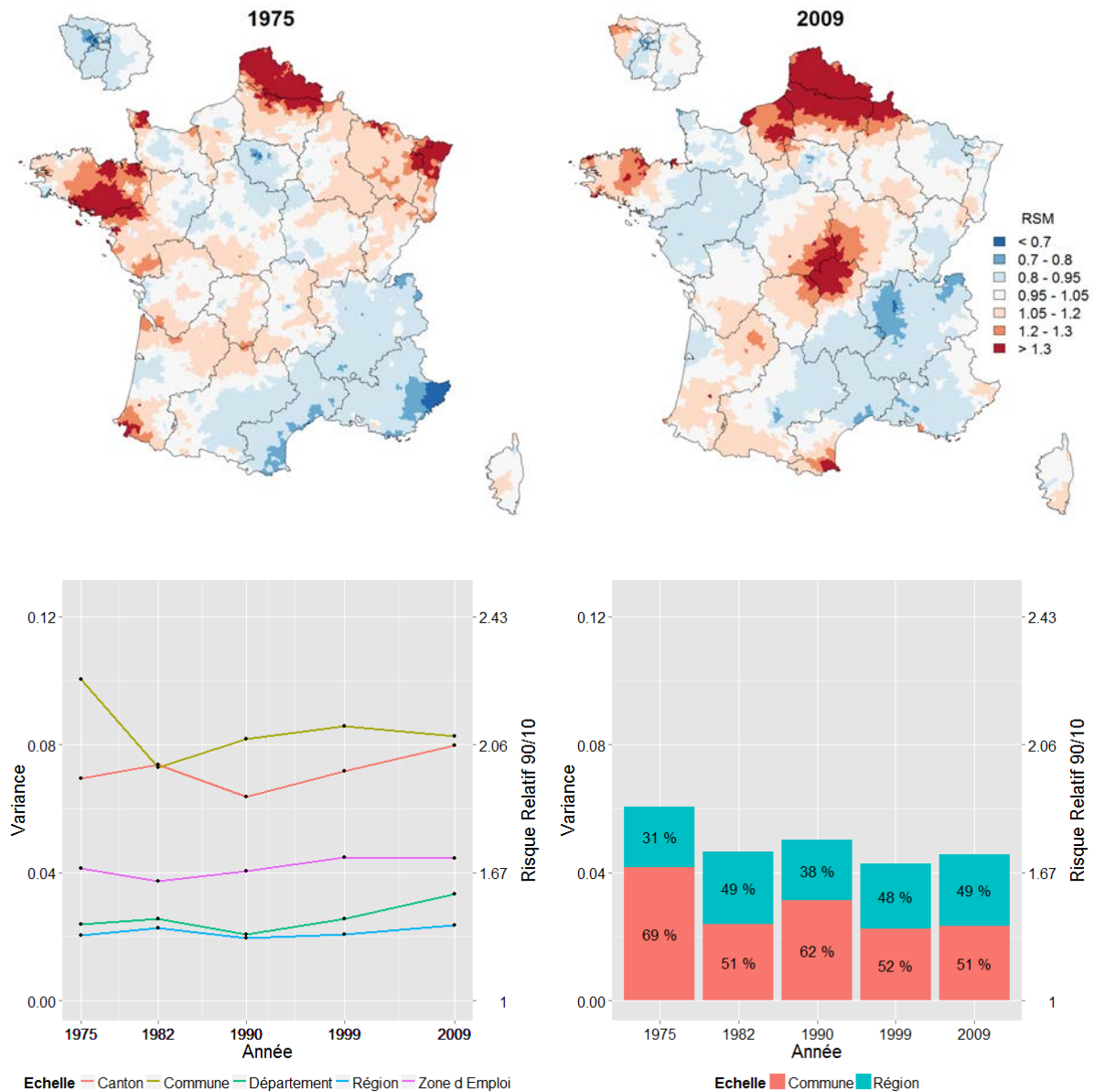
Les RSM représenté sur la cartographie à l'échelle cantonale ont été lissés (haut). Sur le graphique en bas à gauche représenté la variance des effets aléatoires estimé à l'aide d'un modèle mixte de Poisson. Sur le graphique en bas à droite sont représentés la somme des variances aux échelles communales et régionales d'un modèle mixte de Poisson à deux niveaux. La proportion affichée mesure la contribution de chaque échelle à la variance totale (VPC à l'échelle communale et régionale).

Figure I.3: Disparités géographiques de mortalité prématurée des hommes par cardiopathies ischémiques: cartographie, mesure synthétique et décomposition par échelle



Les RSM représenté sur la cartographie à l'échelle cantonale ont été lissés (haut). Sur le graphique en bas à gauche représenté la variance des effets aléatoires estimé à l'aide d'un modèle mixte de Poisson. Sur le graphique en bas à droite sont représentés la somme des variances aux échelles communale et régionale d'un modèle mixte de Poisson à deux niveaux. La proportion affichée mesure la contribution de chaque échelle à la variance totale (VPC aux échelles communale et régionale).

Figure I.4: Disparités géographiques de mortalité prématurée des femmes par maladies cérébrovasculaires: cartographie, mesure synthétique et décomposition par échelle



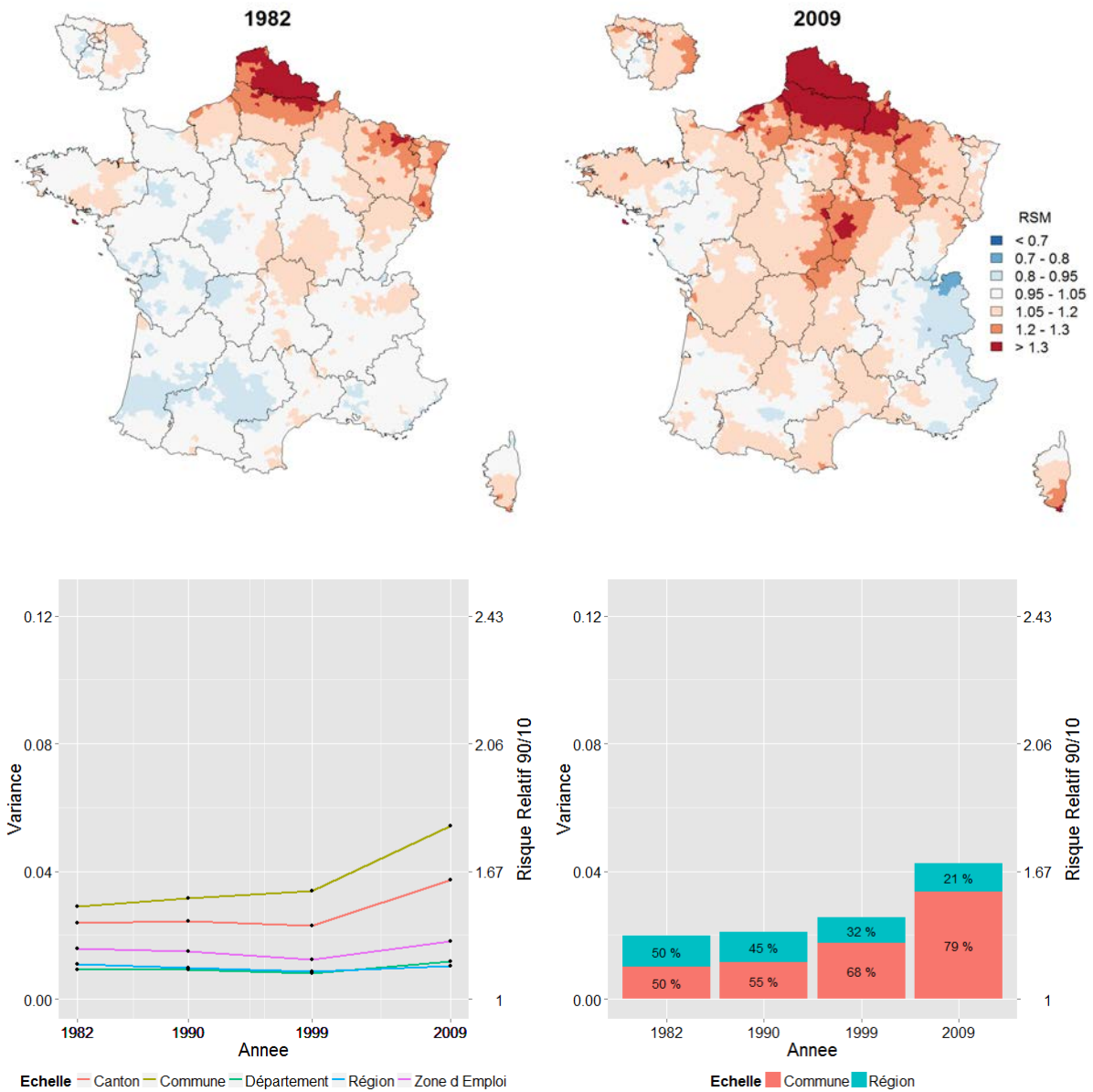
Les RSM représentés sur la cartographie à l'échelle cantonale ont été lissés (haut). Sur le graphique en bas à gauche représenté la variance des effets aléatoires estimé à l'aide d'un modèle mixte de Poisson. Sur le graphique en bas à droite sont représentés la somme des variances aux échelles communale et régionale d'un modèle mixte de Poisson à deux niveaux. La proportion affichée mesure la contribution de chaque échelle à la variance totale (VPC aux échelles communale et régionale).

De manière générale, les mesures de variances spatiales indiquent une légère baisse des disparités spatiales entre 1975 et 1990 aux échelles communales et cantonales, puis une augmentation de celles-ci à partir de 1990 exceptée pour la mortalité prématurée des femmes par maladies cérébrovasculaires qui reste stable à partir de cette période (Figure I.4). La mortalité prématurée évitable des femmes, imputable au système de soins augmente à partir de 1990 (Figure I. 5) tandis que la mortalité prématurée des hommes évitable, imputable à la

prévention primaire augmente très fortement à partir de 1982. La variance spatiale est toujours plus élevée à l'échelle la plus fine, l'échelle communale, et diminue à mesure que l'échelle spatiale augmente suivant l'ordre : commune, canton, Zone d'Emploi, département et région. Concernant la mortalité toutes causes hommes et femmes, la chute de la variance spatiale consécutive au passage de l'échelle communale à l'échelle cantonale est la plus élevée avec une perte d'information allant de 70% à 80%. La variance intercommunale de la mortalité toutes causes est plus grande pour les femmes que les hommes. Toutefois, cet écart disparaît aux échelles plus larges (Figure I.1 et I.2). Dans les modèles à 2 niveaux (région et commune), les disparités spatiales de mortalité expliquées par l'échelle régionale sont plus fortes pour les hommes que pour les femmes, bien que cette échelle explique une proportion assez limitée avec une tendance à la baisse allant de 18% à 8% pour les hommes (Figure I.2). Concernant les disparités spatiales de mortalité prématurée des hommes par cardiopathies ischémiques, la proportion des disparités totales qui s'articulaient entre régions était de 49% en 1975, augmentait jusqu'à 60% en 1990 pour retomber à 19% en 2009. Le pourcentage de variance totale de la mortalité prématurée des femmes par maladies cérébrovasculaires expliqué par l'échelle régionale était de 31% en 1975 et augmentait jusqu'à 49% en 2009. Cette proportion de variance expliquée par l'échelle régionale a diminué pour la mortalité prématurée évitable des hommes et des femmes. Pour les femmes, la mortalité liée au système de soins expliquée par l'échelle régionale passe de 50% en 1982 à 21% en 2009 (Figure I.5). Pour les hommes, la contribution de l'échelle régionale aux disparités spatiales de la mortalité liée à la prévention primaire passe de 35% à 18% (Figure I.6).

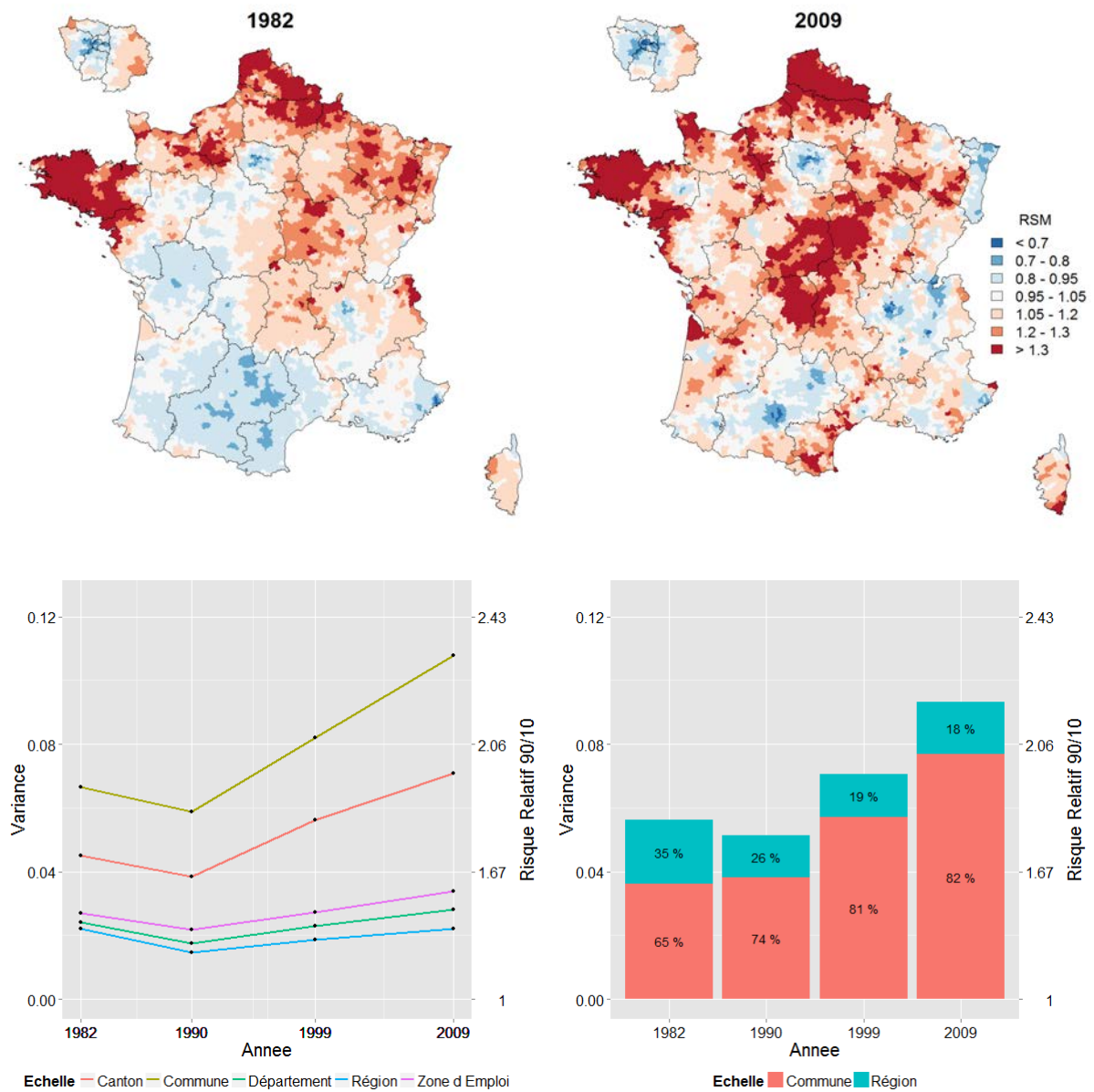


Figure I.5: Disparités Géographiques de Mortalité Evitable Prématuurée des Femmes imputable aux systèmes de soins : Cartographie, Mesure Synthétique et Décomposition par Echelle



Les RSM représentés sur la cartographie à l'échelle cantonale ont été lissés (haut). Sur le graphique en bas à gauche représenté la variance des effets aléatoires estimé à l'aide d'un modèle mixte de Poisson. Sur le graphique en bas à droite sont représentés la somme des variances aux échelles communale et régionale d'un modèle mixte de Poisson à deux niveaux. La proportion affichée mesure la contribution de chaque échelle à la variance totale (VPC aux échelles communale et régionale).

Figure I.6: Disparités Géographiques de Mortalité Evitable Prématuurée des Hommes imputable à l'absence de prévention primaire: Cartographie, Mesure Synthétique et Décomposition par Echelle

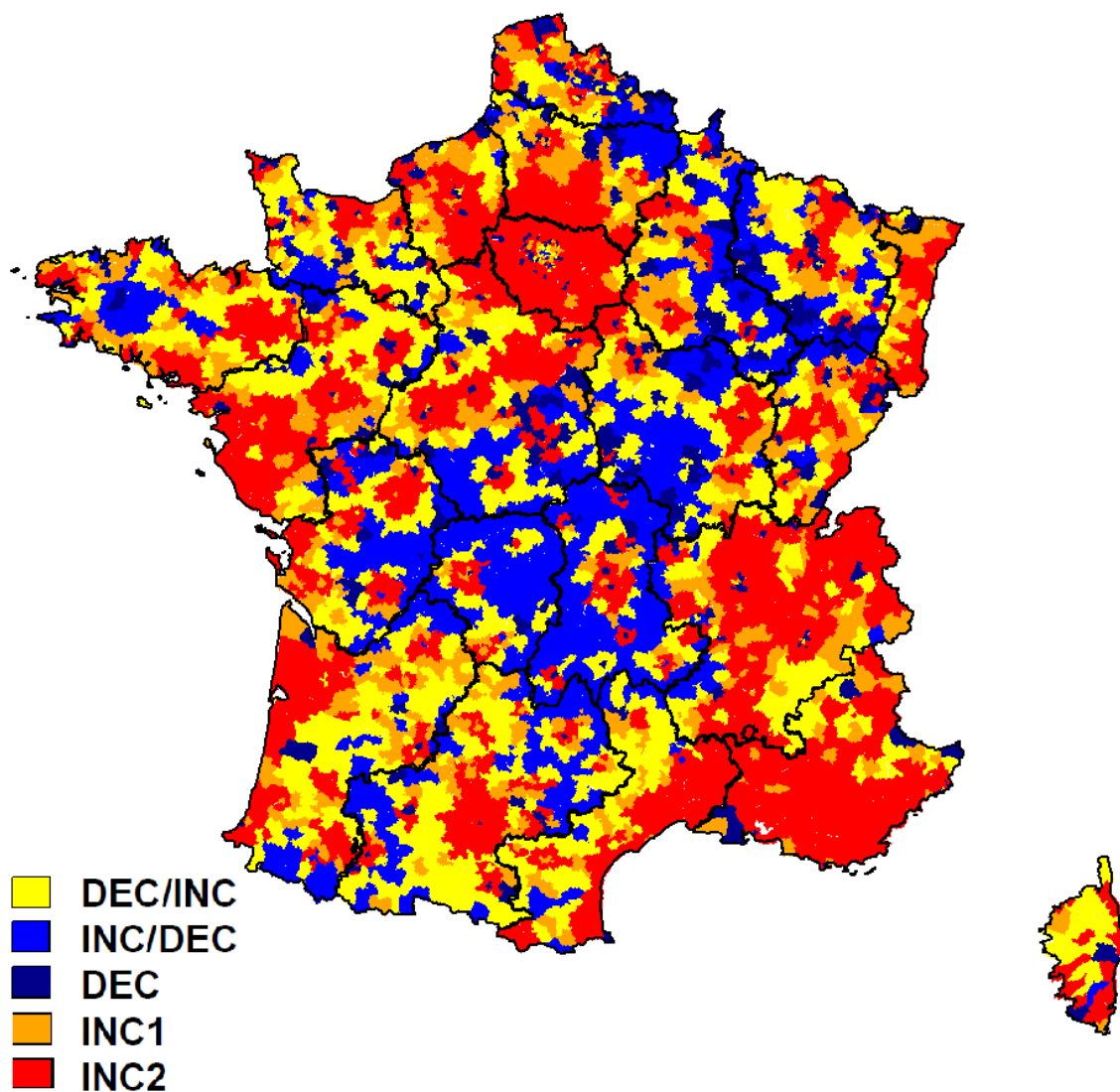


Les RSM représentés sur la cartographie à l'échelle cantonale ont été lissés (haut). Sur le graphique en bas à gauche est représentée la variance des effets aléatoires, estimée à l'aide d'un modèle mixte de Poisson. Sur le graphique en bas à droite est représentée la somme des variances aux échelles communales et régionales d'un modèle mixte de Poisson à deux niveaux. La proportion affichée mesure la contribution de chaque échelle à la variance totale (VPC à l'échelle communale et régionale).



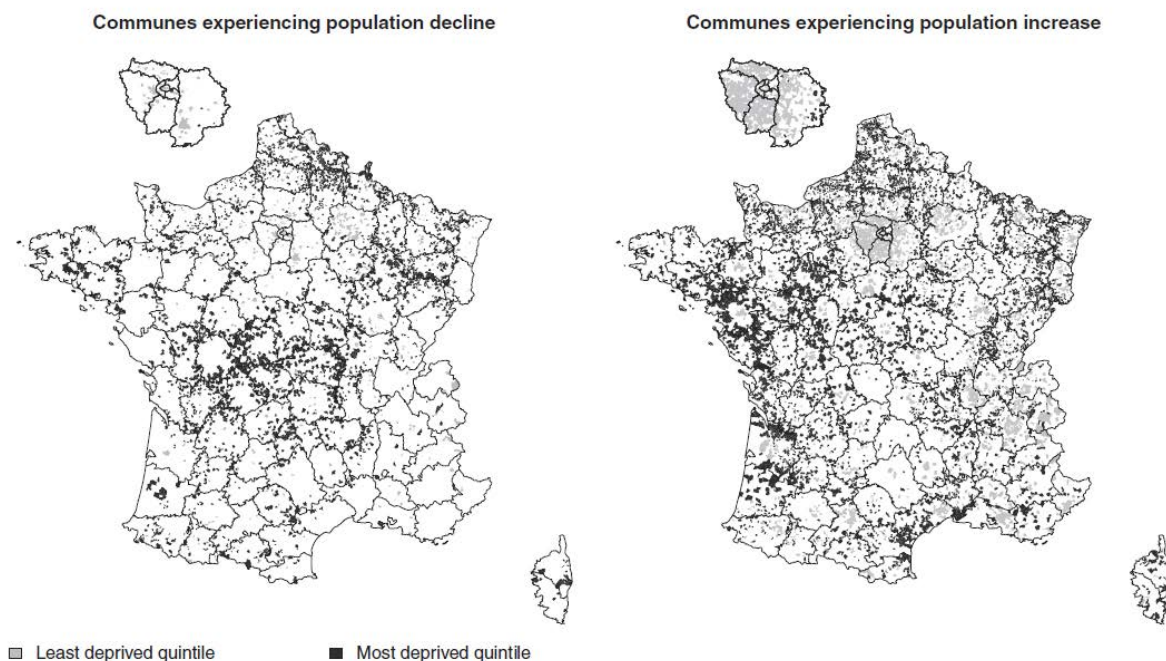
## Association écologique entre changements démographiques et évolution de la mortalité

Figure I.7: Cartographie de la typologie des accroissements démographiques de la population entre 1962 et 2006



- DEC/INC: Population qui décline ( $TAAM < 0$ ) de 1962 à 1990 puis croît ( $TAAM > 0$ ) de 1990 à 2006.
- INC/DEC: Population qui croît entre 1962 et 1990, et décline de 1990 à 2006.
- DEC: Population déclinante ( $TAAM < 0$ ) de 1962 à 1990 et également de 1990 à 2006.
- INC1: Croissance faible entre 1962 et 1990 ( $TAAM > 0$  mais inférieure à la moyenne nationale) et également entre 1990 et 2006.
- INC2: Croissance nette entre 1962 et 1990 ( $TAAM > 0$  supérieure à la moyenne nationale) et également entre 1990 et 2006

Figure I.8: Carte synthétique de la typologie des accroissements démographiques dans les quintiles extrêmes de désavantage social (FDep90).

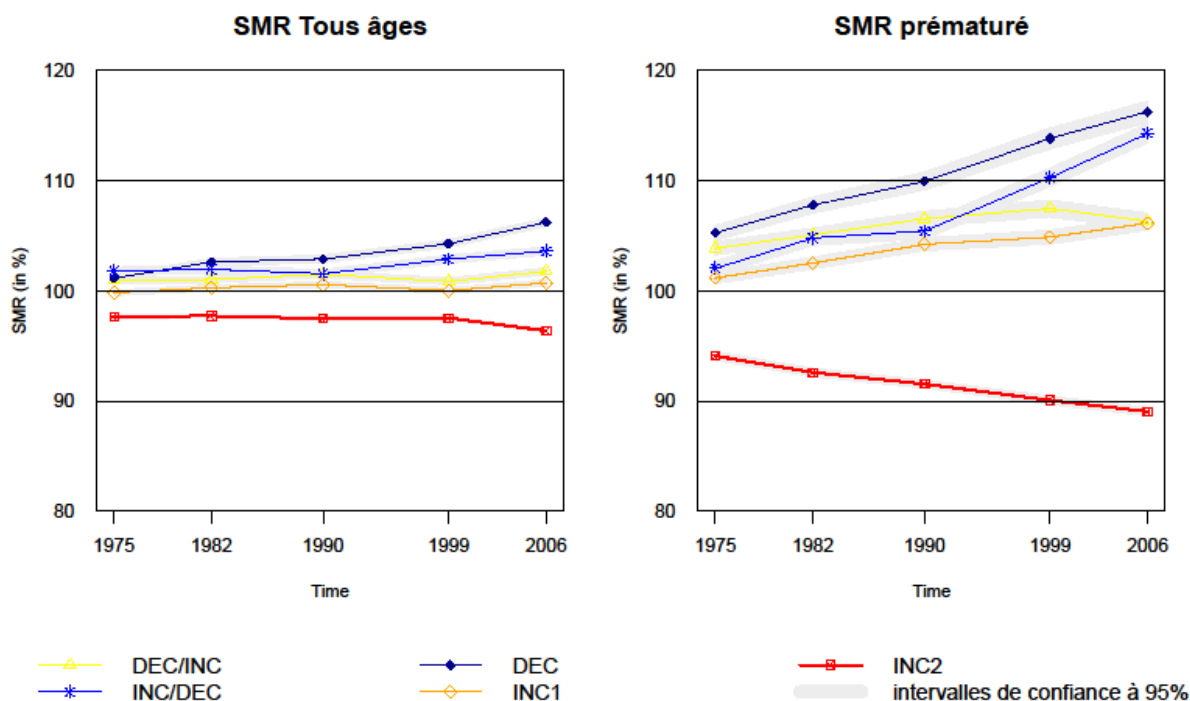


*Les communes dont la population a augmenté durant toute la période d'analyse ou seulement après 1990 (INC1, INC2, DEC/INC) sur la droite. Les communes dont la population a décliné durant toute la période d'analyse ou seulement après 1990 (DEC, INC/DEC) sur la gauche. Le zoom situé en haut à gauche de chaque carte est la région Ile-de-France avec Paris en son centre.*

Les communes les moins socialement désavantagées présentant un déclin de la population étaient principalement les arrondissements parisiens et les banlieues de certaines villes moyennes et grands centres urbains dans lesquels, historiquement, les habitants les plus riches ont eu tendance à se rassembler. Les communes les moins socialement désavantagées présentant une croissance de la population étaient des arrondissements parisiens et certaines communes de la banlieue ouest de Paris. Plusieurs centres villes dont la population dépasse 100 000 habitants font également partie de cette catégorie. Certaines banlieues des principales zones urbaines de l'Est, du Sud-Ouest, ainsi que celles du réseau de villes qui s'étend de Paris aux zones urbaines de la côte atlantique y figurent également (Figure I.8).

Les communes les plus socialement désavantagées dont la population a augmenté étaient principalement situées dans l'arrière-pays du littoral atlantique et méditerranéen. Ces communes sont des zones rurales ou péri-urbaines qui ont bénéficié de l'activité économique des zones urbaines des littoraux les plus proches. Les communes des anciens bassins industriels situés aux frontières de la métropole ont également connu une augmentation de leur population. Les communes situées sur les frontières allemande, belge et suisse ont exercé une certaine attractivité sur les habitants des communes voisines (Figure I.8).

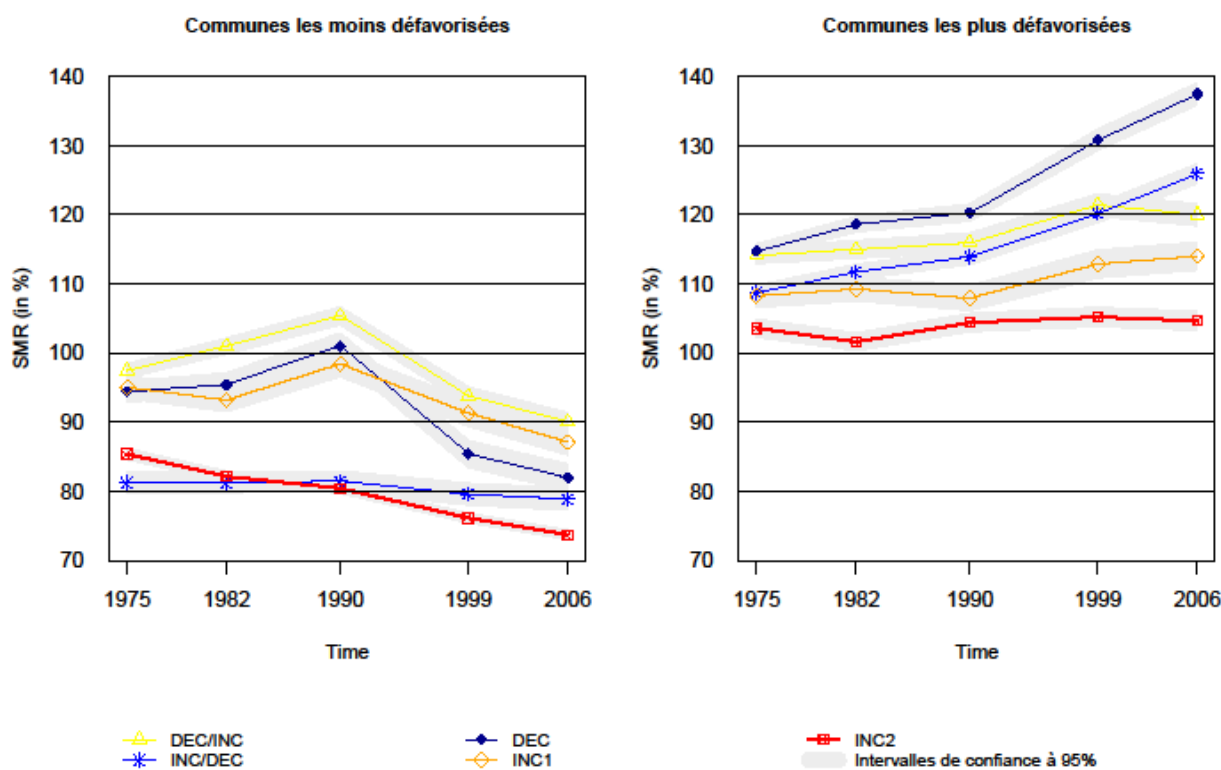
Figure I.9: Evolutions des RSM tous âges et prématurés (moins de 65 ans) entre 1975 et 2006 dans les catégories de la typologie des évolutions démographiques des communes.



- DEC/INC: Population qui décline ( $TAAM < 0$ ) de 1962 à 1990 puis croît ( $TAAM > 0$ ) de 1990 à 2006.
- INC/DEC: Population qui croît entre 1962 et 1990, et décline de 1990 à 2006.
- DEC: Population déclinante ( $TAAM < 0$ ) de 1962 à 1990 et également de 1990 à 2006.
- INC1: Croissance faible entre 1962 et 1990 ( $TAAM > 0$  mais inférieure à la moyenne nationale) et également entre 1990 et 2006.
- INC2: Croissance nette entre 1962 et 1990 ( $TAAM > 0$  supérieure à la moyenne nationale) et également entre 1990 et 2006.

De manière générale, le RSM a augmenté lorsque la population de la commune a décliné et la mortalité relative a baissé lorsque la population de la commune s'est fortement accrue. Le gradient des évolutions différenciées était borné par la catégorie DEC observant la plus grande augmentation du RSM, et la catégorie INC2 avec une mortalité plus faible ainsi qu'une tendance à la baisse. L'amplitude des écarts de mortalité ainsi que des tendances était plus importante pour la mortalité prématurée que pour la mortalité tous âges. Le reste du travail se concentre sur la mortalité prématurée. Bien que les écarts de mortalité fussent déjà importants en 1975, la mortalité de la catégorie DEC a augmenté de plus de 10%, là où la mortalité de la catégorie INC2 a diminué dans des proportions similaires. Le RSM a augmenté sur toute la période dans les catégories INC1 et INC/DEC. La catégorie DEC/INC a observé une légère réduction du RSM durant la dernière période après une augmentation entre 1975 et 1999 (Figure I.9).

Figure I.10: Evolutions des RSM prématurés entre 1975 et 2006 dans les catégories de la typologie des évolutions démographiques des communes les moins et les plus socialement désavantagées.

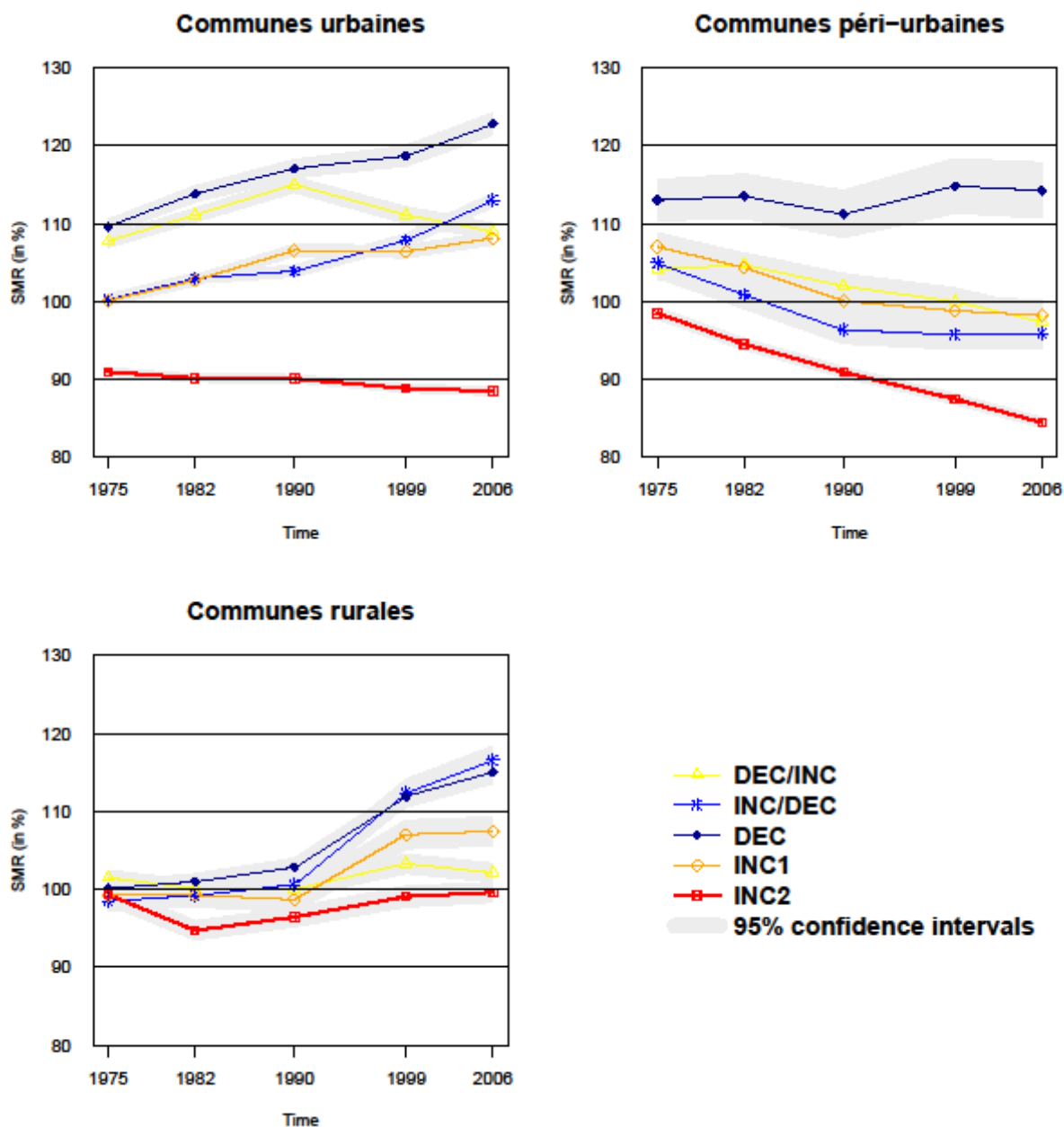


- DEC/INC: Population qui décline ( $TAAM < 0$ ) de 1962 à 1990 puis croît ( $TAAM > 0$ ) de 1990 à 2006.
- INC/DEC: Population qui croît entre 1962 et 1990, et décline de 1990 à 2006.
- DEC: Population déclinante ( $TAAM < 0$ ) de 1962 à 1990 et également de 1990 à 2006.
- INC1: Croissance faible entre 1962 et 1990 ( $TAAM > 0$  mais inférieure à la moyenne nationale) et également entre 1990 et 2006.
- INC2: Croissance nette entre 1962 et 1990 ( $TAAM > 0$  supérieure à la moyenne nationale) et également entre 1990 et 2006.

Les communes les moins socialement désavantagées avaient généralement une mortalité inférieure à celles des communes les plus démunies. Quelle que soit la période et le quintile de désavantage, des écarts de mortalité marqués ont été observés selon le type d'accroissement démographique. Cependant, ces différences étaient moins marquées pour les quintiles les moins défavorisés que pour les quintiles les plus désavantagés. Dans le quintile le moins désavantagé, les différences entre les catégories DEC et INC2 étaient de moins de 20% sur toute la période. Cet écart a par ailleurs augmenté dans le quintile le plus défavorisé : en 2006, l'écart entre les catégories DEC et INC2 a dépassé 30%. Le RSM a diminué dans chaque catégorie du quintile moins désavantagé, mais cette diminution a été plus forte pour la catégorie des communes qui ont décliné avec une baisse entre 15% et 20%. Dans le quintile le plus défavorisé, l'amplitude de l'augmentation du RSM était modulée par la typologie. Le RSM des communes INC2, dont la population s'est fortement accrue, ont observé une légère augmentation sur toute la période, tandis que le RSM de la catégorie DEC a fortement

augmenté après 1990. Dans l'ensemble, les disparités croissantes dans les communes les plus démunies étaient très liées à l'amplitude des changements de population (Figure I.10).

Figure I.11: Evolutions des RSM prématurés entre 1975 et 2006 dans les catégories de la typologie des évolutions démographiques par degré d'urbanisation des communes



- DEC/INC: Population qui décline (TAAM < 0) de 1962 à 1990 puis croît (TAAM > 0) de 1990 à 2006.
- INC/DEC: Population qui croît entre 1962 et 1990, et décline de 1990 à 2006.
- DEC: Population déclinante (TAAM < 0) de 1962 à 1990 et également de 1990 à 2006.
- INC1: Croissance faible entre 1962 et 1990 (TAAM > 0 mais inférieure à la moyenne nationale) et également entre 1990 et 2006.
- INC2: Croissance nette entre 1962 et 1990 (TAAM > 0 supérieure à la moyenne nationale) et également entre 1990 et 2006.

Les changements dans les RSM prématurés étaient très différenciés selon le profil d'évolution démographique et le degré d'urbanisation. Pour les pôles urbains et les banlieues proches, les changements dans le RSM pour la mortalité prématurée ont été associés à la dynamique de la population : le RSM a légèrement diminué pour la catégorie INC2. Cependant, le RSM a augmenté pour toutes les autres catégories. Le RSM a diminué pour toutes les catégories des communes périurbaines, sauf pour la catégorie déclinante DEC, qui a montré une relative stabilité. Dans l'ensemble, dans les zones rurales, la mortalité par rapport au niveau national a augmenté tout au long de la période surtout après 1990, exceptée pour la catégorie INC2. L'augmentation du RSM était plus grande pour les communes DEC et INC/DEC (Figure I.11).

Tableau I.1: Association entre évolution de la mortalité prématurée et taux d'accroissement annuel moyen par période et cause de décès

Période	Tous sexes				Hommes				Femmes			
	1975	1982	1990	1999	1975	1982	1990	1999	1975	1982	1990	1999
	1982	1990	1999	2006	1982	1990	1999	2006	1982	1990	1999	2006
<b>Causes de décès</b>												
Toutes causes	-1.31**	-1.16**	-2.1**	-1.63**	-1.36**	-1.27**	-2.15**	-1.82**	-1.15**	-0.87*	-2.11**	-1.26**
Cardiopathies Ischémiques	-0.44	-1.88*	-0.4	-2.65*	0.01	-1.76*	0.15	-2.25*	-2.89*	-2.61	-4.32	-4.90*
Maladies cérébrovasculaires	-2.03**	-1.28	-2.47	-0.95	-1.95*	-2.26*	-1.84	-1.37	-2.18*	0.59	-3.83	-0.12
Tumeurs	-0.75*	-1.08**	-1.12*	-0.35	-0.87*	-1.10*	-1.35*	-0.19	-0.4	-0.94*	-1.03	-0.79
V.A.D.S	-3.15**	-2.36*	-1.61	1.59	-2.86**	-2.5*	-2.06	1.71	-7.67*	-1.45	0.95	-0.15
Poumons	0.07	-1.08*	-1.18	0.09	0.16	-1.10*	-1.62	0.07	-0.76	-1.6	-0.84	-1.58
Colorectale	0.17	-1.01	-1.72	-1.61	0.45	-1.71	-2.61	-2.2	-0.23	0.04	-0.42	-0.71
Vessie	-0.92	1.02	-4.67	2.96	-0.81	0.52	-2.39	3.45	-1.03	4.51	-19.95*	0.68
Seins									-0.78	-1.21	0.25	-0.28
Prostate					-1.84	0.06	-2.94	-1.23				
Maladies respiratoires	0.4	-1.07	-2.28	-1.26	0.61	-1.04	-2.11	-1.25	-0.08	-1.45	-3.09	-1.35
Morts violentes	-2.01**	-1.43**	-3.62**	-2.16**	-2.24**	-1.15*	-3.92**	-2.13**	-1.29*	-2.28**	-2.72*	-2.11*
Abus d'alcool et maladies chroniques du foie	-2.99**	-3.60**	-3.89**	-2.78*	-3.03**	-3.93**	-4.82**	-2.31*	-2.95**	-2.78*	-1.5	-4.24*

\* significatif à 0.05; \*\* significatif à 0.01

L'association est mesurée par  $\beta_1$ , c'est à dire l'évolution du RSM (en %) pour un taux d'accroissement annuel moyen de 1% (où l'évolution opposée pour -1%). Toutes ces estimations doivent être interprétées comme une variation du RSM (en %) durant une période intercensitaire. Exemple : entre 1975 et 1982, pour un taux d'accroissement annuel moyen de 1% de la population, le RSM a chuté de -1.31% (1982 par rapport à 1975).

De 1975 à 2006, le TAAM national était de 0,5%, avec un minimum de -2,1% et un maximum de 10,7% (résultat non fourni). L'association était globalement négative et significative avec la mortalité toutes causes, par morts violentes et maladies liées à l'abus d'alcool pour les deux sexes, et en particulier pour les hommes (Tableau I.1). L'association

était nettement plus forte pour l'abus d'alcool et les maladies chroniques du foie. Bien que l'association soit négative pour plusieurs causes (tous les cancers, les cancers des voies aéro-digestives supérieures, les maladies cérébrovasculaires), elle n'est pas significative pour certaines périodes. L'association était plus forte durant la période 1990-1999. L'association est non significative ou faiblement significative ( $0,001 < p < 0,05$ ) pour la mortalité par cardiopathies ischémiques, cancer du poumon, cancer colorectal, cancer de la vessie, cancer du sein et cancer de la prostate.

## **Discussion**

Les disparités toutes causes, estimées à l'aide de la mesure synthétique proposée, sont plus fortes pour les femmes que les hommes à l'échelle communale. Le passage à l'échelle cantonale produit une large perte d'information à expliquer pour la mortalité toutes causes. Cette chute de la variance spatiale mesurée est toutefois beaucoup moins importante si on considère la mortalité prématurée des femmes par maladies cérébrovasculaires et la mortalité prématurée des hommes par cardiopathies ischémiques. De manière générale la variance géographique augmente sensiblement à partir de 1990. Très marquée par le fait régional en 1975, les disparités géographiques de mortalité prématurée par cardiopathies et pour la mortalité évitable ne sont plus le fait de différences inter-régionales en 2009. Le fait inverse est observé pour la mortalité prématurée des femmes par maladies cérébrovasculaires où la contribution de disparités régionales à la variance totale augmente au cours du temps pour atteindre une proportion de 50% en 2009. Ces résultats sont à mettre en perspective avec le processus d'urbanisation et de métropolisation amorcé entre 1970 et 1980. Ce dernier a eu pour effet de concentrer les plus riches dans les grandes métropoles en laissant à la marge les populations les plus désavantagées le tout sur fond d'un maintien non négligeable des disparités régionales [93].

Les résultats de la deuxième partie suggèrent l'existence d'une association entre changements démographiques et évolution de la mortalité relative. Bien que la mortalité toutes causes ait fortement diminué entre 1975 et 2006, les inégalités relatives entre les catégories sociodémographiques des communes se sont accentuées, excepté dans les zones les moins désavantagées. Les RSM des communes en déclin démographique se sont détériorés tandis que la mortalité de celles qui connaissaient un accroissement positif de la population s'améliorait plus vite que les autres. L'association était modulée par le degré d'urbanisation et le niveau de désavantage social. Elle était aussi plus forte pour la mortalité prématurée que pour la mortalité tous âges et plus prononcée chez les hommes que chez les femmes. Il en est de même pour les causes de décès liées aux comportements à risques tels que les morts violentes et les maladies liées à la consommation d'alcool. L'association la plus forte était obtenue sans décalage dans le temps.

Les zones socialement désavantagées en déclin démographique ont connu une augmentation rapide de leur mortalité relative, ce qui confirme certains résultats observés dans la littérature [29,82]. Ces zones sont constituées de communes rurales situées à la marge des principaux centres urbains, souvent localisées le long des frontières départementales (Figure I.8). Elles se



trouvent probablement dans une spirale négative sous l'effet des migrations orientées vers les zones urbaines: perte d'attractivité, activité économique réduite, dévitalisation du centre-ville et fermeture progressive des services fournis aux populations les plus désavantagées [77]. Les communes urbaines concernées accumulent problèmes économiques et sociaux. L'arc industriel du Nord-Est et les zones densément peuplées des grandes métropoles en sont des exemples explicites. A l'inverse de ces communes, les moins désavantagées observant un déclin de population ont un indicateur de mortalité qui progresse plus vite que le niveau national. A partir de 1990, ces communes constituées d'arrondissements parisiens ont vécu un changement de population socialement différencié. Pour Paris comme pour certains centres urbains, cette période coïncide avec le lancement de projets de modernisation et de réhabilitation, ainsi que l'amorçage d'un processus de gentrification [94].

Les zones socialement désavantagées dont la population s'est accrue ont une mortalité qui est restée proche du niveau national. Cette catégorie abrite cependant des nouvelles zones résidentielles subissant l'influence d'un pôle urbain ou en renouveau touristique (Figure II.8). Dans certains cas, il est possible qu'elles aient bénéficié d'un flux entrant de personnes plus fortunées ou d'un effet protecteur influençant la mortalité [95,96]. Depuis les années 80, les espaces péri-urbains se caractérisent par l'arrivée d'employés de classes moyennes travaillant dans le centre urbain le plus proche. Selon le même mécanisme, cette population était peut-être en meilleure santé que la population déjà installée, ce qui a possiblement amélioré l'indicateur de mortalité [31]. Cependant, l'inflation des prix de l'immobilier a poussé loin des centres les populations les plus vulnérables. Certaines zones péri-urbaines de la région parisienne se caractérisent par la présence de populations situées en bas de l'échelle sociale [97].

L'analyse par âge et sexe suggère une relation plus forte pour les hommes, ce qui est en accord avec les études précédentes [82,84,86,98]. Bien qu'elles ne se soient pas toutes avérées significatives, les associations par cause spécifique se sont quasiment toutes montrées négatives. Ces associations étaient plus fortes pour les causes de décès généralement très liées au niveau socio-économique comme l'alcool et les causes externes. Une étude écossaise a mis en évidence que les décès par ces causes contribuent à l'excès de mortalité dans les zones désavantagées en déclin démographique [82]. Les résultats de notre étude ont également montré que le RSM pour ces causes dans ces mêmes zones augmentait significativement. Si les migrants ont, en moyenne, un niveau socio-économique plus élevé que les sédentaires [99], ils ont également un risque réduit de mortalité par alcool et par causes externes. Par conséquent, le RSM peut baisser artificiellement pour ces raisons dans les zones quittées par ces populations et baisser de la même manière dans les zones attractives, alors qu'aucune amélioration dans la population sédentaire n'a eu lieu.

Certains auteurs ont mis en lumière le fait que les évolutions du désavantage étaient associées aux évolutions de la mortalité [100,101]. Néanmoins, ces évolutions n'ont pas été prises en compte dans cette étude. Les caractéristiques socio-économiques des communes ont été mesurées en 1990, ce qui a possiblement influé sur la mortalité. Cependant, l'indice de désavantage social de 1990 utilisé ici est corrélé à hauteur de 0.97 avec celui de 1999 [5]. De futures études devraient investiguer dans quelle mesure ces changements liés au niveau de

désavantage social sont associés aux changements dans la mortalité sur une plus longue période, le principal obstacle restant aujourd'hui l'indisponibilité de données fiscales localisées avant 1984. Par ailleurs, l'échelle communale ne permet pas de mettre en évidence les fortes disparités intra-urbaines existantes [4], alors même que les situations les plus extrêmes y cohabitent aux côtés des populations les plus favorisées.

Dans la régression de Poisson, l'association entre la mortalité relative toutes causes et le taux d'accroissement annuel moyen est restée significative après ajustement sur le désavantage social et sur le degré d'urbanisation. Le taux d'accroissement de la population qui mesure strictement la variation des effectifs de population sur une période, ne permet pas de distinguer l'accroissement naturel des mobilités résidentielles. Cependant, le taux d'accroissement est fortement corrélé avec le taux net de migration (de 0.82 à 0.95, selon la période). De fait, le taux d'accroissement de la population estime partiellement l'impact du renouvellement de population sur le risque de mortalité. De futures études devraient permettre l'estimation des poids respectifs du renouvellement de la population (effet de sélection) et de l'effet de contexte non mesuré sur l'évolution de la mortalité. Il serait par ailleurs nécessaire de travailler sur des données individuelles afin de comparer les risques de mortalité entre populations sédentaires et mobiles, ce qui permettrait ensuite de déterminer leur impact sur l'accroissement des disparités de santé [79,102].

Cette étude écologique exploratoire met en évidence la pertinence d'une décomposition des disparités spatiales de mortalité par cause et par échelle. L'approfondissement d'une telle analyse devrait permettre d'identifier des échelles géographiques pertinentes par cause, par âge et par sexe pour mettre en place des interventions en santé publique plus ciblées. La proposition d'une telle mesure des disparités géographiques pourrait également participer à des objectifs de veille sanitaire. L'approche dynamique pour expliquer les disparités de mortalité par des marqueurs pourrait également constituer une option crédible dans un objectif de veille, de prédiction et de planification sanitaire. Finalement, ce travail met en évidence que le taux d'accroissement de la population est un marqueur indirect des caractéristiques spatiales. Cette information vient s'ajouter au désavantage social et au contexte rural-urbain associés aux évolutions de la mortalité.

## II. ANALYSE MULTINIVEAUX DE LA MORTALITE PAR CAUSE - ETUDE DE FAISABILITE

### Introduction

#### **Epidémiologie et effets du contexte**

L'épidémiologie moderne des maladies chroniques était, jusqu'à récemment, uniquement tournée vers l'identification de facteurs de risques individuels [38]. Grâce aux développements des techniques qui ont permis la collecte systématique de données individuelles, leur traitement ainsi que des outils statistiques tels que l'analyse multivariée, les modèles de prédiction individuelle ont prospéré [103]. Dans ce champ d'analyse, le rôle des comportements à risque (le tabac, l'alcool, le régime alimentaire et l'activité physique) sur la santé des individus a souvent été mis en évidence. Progressivement, l'individualisation du risque et sa contrepartie qui est l'idée du libre arbitre dans les choix et modes de vie des populations sont devenus la norme [40]. Cette idée a cependant clairement été remise en question. D'une part, les maladies chroniques sont toujours réductibles à des dérèglements physiopathologiques individuels. Ces mesures sont souvent constatées trop tard, peu avant ou au moment du diagnostic. Même si leur identification permet de traiter la maladie dans certains cas mais aussi d'élaborer de nouvelles stratégies de soins individuelles, les comportements de santé délétères les précèdent souvent dans les processus étiologiques menant aux maladies chroniques. D'autre part, il a été démontré que les comportements de santé des populations étaient largement conditionnés et contraints par les interactions sociales dans lesquels s'insèrent les individus ainsi que par le contexte dans lequel ils évoluent. L'identification d'effets multiniveaux et plus particulièrement du contexte de résidence relève donc de l'action préventive de santé publique [48].

#### **Fiabilité et analyse multiniveaux des événements rares**

Ces dernières années, l'étude simultanée des facteurs de risque intervenant à plusieurs niveaux, notamment au niveau individuel et au niveau du contexte, a pris son essor, portée par le développement des modèles multiniveaux. Cette classe de modèle permet d'identifier les effets du contexte sur la santé puisqu'elle permet de retirer la part des disparités géographiques expliquée par la composition de la population [45,104,105]. Dans les études en santé des populations, l'utilisation du modèle de Poisson multiniveaux est très répandue pour mesurer le rôle du contexte de résidence sur l'incidence d'événements rares de santé tels que la mortalité, indépendamment des facteurs de risques individuels. Plusieurs études ont mis l'accent sur le fait que ce type d'analyse permettrait une meilleure identification des populations à risques, étape préalable à une éventuelle répartition des ressources plus efficace [40,45]. Une investigation des facteurs contextuels de la mortalité par cause spécifique en population générale nécessite une bonne précision des estimations. Cependant, les performances du modèle multiniveaux de Poisson n'ont reçu jusqu'à présent que très peu d'attention, alors même que des erreurs d'estimation et de diagnostics seraient préjudiciables à la qualité de la prise de décision en santé publique.

### **Les effets du contexte sur les comportements, la santé et la mortalité**

Dans la littérature, plusieurs associations contextuelles entre comportements de santé délétère et certaines caractéristiques du lieu de résidence ont été mises en lumière. Ainsi, différents indicateurs socio-économiques contextuels comme le niveau de revenu, le taux de chômage ou des scores de désavantage social sont associés à des risques accrus de consommer du tabac [106,107] et de l'alcool [108]. La présence, densité ou proximité de revendeurs d'alcool sont également positivement associés au risque d'en consommer pour les jeunes adultes [109]. La même association est également observée pour les revendeurs de tabac concernant les individus vivant dans un voisinage socialement désavantagé et les individus d'un niveau socio-économique plus modeste [110]. Certains comportements alimentaires délétères pour la santé sont associés à la densité de fast-food [111] et associés inversement à une offre alimentaire saine [112]. Parmi les indicateurs de santé, un faible niveau socio-économique du lieu de résidence est associé à une mauvaise santé déclarée [113,114] et à un risque plus élevé de contracter des maladies chroniques [115–117]. Une méta-analyse de l'effet contextuel du niveau socio-économique a également mis en évidence un effet négatif et significatif sur la mortalité dans plusieurs pays occidentaux [118]. Il n'existe cependant à ce jour aucune étude de ce type en France en population générale.

### **Définition du contexte, problématique de l'échelle géographique**

Une question inhérente aux analyses multiniveaux est celle de la définition et des frontières géographiques à l'intérieur desquelles le contexte pourrait jouer un rôle sur la santé des individus. Cette problématique méthodologique est un sujet de recherche à part entière au vu du nombre d'éléments à considérer. Le premier est incontournable : la disponibilité de données aux échelles voulues, ainsi que toutes problématiques qui en découlent. En effet, les données individuelles les plus fournies sont collectées à d'autres fins que les études de santé, souvent aux échelles administratives [119]. Dans les modèles multiniveaux, la puissance est déterminée par le nombre d'unités géographiques, le nombre d'individus par unité et la prévalence ou l'incidence de l'évènement à analyser. Dès lors que l'on travaille sur un échantillon de la population, l'analyse multiniveaux ne peut que se faire à des échelles géographiques très agrégées. Le second problème est le « Modifiable Areal Unit Problem » (MAUP) qui résulte de l'imposition d'un découpage artificiel de l'espace pour étudier l'effet d'un facteur de risque contextuel [120]. La dilution de l'information, et par extension la diminution de la variabilité à expliquer, lors du passage d'une échelle plus fine à une échelle plus agrégée, est bien connue. Aussi, une mesure agrégée fait l'hypothèse d'une exposition moyenne des individus qui résident au sein d'une unité géographique donnée alors même que cette exposition n'est pas la même en tous points de l'espace. Par ailleurs, le niveau d'agrégation des données affecte l'amplitude de l'association et parfois le sens de cette dernière. Pour conclure, la décision de considérer une échelle plutôt qu'une autre doit être fondée sur la pertinence de celle-ci pour une intervention en santé publique dont les leviers d'actions sont souvent contraints par les limites politiques et administratives [104]. De plus, la question de la stabilité des estimations dans une analyse contextuelle doit toujours faire l'objet d'un arbitrage équilibré avec l'échelle contextuelle à considérer. Plus l'échelle géographique considérée sera grossière, plus le nombre d'individus dans chaque unité et la puissance seront

élevés et plus le chercheur s'éloignera du contexte « proche » de l'individu ou « l'effet du voisinage » ne pourra être investigué.

### **Source de données pour l'analyse multiniveaux de la mortalité en population générale**

Selon les critères cités plus haut, une analyse contextuelle de la mortalité par cause en population générale nécessiterait des données individuelles sociodémographiques représentatives. Ces données doivent comprendre des informations caractérisant le niveau socio-économique individuel [121], le statut vital et la cause initiale de décès ainsi que le lieu de résidence et des informations caractérisant celui-ci. A ce titre, l'Insee met régulièrement à disposition des informations socio-économiques agrégées qui sont disponibles à plusieurs échelles. Concernant la disponibilité de données individuelles représentatives de la population Française, l'EDP fait figure de référence. Collectées à partir des bulletins individuels de recensement depuis 1968, cette base qui regroupe environ 1% de la population combine données déclaratives et état civil. Ces données ont été chaînées au fil des vagues de recensements afin d'en constituer une base longitudinale individuelle [59]. L'avantage de ce type de base pour étudier la mortalité est d'éviter le biais numérateur-dénominateur puisque les mêmes données sont utilisées pour calculer le nombre de décès et les effectifs de population exposés [122]. Cependant, ces données sont sujettes aux biais de sélection et à une attrition sélective. D'autres parts, le mode de recueil des données et leur traitement ont été informatisés à partir du recensement de 1990, ce qui a pu engendrer des erreurs [123]. L'utilisation de la base du CépiDc implique également certaines limites. Tout d'abord l'information socio-économique, indispensable pour une analyse multiniveaux, est déclarée par le médecin ou la famille au moment du décès. Un des effets néfastes de ce mode de recueil est la possibilité que le statut social du sujet décédé soit influencé par la maladie. Le sens du lien causal entre statut social et mortalité est par conséquent potentiellement bidirectionnel. Les gradients sociaux estimés sur cette base sont d'ailleurs considérés comme biaisés [122].

### **Mesure de concordance des données de mortalité**

Une étude des disparités spatiales de mortalité en population générale faite à partir d'un échantillon, ici l'EDP, doit vérifier différentes hypothèses afin que les associations investiguées soient généralisables à la population entière. Nous ferons donc premièrement l'hypothèse que l'EDP, échantillonné à partir du jour de naissance, est représentatif de la population française dans sa composition sociodémographique à chaque période de recensement (au moins). La seconde hypothèse concerne la concordance de la mortalité par unité géographique des indicateurs calculés sur la base transversale du CépiDc et longitudinale de l'EDP. A l'inverse de l'information socio-économique individuelle, la commune de domicile des défunts est une information a priori fiable et peu sensible aux maladies ayant provoqué le décès dans la base transversale ce qui minimise un certain nombre de biais. Par ailleurs dans la base longitudinale, il est possible qu'un individu déménage entre la période de recensement et son décès. Cependant, l'insensibilité de ce type de données au biais numérateur-dénominateur ainsi que le choix d'une fenêtre temporelle d'observation assez courte nous permet de faire l'hypothèse que les taux de mortalité obtenus avec les deux bases sont proches. Cette hypothèse de cohérence entre les deux bases est capitale pour mesurer et

généraliser l'impact potentiel de la composition et du contexte sur les disparités géographiques de mortalité. Une mesure de concordance de distribution entre deux sources de données pourrait être rapportée à une mesure de corrélation. Une telle mesure entre indicateurs de mortalité, qui ne sont pas considérés comme suivant une loi gaussienne, n'est cependant pas disponible à ce jour.

### **Objectifs**

Ce chapitre se présente comme une étude de faisabilité d'une analyse multiniveaux de la mortalité par cause en population générale. La réalisation d'une telle étude implique tout d'abord que les données spatiales de mortalité des individus issus de la base EDP utilisées pour effectuer une telle analyse soient fiables. Le premier objectif s'applique donc à proposer une mesure de concordance de la mortalité pour ensuite comparer les ratios standardisés de mortalité des individus EDP aux données de mortalité exhaustives issues du CépiDc. Dans un second temps, une analyse de fiabilité du modèle de Poisson multiniveaux qui fait appel à la simulation est présentée. Trois méthodes statistiques d'estimations de ce modèle sont mises à l'essai sur les données de l'EDP et dans le cadre d'un design équilibré où le nombre d'unités géographiques varie ainsi que le nombre d'individus. Certaines recommandations concernant le design et les méthodes statistiques sont énoncées en discussion pour clore cette partie.

### **Méthodes et sources**

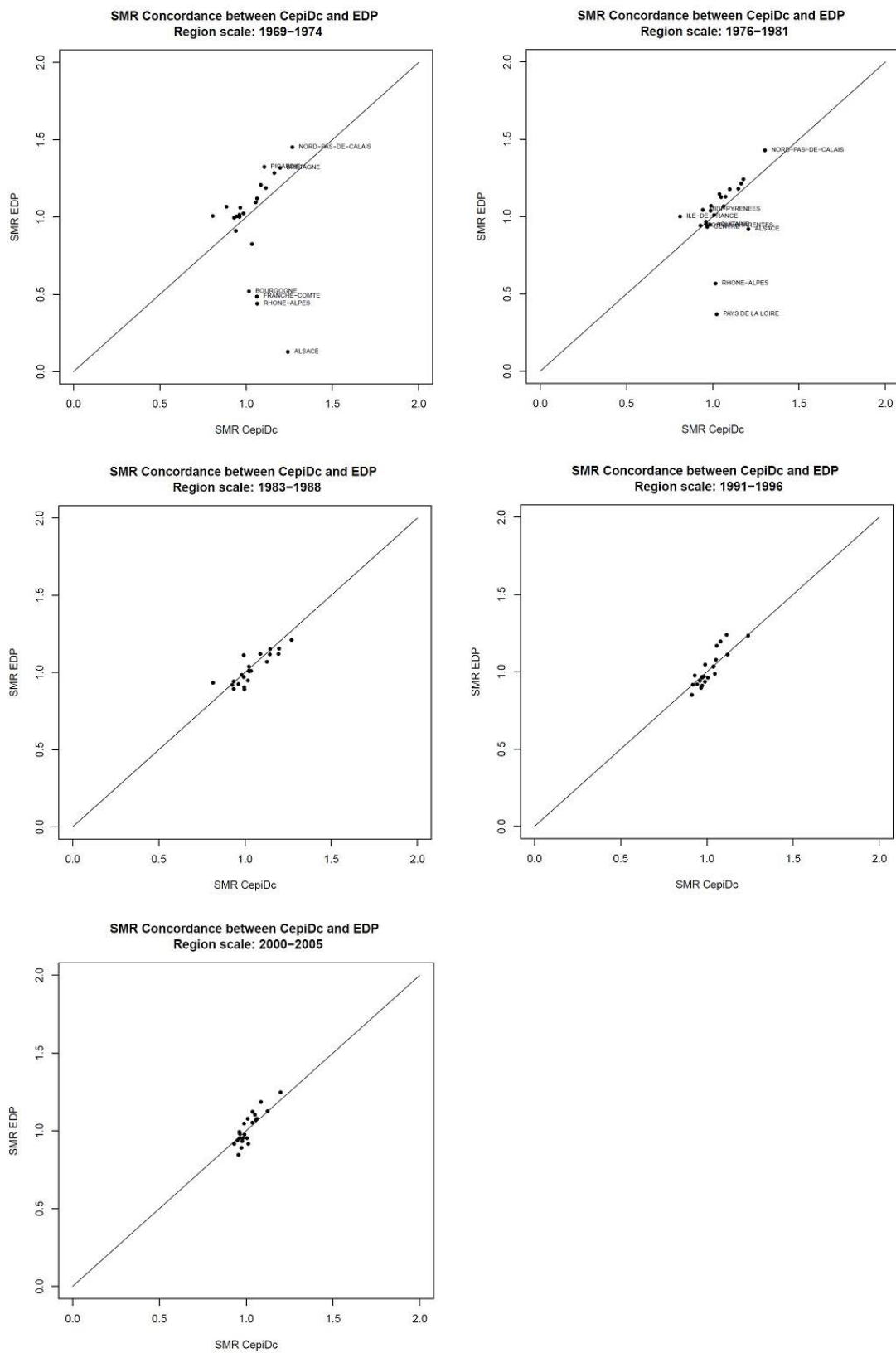
#### **Concordance des données de mortalité entre 2 sources - Analyse graphique**

L'objectif d'effectuer une analyse multiniveaux des disparités géographiques de mortalité implique l'utilisation de données individuelles. Les données utilisées dans ce travail proviennent de l'EDP. Afin de s'assurer que cette source de données est fiable notamment pour la mortalité, une analyse de concordance des RSM entre le CépiDc et l'EDP a été réalisée.

Les RSM tous âges issus du CépiDc et de l'EDP des périodes 1969-1974, 1976-1981, 1983-1988, 1991-1996 et 2000-2005 sont comparés en considérant les taux du CépiDc comme gold standard.

La Figure II.1 compare les RSM calculés à l'échelle régionale sur ces deux sources. On observe que les taux de mortalité EDP de quelques régions sont bien en dessous de ceux du CépiDc pour les deux premières périodes, soit 1969-1974 et 1976-1981. Précisément concernant la première période, les taux divergent pour la Bourgogne, la Franche-Comté, Rhône-Alpes et l'Alsace. Pour la seconde, les taux sont divergents pour la région Rhône-Alpes et pour les Pays de la Loire.

Figure II.1: Comparaison des RSM régionaux entre EDP et CépiDc



### Mesure de concordance des RSM à plusieurs échelles

L'objectif est de déterminer à quelles échelles géographiques et pour quelles périodes les données individuelles de l'EDP sont fiables en utilisant le CépiDc comme gold standard. L'établissement de la mesure de concordance suivante pour la mortalité permet d'effectuer ces comparaisons de manière systématique.

L'utilisation d'un modèle de Poisson à effets mixtes permet d'obtenir une telle mesure. Soit le modèle suivant :

$$\left\{ \begin{array}{l} O_{it} \sim \text{Poisson}(\lambda_{it}) \\ \log \lambda_{iu} = \alpha + \log E_{iu} + U_{iu} \\ U_{iu} \sim N(0, \Sigma) \text{ avec } \Sigma = \begin{pmatrix} \sigma_{\text{spat\_CépiDc}}^2 & \rho \cdot \sigma_{\text{spat\_CépiDc}} \cdot \sigma_{\text{spat\_EDP}} \\ \rho \cdot \sigma_{\text{spat\_CépiDc}} \cdot \sigma_{\text{spat\_EDP}} & \sigma_{\text{spat\_EDP}}^2 \end{pmatrix} \end{array} \right.$$

où  $u \in \{\text{CépiDc}, \text{EDP}\}$  les deux sources de données,  $i \in I$  l'ensemble des unités géographiques.

Ainsi la quantité d'intérêt  $\rho$  constitue l'indice de concordance proposé. Plus cette valeur est proche de 1, plus les deux sources concordent. Si cette valeur est inférieure à 0.80, il est possible d'émettre des doutes quant à la fiabilité du statut vital des individus l'EDP (seuil arbitraire).

Son intervalle de confiance à 95% est obtenu à l'aide d'une méthode bootstrap paramétrique.



Tableau II.1: Indice de concordance des distributions spatiales de la mortalité provenant des 2 sources: CépiDc et EDP

Echelle	Période	Concordance		
		Corrélation	IC à 95%	
Région	1969-1974	-0.19	-0.56	0.25
	1976-1981	0.31	-0.12	0.65
	1983-1988	0.87	0.67	0.97
	1991-1996	0.91	0.78	0.99
	2000-2005	0.94	0.82	1.00
Département	1969-1974	-0.08	-0.28	0.12
	1976-1981	0.21	0.00	0.41
	1983-1988	0.71	0.54	0.85
	1991-1996	0.89	0.80	0.98
	2000-2005	0.95	0.86	1.00
Aire Urbaine	1969-1974	0.13	0.01	0.26
	1976-1981	0.45	0.32	0.57
	1983-1988	0.81	0.69	0.94
	1991-1996	0.98	0.90	1.00
	2000-2005	0.91	0.82	1.00
Zone d'emploi	1969-1974	0.04	-0.04	-0.04
	1976-1981	0.34	0.29	0.29
	1983-1988	0.75	0.63	0.85
	1991-1996	0.92	0.83	1.00
	2000-2005	0.90	0.81	1.00

Le Tableau II.1 met assez nettement en évidence que les données de mortalité des deux premières périodes 1969-1974 et 1976-1981 ne sont pas fiables dans l'EDP. La période suivante, bien qu'apparaissant comme plus fiable, exhibe un indice de concordance que l'on peut considérer comme trop faible à l'échelle du département et de la Zone d'Emploi.

## Fiabilité du Modèle de Poisson Multiniveaux - Définition du modèle

Le modèle multiniveaux à l'essai est le Poisson à constante aléatoire. C'est un modèle de Poisson à deux niveaux, les individus  $i \in I$  sont le niveau 1 et les lieux de résidence ou unités géographiques  $j \in J$  le niveau 2. Soit:

$$\begin{cases} O_{ij} \sim \text{Poisson}(\lambda_{ij}) \\ \log \lambda_{ij} = \alpha + \log E_{ij} + \beta_1^i \cdot x_{ij} + \beta_2 \cdot z_j + u_j + \varepsilon_{ij} \\ u_j \sim N(0, \sigma_{\text{spat}}^2) \text{ et } \varepsilon_{ij} \sim N(0, \sigma_{\text{surd.}}^2) \end{cases}$$

où  $x_{ij}$  est une variable individuelle et  $\beta_1^i$  les effets individuels,  $z_j$  une variable contextuelle et  $\beta_2$  l'effet de contexte,  $u_j$  l'effet aléatoire de l'unité géographique  $j$  et  $\sigma_{\text{spat}}^2$  la variabilité spatiale,  $\varepsilon_{ij}$  l'effet aléatoire de l'observation  $ij$  et  $\sigma_{\text{surd.}}^2$  la variance de la distribution de la surdispersion.

## Protocole de simulation

### a) Population de l'Echantillon Démographique Permanent

La population de référence est tirée de l'échantillon EDP des individus ayant rempli leur bulletin de recensement, nés en France ayant entre 30 et 79 ans au 1er Janvier 1990 et qui était vivant le 1er janvier 1991.

### b) Design équilibré

C'est un cas particulier des données multiniveaux où tous les groupes de niveau 2 (les unités géographiques dans notre cas) ont le même nombre d'individus. La répartition des effectifs dans chaque unité géographique par classe d'âge, sexe et niveau est extraite de l'échantillon de base qui est représentatif de la population. L'incidence de l'évènement de santé est fixée à 1% en taux brut pour mimer un taux brut annuel de mortalité.

Ces données ont été simulées pour 100, 200, 500, 1000, 5000 individus par clusters (ipc), 10, 20, 50, 100, 200 et 500 unités géographiques (ou clusters) pour une variance spatiale  $\sigma_{\text{spat}}^2$  qui varie de 0.005 à 0.1.

### c) Design déséquilibré

Ce design est celui de la population EDP (cf. a)) où l'on considère successivement le niveau 2 (contexte de résidence) comme étant la région, le département et la Zone d'Emploi. Le nombre de décès attendu  $E_{ij}$  a également été multiplié faisant varier l'incidence de l'évènement pour mimer la mortalité par cause spécifique (10%, 20%, 50%, 100% et 200% du nombre de décès attendus). Le suivi des individus s'arrête le 31 Décembre 1996.

### d) Simulation des données

Ces données ont été simulées à l'aide d'une loi de Poisson de paramètre :

$$\lambda_{ij} = \exp(\alpha + \log E_{ij} + \beta_1^i \cdot x_{ij} + \beta_2 \cdot z_j + u_j + \varepsilon_{ij}),$$

où  $\alpha = -0.3$ ,  $\beta_1^1 = 0$ ,  $\beta_1^2 = 0.3$ ,  $\beta_1^3 = 0.4$ ,  $\beta_1^4 = 0.6$  les effets individuels du niveau d'études,  $\beta_2 = 0.6$  l'effet de contexte, les  $u_j$  et  $\varepsilon_{ij}$  étaient simulés à chaque itération selon

deux lois normales centrées indépendantes de paramètres  $\sigma_{\text{spat}}^2$  (qui varie) et  $\sigma_{\text{surd.}}^2=0.07$ . Toutes ces valeurs cibles sont extraites et proches des associations et des variances mesurées sur cette population. Chacun des scénarios a été simulé 1000 fois.

### e) Méthodes statistiques testées

Les performances des estimations des effets fixes et aléatoires ont été comparées pour trois méthodes statistiques disponibles dans le logiciel R [124] :

- La méthode `glmer` du package `lme4` [125] est une méthode fréquentiste qui ajuste des modèles linéaires généralisés mixtes en approximant la vraisemblance à l'aide de la quadrature de Gauss-Hermite (un point de quadrature par défaut).
- La méthode `Inla` du package `INLA` qui ajuste des modèles linéaires généralisés mixtes en utilisant une approche Bayésienne qui approxime la distribution marginale a posteriori à l'aide d'une approximation de Laplace [126]. Cette méthode fournit le paramètre de précision de l'effet aléatoire qui suit une loi log-gamma (loi a priori par défaut) et qui est l'inverse des variances des effets aléatoires, nos quantités d'intérêts. Afin d'obtenir ces dernières, ces deux variances (spatiale et surdispersion) ont été estimées séparément en calculant la moyenne de 10 000 simulations de Monte-Carlo de l'inverse de ces lois de distribution log-gamma à chaque itération.
- La méthode `jags` du package `R2jags` qui utilise une approche Bayésienne et une méthode MCMC (Monte-Carlo par Chaîne de Markov) pour estimer les paramètres du modèle. Le package `R2jags` permet d'appeler le logiciel JAGS dans une session R qui effectue un échantillonnage de Gibbs [127]. Pour chacune des 1 000 simulations, 50 000 itérations MCMC ont été générées, les 15 000 premières furent mises de côté (*burn-in*). Parmi les 35 000 suivantes, seulement 1 sur 50 ont été retenues (*thinning interval*) afin de se prémunir de l'autocorrélation inhérente aux méthodes MCMC. La loi a priori des effets fixes est une loi gaussienne peu informative (*Vague Gaussian prior*) centrée avec un paramètre de précision fixé à  $10^{-5}$ . Pour la variance des effets aléatoires, une loi a priori *half-Cauchy* a été implémentée. Proposée par Andrew Gelman, cette loi permet de mieux estimer les petites variances et les designs ayant un petit nombre de clusters [128].

Les simulations et estimations ont été effectuées avec le logiciel R. Cette opération a été effectuée sur une grille de calcul à l'aide d'une machine virtuelle Ubuntu 32 cœurs. Le temps de calcul total s'élève à environ 3 semaines.

### f) Mesures de fiabilité des estimations

Trois critères de fiabilité ont été choisis afin d'évaluer les méthodes [129] :

- Le pourcentage de biais par rapport à la valeur cible qui doit se situer entre -10% et +10% pour être acceptable.
- L'Ecart Quadratique Moyen (EQM) afin de mesurer la précision.
- La probabilité de couverture à 95% qui est la proportion de fois où l'intervalle de confiance estimé contient la valeur cible. Cette dernière quantité permet d'évaluer la propension de la

méthode statistique à générer des erreurs de type I (*undercoverage*) et de type II (*overcoverage*) et doit avoisiner les 95% pour être considérée comme acceptable.

## **Résultats**

### **Design équilibré**

#### *Effets fixes*

Les estimations des effets ont montré un biais négligeable excepté pour le design le plus épars avec 10 clusters avec 100 ipc. Dans ce scénario, le biais le plus élevé était observé pour la méthode Inla et concerne quasiment tous les effets fixes. Les méthodes Inla et lme4 ont commencé à rendre des estimations avec un biais acceptable à partir de 1000 individus dans un scénario avec 10 clusters, à partir de 500 individus pour 20 clusters et à partir de 200 individus pour 50 clusters. La seule méthode exhibant un biais acceptable dans tous les scénarios était la méthode Jags. Dans tous les scénarios avec 100 individus par clusters, l'estimation de la constante avec la méthode lme4 était biaisée (Tableau II.2).

Dans le scénario avec 10 ou 20 clusters et 100 individus ou 200 individus, Inla a montré une EQM de 2 à 10 fois plus grand que les autres méthodes investiguées jusqu'à atteindre le même niveau que lme4 à partir de 1000 individus. Dans tous les scénarios, l'EQM des estimations obtenues à l'aide de Jags était nettement plus petit jusqu'aux seuils suivant où les 3 méthodes exhibaient des EQM du même ordre, soit à partir de 5000 individus indépendamment du nombre de clusters. Dans tous les scénarios, la méthode Jags montrait un EQM plus petit. Pour chacune des méthodes, l'estimation de l'effet de contexte avait l'EQM le plus élevé (Tableau II.3).

La probabilité de couverture des effets fixes était globalement satisfaisante pour tous les scénarios et méthodes mis à part pour la constante et l'effet du contexte, estimés à l'aide de la méthode Inla dans les scénarios avec 10 ou 20 clusters. Pour ces cas précis, le nombre croissant d'individus par clusters faisait baisser de manière inacceptable la probabilité de couverture. Par ailleurs la méthode lme4 exhibait une probabilité de couverture souvent trop élevée dans les scénarios avec 10 clusters et de manière inattendue trop faible uniquement pour le design avec 500 clusters et 100 individus. Dans tous les scénarios, une bonne et stable probabilité de couverture des estimations a été constatée pour la méthode Jags (Tableau II.4).

#### *Effets aléatoires*

Les estimations de la variance des effets mesurant la surdispersion étaient toutes biaisées pour les scénarios incluant 10 clusters avec toutefois la méthode Jags qui renvoyait un biais plus acceptable (+15 à +25% de la valeur cible). Globalement, la méthode Inla renvoyait des résultats très biaisés. Avec cette méthode, un minimum de 5000 i.p.c était requis jusqu'à 100 clusters et 1000 i.p.c minimum à partir de 200 clusters. La méthode lme4 réclamait un minimum de 20 clusters et 1000 individus pour atteindre un biais acceptable, là où la méthode Jags demandait un minimum de 20 clusters et 100 individus (Tableau II.2). La méthode Jags exhibait l'EQM le plus faible et une probabilité de couverture toujours proche de 95% (Table II.3). La probabilité de couverture de la méthode Inla était toujours trop peu élevée et nulle dans de nombreux cas excepté à partir de 100 clusters avec 5000 i.p.c (Tableau II.4).

Les estimations de la variance spatiale nécessitaient un minimum de 50 clusters pour Jags, et un minimum de 200 i.p.c pour les scénarios avec 50 ou 100 clusters estimés à l'aide de la

méthode lme4. Dans tous les scénarios, l'EQM le plus faible constaté était celui renvoyé par Jags, les deux autres méthodes donnant des EQM comparables à partir de 5000 i.p.c jusqu'à 100 clusters ou à partir de 1000 individus pour 200 clusters. La probabilité de couverture était toujours trop faible pour Inla mais arrivait à un niveau acceptable à partir de 200 clusters avec un minimum de 5000 i.p.c. La méthode Jags montrait une très grande stabilité quant à l'amplitude de la probabilité de couverture toujours acceptable.

## **Design déséquilibré - Population EDP**

### *Effets fixes*

Globalement, très peu de biais était constaté sur les estimations du modèle aux trois échelles géographiques (Tableau II.5). Aussi, l'EQM des estimations effectuées à l'aide des trois méthodes ont montré très peu de différences (Tableau II.6). Quand la prévalence de l'évènement augmentait, l'EQM déclinait. Plus l'échelle géographique était fine, plus l'influence de la prévalence était forte sur la réduction de l'EQM. La méthode Jags a obtenu une probabilité de couverture acceptable dans tous les scénarios alors qu'Inla exhibait une valeur un peu faible mais acceptable à l'échelle régionale (Tableau II.7). La probabilité de couverture de la méthode lme4 était également un peu faible pour tous les effets à l'échelle Zone d'Emploi quand la prévalence était inférieure à 20% de la mortalité de la population EDP analysée. Une variance spatiale qui augmentait n'avait pas d'influence sur les estimations des effets fixes mis à part pour l'effet de contexte dont l'EQM augmentait.

### *Effets aléatoires*

Concernant la variance mesurant la surdispersion, il n'y avait pas de biais excepté pour la plus faible prévalence (0.1) et la méthode lme4 à l'échelle régionale, ainsi qu'à l'échelle Zone d'Emploi pour la méthode Inla. De manière générale, les EQM étaient identiques peu importe la méthode d'estimation. La méthode Inla montrait une probabilité de couverture un peu faible là où celle de la méthode Jags était fiable.

Pour la variance spatiale, les estimations faites à l'aide d'Inla étaient largement biaisées si la variance était inférieure de 0.05. La méthode Jags montrait un biais positif à l'échelle régionale ainsi qu'aux autres échelles si la prévalence était très faible. Globalement, les biais des estimations obtenues avec lme4 sont apparus comme peu affectés par le choix de l'échelle ou de l'amplitude de la prévalence. Les EQM obtenus avec lme4 étaient plus faibles à l'échelle régionale mais comparables aux deux autres méthodes pour l'échelle départementale et Zone d'Emploi. La probabilité de couverture de cette variance était toujours acceptable avec la méthode Jags. A l'inverse la probabilité de couverture de la méthode Inla nécessitait une variance fixée au minimum à 0.05 à l'échelle départementale et Zone d'Emploi uniquement.

Tableau II.2: Biases (%) en fonction du nombre de clusters et d'individus par clusters - Design équilibré et variance fixée à 0.05

Clusters	Individus	Constante $\alpha$			Contexte $\beta_2$			Indiv $\beta_1^2$			Indiv $\beta_1^3$			Indiv $\beta_1^4$			Var. spatiale $\sigma_{\text{spat}}^2$			Surdispersion $\sigma_{\text{urd}}^2$			
		lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	
% Bias	10	100	-71.1	826.8	13.0	26.3	15.2	4.5	-105.3	0.8	3.8	-108.3	461.8	6.2	-68.3	320.1	4.1	94.3	-99.7	91.4	135.1	-99.8	25.0
		200	9.6	229.5	6.9	-1.3	-3.9	0.4	-34.1	80.4	0.8	-19.4	149.3	2.5	-11.9	95.7	3.1	3.7	-99.7	82.4	17.8	-99.8	20.3
		500	35.2	36.5	0.6	8.1	7.9	1.8	7.1	12.3	-0.7	15.1	23.3	0.1	13.0	15.5	0.4	-21.9	-98.8	72.2	-29.4	-99.8	17.5
		1000	22.0	2.3	-0.2	4.2	3.9	-1.2	8.3	11.4	1.4	13.6	8.5	-1.0	8.4	6.5	-0.4	-24.2	-76.2	67.9	-35.8	-94.9	17.7
		5000	1.4	5.3	-0.8	3.1	-0.1	0.5	4.3	7.2	-2.3	0.6	6.1	-0.9	0.9	3.2	-0.7	-16.9	-35.2	75.9	-27.6	-14.2	15.1
	20	100	45.9	235.1	5.7	0.6	1.3	1.1	-23.9	115.6	2.7	-6.4	146.7	2.0	-7.8	101.3	1.2	62.4	-99.7	28.0	195.9	-99.8	12.2
		200	51.1	64.0	4.0	1.9	0.9	1.0	0.7	35.2	2.0	20.3	42.2	1.7	14.3	31.1	1.2	11.7	-99.7	25.8	60.6	-99.8	11.1
		500	21.5	8.6	0.5	2.6	0.9	0.9	8.3	14.5	-0.3	11.1	11.6	0.4	9.5	9.9	0.8	-17.8	-88.5	21.6	-6.5	-99.5	7.8
		1000	11.2	1.4	1.7	1.1	-1.1	-1.2	8.5	6.7	1.1	8.5	6.5	0.1	4.7	5.6	1.1	-17.3	-52.5	22.0	-20.9	-84.0	7.4
		5000	0.3	2.1	-0.9	-0.1	1.6	-2.7	1.3	4.2	-0.2	-0.6	2.8	-0.6	1.0	2.3	-0.6	-8.7	-35.3	22.0	-13.9	0.8	6.8
	50	100	80.6	30.4	2.9	0.4	-2.3	1.6	27.6	31.3	0.9	27.2	27.1	1.4	20.5	19.5	0.6	26.1	-99.7	6.0	280.4	-99.8	5.6
		200	27.6	4.9	1.3	0.8	-0.9	0.5	4.4	10.4	0.3	8.1	13.8	1.6	7.5	6.3	0.4	8.4	-98.7	6.2	76.1	-99.8	3.3
		500	4.9	-9.2	0.0	1.7	-1.2	-0.5	1.9	-1.0	0.0	2.5	3.5	-0.3	1.2	3.1	-0.1	-10.5	-66.1	6.2	3.4	-94.5	2.9
		1000	2.8	-2.3	0.2	-0.1	-0.4	-0.5	3.7	2.5	0.2	2.8	2.0	-0.2	2.0	1.3	0.1	-10.4	-32.9	5.6	-4.1	-47.3	2.5
		5000	0.8	0.7	0.4	-0.1	0.2	-0.2	0.7	0.9	-0.2	-0.3	0.8	0.2	0.1	-0.4	0.5	-8.4	-15.8	6.9	-4.5	1.6	2.4
	100	100	50.7	3.5	1.3	-2.4	-1.5	-0.2	7.1	10.2	0.8	6.6	10.9	0.4	6.6	8.1	0.7	21.0	-99.7	2.9	287.0	-99.8	2.9
		200	14.2	-7.1	0.8	1.9	1.8	-1.0	-1.0	6.5	0.7	1.2	6.0	0.7	2.1	3.1	0.3	-6.9	-95.0	3.9	79.8	-99.8	2.3
		500	6.8	-6.8	0.5	-0.1	1.4	-0.4	3.7	1.0	0.5	3.8	2.3	0.3	2.4	1.9	0.6	-8.5	-37.2	3.9	16.4	-73.0	1.2
		1000	3.3	-2.5	-0.1	-0.1	-0.7	0.9	2.7	-1.2	-0.3	1.9	0.7	-0.3	0.5	0.2	-0.3	-7.6	-22.4	3.0	-0.1	-24.0	1.3
		5000	-0.5	0.4	-0.1	0.5	-0.5	0.2	-0.8	0.0	-0.4	0.2	0.5	0.1	-0.4	0.2	0.2	-4.3	-8.3	3.6	-3.2	-0.6	1.5
	200	100	39.9	-11.2	0.1	0.9	0.1	-0.2	1.0	3.6	-0.1	3.2	5.3	0.1	2.2	3.1	0.0	-4.8	-99.7	1.3	309.4	-99.8	2.1
		200	13.5	-13.8	0.4	0.3	0.7	-0.8	0.6	1.3	0.1	1.2	4.4	0.2	0.9	2.0	0.1	-12.2	-80.9	2.0	96.4	-99.8	0.9
		500	3.7	-1.9	0.0	0.7	-1.1	-0.4	0.3	1.1	0.2	1.1	2.3	-0.2	0.2	1.8	0.2	-8.0	-26.0	1.6	21.5	-45.5	1.0
		1000	2.6	-1.6	-0.1	0.2	0.1	0.0	0.2	0.8	0.0	1.7	0.1	-0.1	0.8	0.2	-0.1	-9.8	-9.0	1.6	9.4	-14.4	0.6
		5000	0.0	-0.4	0.0	-0.3	-0.7	-0.1	0.2	-0.6	0.6	0.0	0.3	0.3	0.1	-0.3	0.1	-0.9	-4.3	2.4	-2.1	-1.6	0.5
500	100	41.5	-16.0	-0.3	0.9	-0.2	0.0	-2.7	2.4	-0.6	-1.1	3.0	-0.2	-0.4	0.9	0.0	-16.4	-94.5	0.3	392.6	-99.8	0.9	
	200	14.1	-11.2	0.0	0.1	-0.4	-0.1	2.5	-1.0	0.0	2.0	-0.3	0.0	1.3	0.0	-0.1	-12.7	-42.1	0.5	106.9	-73.7	0.5	
	500	2.5	-2.3	0.1	-0.5	0.3	-0.1	0.0	0.2	0.1	0.2	0.1	0.1	0.4	-0.1	0.0	-7.5	-7.7	0.6	23.6	-17.4	0.3	
	1000	1.3	-1.0	0.1	-0.4	-0.6	-0.2	0.6	-0.7	0.0	0.4	-0.2	0.1	0.3	0.0	0.0	-8.2	-4.2	0.9	12.0	-4.8	0.2	
	5000	-0.2	-0.2	0.1	0.5	0.0	0.2	-0.2	-0.2	0.1	0.0	-0.1	0.1	-0.1	-0.2	0.1	-1.2	-2.1	0.9	-1.1	-1.1	0.3	

Tableau II.3: EQM ( $10^{-2}$ ) en fonction du nombre de clusters et d'individus par clusters - Design équilibré et variance fixée à 0.05

Clusters	Individus	Constante $\alpha$			Contexte $\beta_2$			Indiv $\beta_1^2$			Indiv $\beta_1^3$			Indiv $\beta_1^4$			Var. spatiale $\sigma_{\text{spat}}^2$			Surdispersion $\sigma_{\text{surd}}^2$		
		lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags
10	100	35.03	3807.63	3.42	180.75	230.88	13.31	57.29	1522.05	4.58	72.57	2926.19	4.16	64.80	2617.33	3.73	8.14	0.25	0.72	19.05	0.49	0.24
	200	32.15	443.84	2.41	89.08	100.49	9.86	60.66	481.82	3.14	54.56	464.78	2.70	47.14	468.83	2.73	1.53	0.25	0.54	3.46	0.49	0.16
	500	25.04	46.31	1.65	37.76	44.22	10.43	37.57	58.02	2.05	31.08	51.75	1.88	30.75	43.55	1.90	0.61	0.26	0.49	1.08	0.49	0.10
	1000	13.85	12.84	1.60	26.88	23.70	10.37	19.64	19.24	1.85	16.83	15.53	1.68	14.36	15.64	1.74	0.34	0.40	0.46	0.54	0.54	0.07
	5000	3.26	3.57	1.50	11.90	14.94	11.74	4.46	4.99	1.50	3.83	4.19	1.48	3.90	3.95	1.52	0.16	0.27	0.73	0.16	0.37	0.06
20	100	36.12	402.82	1.76	70.54	73.33	5.79	60.48	481.76	2.40	46.02	437.02	2.08	47.13	416.67	1.99	3.17	0.25	0.16	19.24	0.49	0.11
	200	32.75	78.76	1.10	36.66	31.78	4.60	46.52	81.95	1.57	38.97	84.72	1.34	36.37	93.11	1.28	1.22	0.25	0.13	3.93	0.49	0.06
	500	13.08	12.56	0.84	17.90	16.49	4.18	17.91	20.00	1.04	15.49	16.16	0.96	14.42	15.29	1.02	0.36	0.32	0.11	0.81	0.49	0.04
	1000	5.96	6.57	0.72	10.89	10.78	4.31	8.51	10.48	0.87	7.41	7.67	0.87	6.94	7.77	0.80	0.22	0.32	0.10	0.39	0.58	0.03
	5000	1.65	1.58	0.66	5.10	6.09	4.21	2.43	2.32	0.72	2.04	1.93	0.70	1.97	2.05	0.78	0.08	0.15	0.10	0.08	0.17	0.03
50	100	38.40	34.87	0.62	22.22	22.25	1.87	45.60	56.34	0.88	37.60	41.20	0.80	37.05	38.50	0.79	1.41	0.25	0.05	16.87	0.49	0.04
	200	13.25	11.35	0.43	12.73	12.34	1.75	18.56	18.60	0.57	15.43	14.11	0.54	14.28	14.48	0.52	0.56	0.27	0.03	2.85	0.49	0.02
	500	4.03	4.24	0.33	5.91	5.53	1.50	6.13	6.78	0.38	5.16	5.43	0.38	4.73	5.02	0.35	0.20	0.30	0.03	0.54	0.53	0.01
	1000	2.37	2.35	0.27	3.66	3.72	1.44	3.58	3.33	0.35	2.90	2.75	0.34	2.93	2.65	0.32	0.11	0.19	0.03	0.24	0.46	0.01
	5000	0.67	0.65	0.26	2.01	1.90	1.39	0.96	0.94	0.29	0.79	0.89	0.31	0.76	0.77	0.29	0.04	0.06	0.03	0.03	0.05	0.01
100	100	15.69	11.00	0.33	9.91	10.96	0.91	17.85	17.38	0.48	13.85	14.14	0.41	14.00	13.48	0.40	0.88	0.25	0.02	13.83	0.49	0.02
	200	5.81	5.64	0.22	6.02	5.12	0.86	7.72	9.01	0.30	6.39	6.93	0.25	6.22	6.36	0.25	0.32	0.28	0.02	1.95	0.49	0.01
	500	2.35	2.32	0.16	2.86	2.79	0.72	3.36	3.25	0.19	2.70	2.96	0.18	2.85	2.80	0.19	0.13	0.21	0.01	0.36	0.54	0.01
	1000	1.13	1.15	0.15	1.87	1.71	0.82	1.71	1.84	0.19	1.43	1.41	0.18	1.37	1.41	0.15	0.06	0.13	0.01	0.15	0.30	0.00
	5000	0.30	0.33	0.12	1.01	0.95	0.73	0.45	0.48	0.15	0.37	0.43	0.14	0.38	0.39	0.15	0.02	0.02	0.01	0.02	0.02	0.00
200	100	7.66	5.37	0.16	4.75	5.35	0.46	7.88	7.99	0.24	6.22	6.58	0.21	5.93	6.49	0.20	0.47	0.25	0.01	11.43	0.49	0.01
	200	2.74	2.65	0.11	2.47	2.67	0.43	3.61	3.97	0.14	3.12	3.11	0.13	2.76	3.19	0.13	0.21	0.28	0.01	1.64	0.49	0.00
	500	1.13	1.12	0.09	1.42	1.36	0.39	1.62	1.56	0.11	1.30	1.32	0.10	1.28	1.31	0.10	0.08	0.15	0.01	0.24	0.41	0.00
	1000	0.56	0.57	0.08	0.85	0.77	0.35	0.79	0.84	0.09	0.66	0.74	0.09	0.65	0.64	0.08	0.03	0.07	0.01	0.08	0.15	0.00
	5000	0.18	0.15	0.06	0.44	0.45	0.36	0.24	0.22	0.07	0.20	0.18	0.07	0.21	0.20	0.07	0.01	0.01	0.01	0.01	0.01	0.00
500	100	4.25	2.19	0.06	2.08	2.11	0.19	3.14	3.10	0.09	2.50	2.47	0.08	2.35	2.41	0.08	0.26	0.27	0.00	11.53	0.49	0.00
	200	1.31	1.16	0.05	1.06	1.03	0.14	1.66	1.67	0.06	1.35	1.20	0.05	1.28	1.24	0.05	0.11	0.19	0.00	1.16	0.53	0.00
	500	0.43	0.43	0.03	0.49	0.53	0.14	0.67	0.63	0.04	0.52	0.51	0.04	0.51	0.52	0.04	0.03	0.07	0.00	0.13	0.21	0.00
	1000	0.21	0.23	0.03	0.33	0.38	0.14	0.32	0.36	0.03	0.25	0.28	0.03	0.26	0.28	0.03	0.02	0.02	0.00	0.04	0.04	0.00
	5000	0.07	0.07	0.03	0.19	0.18	0.14	0.09	0.09	0.03	0.08	0.08	0.03	0.08	0.08	0.03	0.00	0.00	0.00	0.00	0.00	0.00

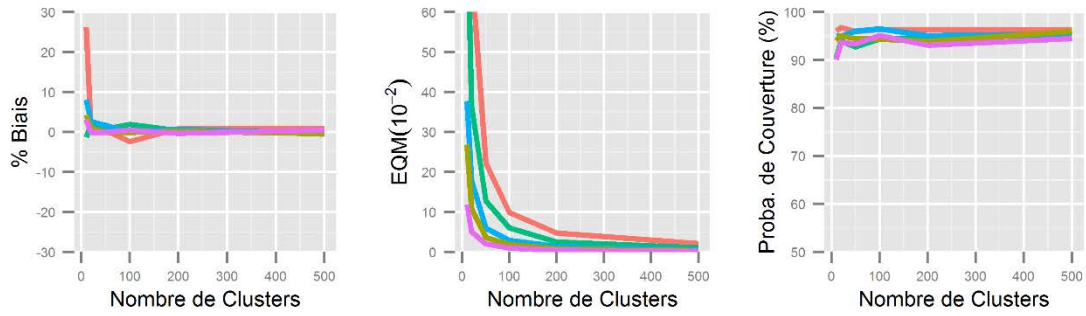


Tableau II.4: Probabilité de couverture (%) en fonction du nombre de clusters et d'individus par clusters - Design équilibré et variance à 0.05

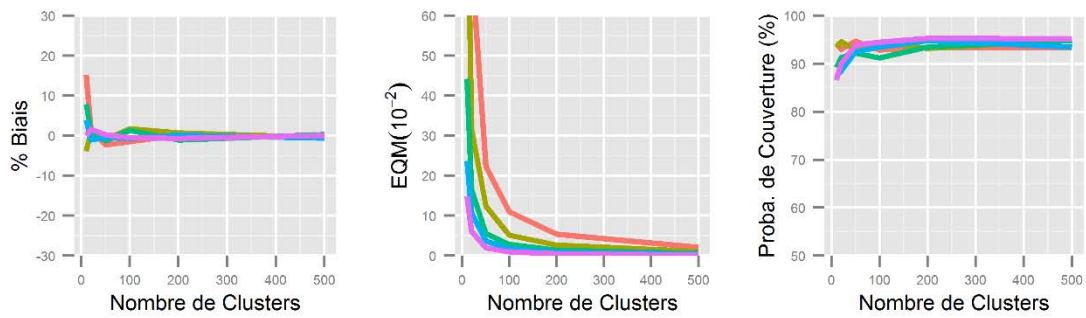
Clusters	Individus	Constante $\alpha$			Contexte $\beta_2$			Indiv $\beta_1^2$			Indiv $\beta_1^3$			Indiv $\beta_1^4$			Var. spatiale $\sigma_{\text{spat}}^2$			Surdispersion $\sigma_{\text{surd}}^2$				
		lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags		
Probabilité de Couverture (%)	10	100	95.9	89.6	96.1	96.1	94.2	95.8	99.8	98.7	94.9	97.5	91.2	94.3	96.9	89.9	95.3	0.0	96.1	0.0	95.1			
		200	95.0	96.9	95.6	90.5	93.4	96.2	96.5	98.7	94.6	95.4	96.2	95.2	96.3	95.2	94.3	0.0	97.0	0.0	92.7			
		500	95.8	95.4	95.9	95.0	89.2	95.9	99.3	93.5	95.0	96.9	95.1	95.4	96.4	94.8	95.0	0.0	95.4	0.0	93.8			
		1000	96.6	92.9	95.8	94.1	87.3	96.0	97.9	93.3	95.4	97.0	93.6	95.6	97.1	93.4	95.2	5.4	95.7	0.6	94.8			
		5000	93.9	92.8	95.0	90.1	86.6	93.4	94.9	92.8	95.6	93.3	92.9	94.9	93.1	93.5	95.0	37.4	94.0	56.0	94.5			
	20	100	94.3	98.1	94.9	96.8	93.0	94.0	95.5	98.6	94.6	97.0	96.2	95.1	95.9	96.8	94.7	0.0	94.7	0.0	93.7			
		200	95.9	96.5	96.3	94.0	94.6	95.7	95.2	95.3	94.3	95.7	95.3	94.7	94.1	95.3	95.0	0.0	94.2	0.0	93.9			
		500	95.6	94.4	94.8	95.0	91.3	95.2	96.4	93.1	94.4	95.8	94.7	94.4	95.4	94.7	93.6	1.6	92.5	0.1	93.3			
		1000	94.1	93.2	95.0	95.1	88.5	95.0	94.8	92.0	95.1	95.4	93.3	94.5	94.3	93.9	95.8	20.1	95.0	3.0	93.9			
		5000	95.6	94.7	94.5	93.7	90.0	93.3	94.8	94.1	94.5	95.7	95.7	95.0	94.0	94.8	94.3	56.7	92.8	76.9	92.5			
	50	100	97.0	95.7	95.6	96.0	94.7	94.6	99.5	95.2	95.4	97.9	95.6	94.6	95.6	97.1	94.3	0.0	91.9	0.0	93.2			
		200	96.5	94.8	94.5	92.8	92.8	93.5	96.0	94.3	95.3	96.1	95.5	94.6	96.0	94.1	94.8	0.0	94.1	0.0	94.6			
		500	96.9	94.3	95.2	96.0	92.2	95.3	98.1	94.0	95.7	97.0	94.5	94.1	97.5	95.3	95.3	12.6	94.1	0.1	94.3			
		1000	96.0	93.0	96.1	94.5	92.5	94.5	96.4	94.8	95.1	95.8	94.6	93.9	97.1	93.8	94.1	45.2	93.9	25.9	95.4			
		5000	93.4	95.4	94.6	93.4	93.9	94.6	92.5	94.8	94.4	93.8	93.6	93.1	93.8	94.6	94.4	83.6	93.4	90.7	95.5			
	100	100	94.8	94.8	94.7	96.4	92.9	94.1	96.1	95.1	94.3	95.9	94.3	94.4	96.5	95.7	94.4	0.0	94.8	0.0	93.9			
		200	94.5	93.8	95.0	94.3	94.4	94.2	96.6	93.4	93.5	94.7	94.2	95.3	95.2	95.4	95.3	0.1	94.8	0.0	94.0			
		500	95.7	93.5	94.4	96.5	91.2	95.1	96.6	95.0	95.3	95.5	93.0	95.4	96.5	93.2	95.0	37.8	93.6	6.9	94.8			
		1000	94.9	93.8	94.8	94.3	93.5	91.9	94.9	93.7	93.3	94.8	94.2	94.3	95.2	94.1	95.6	59.0	93.5	49.6	94.9			
		5000	93.1	93.9	95.4	95.1	94.5	93.8	94.2	94.1	93.9	94.3	93.6	94.2	94.8	95.1	95.1	91.3	94.2	93.4	92.3			
	200	100	97.4	93.5	94.4	96.4	93.4	94.2	97.5	95.4	93.5	97.3	95.5	94.5	97.3	94.0	93.9	0.0	95.4	0.0	94.5			
		200	93.7	93.9	94.6	94.8	93.2	94.3	95.1	94.6	95.2	93.4	96.1	94.6	93.0	94.5	94.6	4.7	95.0	0.0	94.7			
		500	96.6	93.5	92.9	95.1	93.5	93.1	96.2	96.2	94.8	95.8	95.0	93.6	96.2	94.3	94.6	54.1	94.9	27.8	93.4			
		1000	96.6	94.0	94.2	93.9	94.8	94.3	96.6	95.3	93.8	94.8	94.3	93.4	96.7	96.2	94.4	76.3	95.1	69.2	95.5			
		5000	94.0	95.1	93.7	93.1	95.3	93.6	93.9	95.3	94.5	94.9	95.5	95.4	94.6	94.6	94.5	93.8	93.9	94.9	93.9			
500	100	88.1	93.1	94.1	96.3	93.4	94.6	95.6	95.1	93.8	96.5	95.5	94.6	95.5	94.7	93.8	0.0	93.7	0.0	93.7				
	200	95.2	93.5	93.5	95.2	94.9	95.8	95.5	94.3	93.9	95.6	96.1	94.0	95.9	94.3	94.0	37.7	95.0	5.9	94.0				
	500	93.1	94.2	93.0	95.4	94.9	95.5	94.9	93.9	93.8	95.4	95.2	94.2	94.7	95.0	95.0	69.5	94.2	54.6	94.7				
	1000	95.1	94.4	94.3	96.0	93.5	94.0	93.5	94.6	94.9	95.7	94.5	95.3	95.1	93.7	95.4	89.1	93.7	87.0	94.6				
	5000	93.8	94.6	94.8	94.5	95.2	93.7	95.3	94.8	95.4	95.5	94.6	94.5	94.6	95.4	93.1	93.8	94.0	94.2	94.3				

Figure II.2: Estimation de l'effet de contexte en fonction du nombre de clusters et d'individus - Design équilibré et variance fixée à 0.05

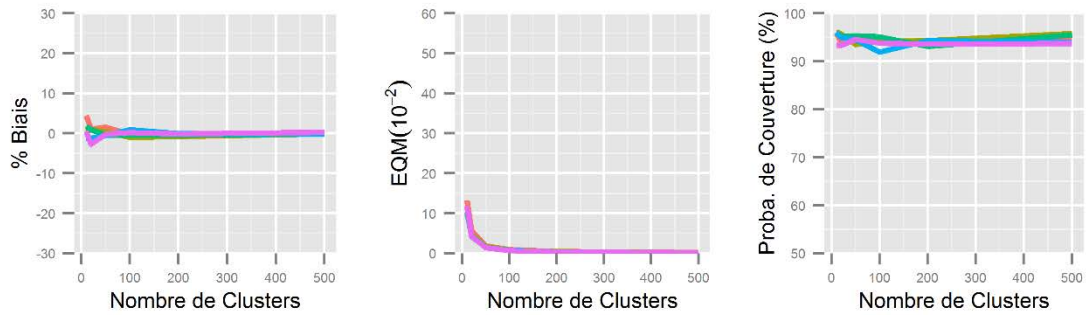
lme4



lmla



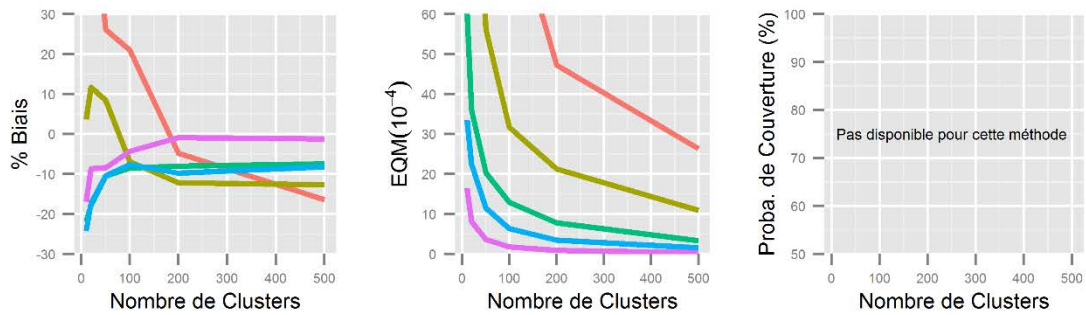
Jags



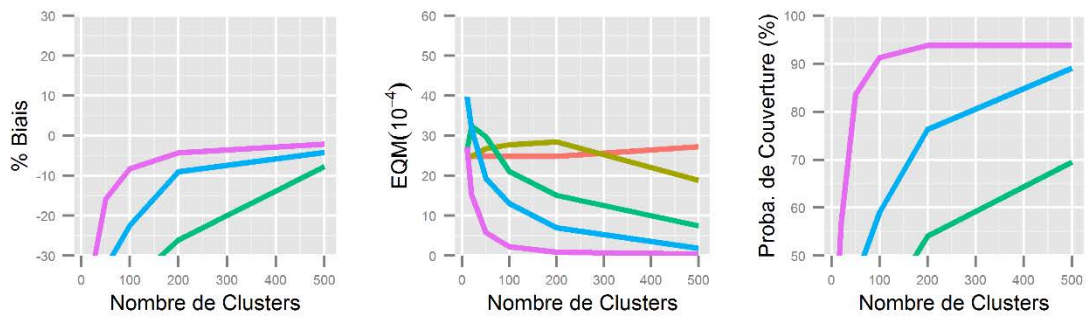
Individus 100 200 500 1000 5000

Figure II.3: Estimation de la variance spatiale en fonction du nombre de clusters et d'individus  
 - Design équilibré et variance fixée à 0.05

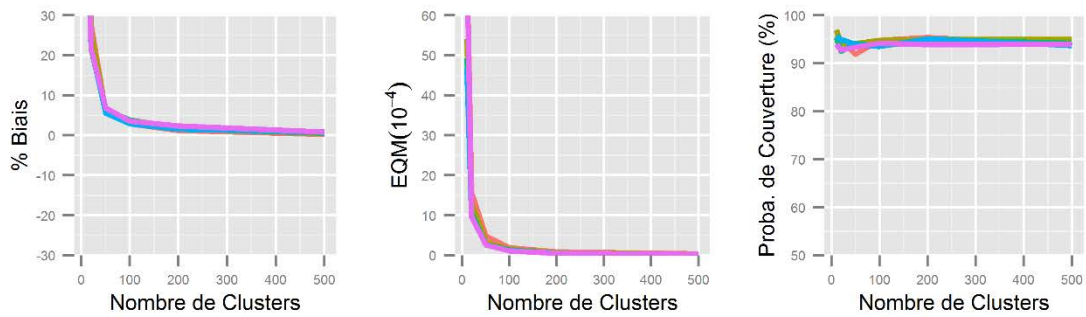
lme4



Inla



Jags



Individus 100 200 500 1000 5000

Tableau II.5: Biais (%) en fonction de la prévalence de l'évènement - Design déséquilibré EDP et variance fixée à 0.01

	Echelle	Prévalence	Constante $\alpha$			Contexte $\beta_2$			Indiv $\beta_1^2$			Indiv $\beta_1^3$			Indiv $\beta_1^4$			Var. spatiale $\sigma_{\text{spat}}^2$			Surdispersion $\sigma_{\text{surd}}^2$			
			lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	
% Bias	Région	0.1	0.6	2.1	4.5	1.0	0.3	2.3	1.3	2.2	2.0	0.3	0.5	3.1	1.3	1.8	2.1	-2.0	-77.8	82.1	-13.4	4.0	3.7	
		0.2	-2.0	-0.1	2.1	-0.3	1.5	3.2	-0.6	-1.4	1.9	-0.8	0.4	1.6	-0.5	0.1	0.7	-1.5	-81.5	62.9	-9.8	6.2	6.2	
		0.5	-0.8	0.6	2.2	2.0	-1.4	-2.7	-1.3	0.8	1.9	-0.6	0.7	0.8	-0.8	0.4	1.1	-11.0	-76.5	46.7	-7.2	7.2	4.8	
		1	-1.1	-0.6	-0.2	3.3	-1.8	0.4	-0.9	0.5	-0.7	-0.9	-1.3	0.3	-0.5	-0.7	0.0	-9.5	-76.3	40.6	-7.2	8.1	4.2	
		2	-0.1	-1.0	-0.1	2.6	-1.1	-0.3	0.0	-1.9	-0.9	-0.1	-0.9	0.3	-0.1	-0.5	-0.4	-13.7	-77.2	40.1	-5.7	7.3	4.9	
	Département	0.1	0.2	0.8	3.1	4.5	-2.6	2.0	-0.7	0.1	-0.2	-0.6	0.9	1.3	0.2	0.9	0.9	-2.4	-71.9	52.4	-3.7	1.7	-1.3	
		0.2	-1.1	-0.4	1.5	-1.0	-0.3	-0.5	-0.3	-1.2	0.6	-0.1	-0.7	0.8	-0.5	-0.5	0.6	-5.7	-70.1	15.2	-3.4	3.7	2.2	
		0.5	-0.5	1.1	1.5	0.3	0.4	0.1	0.8	0.4	0.7	0.5	1.3	1.1	0.1	0.6	0.4	-6.8	-64.2	6.2	-1.6	5.9	2.8	
		1	-0.4	0.4	-0.5	-0.7	1.4	-1.9	-0.3	-0.2	-0.8	-0.1	0.1	0.1	0.1	0.1	0.1	0.0	-6.2	-50.8	1.1	-1.8	4.6	1.9
		2	0.4	0.5	0.8	-2.4	-0.4	-0.3	0.2	0.4	1.1	0.4	0.2	0.8	0.4	0.3	0.5	-8.7	-51.3	-0.1	-1.2	5.5	2.0	
	Zone d'emploi	0.1	1.3	-0.6	2.0	0.0	1.0	0.3	0.0	1.1	-0.4	0.4	0.8	-0.5	0.5	0.6	0.4	15.6	-50.2	60.4	3.9	-15.5	-4.8	
		0.2	0.9	-0.1	0.8	3.9	-0.1	0.7	0.2	-0.3	0.4	0.2	0.3	0.6	0.3	-0.1	0.1	3.0	-70.9	15.1	1.2	4.2	-0.7	
		0.5	0.5	0.2	-0.2	-2.4	1.2	0.7	0.8	0.4	-0.1	0.4	0.1	-0.4	0.0	0.1	-0.1	-6.7	-65.7	-1.3	0.5	6.8	1.3	
		1	-0.1	0.0	0.2	1.9	-0.1	-1.0	0.0	0.3	-0.2	-0.2	-0.1	0.1	0.1	0.1	0.2	-6.2	-52.8	-4.1	-0.7	5.7	1.6	
		2	-0.3	0.3	-0.1	0.0	0.1	-0.2	-0.3	-0.1	-0.2	-0.4	0.4	-0.1	-0.3	0.2	-0.1	-3.4	-31.4	-3.3	-0.4	3.3	1.4	

Tableau II.6: EQM ( $10^{-3}$ ) en fonction de la prévalence de l'évènement - Design déséquilibré EDP et variance fixée à 0.01

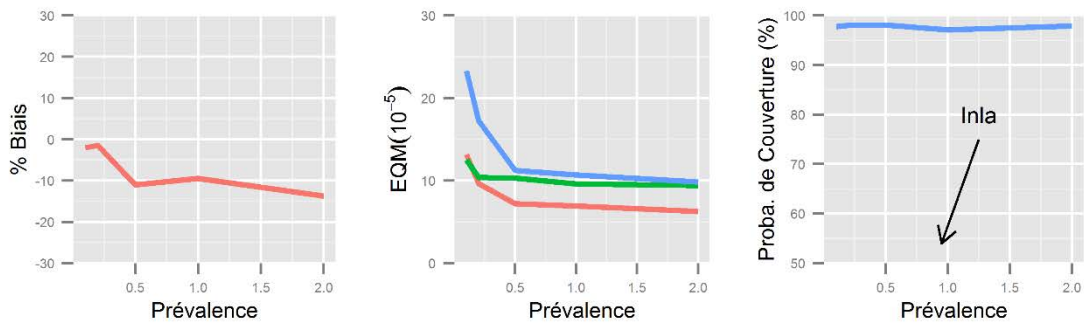
	Echelle	Prévalence	Constante $\alpha$			Contexte $\beta_2$			Indiv $\beta_1^2$			Indiv $\beta_1^3$			Indiv $\beta_1^4$			Var. spatiale $\sigma_{\text{spat}}^2$			Surdispersion $\sigma_{\text{surd}}^2$		
			lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags
Ecart Quadratique Moyen ( $10^{-2}$ )	Région	0.1	1.12	1.11	1.12	31.82	31.21	32.31	1.71	1.73	1.86	1.58	1.45	1.48	1.46	1.55	1.61	0.01	0.01	0.02	0.05	0.06	0.05
		0.2	0.78	0.73	0.74	27.14	29.04	25.97	1.31	1.27	1.14	1.16	1.13	1.15	1.12	1.04	1.12	0.01	0.01	0.02	0.03	0.04	0.03
		0.5	0.56	0.54	0.55	20.86	20.95	21.92	0.93	0.95	0.86	0.92	0.85	0.82	0.84	0.83	0.90	0.01	0.01	0.01	0.02	0.03	0.02
		1	0.44	0.44	0.48	21.00	20.46	19.57	0.81	0.76	0.80	0.75	0.69	0.78	0.75	0.74	0.73	0.01	0.01	0.01	0.02	0.02	0.02
		2	0.42	0.44	0.41	17.30	19.45	19.00	0.71	0.77	0.77	0.72	0.67	0.71	0.72	0.74	0.70	0.01	0.01	0.01	0.02	0.02	0.02
	Département	0.1	0.65	0.69	0.75	18.75	19.44	20.79	1.05	1.11	1.16	0.88	0.92	0.93	0.88	0.90	0.91	0.01	0.01	0.01	0.04	0.05	0.05
		0.2	0.43	0.41	0.42	12.79	12.50	13.16	0.70	0.72	0.64	0.56	0.59	0.55	0.55	0.54	0.55	0.01	0.01	0.01	0.02	0.02	0.02
		0.5	0.23	0.23	0.24	9.24	8.92	9.46	0.37	0.37	0.38	0.33	0.32	0.35	0.33	0.33	0.31	0.00	0.01	0.00	0.01	0.01	0.01
		1	0.15	0.17	0.16	7.92	7.63	8.04	0.26	0.29	0.29	0.22	0.25	0.25	0.24	0.25	0.23	0.00	0.01	0.00	0.01	0.01	0.01
		2	0.13	0.13	0.14	6.95	7.69	7.38	0.22	0.22	0.23	0.19	0.19	0.21	0.20	0.19	0.20	0.00	0.01	0.00	0.01	0.01	0.01
	Zone d'emploi	0.1	0.62	0.62	0.68	13.49	13.88	15.33	0.96	0.95	1.00	0.80	0.78	0.88	0.81	0.74	0.82	0.02	0.02	0.01	0.09	0.13	0.06
		0.2	0.33	0.36	0.34	9.00	8.87	8.75	0.51	0.60	0.51	0.42	0.46	0.47	0.41	0.44	0.45	0.01	0.01	0.00	0.03	0.03	0.02
		0.5	0.19	0.17	0.16	5.12	4.93	5.04	0.28	0.28	0.25	0.23	0.23	0.23	0.23	0.22	0.20	0.00	0.01	0.00	0.01	0.01	0.01
		1	0.11	0.10	0.11	3.44	3.54	3.63	0.17	0.17	0.17	0.15	0.14	0.15	0.14	0.14	0.14	0.00	0.01	0.00	0.01	0.01	0.01
		2	0.07	0.07	0.07	2.48	2.65	2.55	0.12	0.12	0.11	0.10	0.10	0.10	0.10	0.10	0.09	0.00	0.00	0.00	0.00	0.00	0.00

Tableau II.7: Probabilité de couverture (%) en fonction de la prévalence de l'évènement - Design déséquilibré EDP et variance fixée à 0.01

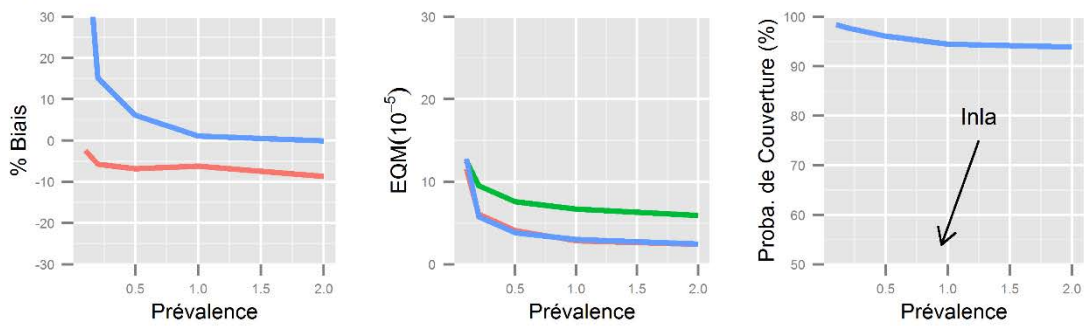
	Echelle	Prévalence	Constante $\alpha$			Contexte $\beta_2$			Indiv $\beta_1^2$			Indiv $\beta_1^3$			Indiv $\beta_1^4$			Var. spatiale $\sigma_{\text{spat}}^2$			Surdispersion $\sigma_{\text{surd}}^2$		
			lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags	lme4	Inla	Jags
Probabilité de Couverture (%)	Région	0.1	93.9	95.4	94.1	93.9	93.9	96	94.1	95.3	93.4	93.5	95.5	94.4	93.8	95.1	93.6	0.0	7.4	97.8	0.0	91.4	95.1
		0.2	93.5	95.2	95.2	93.3	91.1	94.8	94.2	94	95.8	93	95.4	94	94.7	95.9	94.8	0.0	9.3	98	0.0	90.3	95.1
		0.5	94.3	95	94.2	93.6	92.4	95.3	93.6	95.2	94.9	92.8	95	95.3	94.5	95.2	93.7	0.0	10.9	98.1	0.0	90.8	95.7
		1	94.5	94.9	94.5	93	91.1	96.2	93.5	95.3	93.9	93	96.1	94	94.5	95.2	95.5	0.0	13.2	97.1	0.0	90.7	95.2
		2	94.1	92.9	95.5	93.9	91.3	96	94.2	94	92.8	93.7	96.2	95	94.2	94.7	94.9	0.0	13.9	97.9	0.0	91.8	94.1
	Département	0.1	94.3	94.7	92.8	93.2	93.8	93.1	94.3	95.2	94	94	95.5	94.4	92.9	94.6	93.8	0.0	7.6	98.4	0.0	90.8	94.1
		0.2	94.4	95	94.6	94.6	94.9	94.8	93.9	94.6	95.3	94.5	94.4	95.4	94.6	95.1	93.5	0.0	16.4	97.7	0.0	92.6	94.9
		0.5	95.4	95.2	94.5	94.7	94.3	94.9	95.1	95.3	94.2	95.3	96.2	94.2	95.9	95.2	95.3	0.0	28.6	96.1	0.0	88.3	94.1
		1	95.2	94.1	95.3	93.2	94.4	94.7	95.8	95.1	94	96.1	95.4	94.7	94.3	94.9	95.7	0.0	39	94.5	0.0	87.6	94.2
		2	94.3	94	94.5	94	93	93.4	94.2	95	93.7	94.7	95.5	93.6	95.3	95.7	94.3	0.0	42.4	93.9	0.0	86.5	95
	Zone d'emploi	0.1	84.4	95.2	95	85.8	95	94.3	85	95.5	95.3	85.3	95.5	93.9	83.6	95.6	94.2	0.0	7.7	99.2	0.0	76.7	94.8
		0.2	90.7	94.5	95.3	89.1	93.7	94.3	90.6	95.3	95.7	90.7	95.8	94.1	90	94.9	94.5	0.0	12.3	98.6	0.0	89.7	95.1
		0.5	94.1	94.2	95.7	94.8	93.9	93.6	93.9	94.5	95.6	94.3	95.6	94.8	93.7	94.5	95	0.0	25	95.5	0.0	85.9	95.3
		1	94.5	94.8	95.1	94.7	94.2	93	94.6	94.9	94.2	94.1	95.2	94.3	95.6	95	94.1	0.0	41.1	94.5	0.0	81.4	94.7
		2	94.7	95.1	94.1	96.2	93.9	95.6	93.8	95.7	95.4	95.5	96.2	94.9	94.5	95.5	95.1	0.0	64	92.6	0.0	82.5	93.6

Figure II.4: Estimation de la variance spatiale en fonction de la prévalence de l'évènement - Design déséquilibré EDP et variance fixée à 0.01

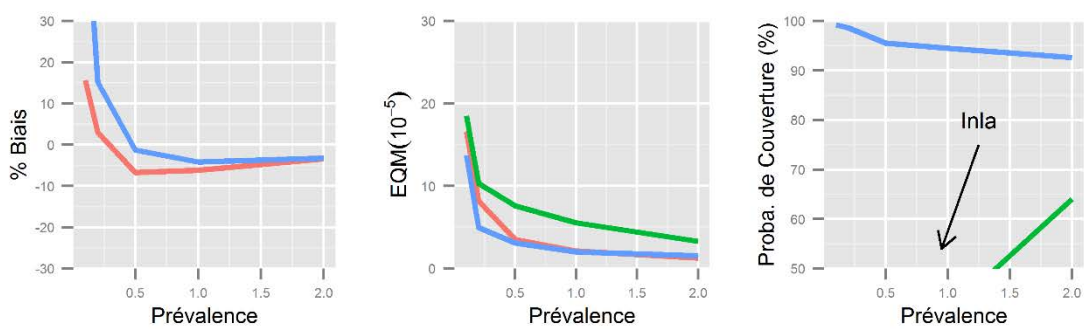
Région



Département



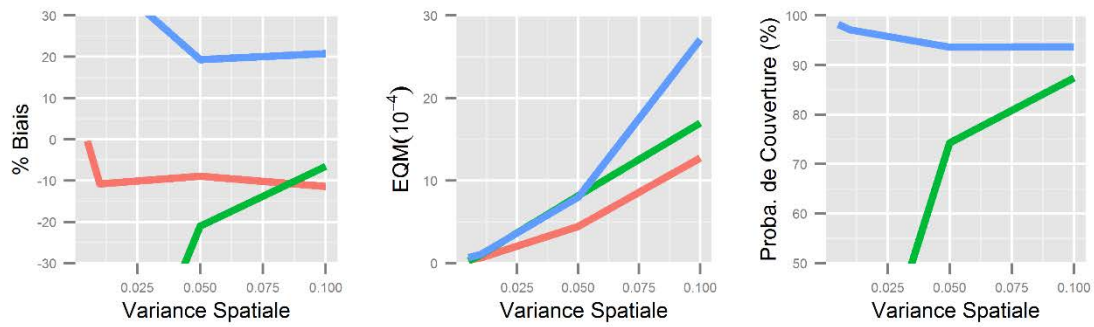
Zone d'Emploi



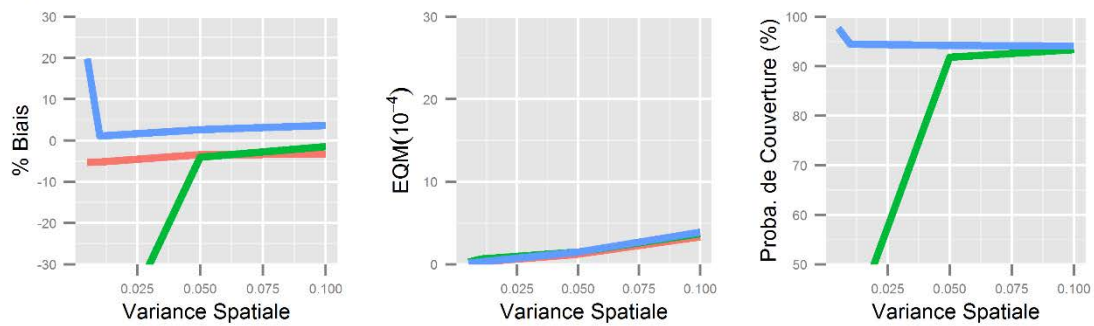
Methode — lme4 — Inla — Jags

Figure II.5: Estimation de la variance spatiale en fonction de son amplitude - Design déséquilibré EDP et prévalence fixée à 0.01

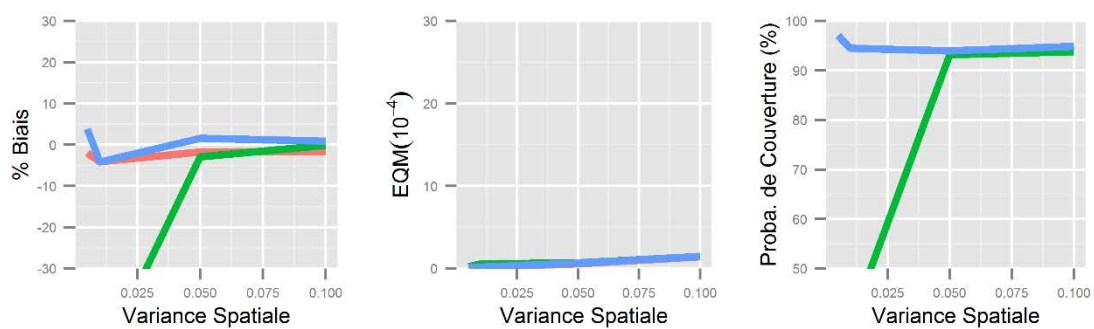
Région



Département



Zone d'Emploi



Methode — lme4 — Inla — Jags



## **Discussion**

### **Synthèse des résultats**

La mesure de concordance EDP-CépiDc des données de mortalité a indiqué une faible cohérence pour les périodes 1975-1981 et 1982-1988. L'estimation des paramètres du modèle de Poisson à constante aléatoire sont sensibles à la méthode d'estimation, au nombre de clusters considérés, au nombre d'i.p.c ou à la prévalence de l'événement et l'amplitude de la variance inter-clusters (variance spatiale par exemple). Dans tous les scénarios, les estimations des effets aléatoires étaient beaucoup plus sensibles aux données éparées que l'estimation des effets fixes. Les méthodes lme4 et Jags ont obtenu une estimation fiable des paramètres aléatoires et fixes pour une prévalence égale à la mortalité toutes causes. Pour ces méthodes, le design minimal requis est de 50 clusters avec un minimum de 500 individus dans chaque groupe tandis qu'Inla requiert un minimum de 100 clusters, un minimum de 5000 individus et une variance spatiale supérieure à 0,05. Dans tous les cas, l'augmentation du nombre de clusters a un impact plus grand sur l'amélioration des estimations que le nombre d'individus, même si elle est également restée un facteur d'amélioration. Dans l'analyse multiniveaux sur l'EDP à l'échelle régionale, lme4 exhibe de bons résultats dans l'estimation de tous les paramètres. Cependant, il lui fallait une prévalence minimale représentant la moitié de la prévalence de la mortalité toutes causes. A l'échelle du département, Jags et lme4 ont obtenu des estimations fiables pour une prévalence cinq fois plus faible que la prévalence de la mortalité toutes causes confondues. A l'échelle de la Zone d'Emploi, lme4 requiert une prévalence minimale de la moitié de la mortalité toutes causes confondues et Jags n'obtient des résultats pertinents qu'à partir d'une cause représentant au moins un cinquième de la mortalité totale.

### **Implications**

Les résultats de cette étude mettent en évidence la difficulté pour effectuer une analyse contextuelle des événements rares de santé. La faible prévalence de l'événement a un impact négatif sur le biais, la précision et la probabilité de couverture des estimations des paramètres. Dans le modèle investigué, le nombre de cas attendus dépend de la prévalence et de la population à risque : plus l'événement de santé est rare, plus la population doit être importante pour obtenir des estimations fiables. Un moyen de surmonter ce problème consiste à prolonger le temps de suivi. Par exemple, une analyse multiniveaux de la mortalité par cancer à l'échelle de la Zone d'Emploi aurait besoin d'une période de suivi deux fois plus longue. Une estimation fiable de la variance spatiale consiste également à bien mesurer la variabilité en excès due à la corrélation des individus vivant au sein des mêmes unités spatiales pour estimer la variance des effets de contexte.

### **Comparaison avec les études précédentes**

La fiabilité des estimations d'un modèle de Poisson à constante aléatoire n'a pas été investiguée jusqu'à présent. Par conséquent, les performances du modèle mixte logistique peuvent y être comparées. À ce jour, deux travaux ont rapporté que la prévalence de

l'évènement a une influence sur la fiabilité [130,131]. Elles sont beaucoup plus élevées que la prévalence de la mortalité (qui est d'environ 1% par an). Dans l'étude où la prévalence a été fixée à 10%, les auteurs recommandent un minimum de 100 clusters avec 50 i.p.c pour obtenir des estimations de paramètres fiables. Nos résultats sont en accord avec ces résultats car ils mettent en évidence la grande sensibilité des estimations des paramètres du modèle à constante aléatoire de Poisson aux données éparses et la nécessité de prendre plus de précaution lorsque l'évènement a une prévalence plus faible (ou une cause précise du décès). En outre, ils montrent que ces estimations dépendent fortement des méthodes statistiques utilisées. La décision d'exprimer cette caractéristique en termes de populations par cluster a été prise pour rendre notre travail comparable aux autres études.

### **Problème des échelles d'analyses**

D'un point de vue épidémiologique, diverses réalités géographiques et facteurs contextuels opèrent à différentes échelles. Cette étude portait uniquement sur deux niveaux : individuel et contextuel. Les résultats reflètent le compromis nécessaire entre tenir compte du contexte sur une large échelle géographique avec quelques clusters mais beaucoup d'individus dans chaque unité géographique, ou travailler sur une échelle plus fine avec beaucoup moins de personnes, mais de nombreuses unités. Dans leur article, Merlo et al. déclaraient que les entités collectives, ou limites géographiques pertinentes définissant le contexte, peuvent être identifiées en mesurant la corrélation intra-classe de l'indicateur de santé estimée aux échelles sociogéographiques a priori [132]. Ayant entamé ce travail dans le premier chapitre, nous avons tenté d'évaluer la fiabilité des estimations des paramètres et d'identifier à quelles échelles la variance spatiale de la mortalité pourrait être mesurée sur un échantillon de la population. Une autre conséquence de la modélisation des disparités spatiales sur une échelle trop large est que certaines informations disponibles à une échelle plus fine seront certainement diluées [120]. La perte d'information due à cette agrégation contrainte peut conduire à la non-identification de certains facteurs de risque contextuels plus proximaux. Dans ce travail, l'espace a été considéré comme étant fragmenté, ignorant ainsi la continuité spatiale des effets de contexte qui définit probablement mieux le contexte de vie des individus [133].

### **Problèmes de modélisation**

Le modèle à constante aléatoire de Poisson est assez simpliste pour décrire l'association complexe entre les caractéristiques individuelles, les caractéristiques contextuelles et des événements rares de santé. Cependant, en plus de permettre une analyse contextuelle, il permet de poser certaines questions importantes telles que : dans quelle mesure les disparités spatiales de santé résultent de l'agrégation spatiale des mêmes groupes sociaux ? (effet de composition). Il est fort probable que les individus ne sont pas égaux face aux effets contextuels. Cette interaction potentielle peut être modélisée à l'aide d'une pente aléatoire dont la fiabilité des estimations n'a pas été investiguée dans cette partie. En outre, le modèle à constante aléatoire de Poisson suppose que la distribution spatiale suive une distribution gaussienne, ce qui n'est pas nécessairement le cas et pourrait avoir des répercussions négatives sur la fiabilité des estimations [134].

### **Problèmes des données**

Une étude antérieure a montré que la présence de clusters peu fournis a une influence très limitée sur le biais et l'erreur standard des estimations de paramètres lorsque le nombre de clusters est élevé [130], bien que la probabilité de couverture de ces estimations n'ait pas été évaluée. Dans de futurs travaux, il sera nécessaire de mesurer l'impact du déséquilibre de répartition des individus entre les clusters. Dans la présente étude, les estimations de la variance spatiale obtenue sur la population réelle ont été biaisées lorsque la population était petite (ou lorsque la prévalence était faible). Par conséquent, la question du taux d'échantillonnage de la population lorsque l'on veut mesurer l'ampleur des disparités spatiales est incontournable.

### **Problèmes de méthodes statistiques**

Trois méthodes statistiques ont été comparées pour estimer les paramètres d'un modèle à constante aléatoire de Poisson. Bien qu'elle ait fourni les estimations les moins biaisées dans les scénarios épars, la méthode de lme4 a également montré une précision légèrement plus faible dans ces situations. Les intervalles de confiance des effets fixes ont été estimés à l'aide d'une approximation normale. L'intervalle de confiance de la variance des effets aléatoires n'est pas fourni par cette méthode. Par conséquent, la probabilité de couverture n'a pas pu être évaluée pour ces paramètres. Néanmoins, lme4 inclut une méthode de bootstrap paramétrique pour résoudre ce problème. Il propose également d'améliorer les estimations en agissant sur les paramètres de la méthode d'approximation de la vraisemblance. Ces deux dernières options ne sont pas testées dans cette partie. Le temps de calcul aurait été trop important. La méthode Inla a obtenu des résultats fiables pour les effets fixes, mais les estimations des paramètres aléatoires dépendaient essentiellement de l'amplitude de la variance spatiale, une information qui est a priori inconnue.

La méthode Jags fournit des estimations fiables en général. En dépit de sa grande dépendance à l'égard des lois a priori [135], les estimations des effets aléatoires ont montré une probabilité de couverture acceptable quel que soit le scénario. En outre, les performances de cette méthode dépendent fortement du nombre d'échantillons qui a été fixé à 50 000 pour obtenir une estimation dans un délai raisonnable. Cependant, les diagnostics de convergence étaient acceptables en moyenne (résultats non communiqués). Il ne fait aucun doute qu'allonger la longueur de la chaîne de Markov permettrait d'améliorer la fiabilité des estimations. Enfin, bien qu'elle soit une mesure de fiabilité fréquentiste, la probabilité de couverture a été considérée sans distinction en utilisant l'intervalle de confiance et l'intervalle de crédibilité [136].

### **Conclusion**

Ce travail met en évidence l'impossibilité de procéder à une analyse multiniveaux des événements rares de santé sans précaution. Les estimations, en particulier des paramètres aléatoires sont très sensibles à tous les aspects du design et de la méthode statistique utilisée. Un choix éclairé de l'échelle contextuelle d'analyse et de l'incidence de l'événement (ou la cause du décès) à étudier est nécessaire. La mesure des disparités spatiales de santé doit être considérée avec prudence. La variance spatiale est le paramètre le plus sensible et les méthodes statistiques convergent facilement sur des estimations biaisées. Plusieurs travaux

ont proposé des recommandations de design minimal requis pour l'utilisation de modèles multiniveaux. À cette étape, les résultats de cette étude suggèrent qu'une analyse multiniveaux de la mortalité sur l'échantillon EDP serait fiable à l'échelle des départements et Zone d'Emploi (niveau 2) pour des causes de décès dont la prévalence représente au moins 20% de la mortalité toutes causes avec un temps de suivi de 6 ans minimum. Les résultats obtenus pour les designs équilibrés montrent qu'un minimum de 50 groupes et 500 i.p.c sont nécessaires pour étudier un événement de santé avec une prévalence fixée à 1%.

### III. ANALYSE MULTINIVEAUX DE LA MORTALITE PAR CAUSE – APPLICATION

#### Introduction

Les disparités géographiques de mortalité se sont accrues de manière générale en France ces dernières années [3,4]. Pour expliquer ces disparités, la théorie relative aux facteurs multiniveaux considère que la santé est aussi le produit de l'environnement social et matériel qu'elle mesure à travers l'effet de contexte [46,47]. La théorie *ecosocial* (en anglais) dresse les contours des effets concomitants sur la santé du niveau socio-économique individuel et des caractéristiques de l'environnement de résidence [47]. En effet, l'environnement social et physique d'un voisinage affecte le niveau de stress et offre des opportunités différenciées aux individus qui peuvent ainsi adopter certains comportements de santé [13,43]. Avant de se pencher sur le processus d'*embodiment* (en anglais) qui renvoie aux mécanismes par lesquelles l'environnement influe sur la santé des individus, l'épidémiologie sociale cherche à identifier les caractéristiques de l'espace associées à des conséquences délétères pour la santé [137]. La méthode communément utilisée pour mettre en évidence des facteurs de risques intervenant à plusieurs niveaux est le modèle multiniveaux car il permet de dénouer les effets individuels des effets de contexte en prenant en compte la corrélation inhérente aux individus vivant dans le même lieu [45,138]. Parmi les indicateurs de santé important pour mesurer la santé des populations, la mortalité est particulièrement adaptée aux études en population générale. Les décès et leur analyse par cause constituent des révélateurs synthétiques des inégalités de santé. Ils constituent en effet un événement final qui résulte d'une chaîne complexe d'évènements touchant aussi bien à l'incidence morbide, les rapports à la maladie, aux soins, la prise en charge par le système de soins, etc. [139].

Analyser l'effet du contexte de résidence requiert une utilisation éclairée des modèles multiniveaux. Le choix de l'échelle du contexte doit faire l'objet d'une investigation, aussi bien pour ne pas nuire à la fiabilité des estimations que pour tenir compte de la pertinence de considérer un découpage de l'espace pour définir un contexte de vie. L'ajustement sur le niveau socio-économique individuel est également une étape importante et sensible. Elle permet de retirer la part des disparités spatiales attribuable à la composition sociale de la population. Deuxièmement, dans la mesure où il est envisageable que le niveau socio-économique individuel soit médiateur de l'effet du contexte sur la santé humaine, il convient d'ajuster sur celui-ci afin de mesurer l'effet direct du contexte de manière appropriée. Cependant, le sens de la causalité entre niveau socio-économique individuel et contextuel est vraisemblablement double puisque des personnes favorisées peuvent avoir tendance à quitter des zones défavorisées. Pour mesurer la position socio-économique individuelle, le niveau d'études atteint à l'âge adulte constitue un choix pratique. Cette mesure possède un certain nombre d'avantages par rapport au niveau de revenu ou à la catégorie socio-professionnelle. Parmi ceux-ci figure le peu de variation de celui-ci au cours de la vie et le fait qu'il synthétise un ensemble de ressources intellectuelles et culturelles acquises qui auront une influence sur l'activité professionnelle, le niveau de revenu, les comportements de santé et les modes de vie [140,141]. Son utilisation limite également la causalité inverse où la santé influencerait sur le niveau socio-économique.

En utilisant une approche multiniveaux, l'association entre niveau socio-économique du lieu de résidence et mortalité a été étudiée dans de nombreux pays. Une méta-analyse a souligné qu'un faible niveau socio-économique d'une zone est associé à une surmortalité indépendamment du niveau socio-économique individuel [118]. Après ajustement sur celui-ci, une mortalité plus forte est observée dans les zones qui exhibent un désavantage social plus prononcé [142,143], un environnement social défavorable [144] et de bas revenus [145]. L'effet est en moyenne plus fort quand l'unité géographique considérée est plus fine [118]. En France, bien que les inégalités sociales de mortalité aient été rapportées comme les plus fortes d'Europe Occidentale [10,146], aucune étude n'a encore été conduite pour investiguer cette association en population générale sur un échantillon représentatif [142,143,147–155].

Une autre caractéristique de l'environnement qui pourrait avoir un impact sur la mortalité et qui a relativement été peu étudiée en France est l'accessibilité géographique aux soins. Une longue distance à parcourir pourrait constituer un effet barrière, décourageant la consommation de soins [17]. Dans la littérature, la distance aux soins est associée avec le type de traitement et d'intervention médicale [156], négativement avec le diagnostic de l'hépatite C [157] et positivement avec le stade du cancer au moment du diagnostic [158]. Des études écologiques ont montré qu'une plus longue distance à parcourir jusqu'aux donneurs de soins est associée à une surmortalité par asthme [159], aux décès néonataux après un accouchement hors-hôpital [160], à une surmortalité des vétérans ayant eu un accident vasculaire cérébral [161], une surmortalité par cancer du poumon et cancer de la prostate [162], et enfin pour les cas urgents transportés par ambulance [163]. De manière générale, cette association était relativement faible. Certaines études ont par ailleurs rapporté une association inverse, ou une plus longue distance à parcourir était liée à une sous-mortalité par certaines causes. Ce dernier résultat révèle la possible présence de facteurs de confusion sous-jacents [160,164,165]. Dans ces études écologiques, l'effet de la distance aux soins était de manière générale très sensible à l'introduction des caractéristiques socio-économiques. Ainsi, elle ne prenait pas en compte le fait que les effets individuels et contextuels sur la mortalité sont additionnels et indépendants. Une seule étude a considéré la distance à l'hôpital comme un effet de contexte sur la survie après un cancer dans une analyse multiniveaux appropriée [166].

Cette dernière partie est une application d'une analyse multiniveaux de l'association entre mortalité par cause et deux caractéristiques du lieu de résidence effectuée sur un échantillon représentatif de la population Française. Les effets contextuels investigués sont le désavantage social du lieu de résidence pris comme une dimension de l'environnement social, et l'accessibilité géographique aux structures de soins considérée comme dimension de l'environnement physique. Dans la littérature, il n'existe à ce jour aucune vue d'ensemble de ces associations par cause de décès. Cette analyse a été effectuée sur la mortalité toutes causes et sur la mortalité par cause spécifique afin d'explorer les différentes voies à travers lesquelles le contexte peut affecter la mortalité. Une autre question qui reste à clarifier est de savoir dans quelle mesure les disparités spatiales de mortalité sont le reflet de la composition sociale ou des contextes de résidence différenciés.

## **Méthodes et sources**

### **Population**

Comme dans la partie II, la population est issue de l'échantillon EDP des individus ayant rempli leur bulletin de recensement de 1990, nés en France, ayant entre 30 et 79 ans au 1er

janvier 1990 et toujours vivant le 1er janvier 1991. Les caractéristiques individuelles et le lieu de résidence étaient mesurés au début de la période de suivi, c'est à dire au recensement de 1990. Le suivi prenait fin le 31 Décembre 2007. Les niveaux d'études étaient regroupés en quatre niveaux selon *la classification internationale des niveaux d'études* (ISCED) : enseignement élémentaire incomplet, Certificat d'Etude Primaire (CEP) (ISCED niveau 1), Diplôme Professionnel (BEP, CAP) et BEPC (ISCED niveau 2), et enfin Baccalauréat général et technique, incluant également les niveaux d'études supérieures (ISCED niveaux  $\geq 3$ ).

### **Causes de décès**

Elles ont été obtenues après un appariement avec la base du CépiDc. Cinq groupes de causes ont été investigués dans ce travail: les cardiopathies ischémiques, les maladies cérébrovasculaires et toutes tumeurs, les décès qui auraient pu être évités grâce à une prise en charge adéquate du système de soins (*Amenable diseases*) et les décès qui auraient pu être évités grâce à la prévention (*Preventable diseases*) [58]. Les codes CIM de ces regroupements par cause spécifique sont disponibles en Annexe (Tableau S1, Tableau S2).

### **Echelle géographique**

La Zone d'Emploi de 1994 (ZE94, 348 unités géographiques) a été considérée comme le niveau 2 du contexte de résidence pour deux raisons. D'après les analyses précédentes, il est possible d'y effectuer une analyse multiniveaux pour des causes assez fines puisque le suivi est environ 3 fois plus long que pour la population utilisée dans les simulations. De plus, cette échelle est construite à partir des déplacements quotidiens domicile-travail des individus y résidant (Insee). Elle reflète une forme de contexte où la majorité de la population active se meut quotidiennement.

### **Analyse statistique**

Les recommandations issues de la partie précédentes poussent vers un allongement de la période de suivi des individus. Cependant, la définition d'un modèle de Poisson n'est pas adaptée à un tel design (17 années de suivi). Pour analyser de telles données, la période de suivi doit être découpée en diverses observations dont la corrélation doit être prise en compte. La fiabilité d'un tel modèle à classification croisée pour analyser les événements rares n'a actuellement pas fait l'objet d'une évaluation. Cette situation nous oriente vers l'utilisation d'un modèle de survie qui prend en compte de façon optimale l'augmentation du risque de mortalité avec l'âge. Le modèle de fragilité de Cox a donc été utilisé pour estimer les associations entre mortalité et facteurs individuels et contextuels. Tous les modèles présentés sont ajustés sur des effets cohortes. Chaque effet regroupe les individus par années de naissances décennales. Ces estimations qui sont d'un intérêt limité et ne modifient pas les autres associations ne sont volontairement pas fournies afin d'améliorer la lisibilité des résultats. Toutes les estimations ont été effectuées à l'aide du package *coxme* du logiciel R avec un seuil de significativité fixé à 5%.

### **Données d'accessibilité géographique aux structures de soins**

Les temps de trajet aux spécialistes et spécialités hospitalières en 2007, données qui sont produites et diffusées par l'Institut de Recherche et Documentation en Economie de la Santé

(IRDES). Le temps de trajet a été estimé à l'aide du logiciel ChronoMap© sur la base des temps de trajet entre le centre de chaque commune avec le centre de la commune la plus proche où la structure de soins était disponible. Dans le cas où le dit spécialiste ou service hospitalier était présent dans la commune, le temps de trajet à cette structure était considéré comme nul. Des précisions sur la méthodologie ainsi qu'une étude descriptive sont disponibles dans le document de travail suivant [167].

### Distribution des Temps de Trajet aux Spécialistes et Spécialités Hospitalières

Tableau III.1: Distribution des temps de trajet aux spécialistes

Spécialistes	moyenne/pers.	Quartiles			maximum
		1er	2ème	3ème	
Omnipraticien	01'23	00'00	00'00	00'00	45'59
Infirmière	01'42	00'00	00'00	00'00	65'38
Dermatologue	11'29	00'00	09'09	19'49	115'57
Gastro-entérologue	13'42	00'00	12'17	22'42	120'56
Gynécologue	10'56	00'00	08'34	19'01	120'00
Dentiste	02'27	00'00	00'00	00'00	69'21
M.E.P	08'36	00'00	01'19	15'20	106'49
Ophtalmologue	09'49	00'00	07'04	17'22	115'57
Radiologue	09'59	00'00	07'31	17'22	106'31
Kinésithérapeute	10'13	00'00	08'14	17'22	110'08
Cardiologue	11'00	00'00	09'01	19'06	115'57
Psychiatre	12'28	00'00	10'19	21'12	115'57
Pédiatre	12'53	00'00	10'21	21'49	115'57
O.R.L	13'07	00'00	11'23	22'10	128'25
Rhumatologue	13'48	00'00	12'03	22'53	120'56
Pneumologue	16'17	00'00	14'44	25'59	128'25
Urologue	21'01	06'36	18'07	32'14	128'25

Les moyennes des temps de trajet par individu sont assez inégales selon la spécialité. Elles varient de 1 minute 30 pour les omnipraticiens à 21 minutes pour consulter un urologue. Cependant la majorité des spécialistes se situent à un temps d'accès moyen entre 10 et 14 minutes environ. Un quart des individus habitent une commune où toutes ces spécialités sont présentes mis à part pour les urologues. Les trois quarts des personnes résident dans une commune où au moins un omnipraticien, un infirmier ou un dentiste réside. Un quart des individus résident à plus de 15 minutes d'un dermatologue, gastro-entérologue, gynécologue, médecin à exercice particulier, ophtalmologue, radiologue, kinésithérapeute, cardiologue, psychiatre, pédiatre O.R.L, rhumatologue, pneumologue ou urologue. Certaines communes les plus isolées se situent entre 1 heure à plus de 2 heures de trajet pour accéder à un spécialiste.



Tableau III.2: Distribution des temps de trajet aux spécialités hospitalières

Spécialités hospitalières	moyenne/pers.	Quartiles			maximum
		1er	2ème	3ème	
Chirurgie orthopédique	14'03	00'00	13'01	22'52	115'57
Endoscopie	14'24	00'00	13'20	23'25	121'39
Maternité06	15'24	00'00	14'25	24'29	128'25
HGE	15'40	00'00	14'20	24'34	120'56
Cardiologie	15'50	00'00	14'52	24'45	128'25
Néphrologie	16'31	00'00	15'03	25'32	120'56
Chirurgie digestive	16'53	00'00	14'59	26'21	122'51
Ophtalmologie	17'00	00'00	15'25	26'35	128'25
Service Pneumologie	17'02	00'00	15'31	26'18	128'25
Neurologie	18'25	05'13	16'35	27'58	128'25
Rhumatologie	19'54	06'46	17'50	29'44	128'25
Urologie	19'57	04'43	17'53	30'45	128'25
Endocrinologie	20'02	06'27	17'42	30'03	137'04
Hématologie	21'13	07'36	18'42	32'01	128'25
Maternité niv 1	23'43	10'23	21'13	35'10	128'25
Dermatologie	24'38	10'20	21'08	35'54	196'17
Maternité niv 2	25'17	09'58	22'26	37'17	146'31
O.R.L	25'39	10'43	23'36	38'13	151'46
Chirurgie vasculaire	29'15	11'40	25'36	43'26	196'17
Chirurgie thoracique	31'42	12'21	26'06	46'37	196'17
Maternité niv 3	40'11	16'43	34'47	59'34	196'17
Chirurgie cardiaque	45'36	17'44	40'10	67'51	219'08
Neurochirurgie	46'10	18'40	41'26	68'54	204'01
Grands brûlés	74'16	30'39	69'49	108'18	303'08

La majorité des spécialités hospitalières sont accessibles en moyenne pour chaque personne entre 14 minutes pour la chirurgie orthopédique et 32 minutes pour la chirurgie thoracique. L'accès à une maternité de niveau 3, la chirurgie cardiaque et la neurochirurgie est en moyenne un peu plus long, entre 40 et 46 minutes, alors qu'il est de 74 minutes vers un service grands brûlés. Un quart des individus habitent une commune où un service de chirurgie orthopédique, service d'endoscopie, maternité06, service d'hépto-gastro-entérologie, service de cardiologie, service de néphrologie, chirurgie digestive, service d'ophtalmologie ou service de pneumologie est présent. Un quart des individus habitent à moins de 30 minutes de toutes ces spécialités. La moitié des individus se situent maximum entre 13 minutes et 41 minutes de toutes les spécialités, mis à part pour le service grands brûlés situé à plus d'une heure de trajet. Un quart des individus vivent à plus de 20 minutes de toutes les spécialités hospitalières voire à plus d'une heure pour accéder à une maternité de niveau 3, chirurgie cardiaque, neurochirurgie et service de grands brûlés. Les communes les plus isolées se situent à plus de 2 heures de trajets de toutes ces spécialités.

### **Corrélations spatiales des temps de trajet**

Il existe une forte corrélation spatiale entre les temps de trajet et les différentes structures de soins mesurées en 2007, que ce soit pour la médecine de ville ou pour les spécialités hospitalières. Sur la Figure III.1 sont représentés les coefficients de corrélation des temps de trajet vers chacune de ces ressources pondérés par les effectifs de population.

Cinq groupes d'accessibilité géographique aux structures de soins se dessinent.

Le groupe 1 contient kinésithérapeute, médecin à exercice particulier, radiologue, ophtalmologue, dermatologue, cardiologue, gynécologue, pédiatre, pneumologue, rhumatologue, O.R.L, gastro-entérologue, service de chirurgie digestive, HGE, service de néphrologie, service de cardiologie, maternité 06, service de chirurgie orthopédique et service d'endoscopie. Cette catégorie regroupe indistinctement spécialités hospitalières et médecines de ville.

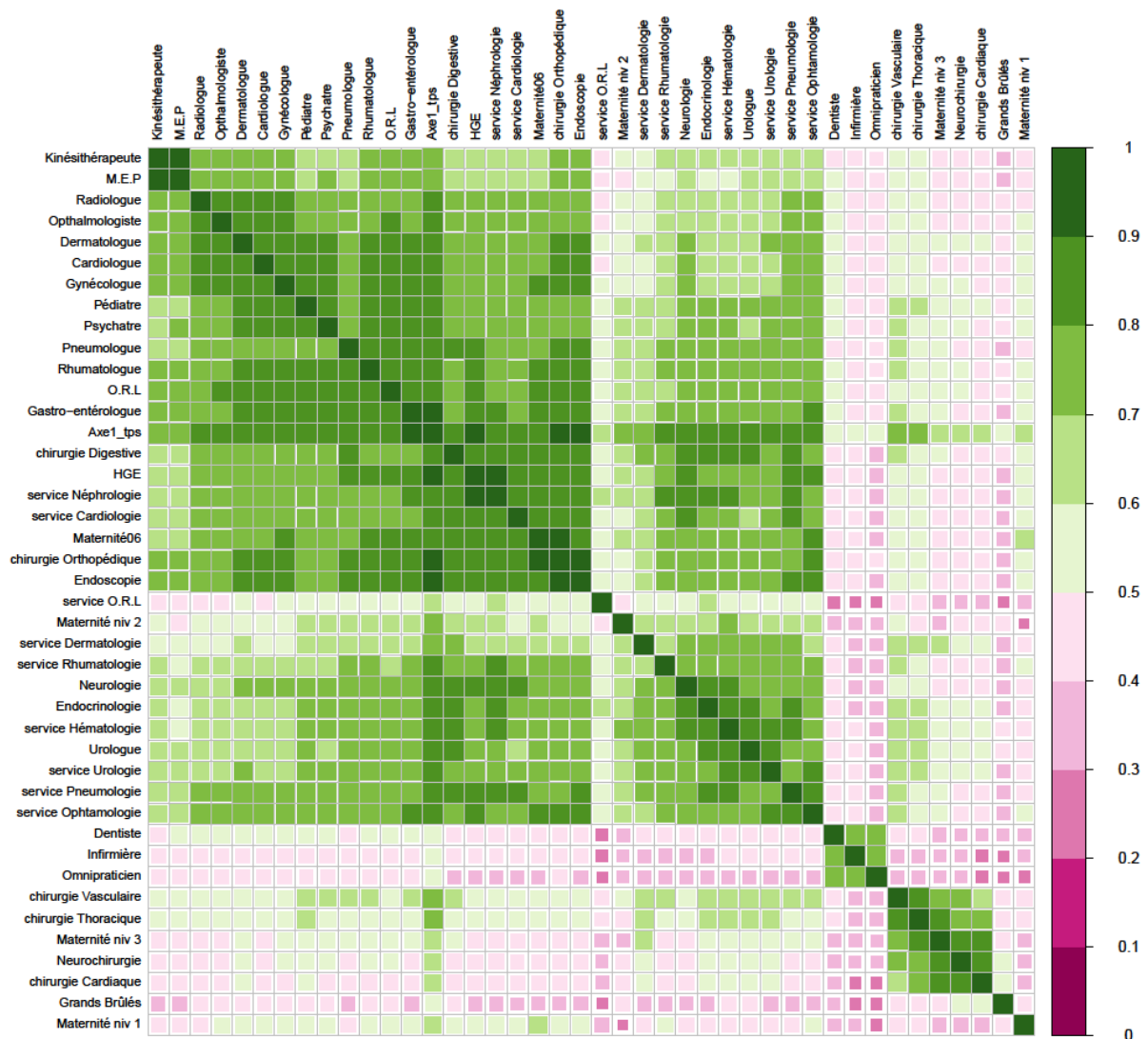
Le groupe 2 contient maternité de niveau 2, service de dermatologie, service de neurologie, service d'endocrinologie, service d'hématologie, urologue, service d'urologie, service de pneumologie et service d'ophtalmologie. Cette catégorie regroupe principalement des services hospitaliers et reste tout de même corrélée spatialement avec le groupe 1 ( $\geq 0.6$ ).

Le groupe 3 contient dentiste, infirmière et omnipraticien. Cette catégorie regroupe les soins de proximité. Il est peu corrélé spatialement avec les autres structures ( $\leq 0.4$ ).

Le groupe 4 contient chirurgie vasculaire, chirurgie thoracique, maternité de niveau 3, service de neurochirurgie et chirurgie cardiaque. Cette catégorie est peu corrélée avec les groupe 1 et 2 ( $\approx 0.5$ ).

Le groupe 5 contient le service O.R.L, le service de grands brûlés et la maternité de niveau 1. Ces trois spécialités ont la particularité d'être très peu corrélées entre elles et avec toutes les autres mis à part pour le service O.R.L peu corrélé avec les groupe 1 et 2.

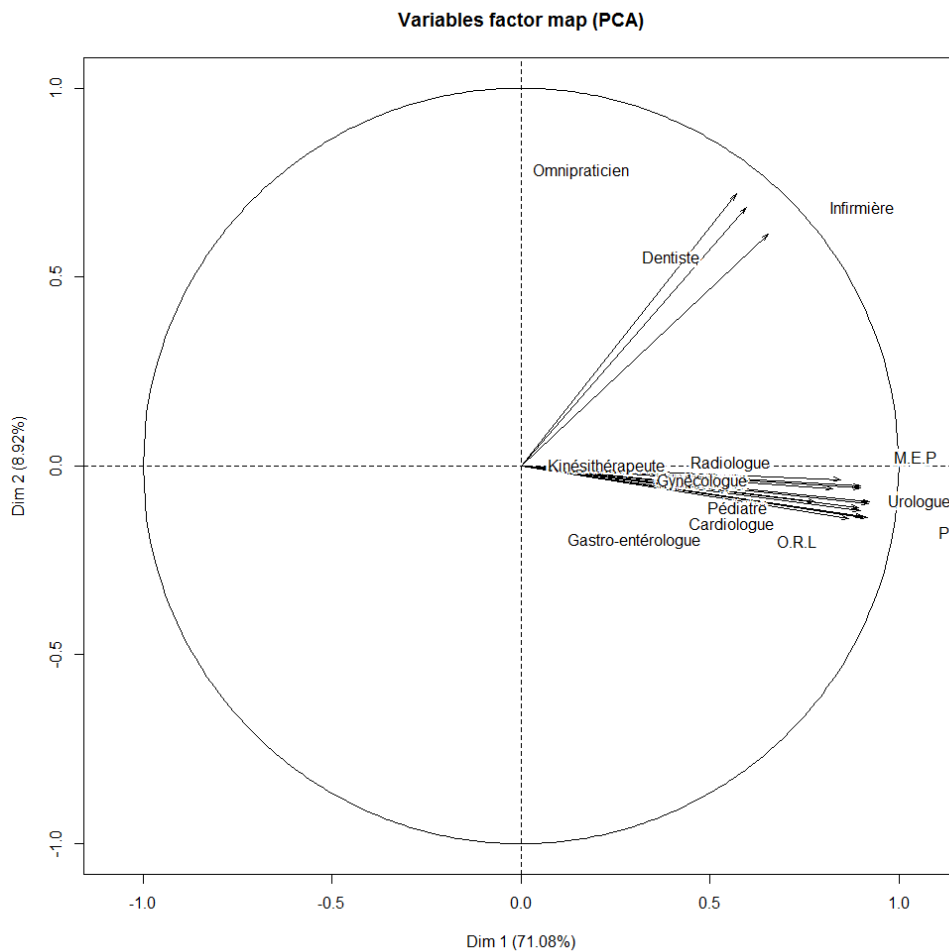
Figure III.1: Corrélations spatiales entre les temps de trajet intercommunaux vers les différentes structures de soins en 2007



### Analyse en Composante Principale des temps de trajet aux structures de soins

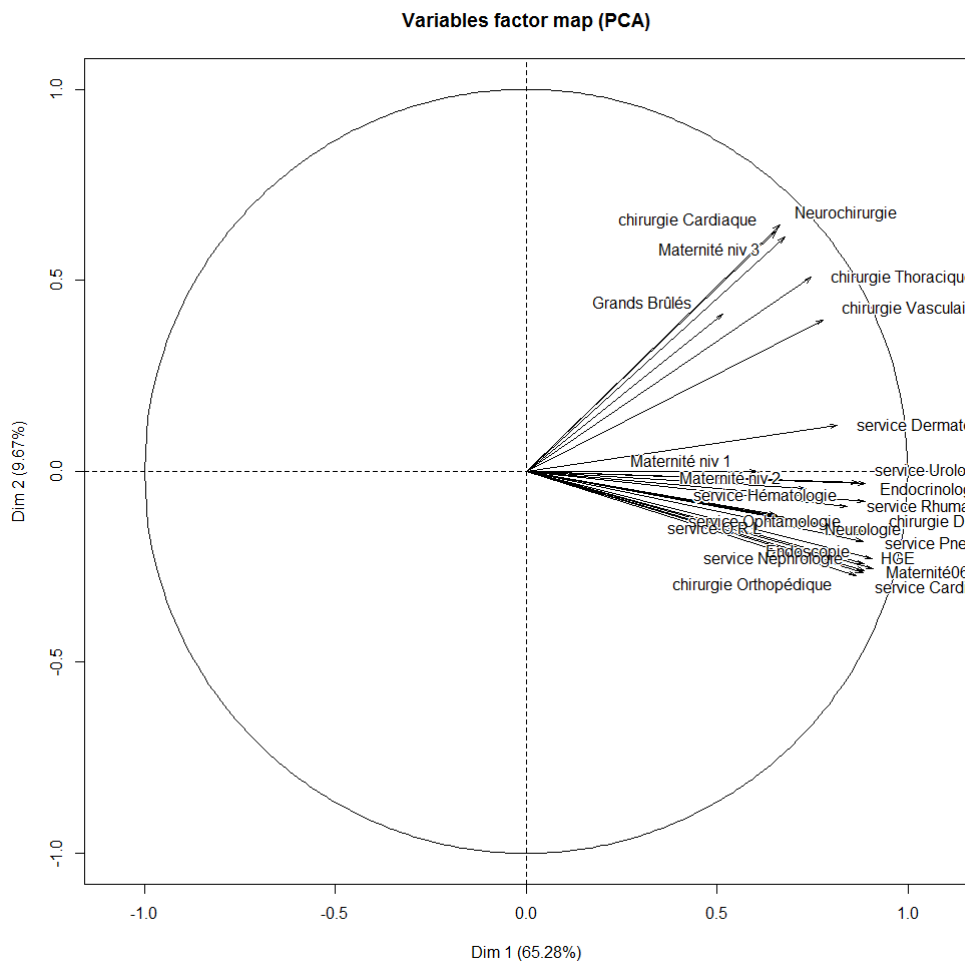
Les corrélations géographiques des temps de trajet, de manière générale assez élevées, poussent à réduire le nombre de variables, afin de rendre l'information moins redondante et de la généraliser. Pour ce faire, trois analyses en composantes principales pondérées par la population ont été effectuées. La première concerne les temps de trajet vers les spécialistes, la seconde les spécialités hospitalières et la troisième tous les spécialistes et spécialités hospitalières.

Figure III.2: Analyse en composantes principales des temps de trajets aux spécialistes



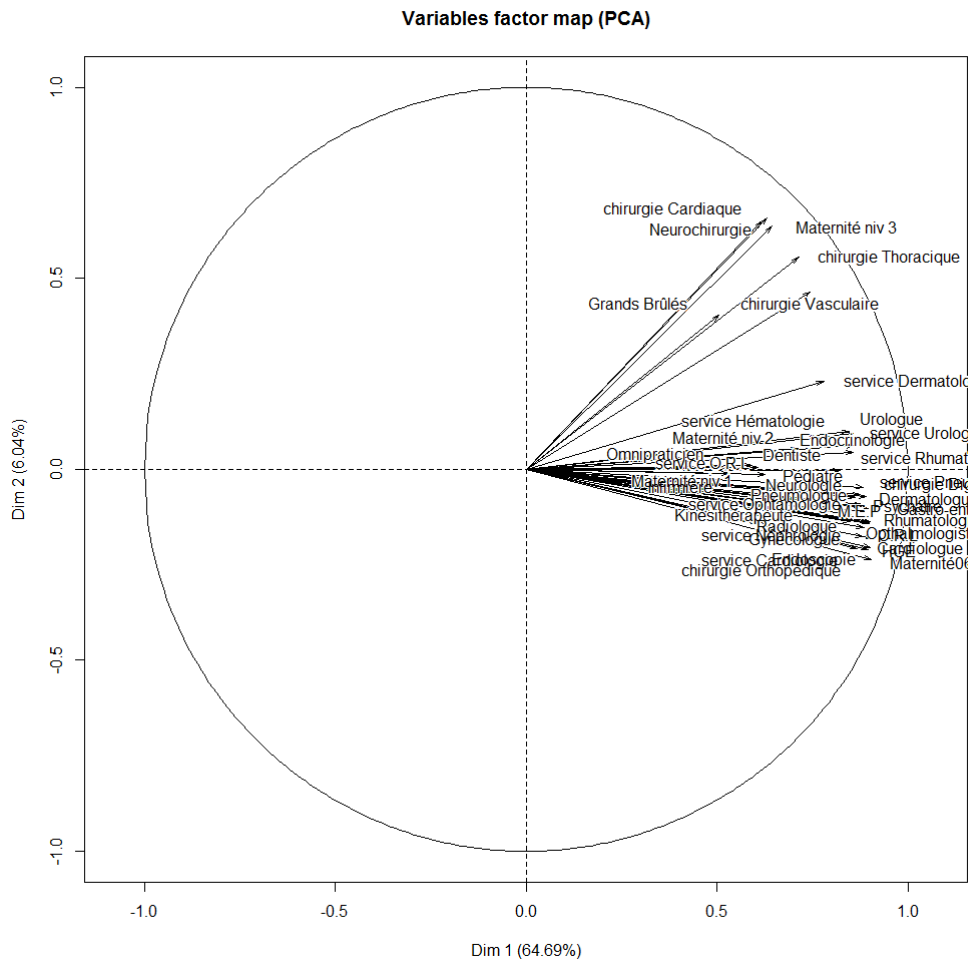
La figure III.2 met en lumière que les temps de trajets aux spécialistes sont très corrélés autour du premier axe à travers lequel 71% de la variance totale est expliquée. La deuxième composante n'explique qu'une faible partie de la variance (8.9%) et reflète principalement ce qui relèverait plutôt de la médecine de proximité. Ce groupe réunit omnipraticiens, dentistes et infirmiers.

Figure III.3: Analyse en composantes principales des temps de trajets aux spécialités hospitalières



La Figure III.3 des spécialités hospitalières montre également une corrélation générale entre variables. Le premier axe explique 65.3% de la variance totale, tandis que le second explique une faible part (9.7%). Sur le premier axe sont corrélées toutes les spécialités hospitalières, alors que le deuxième réunit les services plus rares tels que les maternités de niveau 3, les services de neurochirurgie, de chirurgie thoracique, de chirurgie vasculaire, de chirurgie cardiaque et de grands brûlés.

Figure III.4: Analyse en composantes principales des temps de trajets aux structures de soins



La Figure III.4 qui représente la projection des temps de trajet à toutes les structures de soins montre que l'axe 1 explique une majeure partie de la variance totale. Cette proportion atteint 65.7% de la variance totale, ce qui laisse penser qu'il serait difficilement envisageable de séparer les effets de l'accessibilité à la médecine de ville de ceux de l'accessibilité aux spécialités hospitalières. Le second axe, qui explique une très faible partie de la variance (6%), regroupe uniquement des services hospitaliers très spécifiques et moins accessibles comme les maternités de niveau 3, les services de neurochirurgie, de chirurgie thoracique, de chirurgie vasculaire, de chirurgie cardiaque et de grands brûlés.

### Corrélations spatiales des caractéristiques contextuelles

Les corrélations spatiales entre les trois scores d'accessibilité géographique à l'échelle communale (Tableau III.3) sont très élevées (0.91 à 0.98). La corrélation de ces scores avec l'indice de désavantage social FDep99 est modérée (0.35 environ), celle avec la densité de population est un peu plus élevée (-0.45) et environ du même ordre que celle entre le désavantage social et la densité.

Ces variables de contexte calculées à l'échelle Zone d'Emploi ont des corrélations plus élevées entre elles. Le score d'accessibilité géographique globale est corrélé à hauteur de 0.61 avec le désavantage social à -0.61 avec la densité.

Tableau III.3: Corrélations spatiales entre les caractéristiques contextuelles des communes

<i>Corrélations Géographiques</i>	Score Accessibilité Globale	Score Accessibilité Spécialistes	Score Accessibilité Hôpitaux	FDep99	Densité Pop.
<b>Score Accessibilité Globale</b>	1				
<b>Score Accessibilité Spécialistes</b>	0.97	1			
<b>Score Accessibilité Hôpitaux</b>	0.98	0.91	1		
<b>FDep99</b>	0.36	0.34	0.36	1	
<b>Densité Pop.</b>	-0.45	-0.41	-0.46	-0.43	1

Tableau III.4: Corrélations spatiales entre les caractéristiques contextuelles des Zones d'Emploi

<i>Corrélations spatiales</i>	Score Accessibilité Globale	FDep99	Densité Pop.
<b>Score Accessibilité Globale</b>	1		
<b>FDep99</b>	0.61	1	
<b>Densité Pop.</b>	-0.45	-0.63	1

### Score d'accessibilité géographique aux structures de soins

La forte corrélation constatée entre les scores d'accessibilité pousse à se diriger vers l'utilisation du score global qui reflète l'accessibilité géographique à toutes ces structures de soins de manière générale. Cette corrélation à l'échelle communale, bien que déjà trop fine pour effectuer une analyse multiniveaux d'évènements rares de santé, ne permettrait pas de séparer proprement ces différents effets sur la mortalité. Aussi, la population d'étude décrite plus haut et les recommandations issues de la partie II poussent à analyser les effets du contexte et la variabilité spatiale à l'échelle Zone d'Emploi. Le passage à une échelle moins fine implique également une perte d'information qui se mesure en comparant les variances aux deux échelles.





## Résultats

Tableau III.5: Distribution de la population EDP des plus de 30 ans le 1er Janvier 1990, vivant au 31 Décembre 1991 et ayant rempli leur bulletin de recensement RP90

Variables	Tous sexes			Femmes			Hommes		
	Effectifs	Décès	RSM	Effectifs	Décès	RSM	Effectifs	Décès	RSM
<i>Niveau d'études Indiv.</i>									
Bachelier et plus	56 142	6 176	0.71	27 277	1 972	0.75	28 865	4 204	0.69
Secondaire et technique	72 308	9 824	0.93	34 380	3 551	0.90	37 928	6 273	0.94
Elémentaire complète	67 483	18 088	0.99	39 552	8 258	0.96	27 931	9 830	1.01
Elémentaire incomplète	63 086	20 507	1.21	35 368	9 882	1.17	27 718	10 625	1.25
<i>Désavantage social du lieu</i>									
Q1 (le moins désavantagé)	51 361	9 186	0.91	27 766	4 280	0.94	23 595	4 906	0.88
Q2	52 390	10 477	0.96	27 836	4 564	0.96	24 554	5 913	0.97
Q3	50 970	10 994	1.00	26 781	4 619	0.98	24 189	6 375	1.02
Q4	51 911	11 609	1.03	26 979	4 925	1.03	24 932	6 684	1.03
Q5	52 387	12 329	1.09	27 215	5 275	1.09	25 572	7 054	1.09
<i>Temps de trajets aux soins</i>									
Q1 (temps le plus court)	54 781	10 791	0.96	29 990	5 071	0.99	24 791	5 720	0.95
Q2	50 070	9 464	0.98	26 647	4 131	0.98	23 423	5 333	0.97
Q3	50 179	10 629	1.04	26 248	4 471	1.03	23 931	6 158	1.04
Q4	51 924	11 590	1.04	27 088	4 847	1.02	24 836	6 743	1.06
Q5	52 065	12 121	0.98	26 604	5 143	0.99	25 461	6 978	0.98

La répartition de la population est assez équilibrée entre hommes et femmes selon les variables explicatives individuelles et contextuelles, sauf pour les deux catégories les plus faibles du niveau d'études où les femmes sont surreprésentées (Tableau III.5). Le plus fort gradient de mortalité est observé pour le niveau d'études individuel entre les catégories les plus et les moins instruites avec de plus grandes différences chez les hommes. Parmi les caractéristiques contextuelles, les inégalités de mortalité entre quintile de désavantage social de la zone de résidence étaient plus grandes que pour les temps de trajet aux soins et ces différentiels étaient plus grands pour les hommes que pour les femmes. L'association entre mortalité et temps de trajet aux soins a une forme en U inversé. La mortalité augmente avec le temps de trajet, sauf pour les zones où les temps de trajet sont les plus longs (la mortalité dans ces zones est proche du niveau national (Tableau III.5).

Tableau III.6: Association entre mortalité toutes causes, sexe, niveau d'études individuel, désavantage social et temps de trajets aux soins

Variables	Modèle 1 HR	Modèle 2 HR	Modèle 3 HR	Modèle 4 HR	Modèle 5 HR	Modèle 6 HR	Modèle 7 HR	Modèle 8 HR
<i>Effets Individuels</i>								
Hommes	<i>réf</i>	<i>réf</i>	<i>réf</i>	<i>réf</i>	<i>réf</i>	<i>réf</i>	<i>réf</i>	<i>réf</i>
Femmes	0.49 ***	0.48 ***	0.49 ***	0.49 ***	0.48 ***	0.48 ***	0.49 ***	0.48 ***
Bachelier et plus		<i>réf</i>			<i>réf</i>	<i>réf</i>		<i>réf</i>
Premier cycle secondaire		1.31 ***			1.31 ***	1.31 ***		1.31 ***
Elémentaire complète		1.42 ***			1.42 ***	1.42 ***		1.42 ***
Elémentaire incomplète		1.75 ***			1.74 ***	1.75 ***		1.75 ***
<i>Effets de contexte</i>								
Désavantage social du lieu								
Q1 (le moins désavantagé)			<i>réf</i>		<i>réf</i>		<i>réf</i>	<i>réf</i>
Q2			1.05		1.01		1.08 *	1.05
Q3			1.08 **		1.02		1.14 ***	1.09 **
Q4			1.13 ***		1.04		1.19 ***	1.13 ***
Q5			1.17 ***		1.06 *		1.26 ***	1.17 ***
Tendance			1.10 ***		1.05 ***		1.15 ***	1.10 ***
Temps de trajets aux soins								
Q1 (temps le plus court)				<i>réf</i>		<i>réf</i>	<i>réf</i>	<i>réf</i>
Q2				0.99		0.97	0.96	0.95
Q3				1.04		0.98	0.96	0.93 *
Q4				1.05		0.98	0.94	0.91 **
Q5				0.98		0.91 **	0.87 ***	0.84 ***
Tendance				0.98		0.93 ***	0.89 ***	0.87 ***
Variance Spatiale (ZE94)	0.013	0.011	0.010	0.012	0.011	0.010	0.008	0.008

\* p-value <0.050, \*\* p-value <0.010, \*\*\* p-value <0.001

La tendance est issue d'un modèle séparé dans lequel le contexte est introduit en tant qu'effet linéaire. (log-linéaire précisément).

Tableau III.7: Association entre mortalité par cause, sexe, niveau d'études individuel, désavantage social et temps de trajets aux soins

Variables	Maladies Cérébrovasc.	Cardiopathies Ischémiques	Toutes Tumeurs	Evitable Soins	Evitable Prévention
	HR	HR	HR	HR	HR
Hommes	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Femmes	0.61 ***	0.37 ***	0.43 ***	0.59 ***	0.31 ***
<i>Effets Individuels</i>					
Bachelier et plus	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Premier cycle secondaire	1.27 ***	1.30 ***	1.28 ***	1.26 ***	1.44 ***
Elémentaire complète	1.41 ***	1.46 ***	1.29 ***	1.39 ***	1.57 ***
Elémentaire incomplète	1.71 ***	1.62 ***	1.47 ***	1.72 ***	1.99 ***
<i>Effets de contexte</i>					
Désavantage social du lieu					
Q1 (le moins désavantagé)	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Q2	1.06	1.11	1.04	1.05	1.04
Q3	1.08	1.14 *	1.10 *	1.08	1.13 *
Q4	1.07	1.22 **	1.12 **	1.11 *	1.13 *
Q5	1.21 **	1.29 ***	1.14 ***	1.20 ***	1.24 ***
Tendance	1.10 **	1.15 ***	1.07 ***	1.12 ***	1.13 ***
Temps de trajets aux soins					
Q1 (temps le plus court)	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Q2	0.98	0.88 *	0.98	0.93	0.94
Q3	1.15 *	1.04	0.90 **	0.97	0.93
Q4	1.11	1.01	0.90 **	0.93	0.92
Q5	1.05	0.90	0.83 ***	0.88 **	0.80 ***
Tendance	0.98	0.95 *	0.93 ***	0.95 ***	0.91 ***
Variance Spatiale (ZE94)	0.003	0.007	0.006	0.010	0.017

\* p-value <0.050, \*\* p-value <0.010, \*\*\* p-value <0.001

La tendance est issue d'un modèle séparé dans lequel le contexte est introduit en tant qu'effet linéaire. (log-linéaire précisément).

Dans les modèles séparés Tableau III.6, le gradient éducationnel de mortalité mesuré par un HR varie de 1 à 1,75 entre bacheliers et les individus ayant une éducation élémentaire incomplète (modèle 2). L'association est positive et significative entre le désavantage social contextuel et la mortalité (modèle 3), mais elle est non significative (valeur p = 0,29) avec le temps de trajets aux structures de soins (modèle 4). Dans le modèle avec le niveau d'études individuel et le désavantage social contextuel, l'association avec le niveau d'étude n'est pas affectée, bien que l'association avec le désavantage social soit réduite de moitié tout en restant significative. L'introduction du niveau d'études ou du désavantage contextuel change le sens de l'association entre mortalité et temps de trajet aux soins qui devient négative et significative mais faible (modèles 6 et 7). Dans les modèles incluant le temps de trajet,

l'association avec le désavantage contextuel est amplifiée tandis que les associations avec le niveau d'études individuel restent stables pour tous les ajustements. Dans le modèle complet, le niveau d'études individuel et le désavantage social sont positivement associés à la mortalité tandis que l'association avec le temps de trajets aux structures de soins est négative, significative et plus forte que dans les autres modèles.

La variance spatiale diminue de 11% lorsque le niveau d'étude est introduit par rapport au modèle 1, de 21% avec le désavantage contextuel et seulement de 5% en présence du temps de trajet uniquement. Dans les modèles avec au moins deux variables, le niveau d'études et le désavantage expliquent 16% des variations spatiales, le niveau d'études et le temps de trajet expliquent 18%, alors que le désavantage et le temps de trajets aux soins réduisent de 40% les variations totales, ce qui est proche de la proportion expliquée par le modèle complet (modèle 8).

Toutes les associations par cause vont dans la même direction que celle de la mortalité toutes causes exceptée pour le temps de trajet aux soins et les maladies cérébrovasculaires qui sont parfois positives (U inversé) mais non significative (Tableau III.7). Les associations avec le niveau d'études individuel faible et le désavantage social contextuel sont toutes deux positives et significatives, plus faibles pour la mortalité toutes tumeurs, et plus fortes pour la mortalité évitable liée à la prévention. L'association avec le temps de trajet aux structures de soins est négative, significative et plus forte pour la mortalité évitable par la prévention. Les variances spatiales des modèles complets par cause de décès sont les plus élevées pour la mortalité évitable par la prévention et par une prise en charge par le système de soins (Tableau III.7).

Le désavantage social contextuel est significativement associé à la mortalité toutes causes avec la même amplitude peu importe l'âge ou le sexe (Tableau III.8). Concernant la mortalité par cause et l'association positive avec le désavantage contextuel, elle est légèrement plus forte pour les hommes que pour les femmes, sauf pour les cardiopathies ischémiques. Pour les hommes, l'association avec le désavantage social est un peu plus forte pour la mortalité prématurée que pour les personnes âgées, non significative pour les maladies cérébrovasculaires prématurés et globalement un peu plus forte pour la mortalité évitable par la prévention et les cardiopathies ischémiques, mais ces différences étaient faibles. Pour les femmes, l'association par âge avec le désavantage social est contrastée. Elle est la plus forte pour la mortalité prématurée par cardiopathies ischémiques là où les associations avec la mortalité par maladies cérébrovasculaires, cancers et maladies évitables par la prévention ne sont pas significatives. Pour les femmes plus âgées, le désavantage contextuel est associé à toutes les causes investiguées à l'exception des cancers. L'association avec le temps de trajet aux soins est toujours négative sauf pour la mortalité prématurée des femmes par maladies cérébrovasculaires, une association toutefois non significative. Il n'y a pas de différences nettes selon le sexe sauf pour l'association avec la mortalité par cardiopathies ischémiques, beaucoup plus forte chez les hommes de moins de 65 ans (Tableau III.8).

Tableau III.8: Effets du contexte sur la mortalité par cause (ajusté sur le niveau d'études individuel)

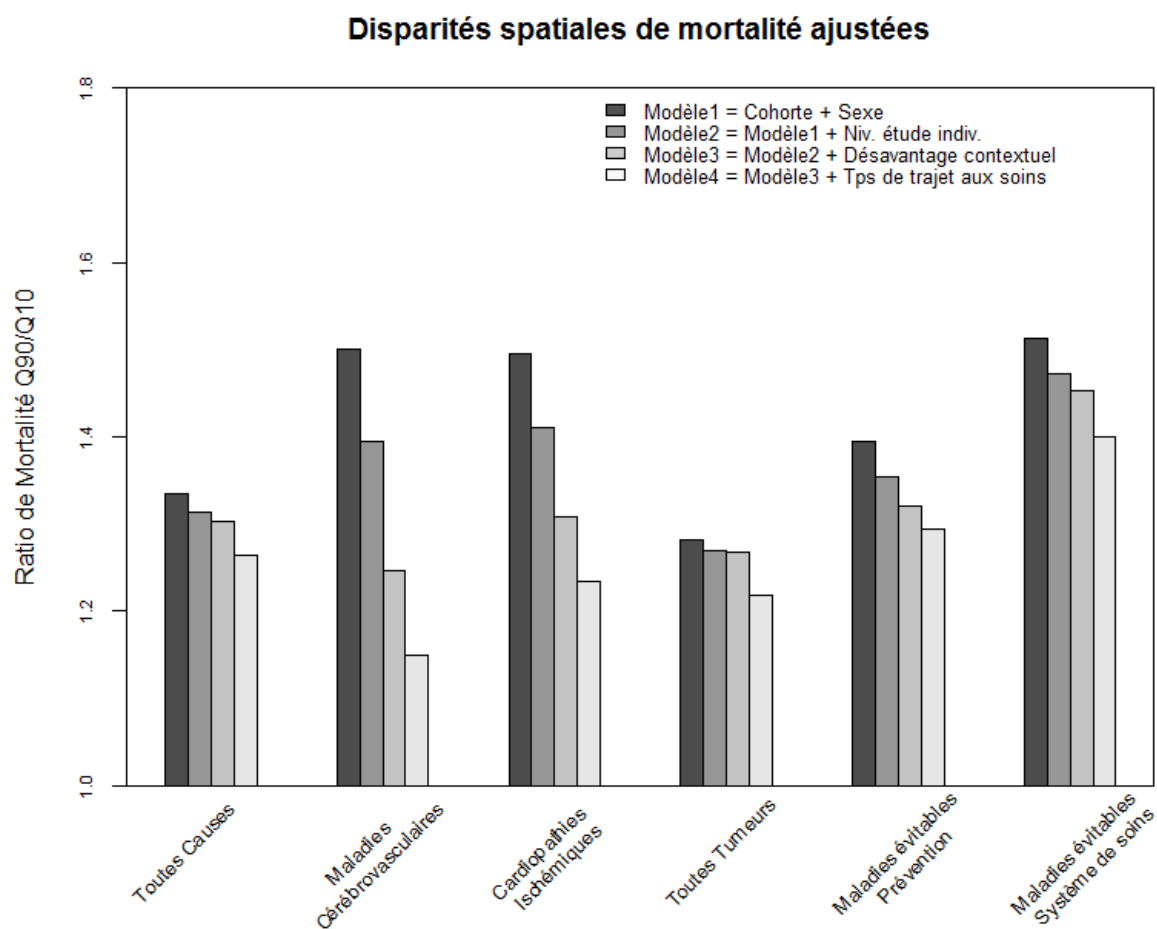
Cause de décès	Tous âges		< 65 ans		≥65 ans	
	Tps aux Soins	Désavant. Soc.	Tps aux Soins	Désavant. Soc.	Tps aux Soins	Désavant. Soc.
<b>Hommes</b>						
<i>Toutes Causes</i>	0.93 ***	1.10 ***	0.91 ***	1.09 **	0.95 ***	1.10 ***
<i>Maladies Cérébrovasc.</i>	0.97	1.11 *	0.88	1.26	0.99	1.08
<i>Cardiopathies Ischémiques</i>	0.94 *	1.13 **	0.89 *	1.18 *	0.96	1.11 *
<i>Toutes Tumeurs</i>	0.91 ***	1.10 ***	0.88 ***	1.14 ***	0.94 ***	1.08 **
<i>Évitable par le syst. de soins</i>	0.94 **	1.12 ***	0.91 *	1.15 **	0.95 **	1.12 ***
<i>Évitable par la prévention</i>	0.91 ***	1.14 ***	0.88 ***	1.15 ***	0.93 ***	1.13 ***
<b>Femmes</b>						
<i>Toutes Causes</i>	0.94 ***	1.11 ***	0.92 **	1.09 *	0.94 ***	1.11 ***
<i>Maladies Cérébrovasc.</i>	1.00	1.10 *	1.08	0.98	0.99	1.11 *
<i>Cardiopathies Ischémiques</i>	0.98	1.18 ***	1.00	1.52 *	1.00	1.17 **
<i>Toutes Tumeurs</i>	0.94 ***	1.02	0.89 **	1.05	0.96 *	1.01
<i>Évitable par le syst. de soins</i>	0.96 **	1.11 ***	0.95	1.16 **	0.96 **	1.11 ***
<i>Évitable par la prévention</i>	0.93 **	1.09 **	0.84 ***	1.05	0.95 *	1.09 *

\* p-value <0.050, \*\* p-value <0.010, \*\*\* p-value <0.001

Après ajustement sur le sexe et l'effet cohorte, les différentiels spatiaux de mortalité sont plus importants pour les maladies cérébrovasculaires, les cardiopathies ischémiques et les deux types de mortalité évitable (Figure III.5). Après ajustement sur la composition sociale (le niveau d'études individuel), les risques relatifs inter-décile de mortalité par cardiopathies ischémiques et maladies cérébrovasculaires chutent d'environ 1,5 à 1,4 tandis que toutes les autres causes sont peu affectées. L'introduction du désavantage social provoque la diminution de ce risque à près de 1.25 pour ces deux mêmes causes. L'effet du temps de trajet aux soins explique la plus grande partie de chaque variance spatiale de la mortalité toutes causes, toutes tumeurs et mortalité évitable par une prise en charge dans le système de soins (Figure III.5).

L'ajustement sur la densité de population a un impact sur les associations contextuelles (Tableau III.9). L'effet positif du désavantage social est légèrement amplifié par l'introduction de la densité pour toutes les causes de décès investiguées. L'ajustement sur la densité de population a un impact fort puisque l'association entre mortalité et temps de trajet aux soins devient significativement positive pour la mortalité par maladies cérébrovasculaires. Pour les autres causes investiguées, cette association prend une forme de U inversé positive et non significative pour les deuxième et troisième quintiles, mais aussi plus forte pour les cardiopathies ischémiques (Tableau III.9).

Figure III.5: Variabilité spatiale de la mortalité : avant/après ajustements



Les déciles sont extraits de la distribution spatiale de la mortalité mesurée par la variance spatiale. Cette variance a été estimée pour chacun de ces modèles par cause de décès.

Tableau III.9: Association entre mortalité par cause, sexe, niveau d'études individuel, désavantage social du contexte, temps de trajets aux soins et densité de population

Variables	Toutes Causes	Cardiopathies Ischémiques	Maladies Cérébrovasc.	Toutes Tumeurs	Evitable Prévention	Evitable Soins
	HR	HR	HR	HR	HR	HR
Hommes	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Femmes	0.48 ***	0.37 ***	0.61 ***	0.43 ***	0.31 ***	0.59 ***
<i>Effets Individuels</i>						
Bachelier et plus	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Premier cycle secondaire	1.31 ***	1.30 ***	1.26 ***	1.28 ***	1.44 ***	1.26 ***
Elémentaire complète	1.42 ***	1.46 ***	1.42 ***	1.29 ***	1.57 ***	1.39 ***
Elémentaire incomplète	1.75 ***	1.62 ***	1.71 ***	1.47 ***	1.99 ***	1.72 ***
<i>Effets de contexte</i>						
Désavantage social du lieu						
Q1 (le moins désavantagé)	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Q2	1.09 **	1.14 *	1.10	1.07 *	1.08	1.09 *
Q3	1.12 ***	1.15 *	1.09	1.11 **	1.16 **	1.10 *
Q4	1.16 ***	1.23 **	1.08	1.13 ***	1.17 **	1.14 **
Q5	1.21 ***	1.31 ***	1.23 **	1.15 ***	1.27 ***	1.23 ***
Temps de trajets aux soins						
Q1 (temps le plus court)	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>	<i>ref</i>
Q2	1.00	0.92	1.07	1.04	1.00	0.97
Q3	1.04	1.13	1.35 ***	1.02	1.06	1.08
Q4	1.06	1.16	1.38 **	1.06	1.10	1.09
Q5	1.00	1.02	1.32 **	0.97	0.98	1.03
Densité de population						
Q1 (le moins dense)	<i>ref</i>	<i>ref</i>				
Q2	1.02	0.97	1.00	0.99	1.01	1.00
Q3	1.09 ***	1.07	1.08	1.04	1.10 *	1.09 *
Q4	1.18 ***	1.13	1.21 *	1.16 ***	1.21 ***	1.17 ***
Q5	1.24 ***	1.15	1.30 *	1.21 ***	1.28 ***	1.22 **
Variance Spatiale (ZE94)	0.007	0.005	0.000	0.003	0.014	0.008

\* p-value <0.050, \*\* p-value <0.010, \*\*\* p-value <0.001

## **Discussion**

### **Résumé des résultats**

Dans cette étude, le désavantage social avait un effet positif et significatif, s'ajoutant au niveau d'études individuel pour toutes les causes de décès investiguées. Il n'y avait pas vraiment de différence entre les hommes et les femmes dans ces associations exceptée pour la mortalité prématurée par cardiopathies ischémiques et par cancer, qui étaient respectivement plus forte et plus faible pour les femmes que les hommes. A l'inverse, l'association avec le temps de trajet allait dans un sens inattendu : la mortalité était plus élevée dans les zones proches des structures de soins et déclinait avec un temps de trajet plus long. Cette association inverse était significative pour toutes les causes investiguées, exceptée pour la mortalité par accidents vasculaires cérébraux. Cette association négative avec la mortalité était dépendante de l'introduction du niveau d'études ou du désavantage social de la zone. De façon similaire, l'introduction du temps de trajet dans les modèles ajustés sur le niveau socio-économique individuel et contextuel, renforçait l'association contextuelle avec le désavantage social sur la mortalité toutes causes. Dans tous les modèles, l'effet du niveau d'études individuel est resté significatif, son amplitude était la plus élevée et elle n'était pas sensible à l'introduction des facteurs contextuels. Les disparités spatiales de mortalité par cardiopathies ischémiques et accidents vasculaires cérébraux étaient largement expliquées par les caractéristiques individuelles et contextuelles. Chacun de ces facteurs a quasiment expliqué la même proportion de variabilité spatiale pour ces causes, alors même qu'une quantité négligeable de celle-ci était expliquée pour les autres causes de décès.

### **Comparaison avec la littérature**

Ce travail confirme l'existence d'un effet de contexte du désavantage social sur la mortalité pour plusieurs causes de décès. Comparativement aux autres études, son amplitude était parmi les plus élevées bien qu'elle était attendue comme plus faible au regard de la large échelle choisie pour définir le contexte [142,143]. Les comparaisons avec l'existant sont difficiles compte tenu du peu d'études multiniveaux sur l'association contextuelle entre distance aux structures de soins et mortalité. Passant outre ces restrictions, une étude écologique a également mis en évidence une association inverse entre distance aux soins et mortalité par maladies du cœur après introduction des caractéristiques socio-économiques [164]. La distance aux soins affectait la survie des vétérans américains ayant un accident vasculaire cérébral [161]. Ce dernier résultat est compatible avec l'association positive, bien que non significative pour les femmes de moins de 65 ans dans notre étude. Cependant, l'étude sur les vétérans n'ajustait pas sur les niveaux socio-économiques individuel et contextuel. Une autre observation importante est que le temps de trajet aux soins est en moyenne bien plus élevé aux Etats-Unis qu'en France [168]. En France, une association positive fut trouvée entre distance aux soins médicaux et survie des patients avec certains cancers de l'appareil digestif [169]. Cependant, ces résultats doivent être pris avec précaution en raison du manque de puissance statistique et de choix méthodologiques importants qui diffèrent. Parmi ces différences notables, on peut citer le choix d'un autre indice de désavantage, l'absence d'ajustement sur le niveau socio-économique individuel, l'utilisation d'une échelle géographique plus fine mais sur une population largement rurale.



## Interprétation

L'effet du niveau d'études sur la mortalité influe sur d'autres caractéristiques socio-économiques et sur leurs évolutions au cours de la vie [140]. Un niveau d'études bas affecte les ressources matérielles à travers un revenu plus bas, un risque de chômage plus élevé est associé à un risque plus fort de faire un travail pénible et d'être exposé à des risques psychosociaux [170]. Il peut également être une porte d'entrée vers des comportements de santé délétères comme la consommation de tabac, l'abus d'alcool, une consommation plus faible de fruits et légumes et une consommation de soins moins élevées, possiblement à l'origine d'un diagnostic tardif d'une maladie chronique [171–173]. En comparaison à l'effet du contexte, les effets du niveau d'études individuel étaient largement plus forts pour toutes les causes de décès investiguées.

L'effet de contexte du désavantage social dans la littérature est associé à un plus grand accès aux fast-food, aux revendeurs de tabac et d'alcool, à une exposition plus intense à la pollution et globalement à des caractéristiques qui contribuent à accroître le niveau de stress et qui offre plus d'opportunités à adopter certains comportements délétères pour la santé [13,112,118,174–176]. Cependant, l'échelle des Zones d'Emploi est relativement large comparée aux échelles des « wards » ou des « census » (échelles du Royaume-Uni). La variation de la densité des revendeurs de tabac, d'alcool et de magasin d'alimentation n'explique sans doute pas à elle seule l'association avec le désavantage social. L'interprétation devrait également se situer dans les constructions locales des modes de vie et des facteurs culturels puisque l'introduction d'un effet régional dans le modèle a réduit l'amplitude de l'association qui est toutefois restée significative (résultats non fournis). Mises à part ces explications multifactorielles qui requièrent plus d'attention et d'investigations en France, il est aussi possible que l'effet des mobilités résidentielles aboutisse à un effet de sélection de population. En effet, si les zones désavantagées socialement voient leur population en meilleure santé changer de lieu de résidence (« healthy migrant effect »), la mortalité de ces zones augmentera, et celle des zones dans lesquelles ces mêmes individus s'installent, souvent moins désavantagées, diminuera.

Concernant l'effet du temps de trajet aux structures de soins, l'association négative et globalement faible est difficile à interpréter. Il n'est pas envisageable qu'une distance plus longue ait un effet positif sur la consommation de soins qui réduirait le risque de mortalité. Cette association inverse est probablement confondue par une mortalité plus élevée dans les zones urbaines par rapport aux zones rurales une fois le désavantage social pris en compte. Le gradient rural-urbain regroupe plusieurs caractéristiques spatiales hautement corrélées entre elles, ce qui rend difficile la décomposition en facteurs de risques indépendants (la corrélation avec la densité était de -0.45 à l'échelle ZE94, Tableau III.4). L'introduction de la densité de population a également fait changer le sens de l'association qui est devenue positive dans certains cas, mais non significative à part pour la mortalité par maladies cérébrovasculaires (Tableau III.8). Cependant la notion d'urbanité reste floue à cette échelle géographique. Une autre interprétation pour cette association est la causalité inverse. Il est possible que les populations ayant des besoins en soins plus élevés vivent ou déménagent proches des soins tandis que ceux en bonne santé en vivent éloignés. Dans la littérature, le renoncement aux soins est plus motivé pour des raisons financières que par la distance à parcourir pour accéder

aux structures de soins en région parisienne [177] et dans la région Nord-Pas-de-Calais [178]. Par conséquent, on pourrait s'attendre à ce que la majorité des différentiels de santé dus au renoncement aux soins se retrouvent dans les inégalités sociales de santé [19]. Les introductions successives de la composition sociale et de l'effet de contexte du désavantage social ont provoqué la chute des disparités spatiales de mortalité par accidents vasculaires cérébraux et cardiopathies ischémiques. Ces résultats indiquent que ces caractéristiques se situant aux niveaux individuel et contextuel sont essentielles pour toutes interventions visant à réduire ces différentiels géographiques.

### **Limites**

Dans cette partie, l'échelle Zone d'Emploi était considérée comme étant constitutive d'un contexte de vie dans son ensemble puisqu'une majorité de la population active y vit et s'y déplace quotidiennement. Bien qu'une bonne partie de la dispersion de la mortalité et 40% de la variabilité du désavantage social [179] aient été diluées par agrégation de ces mesures disponibles à l'échelle communale, la variance spatiale a pu être estimée avec un certain niveau de fiabilité (Chapitre II). De plus, les estimations à l'échelle communale (36 000 unités, obtenues sans modélisation multiniveaux) allaient dans le même sens avec la même amplitude, excepté un effet plus fort du désavantage social contextuel (résultat non communiqué), ce qui confirme en partie le résultat de la méta-analyse de Meijer [118].

Le score d'accessibilité géographique aux structures de soins reflète l'accès aux hôpitaux et à la médecine de ville qui ne partage pas les mêmes mécanismes avec la mortalité. Puisque les corrélations spatiales étaient très élevées (Figure III.1), le temps de trajet a d'abord été estimé à l'échelle communale et ensuite agrégé à l'échelle Zone d'Emploi en calculant une moyenne pondérée par la population. Ce score s'exprime donc comme un temps de trajet moyen par habitant pour accéder aux structures de soins. Ainsi, une part non négligeable de l'hétérogénéité spatiale a été diluée. Une limite sérieuse de ce score se situe dans le manque d'information sur la demande en soins de la population, le volume d'activité des personnels de soins et des hôpitaux. Ces caractéristiques devront être prises en compte dans de futures études qui évalueront la demande imputable à la composition sociodémographique de la population et la zone de chalandise des structures de soins en utilisant par exemple la méthode « enhanced two-step floating catchment area » [180]. Cependant, bien que le volume d'activité puisse être mesuré, la demande en soins ne reflète pas forcément le besoin. Cette observation est particulièrement vraie pour les populations socialement désavantagées qui consultent les médecins généralistes en premier lieu et très peu les spécialistes comparativement aux populations mieux loties [19]. De prochaines études devraient aussi prendre en compte le temps d'attente pour obtenir une consultation chez un spécialiste qui pourrait constituer une autre barrière potentielle à la consommation de soins. Une autre limite relative aux données de temps de trajet disponibles est qu'elles datent de 2007, ce qui est relativement éloigné du début du suivi. Cependant, la corrélation des distances (en km) aux soins entre 1990 et 2007 est restée très élevée (de 0.97 à 0.99, résultats non fournis), confirmant qu'aucun changement majeur n'a eu lieu pendant cette période à cette échelle d'analyse. Toutes les interactions entre les facteurs contextuels et individuels ont été testées et n'étaient pas significatives (résultats non communiqués).

Les causes de décès considérées agrègent un certain nombre de maladies qui ne partagent pas

nécessairement les mêmes facteurs de risques. Cependant les catégories des décès évitables ont été construites pour regrouper, autant que possible, les décès qui aurait pu être évités par une prise en charge à temps par les structures de soins ou un meilleur accès à la prévention (respectivement *amenable* et *preventable*) [58]. Ces regroupements nous permettent donc de questionner si les environnements physiques et sociaux investigués ici ont un lien avec ces mécanismes menant au décès.

Une dernière limite concerne l'amplitude de la variance spatiale qui pourrait résulter d'un effet de composition résiduel non mesuré. D'un autre côté, l'effet du contexte aurait pu être masqué par des caractéristiques individuelles médiatrices de l'effet du contexte sur la mortalité [40,51,144].

### **Conclusion**

De futures travaux devront s'atteler à quantifier la part de ces associations contextuelles attribuable aux mobilités résidentielles, à analyser les tendances de ces dernières ainsi que leurs répercussions sur la future distribution géographiques de la mortalité. Le problème de la causalité inverse peut également se poser dans les associations contextuelles. Il est possible que les individus malades subissent un déclassement et doivent déménager dans une zone désavantagée. De la même manière, les individus malades peuvent avoir tendance à se rapprocher des structures de soins avant de décéder proche de celles-ci. Cette dernière hypothèse n'est probablement pas la seule à expliquer le sens non attendu de l'association entre distance aux soins et mortalité. Cependant, il est exclu de penser qu'un éloignement géographique aux structures de soins aurait un effet bénéfique sur la santé. Pour comprendre cette association, il faut rappeler que les caractéristiques contextuelles sont spatialement corrélées. Le degré d'urbanisation et le niveau socio-économique conditionne l'offre de soins à travers la liberté d'installation des professionnels de santé. Ceci a abouti à la création de déserts médicaux dans certaines zones à population rurale peu attractives ayant une population vieillissante.

Ce travail est le premier à mettre en évidence un effet contextuel du désavantage social sur la mortalité en France en population générale. Après ajustement sur le niveau socio-économique individuel et le désavantage social du lieu de résidence, il persiste de larges disparités spatiales de mortalité. Cela implique que la composition de la population ne peut pas expliquer à elle seule les disparités géographiques constatées. Les résultats ne permettent pas de conclure que le temps de trajet aux structures de soins affecte la mortalité à cette échelle. Plusieurs facteurs spatiaux associés à la densité urbaine restent à investiguer pour avancer sur cette question.

Les études futures devraient examiner le processus médiateur à travers lequel le désavantage social du lieu de résidence est lié à une surmortalité. La variabilité spatiale substantielle après ajustement sur les facteurs socio-économiques individuels et contextuels suggère que d'autres caractéristiques contextuelles pourraient influencer sur la santé. Au final, outre la nécessité d'un effort commun de se focaliser sur les individus avec un faible niveau d'études, cibler les zones socialement désavantagées dans les politiques de santé publique est une question fondamentale pour réduire les inégalités sociales de mortalité.



## DISCUSSION GENERALE

### **Mesure synthétique des disparités géographiques**

L'analyse des disparités géographiques de mortalité nous a amené à utiliser plusieurs approches. La première partie du travail s'est appliquée à décrire cette dispersion de manière synthétique afin d'identifier les échelles géographiques les plus discriminantes ainsi que les causes de décès pour lesquelles les écarts de mortalité sont les plus élevés. Une telle mesure synthétique des disparités géographiques en population générale n'avait pas été proposée jusqu'ici. Elle a permis de suivre l'évolution des disparités géographiques au cours du temps et d'interroger l'articulation d'échelles géographiques en jeu dans ces différences. Un résultat important est que les disparités géographiques relatives de mortalité ont augmenté de manière générale. Ces écarts se sont creusés plus rapidement à partir de 1990 aux échelles plus fines (communale et cantonale), alors que les disparités entre régions ou départements sont restées plutôt stables. Ceci confirme les observations des travaux issus de la géographie de la santé [3,4,93]. Si la méthode proposée n'est pas sans limite, elle permet de valider à l'aide d'un outil quantitatif une hypothèse souvent mise en avant, en la testant sur l'ensemble du territoire métropolitain. La mesure synthétique de la variabilité géographique de la mortalité, bien qu'étant l'estimation la plus sensible aux design des données, est une quantité intéressante à bien des égards. Ces quantités pourraient renseigner les prises de décision concernant l'allocation de ressources et la mise en œuvre d'interventions en santé publique. L'application des politiques de santé s'effectuant finalement au niveau des institutions et des territoires qu'elles couvrent, la décomposition par échelle de la variance pourraient renseigner sur la pertinence à conduire une politique à l'échelle régionale, départementale ou plus locale. Ainsi, l'identification des déterminants associés à chaque niveau géographique pourrait constituer autant de leviers d'action pour lutter contre les inégalités de santé.

### **Analyse écologique des tendances de la distribution géographique**

L'analyse écologique confrontant la croissance démographique et l'évolution de la mortalité vient appuyer l'idée que les processus spatiaux ont un rôle dans les changements socio-sanitaires. A l'aide d'un marqueur des dynamiques territoriales, ici le taux d'accroissement, les résultats de cette étude mettent l'accent sur l'importance de la position occupée par les communes dans la hiérarchie urbaine. Leur inscription dans les dynamiques régionales et nationales est liée à leur trajectoire sanitaire. Au-delà de « l'effet marqueur » du taux d'accroissement, il est important de souligner que les mobilités résidentielles ont probablement un impact sur les indicateurs de mortalité. En effet, les échanges migratoires modifient la composition de la population et peuvent de ce fait avoir un impact sur les niveaux de mortalité. Quelques études Européennes font état de leur impact plus ou moins fort sur le renforcement des inégalités socio-spatiales de santé et dans certains cas sur les disparités spatiales aux échelles fines [181–185]. Il conviendra d'effectuer ce type d'investigation pour la France ce qui n'a pas été fait jusqu'ici. Le principal frein à un tel travail est le manque de puissance sur la base longitudinale actuellement disponible (EDP). Une telle analyse ne pourrait actuellement être faite qu'à une échelle agrégée ce qui limite ses retombées potentielles. A défaut d'accéder à un échantillon plus fourni du recensement,

l'historique de la commune de résidence pourrait être extrait à partir des bases médico-administratives. Pour cela, un appariement entre l'EDP enrichi des individus issus de l'EAR (représentatif à 4%) et les données du Système National des Données de Santé (SNDS) créé par la loi de modernisation de notre système de santé de janvier 2016, offrirait un potentiel d'étude sans précédent en France.

### **Fiabilité de l'analyse multiniveaux des effets du contexte**

La maîtrise de l'outil d'analyse des facteurs de risques multiniveaux est importante pour avancer dans l'identification des liens potentiels entre contexte et santé [186]. Dans le cadre d'une analyse de la mortalité qui est un évènement rare, cette méthode doit être employée avec certaines précautions. Les recommandations quant aux design des données et aux méthodes à utiliser ne doivent en aucun cas être négligées. Aussi, le déséquilibre des designs de données a également un impact sur la fiabilité des estimations. De futurs travaux devraient s'appliquer à fournir des recommandations quant au niveau de déséquilibre minimum pour l'analyse multiniveaux d'évènements rares de santé. La fiabilité des modèles de survie à effets aléatoires et leurs méthodes d'estimations associées devraient également faire l'objet d'investigations. Une mesure fiable de la variance spatiale de la mortalité est incontournable quand l'intérêt est d'investiguer les effets du contexte sur la mortalité. Après ajustement sur la composition sociodémographique, l'amplitude de cette quantité révèle la plausibilité d'un effet du contexte sur la santé [40,186,187]. La construction d'une échelle géographique d'analyse pourrait également constituer une suite pertinente à ce travail. Elle devrait répondre simultanément à plusieurs critères. Parmi ceux-ci, construire un découpage de l'espace qui répartie la population de manière plus équilibrée tout en maximisant la variabilité géographique d'une mesure d'exposition contextuelle.

### **Sources de données pour l'analyse multiniveaux en population générale**

Mettre en œuvre ce type d'investigation requiert au préalable que l'information socio-économique individuelle soit disponible et précède l'évènement de santé [122,188]. Dans de nombreux cas, l'accès à ces données peut être fastidieux d'un point de vue réglementaire. Dans ce travail de thèse nous avons eu accès à un échantillon représentatif à 1% de la population ce qui pour l'étude des disparités spatiales de la mortalité par cause s'est révélé trop peu fourni. Une population d'étude plus large pourrait être constituée à partir des données des caisses de retraites dans lesquelles le revenu annuel des cotisants est disponible. Un chaînage en routine des données de CépiDc avec l'EDP enrichi des individus issus de l'EAR est également à espérer. En présence de données exhaustives de mortalité par cause et de données individuelles longitudinales échantillonnées, une piste importante à suivre pour effectuer une analyse multiniveaux se situe sur le plan méthodologique. Les modèles dits « hybrides » et plus spécifiquement la *Hierarchical Related Regression* permet de combiner données agrégées et données individuelles [189,190], pour combiner à la fois la puissance statistique des premières et la précision des informations obtenues des secondes.

### **Définition du contexte - problématique méthodologique et statistiques**

Une limite majeure de l'analyse des facteurs de risques multiniveaux sur la santé repose dans le choix des niveaux à considérer. Premièrement, les limites du contexte (ou des contextes) ne sont pas clairement définies. Il est fort probable que la santé de l'individu n'est pas

uniquement influencée par son environnement de résidence. Plus largement, il faudrait considérer les environnements dans lesquels il se meut au quotidien ainsi que les durées d'exposition : lieu de travail, établissement scolaire, commerces fréquentés, etc. Dans la littérature ce problème méthodologique fait référence au *problème d'incertitude du contexte géographique (UGCoP)* [191].

Deuxièmement, les associations contextuelles sont potentiellement biaisées par un découpage arbitraire de l'espace ; ce problème est nommé *problème des unités spatiales modifiables (MAUP)* [120]. Si une caractéristique contextuelle a un impact sur la santé individuelle, son effet n'opère pas forcément qu'à l'intérieur des limites administratives. Par exemple, la qualité de l'eau va affecter la santé des individus à l'échelle du réseau de distribution. Ce réseau peut dépasser certains découpages fins ou au contraire être dilué au sein d'une métropole qui en possède parfois plusieurs. Pour faire face à ce type de problème, il est possible d'utiliser des modèles statistiques prenant en compte l'autocorrélation spatiale de la variable de santé et de construire des mesures d'exposition contextuelle qui capture la dimension continue de celle-ci [192]. Plus simplement, il conviendra d'explorer la variabilité spatiale du phénomène étudié en estimant sa variance selon plusieurs découpages géographiques. L'amplitude de cette estimation après ajustement sur la composition sociodémographique indique le potentiel d'un effet du contexte à avoir un impact direct ou indirect sur la santé à l'échelle considérée.

Dans le cadre de l'analyse multiniveaux d'un événement rare de santé, ce travail a mis en évidence que le choix de l'échelle géographique a également un impact fort sur la fiabilité des estimations. Dès lors, un arbitrage est nécessaire entre une définition fine du contexte impliquant une faible puissance statistique contre une définition plus large du contexte mais des estimations fiables. Pour finir, il est envisageable que la santé individuelle soit influencée par un niveau « méso » qui se situerait entre le contexte local et l'échelle nationale. Ce niveau pourrait être celui de la région qui pour la France refléterait les politiques institutionnelles, le niveau de richesse moyen, la spécialisation et la culture régionale dont le régime alimentaire est une composante importante.

### **Les associations contextuelles en population générale**

Ce travail s'est appliqué à estimer l'association contextuelle de l'environnement social et physique avec la mortalité en suivant les recommandations nécessaires à sa fiabilité. Nous avons établi qu'il existe une association contextuelle positive, significative et régulière du désavantage social avec la mortalité tous âges, prématurée, hommes et femmes pour les grands groupes de causes de décès. Jusqu'ici, ce résultat n'avait pas été mis en évidence à l'échelle nationale (territoire métropolitain). Cette association nous indique l'existence potentielle d'un effet du contexte de l'environnement socio-économique. Cependant, elle ne nous renseigne pas sur les principaux mécanismes par lesquels celui-ci affecte la santé de l'individu ce que de futurs travaux devraient entreprendre. Les études sur ce sujet se sont jusqu'à présent interrogées sur les risques d'adopter des comportements délétères pour la santé dans un environnement socio-économique désavantagé [106–108,193]. En France, une étude sur l'incidence de l'infarctus du myocarde a toutefois mis en évidence que les facteurs de risques individuels comme la consommation de tabac et des mesures physiopathologiques

supposés être des facteurs médiateurs expliquaient une faible part de l'association contextuelle avec l'environnement socio-économique [191]. Par ailleurs, les associations contextuelles avec la mortalité sont en général plus faibles qu'avec l'incidence ou les comportements. De par sa nature, l'indicateur de mortalité intègre l'incidence et la survie ce qui en fait une mesure synthétique de l'état de santé de la population. Certaines limites évoquées dans la littérature sur les liens entre contexte et santé mettent en avant l'argument selon lequel il s'agirait de caractéristiques individuelles non mesurées. D'un point de vue étiologique, il est toujours possible de réduire une exposition contextuelle à une mesure individuelle qui peut jouer le rôle d'effet médiateur. Dans la mesure où la question de recherche se situe au niveau de l'environnement, ajuster sur les effets médiateurs peut servir à comprendre les mécanismes à travers lesquels le contexte affecte la santé. L'environnement se situant en amont dans la chaîne causale reliant le contexte à la santé, un modèle ajusté sur les effets médiateurs s'expose au sur-ajustement [40,51,144]. D'un point de vue interventionnel, il est utile de démêler ce qui peut être traité au niveau individuel de ce qui pourrait être traité à l'échelle territoriale. Il est aussi important d'avancer sur la question des niveaux à considérer pour réduire les disparités géographiques de santé. Enfin ces analyses macro visent à identifier de grandes tendances, mais des particularités locales liées notamment à des aménagements du territoire, des flux migratoires, des organisations spécifiques, doivent faire l'objet d'analyses plus poussées sur des territoires plus restreints.

Mis à part ceux évoqués dans le paragraphe précédent, les associations contextuelles peuvent faire l'objet d'un biais qui relève de la sélection de population. En effet, indépendamment du niveau socio-économique individuel, les échanges de population dus aux mobilités résidentielles entre zones avantagées et désavantagées peuvent aboutir aux regroupements d'individus respectivement en meilleure santé dans les premières et en moins bonne santé dans les secondes. A ce stade, la frontière tenue entre, respectivement, la recherche sur l'étiologie des maladies et l'identification de zones à risques devrait être établie. Dans le premier cas, les futurs travaux devront prendre en compte le temps d'exposition au contexte à l'aide d'un modèle spatio-temporel. Dans le second cas, il importera d'identifier les caractéristiques contextuelles associées à un risque plus élevé de mortalité dans un but interventionnel. Ceci permettrait de cibler conjointement les individus et les zones à risques.

Dans toute étude traitant des effets du contexte sur la santé en France, nous conseillerons de considérer le niveau socio-économique contextuel et le degré d'urbanisation comme des variables de confusion potentielles. Par ailleurs, de la même manière que pour une mesure du statut social individuel, il est possible que le niveau socio-économique du lieu de résidence reflète d'autres caractéristiques potentiellement liées à la santé : habitat dégradé, espaces verts, pollution, violence, transports, bruit, etc. Ceci se traduirait par une corrélation spatiale des variables contextuelles. Cette corrélation pourrait engendrer des problèmes méthodologiques quant à l'identification d'effets indépendants. Enfin, une piste de réflexion complémentaire interroge la pertinence de caractériser un espace non pas en fonction de ses caractéristiques propres, mais en fonction de ses interactions avec les autres espaces. Parmi celles-ci, les mobilités quotidiennes et résidentielles pourraient aider à cerner la position et le rôle de chacun des espaces dans les hiérarchies urbaines, régionales et nationales. Par ailleurs, il est



possible que certains individus soient plus vulnérables à certaines expositions contextuelles ce qui est appelé communément une interaction inter-niveaux. Bien qu'aucune de ces interactions n'ait été identifiée dans ce travail, ce type d'association est courante en analyse multiniveaux et doit être systématiquement investiguée.

### **Conclusion**

Ce travail avait pour objectif d'expliquer les disparités géographiques de mortalité par cause. L'hypothèse réductionniste des facteurs de risques multiniveaux et son outil associé, le modèle multiniveaux, nous ont permis de poser un certain nombre de questions. Parmi celles-ci, le choix de l'échelle géographique à considérer reste un problème non résolu. Dans un souci d'aide à la décision, il est toutefois important de considérer les échelles administratives ou leurs agrégats qui sont les seules portes d'entrées vers les politiques territoriales. Durant tout le long de ce travail, la tension entre recherche pour la connaissance et recherche pour l'action en santé publique s'est ressentie. Les choix méthodologiques ont été effectués dans un but pragmatique, acceptant la contradiction entre la visée savante et la posture réaliste. La régularité constatée des associations entre niveau socio-économique individuel, contextuel et mortalité pour toutes les causes investiguées est conforme aux résultats publiés dans d'autres pays développés. Sans apporter de preuve supplémentaire, nos résultats appuient l'idée que la condition sociale est une cause fondamentale de la mortalité [195]. Les répercussions sur la santé des inégalités de ressources matérielles, d'environnement de résidence, de capital humain et social ne pourrait être régularisées par l'unique action réparatrice du système de soins. Une politique nationale efficace et volontariste de réduction des inégalités sociales de santé nécessiterait d'impliquer plusieurs acteurs se situant également hors du champ de la santé publique.

## BIBLIOGRAPHIE

- [1] Van de Walle E. La mortalité des départements français ruraux au XIXe siècle. *Ann. Démographie Hist.*, vol. 1973, Persée-Portail des revues scientifiques en SHS; 1973, p. 581–589.
- [2] Tonnelier F. Évolution des inégalités géographiques de la mortalité en France depuis 1911. *Espace Popul Sociétés* 1991;9:29–36.
- [3] Salem G, Rican S, Jouglu E. Atlas de la santé en France. Vol 1 : Les causes de décès. Paris: John Libbey Eurotext; 2000.
- [4] Rican S, Salem G, Vaillant Z, Jouglu E, DIACT D à l'aménagement du territoire et à l'action régionale (DATAR)-. Dynamiques sanitaires des villes françaises. La Documentation française; 2010.
- [5] Windenberger F, Rican S, Jouglu E, Rey G. Spatiotemporal association between deprivation and mortality: trends in France during the nineties. *Eur J Public Health* 2012;22:347–53. doi:10.1093/eurpub/ckr029.
- [6] Trugeon A. Inégalités socio-sanitaires en France: de la région au canton. Elsevier Health Sciences; 2010.
- [7] mondiale de la Santé O. Rapport sur la santé dans le monde 2000. Améliorer la performance des systèmes de santé. Genève, Organisation mondiale de la Santé; 2000.
- [8] Mackenbach JP, Bos V, Andersen O, Cardano M, Costa G, Harding S, et al. Widening socioeconomic inequalities in mortality in six Western European countries. *Int J Epidemiol* 2003;32:830–7.
- [9] Blanpain N, Chardon O. Les inégalités sociales face à la mort. *Doc Trav* 2011;6549.
- [10] Menvielle G, Chastang J-F, Luce D, Leclerc A. [Changing social disparities and mortality in France (1968-1996): cause of death analysis by educational level]. *Rev Dépidémiologie Santé Publique* 2007;55:97–105. doi:10.1016/j.respe.2006.10.001.
- [11] Leclerc A, Fassin D, Grandjean H, Kaminski M, Lang T. Les inégalités sociales de santé. La découverte; 2000.
- [12] Kunst A. Cross-national comparisons of socio-economic differences in mortality. 1997.
- [13] Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav* 1995;Spec No:80–94.
- [14] Valkonen T. Les inégalités sociales devant la mort. *Démographie Anal. Synthèse Vol. 3 Déterm. Mortal.*, Paris: INED; 2002.
- [15] Moleux M, Schaetzel F, Scotton C. Les inégalités sociales de santé: Déterminants sociaux et médiateurs d'action. Inspection Generale des Affaires Sociales; 2011.
- [16] Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. *Med Care* 1981;19:127–40.
- [17] Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav* 1995;36:1–10.
- [18] Or Z, Jusot F, Yilmaz E. Inégalités de recours aux soins en Europe: Quel rôle attribuable aux systèmes de santé ? *Rev Économique* 2009;60:521. doi:10.3917/reco.602.0521.
- [19] Jusot F. Les inégalités de recours aux soins : bilan et évolution. *Rev Dépidémiologie Santé Publique* 2013;61, Supplement 3:S163–9. doi:10.1016/j.respe.2013.05.011.
- [20] Déserts médicaux: agir vraiment n.d. [https://www.senat.fr/rap/r12-335/r12-335\\_mono.html](https://www.senat.fr/rap/r12-335/r12-335_mono.html) (accessed December 19, 2016).
- [21] Tonnelier F, Lucas V, Grandfils N. Health Care Reforms in France: between centralism and local authorities. *CreDES*; 1996.

- [22] Couty E. La loi HPST: regards sur la réforme du système de santé. Presses de l'école des hautes études en santé publique; 2009.
- [23] Touraine M. Health inequalities and France's national health strategy. *The Lancet* 2014;383:1101–2. doi:10.1016/S0140-6736(14)60423-2.
- [24] Aouba A, Jouglu E, Eb M, Rey G. L'évolution de la mortalité et des causes de décès entre 1990 et 2009. *ADSP* 2012;80:24–8.
- [25] Marmot MG, McDowall ME. Mortality decline and widening social inequalities. *Lancet Lond Engl* 1986;2:274–6.
- [26] Cambois E, Jusot F, others. Ampleur, tendance et causes des inégalités sociales de santé et de mortalité en Europe: une revue des études comparatives. *Bull Épidémiologique Hebd* 2007;10–14.
- [27] Rey, Jouglu E, Anne Fouillet, Hémon D. Ecological association between a deprivation index and mortality in France over the period 1997 – 2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health* 2009;9:33. doi:10.1186/1471-2458-9-33.
- [28] Windenberger F, Rican S, Jouglu E, Rey G. Spatiotemporal association between deprivation and mortality: trends in France during the nineties. *Eur J Public Health* 2011.
- [29] Boyle P, Exeter D, Flowerdew R. The role of population change in widening the mortality gap in Scotland. *Area* 2004;36:164–173.
- [30] O'Reilly D, Stevenson M. Selective migration from deprived areas in Northern Ireland and the spatial distribution of inequalities: implications for monitoring health and inequalities in health. *Soc Sci Med* 2003;57:1455–62.
- [31] O'Reilly D, Browne S, Johnson Z, Kelly A. Are cities becoming more unhealthy? An analysis of mortality rates in Belfast and Dublin between 1981 and 1991 to illustrate a methodological difficulty with ecological studies. *J Epidemiol Community Health* 2001;55:354–5.
- [32] Connolly S, O'Reilly D, Rosato M. Increasing inequalities in health: is it an artefact caused by the selective movement of people? *Soc Sci Med* 2007;64:2008–15.
- [33] Brown D, O'Reilly D, Gayle V, Macintyre S, Benzeval M, Leyland AH. Socio-demographic and health characteristics of individuals left behind in deprived and declining areas in Scotland. *Health Place* 2012;18:440–4. doi:10.1016/j.healthplace.2012.01.002.
- [34] Durkheim E. *Le suicide: étude de sociologie*. F. Alcan; 1897.
- [35] Robinson WS. Ecological correlations and the behavior of individuals. *Int J Epidemiol* 2009;38:337–341.
- [36] Estève J, Benhamou E, Raymond L. *Méthodes statistiques en épidémiologie descriptive*. Ed. INSERM; 1993.
- [37] Greenland S. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *Int J Epidemiol* 2001;30:1343–50. doi:10.1093/ije/30.6.1343.
- [38] Schwartz S, Susser E, Susser M. A future for epidemiology? *Annu Rev Public Health* 1999;20:15–33. doi:10.1146/annurev.publhealth.20.1.15.
- [39] Krieger N. Epidemiology and the web of causation: has anyone seen the spider? *Soc Sci Med* 1982 1994;39:887–903.
- [40] Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;88:216–22.
- [41] O'Campo P. Invited commentary: Advancing theory and methods for multilevel models of residential neighborhoods and health. *Am J Epidemiol* 2003;157:9–13.

- [42] Behavior and the D of the HC for the S and HIKP of H and S, Health LFBTDCP of PP and E and C of the D of H and SBB at HS of P. *Neighborhoods and Health*. Oxford University Press; 2003.
- [43] Duncan C, Jones K, Moon G. Health-related behaviour in context: a multilevel modelling approach. *Soc Sci Med* 1982 1996;42:817–30.
- [44] Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health* 2001;55:111–22.
- [45] Chaix B, Chauvin P. [The contribution of multilevel models in contextual analysis in the field of social epidemiology: a review of literature]. *Rev Dépidémiologie Santé Publique* 2002;50:489–99.
- [46] Kawachi I, Subramanian SV, Almeida-Filho N. A glossary for health inequalities. *J Epidemiol Community Health* 2002;56:647–52. doi:10.1136/jech.56.9.647.
- [47] Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. *Int J Epidemiol* 2001;30:668–77. doi:10.1093/ije/30.4.668.
- [48] Rose G. Sick individuals and sick populations. *Int J Epidemiol* 2001;30:427–32. doi:10.1093/ije/30.3.427.
- [49] Snijders TABAB, Bosker RJ. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage Publications Limited; 2011.
- [50] De Leeuw J, Meijer E, Goldstein H. *Handbook of multilevel analysis*. Springer; 2008.
- [51] Chaix, B. [Modelling contextual effects on health and healthcare]. Université Paris VI, 2004.
- [52] DiPrete TA, Forristal JD. *Multilevel Models: Methods and Substance*. *Annu Rev Sociol* 1994;20:331–57.
- [53] Kreft IG, Kreft I, de Leeuw J. *Introducing multilevel modeling*. Sage; 1998.
- [54] Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *J Epidemiol Community Health* 2003;57:550–2. doi:10.1136/jech.57.8.550.
- [55] Bingenheimer JB. Multilevel models and scientific progress in social epidemiology. *J Epidemiol Community Health* 2005;59:438–9. doi:10.1136/jech.2004.028456.
- [56] Spira A. L'épidémiologie, une discipline carrefour: Epidemiology, an interdisciplinary crossroads. *Rev Dépidémiologie Santé Publique* 2004;52:489–91. doi:10.1016/S0398-7620(04)99088-8.
- [57] Gatrell AC, Elliott SJ. *Geographies of health: An introduction*. John Wiley & Sons; 2014.
- [58] *Definitions of avoidable mortality*. Office for National Statistics; 2011.
- [59] Couet C. INSEE's Permanent Demographic Sample (EDP). *Courr Stat Engl Ser N* 13 2007.
- [60] Rey G, Jouglé E, Fouillet A, Hémon D. Ecological association between a deprivation index and mortality in France over the period 1997 - 2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health* 2009;9:33. doi:10.1186/1471-2458-9-33.
- [61] Cislighi C, Mezzanotte G, La Vecchia C, De Carli A, Vigotti MA. Geografia del Cancro: analisi dei trend territoriali e individuazione delle zone anomale. *Epidemiol E Prevenzione* 1986;29:15–49.
- [62] *La santé en France H*. La santé en France. Paris Doc Fr 2002:167.
- [63] Jerrett M, Burnett RT, Ma R, Pope CA 3rd, Krewski D, Newbold KB, et al. Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiol Camb Mass* 2005;16:727–36.
- [64] Cambra K, Martínez-Rueda T, Alonso-Fustel E, Cirarda FB, Ibáñez B, Esnaola S, et al. Mortality in small geographical areas and proximity to air polluting industries in the

- Basque Country (Spain). *Occup Environ Med* 2010;oem.2009.048215.  
doi:10.1136/oem.2009.048215.
- [65] Pless-Mulloli T, Phillimore P, Moffatt S, Bhopal R, Foy C, Dunn C, et al. Lung cancer, proximity to industry, and poverty in northeast England. *Environ Health Perspect* 1998;106:189.
- [66] Grant WB. An ecological study of cancer incidence and mortality rates in France with respect to latitude, an index for vitamin D production. *Dermatoendocrinol* 2010;2:62–67.
- [67] Norström T, Norström T, Ramstedt M, Norström T, Ramstedt M. Mortality and population drinking: a review of the literature. *Drug Alcohol Rev* 2005;24:537–547.
- [68] Pearce J, Hiscock R, Moon G, Barnett R. The neighbourhood effects of geographical access to tobacco retailers on individual smoking behaviour. *J Epidemiol Community Health* 2009;63:69–77.
- [69] Subramanian SV, Jones K, Kaddour A, Krieger N. Revisiting Robinson: The perils of individualistic and ecologic fallacy. *Int J Epidemiol* 2009:dyn359.  
doi:10.1093/ije/dyn359.
- [70] Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Lippincott Williams & Wilkins; 2008.
- [71] Townsend P. Deprivation. *Int Soc Pol* 1987;16:125–48.
- [72] Carstairs V, Morris R. Deprivation: explaining differences in mortality between Scotland and England and Wales. *Bmj* 1989;299:886–9.
- [73] Pampalon R, Raymond G. A deprivation index for health and welfare planning in Quebec. *Chronic Can* 2000;21:104–13.
- [74] Ocana-Riola R, Saurina C, Fernandez-Ajuria A, Lertxundi A, Sanchez-Cantalejo C, Saez M, et al. Area deprivation and mortality in the provincial capital cities of Andalusia and Catalonia (Spain). *J Epidemiol Community Health* 2008;62:147–52.
- [75] Turrell G, Mathers C. Socioeconomic inequalities in all-cause and specific-cause mortality in Australia: 1985-1987 and 1995-1997. *Int J Epidemiol* 2001;30:231–9.
- [76] Damette F, Scheibling J. *Le territoire français: permanences et mutations*. Hachette supérieur; 2011.
- [77] Laurent E. *Vers l'égalité des territoires*. Dyn Mes Polit 2013.
- [78] Baccaïni B. Observation et concepts en matière de migrations internes. *Démographie Anal Synthèse* 2006;8:65–88.
- [79] Boyle PJ, Norman P, Popham F. Social mobility: evidence that it can widen health inequalities. *Soc Sci Med* 2009;68:1835–42. doi:S0277-9536(09)00146-4 [pii]  
10.1016/j.socscimed.2009.02.051.
- [80] Chaix B, Rosvall M, Merlo J. Neighborhood socioeconomic deprivation and residential instability: effects on incidence of ischemic heart disease and survival after myocardial infarction. *Epidemiology* 2007;18:104–11.
- [81] O'Reilly D. Health and social inequality in Europe. Migration from deprived areas may be a factor. *BMJ* 1994;309:57–8.
- [82] Brown D, Leyland AH. Scottish mortality rates 2000-2002 by deprivation and small area population mobility. *Soc Sci Med* 2010;71:1951–7.
- [83] Davey Smith G, Hart C, Hole D, MacKinnon P, Gillis C, Watt G, et al. Education and occupational social class: which is the more important indicator of mortality risk? *J Epidemiol Community Health* 1998;52:153–60.
- [84] Molarius A, Janson S. Population change and mortality in men and women. *J Epidemiol Community Health* 2000;54:772.
- [85] Regidor E, Calle ME, Domínguez V, Navarro P. Inequalities in mortality in shrinking and growing areas. *J Epidemiol Community Health* 2002;56:919–21.

- [86] Davey Smith G, Shaw M, Dorling D. Population change and mortality in men and women. *J Epidemiol Community Health* 2001;55:9.
- [87] McLoone P, Boddy FA. Deprivation and mortality in Scotland, 1981 and 1991. *Bmj* 1994;309:1465–70.
- [88] Davey Smith G, Shaw M, Dorling D. Shrinking areas and mortality. *Lancet* 1998;352:1439–40.
- [89] Exeter DJ, Boyle PJ, Feng Z, Boyle M. Shrinking areas and mortality: an artefact of deprivation effects in the West of Scotland? *Health Place* 2009;15:399–401.
- [90] Exeter DJ, Feng Z, Flowerdew R, Boyle PJ. Shrinking areas and mortality: an artefact of deprivation effects? *J Epidemiol Community Health* 2005;59:924–6.
- [91] Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Stat Math* 1991;43:1–59.
- [92] Yan J, Fine J. Estimating equations for association structures. *Stat Med* 2004;23:859-874-877-880. doi:10.1002/sim.1650.
- [93] Rican S, Salem G, Jouglu E. Disparités spatiales de mortalité par cancers broncho-pulmonaires selon différentes échelles spatiales en France : intérêt des changements d'échelle en géographie. *Rev Dépidémiologie Santé Publique* 1999;47:411–20.
- [94] Clerval A. Les dynamiques spatiales de la gentrification à Paris. Une carte de synthèse. *Cybergeog Eur J Geogr* 2010.
- [95] Tunstall H, Mitchell R, Gibbs J, Platt S, Dorling D. Is economic adversity always a killer? Disadvantaged areas with relatively low mortality rates. *J Epidemiol Community Health* 2007;61:337–43.
- [96] Mitchell R, Gibbs J, Tunstall H, Platt S, Dorling D. Factors which nurture geographical resilience in Britain: a mixed methods study. *J Epidemiol Community Health* 2009;63:18–23.
- [97] Guilly C, Noyé C. Atlas des nouvelles fractures sociales en France (Atlas of the new social fractures in France). Paris: Autrement; 2004.
- [98] Regidor E, Calle ME, Domínguez V, Navarro P. Inequalities in mortality in shrinking and growing areas. *J Epidemiol Community Health* 2002;56:919–21.
- [99] Bentham G. Migration and morbidity: implications for geographical studies of disease. *Soc Sci Med* 1988;26:49–54.
- [100] Exeter DJ, Boyle PJ, Norman P. Deprivation (im)mobility and cause-specific premature mortality in Scotland. *Soc Sci Med* 2011;72:389–97. doi:S0277-9536(10)00744-6 [pii] 10.1016/j.socscimed.2010.10.009.
- [101] Norman P, Boyle P, Exeter D, Feng Z, Popham F. Rising premature mortality in the U.K.'s persistently deprived areas: only a Scottish phenomenon? *Soc Sci Med* 2011;73:1575–84. doi:S0277-9536(11)00606-X [pii] 10.1016/j.socscimed.2011.09.034.
- [102] Riva M, Curtis S, Norman P. Residential mobility within England and urban-rural inequalities in mortality. *Soc Sci Med* 2011;73:1698–706. doi:S0277-9536(11)00602-2 [pii] 10.1016/j.socscimed.2011.09.030.
- [103] Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med* 2002;55:125–39.
- [104] Diez Roux A. Investigating neighborhood and area effects on health. *Am J Public Health* 2001;91:1783–9.
- [105] Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *J Epidemiol Community Health* 2003;57:550–2.
- [106] Kleinschmidt I, Hills M, Elliott P. Smoking behaviour can be predicted by neighbourhood deprivation measures. *J Epidemiol Community Health* 1995;49 Suppl 2:S72-77.

- [107] Reijneveld SA. The impact of individual and area characteristics on urban socioeconomic differences in health and smoking. *Int J Epidemiol* 1998;27:33–40.
- [108] Karvonen S, Rimpelä A. Socio-regional context as a determinant of adolescents' health behaviour in Finland. *Soc Sci Med* 1996;43:1467–74.
- [109] Scribner RA, Cohen DA, Fisher W. Evidence of a structural effect for alcohol outlet density: a multilevel analysis. *Alcohol Clin Exp Res* 2000;24:188–95.
- [110] Chuang Y-C, Cubbin C, Ahn D, Winkleby MA. Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking. *J Epidemiol Community Health* 2005;59:568–73. doi:10.1136/jech.2004.029041.
- [111] Morland K, Wing S, Diez Roux A. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. *Am J Public Health* 2002;92:1761–7.
- [112] Mason KE, Bentley RJ, Kavanagh AM. Fruit and vegetable purchasing and the relative density of healthy and unhealthy food stores: evidence from an Australian multilevel study. *J Epidemiol Community Health* 2013;67:231–6. doi:10.1136/jech-2012-201535.
- [113] Reijneveld SA, Verheij RA, de Bakker DH. The impact of area deprivation on differences in health: does the choice of the geographical classification matter? *J Epidemiol Community Health* 2000;54:306–13.
- [114] Blakely TA, Lochner K, Kawachi I. Metropolitan area income inequality and self-rated health--a multi-level study. *Soc Sci Med* 2002;54:65–77.
- [115] Malmström M, Johansson SE, Sundquist J. A hierarchical analysis of long-term illness and mortality in socially deprived areas. *Soc Sci Med* 2001;53:265–75.
- [116] Shouls S, Congdon P, Curtis S. Modelling inequality in reported long term illness in the UK: combining individual and area characteristics. *J Epidemiol Community Health* 1996;50:366–76.
- [117] Silhol R, Zins M, Chauvin P, Chaix B. Investigating the spatial variability in incidence of coronary heart disease in the Gazel cohort: the impact of area socioeconomic position and mediating role of risk factors. *J Epidemiol Community Health* 2011;65:137–43. doi:10.1136/jech.2009.087379.
- [118] Meijer M, Röhl J, Bloomfield K, Grittner U. Do neighborhoods affect individual mortality? A systematic review and meta-analysis of multilevel studies. *Soc Sci Med* 2012;74:1204–12. doi:10.1016/j.socscimed.2011.11.034.
- [119] Diez Roux AV. The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. *Epidemiol Rev* 2004;26:104–11. doi:10.1093/epirev/mxh006.
- [120] Fotheringham AS, Wong DW. The modifiable areal unit problem in multivariate statistical analysis. *Environ Plan A* 1991;23:1025–1044.
- [121] Merlo J, Chaix B, Yang M, Lynch J, Råstam L. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health* 2005;59:443–9.
- [122] Rey G, Rican S, Luce D, Menvielle G, Jouglu E. Measuring social inequalities in cause-specific mortality in France: comparison between linked and unlinked approaches. *Rev Épidémiologie Santé Publique* 2013;61:221–31. doi:10.1016/j.respe.2012.11.004.
- [123] INSEE. Echantillon Démographique Permanent. Manuel de l'utilisateur en statistiques et études 2003.
- [124] R Core Team, R Foundation for Statistical Computing. *R : A Language and Environment for Statistical Computing* 2014.
- [125] Bates D, Maechler M, Bolker B. *lme4: Linear mixed-effects models using Eigen and Eigen++* 2012.

- [126] Rue H avar, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol* 2009;71:319–392.
- [127] Yu-Sung Su and Masanao Yajima. *R2jags: A Package for Running jags from R* 2014.
- [128] Gelman A, others. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal* 2006;1:515–534.
- [129] Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med* 2006;25:4279–92. doi:10.1002/sim.2673.
- [130] Theall KP, Scribner R, Broyles S, Yu Q, Chotalia J, Simonsen N, et al. Impact of small group size on neighbourhood influences in multilevel models. *J Epidemiol Community Health* 2011;65:688–95. doi:10.1136/jech.2009.097956.
- [131] Moineddin R, Matheson FI, Glazier RH. A simulation study of sample size for multilevel logistic regression models. *BMC Med Res Methodol* 2007;7:34. doi:10.1186/1471-2288-7-34.
- [132] Merlo J, Ohlsson H, Lynch KF, Chaix B, Subramanian SV. Individual and collective bodies: using measures of variance and association in contextual epidemiology. *J Epidemiol Community Health* 2009;63:1043–8. doi:10.1136/jech.2009.088310.
- [133] Chaix B, Merlo J, Subramanian SV, Lynch J, Chauvin P. Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in Malmo, Sweden, 2001. *Am J Epidemiol* 2005;162:171–82. doi:10.1093/aje/kwi175.
- [134] Austin PC. Bias in Penalized Quasi-Likelihood Estimation in Random Effects Logistic Regression Models When the Random Effects Are not Normally Distributed. *Commun Stat - Simul Comput* 2005;34:549–65. doi:10.1081/SAC-200068364.
- [135] Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Anal* 2006;1:473–514.
- [136] Austin PC. Estimating Multilevel Logistic Regression Models When the Number of Clusters is Low: A Comparison of Different Statistical Software Procedures. *Int J Biostat* 2010;6. doi:10.2202/1557-4679.1195.
- [137] Krieger N. Embodiment: a conceptual glossary for epidemiology. *J Epidemiol Community Health* 2005;59:350–5. doi:10.1136/jech.2004.024562.
- [138] Diez Roux AV. A glossary for multilevel analysis. *J Epidemiol Community Health* 2002;56:588–94. doi:10.1136/jech.56.8.588.
- [139] Goldacre MJ. Cause-specific mortality: understanding uncertain tips of the disease iceberg. *J Epidemiol Community Health* 1993;47:491–6. doi:10.1136/jech.47.6.491.
- [140] Ribet C, Melchior M, Lang T, Zins M, Goldberg M, Leclerc A. [Characterisation and measurement of social position in epidemiologic studies]. *Rev Epidemiol Sante Publique* 2007;55:285–95. doi:S0398-7620(07)00129-0 [pii] 10.1016/j.respe.2007.04.001.
- [141] Liberatos P, Link BG, Kelsey JL. The measurement of social class in epidemiology. *Epidemiol Rev* 1988;10:87–121.
- [142] Blakely T, Atkinson J, Ivory V, Collings S, Wilton J, Howden-Chapman P. No association of neighbourhood volunteerism with mortality in New Zealand: a national multilevel cohort study. *Int J Epidemiol* 2006;35:981–9. doi:10.1093/ije/dyl088.
- [143] Turrell G, Kavanagh A, Draper G, Subramanian SV. Do places affect the probability of death in Australia? A multilevel study of area-level disadvantage, individual-level socioeconomic position and all-cause mortality, 1998-2000. *J Epidemiol Community Health* 2007;61:13–9. doi:61/1/13 [pii] 10.1136/jech.2006.046094.
- [144] Yen IH, Kaplan GA. Neighborhood Social Environment and Risk of Death: Multilevel Evidence from the Alameda County Study. *Am J Epidemiol* 1999;149:898–907.



- [145] Naess O, Piro FN, Nafstad P, Smith GD, Leyland AH. Air pollution, social deprivation, and mortality: a multilevel cohort study. *Epidemiol Camb Mass* 2007;18:686–94. doi:10.1097/EDE.0b013e3181567d14.
- [146] Mackenbach JP, Stirbu I, Roskam AJ, Schaap MM, Menvielle G, Leinsalu M, et al. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med* 2008;358:2468–81.
- [147] Malmström M, Johansson S, Sundquist J. A hierarchical analysis of long-term illness and mortality in socially deprived areas. *Soc Sci Med* 2001;53:265–75. doi:10.1016/S0277-9536(00)00291-4.
- [148] Anderson RT, Sorlie P, Backlund E, Johnson N, Kaplan GA. Mortality effects of community socioeconomic status. *Epidemiology* 1997;8:42–7. doi:10.1097/00001648-199701000-00007.
- [149] Jaffe DH, Eisenbach Z, Neumark YD, Manor O. Individual, household and neighborhood socioeconomic status and mortality: a study of absolute and relative deprivation. *Soc Sci Med* 1982 2005;60:989–97. doi:10.1016/j.socscimed.2004.06.047.
- [150] Curtis S, Southall H, Congdon P, Dodgeon B. Area effects on health variation over the life-course: analysis of the longitudinal study sample in England using new data on area of residence in childhood. *Soc Sci Med* 1982 2004;58:57–74.
- [151] Kravdal Ø. A fixed-effects multilevel analysis of how community family structure affects individual mortality in Norway. *Demography* 2007;44:519–37.
- [152] Dahl E, Ivar Elstad J, Hofoss D, Martin-Mollard M. For whom is income inequality most harmful? A multi-level analysis of income inequality and mortality in Norway. *Soc Sci Med* 1982 2006;63:2562–74. doi:10.1016/j.socscimed.2006.06.002.
- [153] Martikainen P, Kauppinen TM, Valkonen T. Effects of the characteristics of neighbourhoods and the characteristics of people on cause specific mortality: a register based follow up study of 252,000 men. *J Epidemiol Community Health* 2003;57:210–7.
- [154] Jones K, Gould MI, Duncan C. Death and deprivation: an exploratory analysis of deaths in the health and lifestyle survey. *Soc Sci Med* 1982 2000;50:1059–79.
- [155] Lochner K, Pamuk E, Makuc D, Kennedy BP, Kawachi I. State-level income inequality and individual mortality risk: a prospective, multilevel study. *Am J Public Health* 2001;91:385–91.
- [156] Pierce RP, Williamson HA, Kruse RL. Distance, Use of Resources, and Mortality Among Rural Missouri Residents With Acute Myocardial Infarction. *J Rural Health* 1998;14:28–35. doi:10.1111/j.1748-0361.1998.tb00859.x.
- [157] Monnet E, Ramée C, Minello A, Jooste V, Carel D, Di Martino V. Socioeconomic context, distance to primary care and detection of hepatitis C: a French population-based study. *Soc Sci Med* 1982 2008;66:1046–56. doi:10.1016/j.socscimed.2007.11.044.
- [158] Campbell NC, Elliott AM, Sharp L, Ritchie LD, Cassidy J, Little J. Rural and urban differences in stage at diagnosis of colorectal and lung cancers. *Br J Cancer* 2001;84:910–4. doi:10.1054/bjoc.2000.1708.
- [159] Jones AP, Bentham G, Horwell C. Health service accessibility and deaths from asthma. *Int J Epidemiol* 1999;28:101–5.
- [160] Pilkington H, Blondel B, Drewniak N, Zeitlin J. Where does distance matter? Distance to the closest maternity unit and risk of foetal and neonatal mortality in France. *Eur J Public Health* 2014;24:905–10. doi:10.1093/eurpub/ckt207.
- [161] Ripley DCC, Kwong PL, Vogel WB, Kurichi JE, Bates BE, Davenport C. How does geographic access affect in-hospital mortality for veterans with acute ischemic stroke? *Med Care* 2015;53:501–9. doi:10.1097/MLR.0000000000000366.

- [162] Campbell NC, Elliott AM, Sharp L, Ritchie LD, Cassidy J, Little J. Rural factors and survival from cancer: analysis of Scottish cancer registrations. *Br J Cancer* 2000;82:1863–6. doi:10.1054/bjoc.1999.1079.
- [163] Nicholl J, West J, Goodacre S, Turner J. The relationship between distance to hospital and patient mortality in emergencies: an observational study. *Emerg Med J* 2007;24:665–8. doi:10.1136/emj.2007.047654.
- [164] Yamashita T, Kunkel SR. The association between heart disease mortality and geographic access to hospitals: county level comparisons in Ohio, USA. *Soc Sci Med* 1982 2010;70:1211–8. doi:10.1016/j.socscimed.2009.12.028.
- [165] Souza VC de, Strachan DP. Relationship between travel time to the nearest hospital and survival from ruptured abdominal aortic aneurysms: record linkage study. *J Public Health Oxf Engl* 2005;27:165–70. doi:10.1093/pubmed/fdi001.
- [166] Kravdal Ø. Does place matter for cancer survival in Norway? A multilevel analysis of the importance of hospital affiliation and municipality socio-economic resources. *Health Place* 2006;12:527–37. doi:10.1016/j.healthplace.2005.08.005.
- [167] Coldefy M, Com-Ruelle L, Lucas-Gabrielli V, Marcoux L. Les distances d'accès aux soins en France métropolitaine au 1er janvier 2007. *Irdes Rapp* 2011.
- [168] Lucas-Gabrielli V, Nabet N, Tonnelier F. Les soins de proximité : une exception française 2001.
- [169] Dejardin O, Remontet L, Bouvier AM, Danzon A, Trétarre B, Delafosse P, et al. Socioeconomic and geographic determinants of survival of patients with digestive cancer in France. *Br J Cancer* 2006;95:944–9. doi:10.1038/sj.bjc.6603335.
- [170] Singh-Manoux A, Clarke P, Marmot M. Multiple measures of socio-economic position and psychosocial health: proximal and distal measures. *Int J Epidemiol* 2002;31:1192–1199-1200.
- [171] Lynch JW, Kaplan GA, Salonen JT. Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. *Soc Sci Med* 1982 1997;44:809–19.
- [172] van Oort FVA, van Lenthe FJ, Mackenbach JP. Material, psychosocial, and behavioural factors in the explanation of educational inequalities in mortality in The Netherlands. *J Epidemiol Community Health* 2005;59:214–20. doi:10.1136/jech.2003.016493.
- [173] Chaix B, Chauvin P. Tobacco and alcohol consumption, sedentary lifestyle and overweightness in France: a multilevel analysis of individual and area-level determinants. *Eur J Epidemiol* 2003;18:531–8.
- [174] Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci* 2010;1186:125–45. doi:10.1111/j.1749-6632.2009.05333.x.
- [175] Ross CE. Walking, exercising, and smoking: does neighborhood matter? *Soc Sci Med* 1982 2000;51:265–74.
- [176] Duncan C, Jones K, Moon G. Smoking and deprivation: are there neighbourhood effects? *Soc Sci Med* 1999;48:497–505. doi:10.1016/S0277-9536(98)00360-8.
- [177] Renahy E, Parizot I, Vallée J, Chauvin P. Le renoncement aux soins pour raisons financières dans l'agglomération parisienne: déterminants sociaux et évolution entre 2005 et 2010. 2011.
- [178] Lacoste O, Spinosi L. Distance, proximité, accessibilité attraction et recours de la population vis-à-vis du système de soins. *Rhcp2s IIIe Extr Loos Nord--Calais* 2002.
- [179] Rey G. Surmortalité liée aux vagues de chaleur : Facteurs de vulnérabilité socio-démographique et causes médicales de décès. Université Paris XI, 2007.

- [180] Luo W, Qi Y. An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health Place* 2009;15:1100–7. doi:10.1016/j.healthplace.2009.06.002.
- [181] Brimblecombe N, Dorling D, Shaw M. Mortality and migration in Britain, first results from the British Household Panel Survey. *Soc Sci Med* 1999;49:981–8.
- [182] Martikainen P, Sipilä P, Blomgren J, van Lenthe FJ. The effects of migration on the relationship between area socioeconomic structure and mortality. *Health Place* 2008;14:361–6.
- [183] Norman P, Boyle P, Rees P. Selective migration, health and deprivation: a longitudinal analysis. *Soc Sci Med* 2005;60:2755–71.
- [184] van Lenthe FJ, Martikainen P, Mackenbach JP. Neighbourhood inequalities in health and health-related behaviour: results of selective migration? *Health Place* 2007;13:123–37.
- [185] Verheij RA, van de Mheen HD, de Bakker DH, Groenewegen PP, Mackenbach JP. Urban-rural variations in health in The Netherlands: does selective migration play a part? *J Epidemiol Community Health* 1998;52:487–93.
- [186] Bingenheimer JB, Raudenbush SW. Statistical and Substantive Inferences in Public Health: Issues in the Application of Multilevel Models. *Annu Rev Public Health* 2004;25:53–77. doi:10.1146/annurev.publhealth.25.050503.153925.
- [187] Chaix B. Geographic life environments and coronary heart disease: a literature review, theoretical contributions, methodological updates, and a research agenda. *Annu Rev Public Health* 2009;30:81–105. doi:10.1146/annurev.publhealth.031308.100158.
- [188] Merlo J, Chaix B, Yang M, Lynch J, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *J Epidemiol Community Health* 2005;59:1022–8.
- [189] Jonker MF, Donkers B, Chaix B, van Lenthe FJ, Burdorf A, Mackenbach JP. Estimating the Impact of Health-related Behaviors on Geographic Variation in Cardiovascular Mortality: A New Approach Based on the Synthesis of Ecological and Individual-level Data. *Epidemiol Camb Mass* 2015;26:888–97. doi:10.1097/EDE.0000000000000378.
- [190] Jackson C, Richardson S, others. Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors. *J R Stat Soc Ser A Stat Soc* 2008;171:159–178.
- [191] Kwan M-P. The Uncertain Geographic Context Problem. *Ann Assoc Am Geogr* 2012;102:958–68. doi:10.1080/00045608.2012.687349.
- [192] Chaix B, Merlo J, Chauvin P. Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilisation in France. *J Epidemiol Community Health* 2005;59:517–26. doi:10.1136/jech.2004.025478.
- [193] Monden CWS, van Lenthe FJ, Mackenbach JP. A simultaneous analysis of neighbourhood and childhood socio-economic environment with self-assessed health and health-related behaviours. *Health Place* 2006;12:394–403. doi:10.1016/j.healthplace.2005.03.003.
- [194] Silhol R, Zins M, Chauvin P, Chaix B. Investigating the spatial variability in incidence of coronary heart disease in the Gazel cohort: the impact of area socioeconomic position and mediating role of risk factors. *J Epidemiol Community Health* 2011;65:137–43. doi:10.1136/jech.2009.087379.

[195] Phelan JC, Link BG, Diez-Roux A, Kawachi I, Levin B. “Fundamental causes” of social inequalities in mortality: a test of the theory. *J Health Soc Behav* 2004;45:265–85.

## ANNEXE

Table S1 : Codes CIM de la Mortalité Evitable

Causes de décès	CIM9	CIM10
<b>Mortalité imputable au système de soins</b>		
Maladies intestinales infectieuses	001-009	A00-A09
Tuberculose	010-018,137	A15-A19, B90
Autres infections	037,032,045	A35-A36, A80
Coqueluche	033	A37
Bactéries invasives et infections protozoaires	034,035,036,038,4828, 084,3200-3205,3208, 3209,322,682	A38-A41, A46, A481, B50-B54, G00, G03, L03
Rougeole	55	B05
Tumeur maligne colorectale	153-154	C18-C21
Mélanome malin de la peau	172	C43
Autres tumeurs malignes de la peau	173	C44
Tumeur maligne du sein	174-175	C50
Tumeur maligne du cervix uteri	180	C53
Tumeur maligne corps utérin et non-spécifié de l'utérus	179,182	C54-C55
Tumeur maligne des testicules	186	C62
Tumeur maligne de la vessie	188	C67
Tumeur maligne de la glande thyroïde	193	C73
Maladie de Hodgkin	201	C81
Leucémie	204-208	C91-C95
Tumeurs bénignes	210-229	D10-D36
Affections de la glande thyroïde	240-246	E00-E07
Diabète sucré	250	E10-E14
Epilepsie et état de mal épileptique	345	G40-G41
Rhumatisme articulaire aigu et autres path. valculaires cardiaques	391,3920,393-398	I01-I09
Maladies hypertensives	401-405	I10-I15
Cardiopathies ischémiques	410-414	I20-I25
Maladies cérébrovasculaires	430-438	I60-I69
Autres maladies respiratoires	460-466,470-478 494,500-508, 510-519	J00-J08, J20-J39, J47-J99
Grippe (incluant grippe porcine)	487	J09-J11
Pneumonie	480-486	J12-J18
Maladies chroniques obstructives des voies respiratoire inf.	490-492,496	J40-J44
Asthme	493	J45-J46
Ulcères gastrique et duodenal	531-534	K25-K28
Douleur abdominale aigüe, appendicite, obstruction intestinale, Cholécystite / lithiases, pancréatite, hernie	540-543,550-553, 574-577	K35-K38, K40- K46, K80-K83, K85- K86, K91.5
Néphrite et néphrose	580-583,584,586, 588,589	N00-N07, N17- N19, N25, N27

Uropathie obstructive et hyperplasie de la prostate	593,5920,5921, 598,600	N13, N20-N21, N35, N40, N99.1
Grossesse, accouchement et puerpéralité	630-677	O00-O99
Complications de la période périnatale	760-779	P00-P96, A33
Malformations Congenitales et anomalies chromosomales	740-759	Q00-Q99
Accidents et complications au cours d'actes médicaux et chirurg.	870-876	Y60-Y69, Y83-Y84

#### **Mortalité imputable à l'absence de prévention**

Tuberculose	010-018,137	A15-A19, B90
Autres infections	037,032,045	A35-A36, A80
Coqueluche	33	A37
Hépatite virale	70	B15-B19
HIV	42	B20-B24
Tumeurs maligne de la lèvre, cavité orale, et pharynx	140-149	C00-C14
Tumeurs maligne de l'œsophage	150	C15
Tumeur maligne de l'estomac	151	C16
Tumeur maligne du foie	155	C22
Tumeur maligne de la trachée, des bronches et du poumon	162	C33-C34
Mélanome malin de la peau	172	C43
Autres tumeurs malignes de la peau	173	C44
Maladies liées à l'alcool excluant les causes externes	291,303,3050,3575, 4255,5353,5710-5713	F10, G312, G621, I426, K292, K70, K73, K741, K742, K746, K747, K748, K749, K860
Troubles liés à la consommation de drogues	304-305	F11-F16, F18-F19
Cardiopathies ischémiques	410-414	I20-I25
Thromboses et embolies pulmonaires	415,4511- 4513,4519,4539	I26, I80.1-I80.3, I80.9, I82.9
Anévrisme et dissection de l'aorte	441	I71
Grippe (incluant grippe porcine)	487	J09-J11
Pneumonie	480-486	J12-J18
Maladies chroniques obstructives des voies respiratoire inf.	490-492,496	J40-J44
Accidents de transport	800-848,9290,9291	V01-V99
Blessures accidentelles	880-888,910,850-869, 870-879,890-909, 911-928,9292-9299	W00-X59
Suicide	950-959	X60-X84, Y10-Y34
Homicide & agression	960-969	X85-Y09, U50.9

Table S2: Codes CIM des causes de décès investiguées

Causes de décès	CIM 8	CIM 9	CIM 10
Cardiopathies ischémiques	410-414	410-414	I20-I25
Maladies cérébrovasculaires	430-438	430-438	I60-I69
Tumeurs	140-239	140-239	C00-D48
Voies aéro-digestives supérieures	140-150	140-150	C00-C15
Poumon	161-162	161-162	C32-C34
Colorectale	153	153	C18
Vessie	188	188	C67
Sein	174	174-175	C50
Prostate	185	185	C61
Maladies respiratoires	460-519	460-519	J00-J99
Morts violentes	800-999	800-999	V01-Y89
Abus d'alcool et maladies du foie	291, 303, 571	291, 303, 571	F10, K70, K73, K74

## **Document de travail, rédigé en 2014 par Karim Bounebaché<sup>1</sup> et Grégoire Rey<sup>1</sup> :**

<sup>1</sup>*Inserm CépiDc, Centre Epidémiologique sur les Causes Médicales de Décès, le Kremlin-Bicêtre*

### **Méthodes et Résultats de l'appariement de la base du CépiDc avec l'EDP**

#### **Introduction :**

Ce document présente brièvement les différentes étapes réalisées lors de l'appariement entre la base du CépiDc et la base EDP\_BE2011\_DECES. On nommera dans la suite de ce document « baseaa » la base de donnée EDP\_BE2011\_DECES que nous cherchons à apparié avec la base des causes médicales de décès.

#### **L'EDP:**

La base EDP\_BE2011\_DECES contient un identifiant, la date de naissance, la date de décès, le département de domicile et le département de naissance. Cependant, dans l'EDP il y a eu d'importants trous de collecte des bulletins de décès, et ceux de 1974 à 1990 n'ont pas été intégrés à l'EDP. Par ailleurs seuls les bulletins de décès de personnes nées entre 1<sup>er</sup> et le 4 octobre, de 1991 à 1997 sont intégrés. Pour combler les trous, un rapprochement a été fait avec le RNIPP, par le biais de la création de bulletins fantômes partiellement remplis et ne contenant que la date et le lieu de décès.

Pour trouver les informations manquantes nous avons utilisé la base EDP\_BE2011\_INDIVIDU contenant les informations suivantes : l'identifiant cité précédemment, le sexe, la date de naissance, le lieu de naissance, la date de décès et le lieu de décès. Les individus décédés à l'étranger ont aussi été retirés de la base (puisque la base des causes médicales de décès ne contient que les décès ayant eu lieu en France). Les individus dont il n'a pas été possible de retrouver la date de décès ont été retirés.

Enfin, nous avons standardisé les variables d'appariement au format de la base de Décès en décomposant les dates en trois variables jour, mois et année au format texte ainsi que le code commun en deux variables département et commune.

#### **Les causes médicales de décès :**

La base du CépiDc contient donc les causes médicales de décès (de la cause immédiate à la cause initiale), mais aussi les informations suivantes : le sexe, la date de naissance, la date de décès, le département et la commune de décès, le département et la commune de domicile, et enfin le département et la commune de naissance (pour les décès ayant eu lieu après 2000). Ces deux dernières variables concernant le lieu de naissance ont été retirées lors du dépôt de la base dans le CASD. Par ailleurs cette base ne contient aujourd'hui que les décès de 1968 à 2010.



### **La baseaa (base à appareiller) :**

Au final la base nommée « baseaa » contient les décès de 1968 à 2010 de la base EDP\_BE2011\_DECES décédés en France, ce qui représente un effectif de 223 162 individus.

### **Règle de décision pour l'appariement :**

Les variables d'appariement retenues sont : sexe, date de naissance, date de décès, département de décès et commune de décès.

L'appariement réalisé a suivi les règles suivantes :

Dans un premier temps les doublons de la base « baseaa » sont marqués « doublons<sup>1</sup> » (dans la colonne « résultat » de la base de restitution), puis la base est soumise à un premier appariement avec la base du CépiDc. Les associations trouvées lors de cette étape sont dites « Totales ».

Les non trouvés dans l'étape précédente sont soumis à une succession d'appariements où l'on a retiré un des champs de la liste (ce champ est marqué par la colonne « vardif » de la table de restitution). Les associations trouvées lors de ces étapes successives sont dites « Partielles ».

Si un individu de la base « baseaa » est associé à plusieurs individus de la base du CépiDc, l'appariement est dit « multiple » sinon il est « unique » (cette information est contenue dans la colonne « résultat »). Si un individu de la base « baseaa » est associé à un individu de la base du CépiDc qui a déjà de multiples associations dans la base baseaa, l'appariement est dit « inter ».

### **Résultat :**

Le tableau ci-dessous résume les résultats obtenus lors de l'appariement

Résultat	Fréquence	Pourcentage
Non trouvé	1568	0,70%
Partiel doublon multiple	26	0,01%
Partiel doublon unique	5	0,00%
Partiel multiple	180	0,08%
Partiel unique	12497	5,59%
Total doublon multiple	92	0,04%
Total doublon unique	66	0,03%
Total multiple	106	0,05%
Totale unique	208828	93,49%
Taux d'appariement	99,18%	

---

<sup>1</sup> Les doublons sont les individus présents plusieurs fois dans la base au regard de la liste de variables utilisées pour l'appariement.

Les pourcentages sont calculés sur la base de l'ensemble des associations trouvées et des individus non trouvés lors de l'appariement. Le taux d'appariement est le ratio du nombre des appariements uniques (total et partiel) sur la taille de la base « baseaa ».

## ARTICLES