



HAL
open science

Etude transversale et longitudinale des relations entre les environnements géographiques de vie et l'excès de poids et le risque cardiovasculaire

Antoine Lewin

► **To cite this version:**

Antoine Lewin. Etude transversale et longitudinale des relations entre les environnements géographiques de vie et l'excès de poids et le risque cardiovasculaire. Santé publique et épidémiologie. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066072 . tel-01159542

HAL Id: tel-01159542

<https://theses.hal.science/tel-01159542>

Submitted on 3 Jun 2015

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.

THESE DE DOCTORAT DE L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité : Epidémiologie

École Doctorale 393 Pierre Louis de Santé Publique – Epidémiologie et Sciences de l'Information Biomédicale

Présentée par

Antoine LEWIN

Pour obtenir le grade de

DOCTEUR

de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse

ÉTUDE TRANSVERSALE ET LONGITUDINALE DES RELATIONS ENTRE LES
ENVIRONNEMENTS GEOGRAPHIQUES DE VIE ET L'EXCES DE POIDS ET LE RISQUE
CARDIOVASCULAIRE

Soutenue le 13 mai 2015

Devant le jury de soutenance composé de :

Basile CHAIX	Directeur de Recherche	Directeur de thèse
Chantal SIMON	Professeur des Universités – Praticien Hospitalier	Rapporteur
Cyrille DELPIERRE	Chargé de Recherche	Rapporteur
Grégoire REY	Ingénieur de Recherche	Examineur
Jean-Michel OPPERT	Professeur des Universités – Praticien Hospitalier	Examineur

Thèse réalisée dans l'unité INSERM UMR-S 1136 : « Institut Pierre Louis d'Epidémiologie et de Santé Publique » au sein de l'équipe sur les déterminants sociaux de la santé et du recours aux soins, groupe RECORD « Residential Environment and CORonary heart Disease

A mes parents...

REMERCIEMENTS

Je tiens tout d’abord à remercier très chaleureusement mon directeur de thèse, Basile Chaix, pour m’avoir permis de réaliser ce travail. Je le remercie également pour ses qualités tant professionnelles que personnelles, pour son engagement, sa disponibilité ainsi que sa patience. Toutes nos discussions enrichissantes auront permis de faire que cette thèse soit, pour moi, un réel accomplissement.

Je souhaite remercier Chantal Simon ainsi que Cyrille Delpierre, qui me font l’honneur de me consacrer de leur temps afin d’être les rapporteurs de cette thèse. Je remercie aussi Grégoire Rey et Jean-Michel Oppert qui ont accepté, en faisant partie des membres du jury, d’évaluer ce travail.

Je remercie Dominique Costagliola, directrice de l’Unité Inserm 1136, et Pierre Chauvin, responsable d’équipe au sein de cette Unité, pour leur accueil ainsi que les supports matériels et logistiques mis à ma disposition, qui m’ont permis de réaliser ce travail dans les meilleures conditions possibles.

Je tiens également à remercier toute l’équipe ainsi que les partenaires de l’Etude RECORD. Merci également au Centre d’Investigations Préventives et Cliniques et spécialement à Frédérique et Bruno pour leur aide logistique et méthodologique dans l’élaboration de ce travail.

Je remercie tout particulièrement mes collègues de travail, qui ont su me faire profiter de chaque instant durant ces trois années dans cette ville nouvelle. Un merci spécialement à Audrey, Camille, Julie, Noëlla, Ruben et Valérie qui ont su rendre mes journées de travail (ainsi que les soirées sans travail) très agréables. Merci pour tous ces bons moments...

Mes remerciements seraient incomplets si je ne m'adressais pas à ma famille, mon père, Jean, et ma mère, Michèle, qui m'ont inconditionnellement soutenu et encouragé, par tous les moyens possibles, dans mes choix. Je ne vous remercierai jamais assez même si vous ne me l'avez jamais demandé. Merci également à mes frères, Mathieu et Rémy, qui ont toujours été là pour moi. Je remercie aussi Martine et Fred, ma famille, qui m'ont accueilli comme des parents pendant ces trois années à Paris. C'était un régal de vous voir si souvent, je vous attends à Montréal. Merci à Léonore et Pierre pour tous les moments que nous avons pu partager ensemble et ceux à venir ! En ce moment, je pense aussi bien fort à mes grands-parents. Ce travail est également le leur. Je remercie aussi ma Marie-Lyne qui même loin a toujours été près de moi, dans mon cœur. Merci pour ta patience et ton soutien et merci à Louise aussi ! Merci à mes amis qui m'ont soutenu dans toutes les situations ainsi que durant ces trois années.

J'exprime ma gratitude aux relecteurs de ce manuscrit : Antoine, Camille, Jean, Julie, Martine, Michèle, Noëlla, Rémi et Ruben.

Un grand merci aussi à tous ceux que j'aurais pu oublier.

RÉSUMÉ

Face aux limites de la littérature internationale sur les disparités environnementales de prévalence d'obésité (focus sur le lieu de résidence, etc.) et considérant le manque d'études empiriques réalisées en France, nous nous sommes intéressés tout d'abord aux relations entre, d'une part, les caractéristiques des environnements résidentiels et professionnels et le secteur d'activité économique et, d'autre part, deux indicateurs d'excès de graisse corporelle : l'indice de masse grasse (IMG) et le pourcentage de masse grasse (%MG). Ensuite, nous avons comparé les effets des secteurs d'activité économique sur plusieurs facteurs de risque cardiovasculaire. Finalement, à l'aide d'une étude longitudinale, nous avons évalué les liens entre les variables socio-économiques et l'évolution des paramètres anthropométriques chez les participants de l'Étude RECORD. Parallèlement à l'étude de ces relations, notre objectif méthodologique était d'explorer les biais d'attrition liés aux données manquantes dans les études longitudinales.

Dans la première partie de ce travail, les données de 4331 participants de l'Étude RECORD, géolocalisés à leur résidence et à leur lieu de travail, ont été analysées. La composition corporelle a été évaluée par pèse personne à impédance bioélectrique. Des régressions linéaires multiniveaux ont été utilisées pour étudier les déterminants de l'IMG et du %MG. Les résultats montrent qu'après ajustement, chez les hommes, l'IMG et le %MG augmentaient indépendamment à mesure que la densité de population et le niveau d'éducation du quartier de résidence diminuaient. Chez les femmes, le niveau d'éducation résidentiel était lié négativement au %MG. Chez les hommes, un IMG et un %MG plus faibles étaient observés chez les participants qui travaillent dans les secteurs de la construction et des

transports/communications par rapport à ceux travaillant dans le secteur de l'éducation. Pour les femmes, l'IMG était plus élevé chez les personnes travaillant dans le secteur de l'administration publique et le secteur de la santé et du travail social que dans le secteur des transports et des communications. Une longue distance domicile-travail était associée à un IMG plus élevé chez les femmes. De plus, nous avons pu montrer que l'Indice de Masse Corporelle (IMC) ne permettait pas de saisir pleinement les effets des secteurs d'activité économique sur la masse grasse.

Dans une seconde partie, les données de 4360 participants de l'Étude RECORD ayant un travail ont été analysées. Dix facteurs de risque cardiovasculaire ont été évalués : IMC, le tour de taille, la pression artérielle systolique (PAS) et diastolique (PAD), la pression pulsée, le cholestérol total, la glycémie, le cholestérol des lipoprotéines de haute densité (HDL), le cholestérol des lipoprotéines de basse densité (LDL), et la fréquence cardiaque au repos. Des modèles de régression linéaire multiniveaux stratifiés par sexe et ajustés sur les caractéristiques sociodémographiques individuelles et résidentielles ont été estimés. Les résultats de ces analyses montrent que, chez les hommes, le secteur de la santé et du travail social était le secteur le plus protecteur pour l'IMC, le tour de taille et la glycémie (tandis que le secteur de la construction et le secteur des transports et des communications avaient tendance à être les plus défavorables pour ces facteurs de risques). Le secteur de la santé et du travail social était également associé à un taux de cholestérol HDL plus élevé chez les hommes. Cependant, à l'opposé des résultats pour l'IMC et le cholestérol, les hommes qui travaillent dans le secteur de la santé et du travail social avaient les niveaux de pression sanguine systolique et de pression pulsée les plus élevés. En ce qui concerne les femmes, celles qui travaillent dans le secteur de la santé et du travail social avaient un IMC et un tour de taille plus élevés ainsi que des pressions systolique et diastolique

plus élevées. Le secteur du commerce et de la réparation de véhicules, le secteur des transports et des communications et le secteur des services collectifs, sociaux et personnels étaient associés à un profil plus favorable pour ces facteurs de risque chez les femmes. Au total, il n'existait pas de secteurs d'activité uniformément associés à un surcroît de risque chez les hommes et chez les femmes pour l'ensemble ou la plupart des facteurs de risques examinés.

Finalement, dans la troisième partie de ce travail, nous avons utilisé les données longitudinales de l'Étude de cohorte RECORD comprenant deux vagues (N = 3693 participants, taux de rétention = 51%). Nos premières analyses ont étudié l'effet du niveau d'éducation individuel sur le changement d'IMC. Parallèlement, d'un point de vue méthodologique, afin d'évaluer l'ampleur du biais d'attrition, différents scénarios d'attrition ont été simulés en générant des données manquantes sur la variable dépendante (changement d'IMC) selon les mécanismes manquant au hasard (Missing At Random, MAR) et manquant non au hasard (Missing Not At Random, MNAR). Les analyses des cas complet (Complete Case Analyses, CCA), la pondération par l'inverse de probabilité (Inverse Probability Weighting, IPW) et l'imputation multiple (Multiple Imputation, MI) ont été évaluées afin de voir leur capacité à tenir compte du biais d'attrition. Lors des simulations, les estimations finales ont été calculées à partir de 200 répliques de nos scénarios d'attrition. Nos résultats montrent que quand aucune observation manquante n'a été ajoutée à celles déjà présentes dans les données observées, un faible niveau d'éducation est associé à un changement d'IMC plus important pour les trois méthodes (CCA, IPW et MI). De plus nous montrons qu'à ce stade, la correction avait peu d'impact. Lors de la génération supplémentaire de données manquantes sur la variable dépendante suivant un mécanisme d'attrition MAR, le biais augmentait avec le niveau d'attrition lorsque les analyses étaient conduites avec les méthodes CCA ou MI. La méthode statistique IPW semblait

correctement gérer le biais d'attrition. Finalement, ni MI ni IPW ne semblaient pouvoir surmonter le biais introduit par des données manquantes sur la variable dépendante quand le mécanisme d'attrition était MNAR.

Pour conclure, nous avons pu par ces travaux empiriques mettre en évidence des effets des caractéristiques individuelles, de l'environnement résidentiel, et des secteurs d'activité économique sur l'accumulation de la masse grasse corporelle ainsi que sur les facteurs de risque cardiovasculaire. En ce qui concerne le volet méthodologique de ce travail, d'une part nous avons été en mesure de tenir compte dans les analyses de l'environnement géographique de résidence communément pris en compte. D'autre part, nous démarquant d'une littérature sur les déterminants sociaux de l'obésité essentiellement transversale, nous montrons que selon le mécanisme des données manquantes, les méthodes statistiques communément utilisées ne permettent pas de contourner les biais potentiels induits par l'attrition dans les analyses longitudinales.

Mots clés : Environnements géographiques de vie, secteur d'activité économique, masse grasse, facteur de risque cardiovasculaire, système d'information géographique, biais d'attrition, méthodes statistiques.

ABSTRACT

Given the limits of the international literature on the environmental disparities in obesity prevalence (focused on the residential neighborhood) and considering the lack of empirical studies conducted in France, we were interested in the relations between, on the one hand, the characteristics of residential and professional neighborhood characteristics and the work economic sector, and on the other hand, two body fat indicators: the Fat Mass Index (FMI) and the percentage of Fat Mass (%FM). Then we compared the effects of the work economic sectors on several cardiovascular risk factors. Finally using a longitudinal study, we evaluated the links between socioeconomic variables and the evolution of the anthropometric parameters among participants of the RECORD Study. Simultaneously, our methodological objective was to explore the attrition biases due to missing data in longitudinal studies.

In the first part of this work, data of 4331 participants from the RECORD Study, localized at their residence and at their workplace, were analyzed. Body composition was assessed using bioelectrical impedance analyzer. Multilevel regression analyses were used to study the determinants of FMI and %FM. Results showed that, after adjustment, in men, the FMI and the %FM increase independently while the residential neighborhood population density and education level decrease. For the women, education level was negatively associated with the %FM. For the men, a weaker FMI and %FM were observed on participants working in the construction and in the communication/transport sectors compared to those working in the education sector. For the women the FMI was higher for those working in the public and in the

health and social sectors rather than those working in the transport and communication sector. A long home-work distance was associated with a high FMI for women. Moreover we showed that the BMI was not able to fully capture the effects of the work economic sectors on the fat mass.

In a second part, data of 4360 employed participants from the RECORD survey have been analyzed. Ten cardiovascular risk factors were assessed: body mass index (BMI), waist circumference, systolic and diastolic blood pressure (BP), pulse pressure, total cholesterol, glycaemia, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and resting heart rate. Multilevel linear regression models stratified by sex and adjusted for individual and neighborhood sociodemographic characteristics were estimated. Results shown that among men, the Health and social work sector was found to be the most protective sector for BMI, waist circumference, and glycaemia (while the Construction sector and the Transport and communications sector tended to be unfavorable for these outcomes). The Health and social work sector was also associated with a higher HDL cholesterol among men. However, men working in the Health and social work sector showed the highest systolic BP and pulse pressure. Women working in the Health and social work sector had the highest BMI, the largest waist circumference, and the most elevated systolic and diastolic BP. The Commercial and repair of vehicles sector, the Transport and communication sector, and the Collective, social, and personal services sector were associated with a more favorable profile for these risk factors among women. In conclusion, there were no work economic sectors associated with an increase risks for both women and men for all or most of the examined risk factors.

Finally, in the third part of this work, we used longitudinal data from the first and second waves of the French RECORD Cohort Study (N = 3,693 participants, retention rate = 51%). Our

first analyses studied the effect of the individual education level on the change in BMI. After analyzing attrition bias in the observed data, different scenarios of attrition were simulated by generating additional missing data in BMI at the follow-up under Missing At Random (MAR) and Missing Not At Random (MNAR) mechanisms. Complete Case Analyses (CCA), Inverse Probability Weighting (IPW), and Multiple Imputation (MI) analyses were assessed in their ability to account for attrition bias. In the simulation part, the final estimates were derived from 200 replications of our scenarios of attrition. Our results shown that when no missing observations were added to those already present in the data, a negative association was found between individual education and the change in BMI, in CCA as well as with IPW and MI (the correction had little impact). When generating additional observations with unobserved follow-up outcomes under a MAR pattern, the bias increased with the attrition level when the analyses were based on CCA or MI. IPW correctly handled the attrition bias. MI and IPW were not useful to overcome the bias introduced by missing outcome data under MNAR.

To conclude, we could by those empirical works highlight the effect of the individual and the residential characteristics, and the work economic sectors on the accumulation of body fat as well as on cardiovascular risk factors. Regarding the methodological part of this work, firstly we were able to take into account in the analysis the residential geographical environment commonly considered. Secondly, differentiating us from the transversal literature on the social determinants of obesity, we show that depending on the missing data mechanism, the statistical methods commonly used can not circumvent the potential bias introduced by attrition in longitudinal analyses.

Keywords: *Environmental characteristics, work economic sector, fat mass, cardiovascular risk factor, geographic information system, attrition bias, statistical methods.*

VALORISATION DES TRAVAUX DE THESE

ARTICLES PUBLIES

Chaix B, Kestens Y, Duncan S, Merrien C, Thierry B, Pannier B, Brondeel R, **Lewin A**, Karusisi N, Perchoux C, Thomas F and Méline J; 2013, Active transport and public transport use to achieve physical activity recommendations? A combined GPS, accelerometer, and mobility survey study, Intern J of Behav Nutr and Phys Act, 2014, **11**:124; doi:10.1186/s12966-014-0124-x

Lewin A, Thomas F, Meline J, Chaix B; 2013, Work economic sectors and cardiovascular risk factors: the RECORD Study, BMC Public Health. 2014 Jul; 14:750 doi:10.1186/1471-2458-14-750

Chaix B, Méline J, Duncan S, Jardinier L, Perchoux C, Vallée J, Merrien C, Karusisi N, **Lewin A**, Brondeel R, Kestens Y.; 2013, Neighborhood environments, mobility, and health: Towards a new generation of studies in environmental health research. Rev Epidemiol Sante Publique. 2013 Aug; 61 Suppl 3:S139-45. doi: 10.1016.

Chaix B, Méline J, Duncan S, Merrien C, Karusisi N, Perchoux C, **Lewin A**, Labadi K, Kestens Y; 2013, GPS tracking in neighborhood and health studies: a step forward for environmental exposure assessment, a step backward for causal inference? Health Place. 2013 May; 21:46-51. doi: 10.1016.

Lewin A, Pannier B, Meline J, Karusisi N, Thomas F, Chaix B. Residential neighborhood, geographic work environment, and work economic sector: associations with body fat measured by bio-electrical impedance in the RECORD Study. *Ann Epidemiol.* 2014 Mar; 24(3):180-6. doi: 10.1016

Lewin A, Buteau S, Brand A, Kosatsky T, Smargiassi A.; 2013, Short-term risk of hospitalization for asthma or bronchiolitis in children living near an aluminum smelter. *J Expo Sci Environ Epidemiol.* 2013 Sep-Oct; 23(5):474-80. doi: 10.1038

ARTICLES SOUMIS

Lewin A, Brondel R, Thomas F, Chaix B, Accounting for attrition bias in the association between education and change in body mass index: a longitudinal simulation study based on the RECORD Cohort Study.

COMMUNICATIONS

47th SER (Society for Epidemiological Research) annual meeting, **Lewin A**, Thomas F, Pannier B, Chaix B. Work economic sectors and cardiovascular risk factors: the RECORD Study. Seattle, Washington, USA, June 24-28, 2014. Poster

Urban Dynamics and Health, **Lewin A**, Pannier B, Méline J, Karusisi N, Thomas F, Chaix B. Residential neighborhood, geographic work environment, and work economic sector: associations with body fat measured by bio-electrical impedance in the RECORD Study. Nanterre, France, September 18-21, 2013.

46th SER (Society for Epidemiological Research) annual meeting, **Lewin A**, Pannier B, Méline J, Karusisi N, Thomas F, Chaix B. Residential neighborhood, geographic work environment, and work economic sector: associations with body fat measured by bio-electrical impedance in the RECORD Study. SER conference, Boston, Massachusetts, USA, June 18-21, 2013. Poster

LISTE DES ABRÉVIATIONS

AIC	Akaike information criterion
CNIL	Commission Nationale de l'Informatique et des Libertés
CNAV	Caisse Nationale d'Assurance Vieillesse
IAU	Institut d'Aménagement et d'Urbanisme
INSEE	Institut National de la Statistiques et des Etudes Economiques
IPC	Centre d'Investigations Préventives et Cliniques
IRIS	Îlots Regroupés pour l'Information Statistique
NAF	Nomenclature d'Activités Française
OMS	Organisation Mondiale de la Santé
RECORD	Residential Environment and CORonary heart Disease
SIG	Système d'Informations Géographiques
SIRET	Système d'Identification du Répertoire des ÉTablissements
TRIRIS	Découpage géographique infra-communal qui correspond à trois IRIS

LISTE DES FIGURES ET TABLEAUX

Figure 1 : Evolution de l'obésité aux Etats-Unis dans la population Adulte de 1990 à 2013.

Figure 2 : Toile causa de l'obésité.

Figure 3 : Localisation des 112 communes sélectionnées dans l'Etude RECORD et dans les 4 sites du Centre IPC participant au recrutement.

TABLE DES MATIÈRES

REMERCIEMENTS	4
RÉSUMÉ	6
ABSTRACT	10
VALORISATION DES TRAVAUX DE THESE	14
ARTICLES PUBLIES	14
ARTICLES SOUMIS	16
COMMUNICATIONS	16
LISTE DES ABRÉVIATIONS	17
LISTE DES FIGURES ET TABLEAUX	18
TABLE DES MATIÈRES	19
ORGANISATION DU MÉMOIRE	21
CHAPITRE I - INTRODUCTION	22
I.1. L'OBESITE - DEFINITION ET CONTEXTE GENERAL.....	22
I.2. MESURES DE L'OBESITE.....	26
<i>I.2.1 - Mesures anthropométriques</i>	26
<i>I.2.2 - Mesures d'adiposité</i>	27
I.3. DETERMINANTS DE L'OBESITE.....	32
<i>I.3.1 - Niveau individuel</i>	33
<i>I.3.2 - Niveau environnemental</i>	35
<i>I.3.3 - Niveau des secteurs d'activité économique</i>	41
I.4. ANALYSES TRANSVERSALES ET LONGITUDINALES DES DETERMINANTS DE L'OBESITE.....	44
<i>I.4.1 – Schémas d'analyses utilisés</i>	44
<i>I.4.2 – Données manquantes</i>	46
OBJECTIFS	49
CHAPITRE II - METHODES	51
II.1. ÉTUDE RECORD : CONTEXTE ET OBJECTIFS.....	51

II.2. TERRITOIRE D'ÉTUDE ET POPULATION.....	52
II.3. LES DONNEES COLLECTEES	54
<i>II.3.1 - Les données individuelles.....</i>	<i>54</i>
<i>II.3.2 - Informations sur le lieu de travail des participants RECORD</i>	<i>56</i>
<i>II.3.3 - Secteurs d'activité économique et catégories légales des établissements.....</i>	<i>57</i>
<i>II.3.4 - Les données environnementales.....</i>	<i>58</i>
II.4. ANALYSES STATISTIQUES	61
CHAPITRE III - RELATIONS ENTRE LES CARACTERISTIQUES ENVIRONNEMENTALES RESIDENTIELLES ET PROFESSIONNELLES, LES SECTEURS D'ACTIVITE ECONOMIQUE ET LA MASSE GRASSE.....	65
III.1. PRESENTATION DE L'ARTICLE.....	65
III.2. ARTICLE INTEGRAL	70
CHAPITRE IV - RELATIONS ENTRE LES SECTEURS D'ACTIVITE ECONOMIQUE ET LES FACTEURS DE RISQUE CARDIOVASCULAIRE.....	88
IV.1. PRESENTATION DE L'ARTICLE.....	88
IV.2. ARTICLE INTEGRAL	93
CHAPITRE V - LES BIAIS D'ATTRITION DANS LA RELATION ENTRE LE NIVEAU D'EDUCATION ET LE CHANGEMENT D'IMC : ETUDE LONGITUDINALE ET ETUDE DE SIMULATION.....	114
V.1. PRESENTATION DE L'ARTICLE.....	114
V.2. ARTICLE INTEGRAL	119
CHAPITRE VI - DISCUSSION	155
VI.1. LES PRINCIPAUX RESULTATS	155
VI.2. LIMITES	161
VI. 3. INTERETS DES TRAVAUX REALISES	163
CONCLUSION ET PERSPECTIVES	166
ANNEXE.....	169
RÉFÉRENCES BIBLIOGRAPHIQUES.....	170
CITES DANS LE MANUSCRIT	170
CITES DANS LES ARTICLES.....	186

ORGANISATION DU MÉMOIRE

Conformément aux attentes de l'école doctorale, la structure de ce mémoire de thèse a été faite sous la forme d'une « thèse sur articles ».

Ce mémoire commence par une introduction générale qui structure notre thématique de recherche selon un état de l'art concernant la définition et les causes de l'obésité ainsi que les recherches basées sur les effets des environnements sur l'obésité. Cette introduction se termine par la description des objectifs de ce travail de thèse.

Le deuxième chapitre décrit les principaux éléments méthodologiques qui composent notre base de données ainsi que les méthodes statistiques mises en œuvre dans le cadre de ce travail.

Les troisième, quatrième et cinquième chapitres sont consacrés aux travaux empiriques menés dans le cadre de cette thèse. Ces chapitres commencent chacun par une présentation synthétique de l'article, suivie du texte intégral. Le premier article, présenté dans le troisième chapitre, a été publié dans la revue *Annals of Epidemiology*. Le deuxième article, présenté dans le quatrième chapitre, a été publié dans la revue *BMC Public Health* et le troisième article, présenté dans le cinquième chapitre, a été soumis à la revue *Epidemiology*.

Le sixième chapitre est consacré à une discussion générale qui reprend l'ensemble des résultats obtenus dans le cadre de ce travail et les compare à la littérature existante.

Enfin, ce mémoire de thèse se termine par une partie générale qui conclut et suggère des perspectives pour de futures recherches.

CHAPITRE I - INTRODUCTION

I.1. L'obésité - définition et contexte général

L'obésité est définie comme une accumulation anormale ou excessive de graisse corporelle (tissus adipeux), qui représente un risque pour la santé (1). L'obésité résulte d'un déséquilibre dû au fonctionnement du corps humain entre les apports énergétiques (ou énergie des entrées) et l'énergie dépensée (ou énergie des sorties) d'un individu (2). L'obésité peut être catégorisée de source endogène (troubles métaboliques, génétiques ou endocriniens) et/ou exogène (bilan énergétique positif) (3, 4). Les conséquences d'un déséquilibre du bilan énergétique peuvent être un accroissement des réserves stockées sous forme de triglycérides dans le tissu graisseux qui entraîne de nombreuses complications avec des conséquences somatiques, psychologiques, sociales et un retentissement sur la qualité de vie des personnes concernées (5, 6). En effet, la masse grasse est associée à un certain nombre de comorbidités (7, 8) telles que les maladies cardiovasculaires (8), l'hypertension (8), le diabète de type 2 (8), les maladies respiratoires (8) et certains cancers (8).

Avant le 19^e siècle, chez les hommes, l'accroissement excessif de graisse corporelle et la forte corpulence étaient signes de bonne santé et de bonheur tant qu'elles ne dépassaient pas l'incapacité physique à réaliser des activités de tous les jours telles que, par exemple, celle de monter à cheval. Cependant, chez la femme la « normalité » corporelle, liée à la beauté, a toujours été recherchée ; pour le haut du corps à la Renaissance et sur l'ensemble du corps au début de l'ère industrielle (9). Par conséquent, l'impact de l'obésité sur la qualité de vie n'a pas

été largement envisagé avant le 19^e siècle. Ce n'est qu'au 20^e siècle qu'une préoccupation de Santé publique concernant l'obésité est née (10). Par la suite, il a fallu attendre les années 1970, d'abord aux États-Unis, pour voir apparaître les premières études épidémiologiques portant sur l'obésité qui reconnaissent que l'accumulation de graisse corporelle est un problème sanitaire (11-13). Compte tenu du développement croissant et global du phénomène (14), en 1997, par le biais de l'Indice de Masse Corporelle (IMC), l'Organisation Mondiale de la Santé (OMS) a reconnu le surpoids et l'obésité (IMC \geq 25 et IMC \geq 30, respectivement) comme une maladie et a défini celle-ci comme : « une accumulation anormale ou excessive de graisse corporelle qui peut nuire à la santé » (15). Depuis une dizaine d'années, l'obésité n'a cessé de croître dans la plupart des pays développés et en voie de développement (16-19). Ce phénomène a conduit l'OMS à caractériser l'obésité comme un enjeu majeur de Santé publique et à qualifier d'épidémie l'augmentation de la prévalence de l'obésité et des problèmes de surpoids (15). Cela classe cette épidémie comme la première de l'histoire de l'humanité causée par une maladie non infectieuse (20). En 2014, l'OMS associe l'obésité et le surpoids à 3,4 millions de décès d'adultes par année (44% liés au diabète, 23% aux cardiopathies ischémiques et de 7 à 41% à certains cancers) plaçant ainsi cette maladie au cinquième rang des facteurs de risque de décès au niveau mondial (1).

Situation de l'obésité dans le monde¹

Depuis des dizaines d'années, partout dans le monde, on observe un développement épidémique de l'obésité (15). A titre d'exemple, la figure 1 montre l'évolution de l'obésité aux États-Unis sur une période de 20 ans. En 1990, 10 à 15% de la population des États-Unis était obèse. Dix ans

¹ Mesuré par l'Indice de Masse Corporelle (IMC)

après, la prévalence de l'obésité était au moins de 20% pour la moitié des états des États-Unis. Finalement en 2013, les trois quarts des états ont une prévalence de l'obésité supérieure à 25% (21, 22).

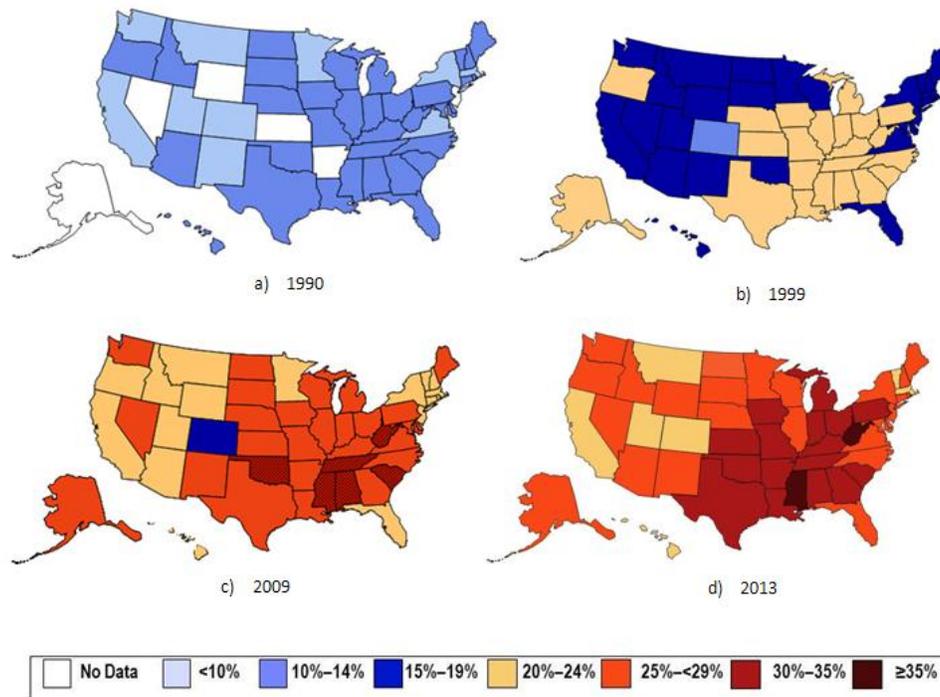


Figure 1 : Evolution de l'obésité (IMC \geq 30) aux États-Unis dans la population adulte entre 1990 et 2013 (22)

Plusieurs études montrent que la progression de l'obésité aux États-Unis est représentative et transposable à l'ensemble des pays du monde (mis à part le Japon et la Finlande où la prévalence reste stable) avec pour seule différence un décalage temporel (15, 23-26). En Europe, la moyenne de la prévalence de l'obésité adulte est de 16,6%. La Hongrie, le Royaume-Uni, l'Irlande, Malte et le Luxembourg sont les Etats les plus fortement touchés par cette épidémie,

avec plus de 22% de la population obèse à l'âge adulte. Les pourcentages d'adultes obèses les plus bas sont observés en Roumanie (7,9%), en Italie (10,3%) et au Pays-Bas (11,4%) (27-29).

Situation de l'obésité en France²

En 2012, la France se situe en dessous de la moyenne européenne avec une prévalence de l'obésité de 15% et du surpoids de 32,2% chez les adultes de plus de 18 ans (30). L'enquête épidémiologique nationale sur le surpoids et l'obésité (ObEpi) mise en place en 1997 sur une base trisannuelle montre que, depuis une quinzaine d'années, l'obésité en France est en forte augmentation (30). En effet, en 1997, l'étude Obépi mettait en évidence une prévalence de l'obésité de 8,5% chez les adultes de plus de 18 ans. Depuis, les cinq études successives ont révélé une augmentation graduelle de la prévalence de l'obésité, soit 10,1% en 2000 ; 11,9% en 2003 ; 13,1% en 2006 ; 14,5% en 2009 pour finalement atteindre 15% en 2012 soit une augmentation moyenne trisannuelle de 1,3 point de pourcentage (30). Cette augmentation concerne également l'obésité dite sévère ($IMC \geq 35$) : en 1997 elle était de 1,5% pour atteindre 4,3% en 2012 (30). En 15 ans, la prévalence de l'obésité chez les adultes de plus de 18 ans a augmenté de façon relative de 76% passant de 3,56 à 6,92 millions d'individus obèses. Le tour de taille moyen a quant à lui augmenté de 5,3 cm et le poids moyen de 3,6 kg (30). La distribution de l'obésité est différente selon l'âge et le sexe. La prévalence de l'obésité chez les femmes est globalement supérieure (15,7%) à la prévalence chez les hommes (14,3%). Cette supériorité est d'autant plus marquée chez les femmes de 45 à 54 ans : 17,5% contre 14,5% pour les hommes de 45 à 54 ans (30). S'agissant des enfants, une étude portant sur les 7 à 9 ans constate une stabilisation de la prévalence de l'obésité en France depuis 2007 (31).

² Mesuré par l'Indice de Masse Corporelle (IMC)

I.2. Mesures de l'obésité

L'obésité est une maladie chronique, multifactorielle et complexe. Afin d'agir sur cette maladie, le diagnostic et le traitement de l'obésité sont des enjeux majeurs de Santé publique (32, 33). Dans ce contexte, afin de mesurer l'excès de masse grasse corporelle, des méthodes de mesures anthropométriques telles que l'Indice de Masse Corporelle (IMC), ont été développées. Devant certaines limites liées à cet indice (que nous verrons plus tard), d'autres indicateurs permettent de prendre en compte le corps humain de façon plus fine lors de la mesure de l'obésité (34).

I.2.1 - Mesures anthropométriques

Afin de comparer les mesures individuelles aux valeurs de référence, la masse grasse est souvent mesurée de façon indirecte à l'aide de 4 variables communément disponibles et facilement obtenues, à savoir le poids, la taille, le tour de taille et le tour de hanches.

Le poids et la taille sont nécessaires pour calculer l'IMC d'un individu. En effet, l'IMC correspond au poids (kg) divisé par la taille (m) au carré. Il permet de calculer rapidement et facilement la corpulence et donc, de manière indirecte, l'importance de la masse grasse dans le corps. L'utilisation de l'IMC est simple et pratique à mettre en place au niveau de la population. C'est pourquoi, malgré de nombreuses critiques (35, 36), sa mesure est devenue la référence internationale dans la pratique clinique et dans les études épidémiologiques portant sur l'obésité. L'OMS fixe des recommandations concernant les seuils des états pondéraux à partir de cet indice. En ce qui concerne l'excès de poids, le surpoids est défini par un IMC compris entre 25 et 29.9, l'obésité par un IMC supérieur ou égal à 30 et l'obésité morbide par un IMC supérieur ou égal à

40 (15, 35). Néanmoins, certaines études démontrent qu'une norme de l'IMC à 24 kg/m² chez la femme et de 28 kg/m² chez l'homme serait plus adaptée pour estimer l'obésité (37). De nombreuses études épidémiologiques montrent une relation entre l'IMC et la mortalité, toutes causes confondues. En effet, les personnes avec un IMC plus élevé meurent plus jeunes que les personnes de poids moyen (26, 38, 39). En ce qui concerne la morbidité, l'augmentation de l'IMC est associée à des pathologies cardiovasculaires (8), cancéreuses (40) et endocriniennes (41).

Les mesures du tour de taille et du tour de hanche permettent d'évaluer l'accumulation de masse grasse de façon spécifique au niveau abdominal. De nombreuses études montrent que, indépendamment de la quantité globale de graisse corporelle, l'accumulation de graisse abdominale est associée à une prévalence accrue de maladies cardiaques et métaboliques (30, 35). La mesure du tour de taille est un estimateur facile à mettre en œuvre pour la détection de la masse grasse abdominale, car il est une mesure unique (35, 42, 43). Cette mesure permet aussi d'identifier les variations entre les taux de masse grasse qui existent entre individus. Par exemple, indépendamment de l'accumulation de masse grasse totale, les hommes ont en moyenne deux fois plus de graisse abdominale que les femmes non-ménopausées (44, 45).

I.2.2 - Mesures d'adiposité

En plus de ces mesures anthropométriques, il existe des outils pour mesurer la masse grasse de façon plus directe. Ces méthodes sont particulièrement utiles lorsque l'on s'intéresse aux déterminants de l'obésité. Les sujets obèses pourront alors être caractérisés selon leur composition corporelle et la répartition des graisses corporelles (45). En effet, la répartition des

grasses est un facteur de risque de maladie aussi important que l'excès de masse grasse en soi (45).

Composition corporelle

L'étude de la composition corporelle est une analyse du corps humain qui le modélise en compartiments. Différents modèles définissent et mesurent les compartiments de façons distinctes (46, 47). *Le Modèle anatomique* (i) est le modèle le plus ancien. Il sépare le corps selon le type de tissus (musculaires, adipeux, organiques, etc.) afin de permettre la compréhension de l'organisation spatiale et l'interconnexion des différents constituants du corps. *Le Modèle biochimique* (ii) sépare les composants de l'organisme en fonction de leurs propriétés chimiques : eau, lipides (carbone), protéines (azote), os (calcium, phosphore), potassium (quasi uniquement intracellulaire) et sodium (extracellulaire). La limite de ce modèle est liée au manque de données biochimiques directes sur la composition corporelle de l'organisme humain. *Le Modèle physiologique* (iii) permet d'introduire la notion de compartiments (deux ou trois selon la méthode) qui regroupent des composants corporels liés entre eux selon leurs fonctions, indépendamment de leurs localisations anatomiques et de leurs natures chimiques. La modélisation complète du corps humain est définie par deux types de compartiments, les compartiments hydriques (divisés en deux catégories : intracellulaire et extracellulaire, composés de liquide corporel et représentant entre 43 et 59% du poids corporel pour un homme et entre 50 et 70% pour une femme) et les compartiments tissulaires (46).

Les compartiments hydriques et tissulaires sont essentiellement définis selon 2 modes de représentation : le modèle à 2 compartiments et le modèle à 3 compartiments. Le modèle à 2 compartiments considère le corps comme étant composé de masse grasse et de masse non grasse

(appelée aussi masse maigre). *La masse grasse* correspond à la quantité totale de graisse corporelle (triglycérides stockés dans les adipocytes) répartie dans l'organisme. Ce compartiment est dépourvu d'eau. Il représente environ 15% du poids de l'homme jeune et 23% du poids de la femme jeune. Il faut le distinguer de la masse non grasse. *La masse non grasse* correspond à la masse totale de la peau, des os, des muscles, des organes et des liquides du corps humain. La masse non grasse est essentiellement constituée d'eau (46). En plus de *la masse grasse*, le modèle à 3 compartiments subdivise *la masse non grasse* en deux parties : *La masse cellulaire active (MCA)* qui correspond à la masse squelettique et à la masse protéinique du corps ; *le volume d'eau extracellulaire* qui correspond à la masse hydrique facilement échangeable par le corps humain (46).

Mesure par bio-impédancemétrie

Les mesures d'impédancemétrie bioélectrique (bio-impédancemétrie) utilisent le modèle physiologique de la composition corporelle. Ces mesures ont été initialement proposées par le Dr Thomasset en 1962 et étaient préconisées pour la mesure des secteurs liquidiens de l'organisme (48). Les mesures de bio-impédancemétrie sont simples, rapides, peu coûteuses et non invasives mais difficiles à mettre en place d'un point de vue populationnel. Ces méthodes consistent à évaluer et mesurer les propriétés électriques à l'intérieur d'un milieu biologique tel que le corps humain. Ces méthodes sont basées sur la capacité des tissus hydratés à conduire l'énergie électrique (49). Au travers d'électrodes, on fait passer un courant alternatif de faible intensité. Le courant passera d'abord dans les compartiments où la conductivité électrique est plus élevée, tels que les compartiments hydriques (eau, électrolytes) et ensuite dans les compartiments plus isolants (membranes cellulaires, graisse, peau). Le résultat de la résistance de ces tissus au courant alternatif correspondra à l'impédance. Un individu avec une plus grande masse

graisseuse aura une plus grande résistance, donc une impédance plus importante qu'une personne qui a plus de muscles, tissus qui contiennent plus d'eau (46).

Les Pèse-Personnes Impédancemètres (PPI) ont fait leur apparition dans le commerce depuis quelques années, ce qui rend les mesures de bio-impédancemétrie accessibles au grand public. Dans le cas des PPI présents dans le commerce, les électrodes sont placées sous les pieds et le courant électrique ne parcourt que les jambes et le bas du tronc. Les mesures obtenues sont ensuite extrapolées au corps entier. Ces mesures de PPI donnent des valeurs des compartiments tissulaires (2 compartiments : masse grasse - masse maigre). Certaines donnent aussi la valeur de la masse hydrique du corps (50). Les mesures des taux de masse grasse et de masse maigre sont souvent exprimées en pourcentage afin de pouvoir les comparer avec les valeurs médicales ou diététiques (50). Outre le pourcentage de masse grasse, des indicateurs tels que l'Indice de Masse Grasse (IMG) et l'Indice de Masse Maigre (IMM) permettent de prendre en compte la répartition interne des masses corporelles. Ces indicateurs correspondent au poids de la masse grasse ou de la masse maigre (en kilogramme) divisé par la taille au carré (en mètres) pour l'IMG ou l'IMM respectivement. Le total de ces deux indicateurs correspond à l'IMC. Par conséquent, ils permettent une décomposition de l'IMC.

Les normes définissant une mesure et un seuil de l'obésité selon le taux de masse grasse sont mal définies dans la population générale. En effet, un rapport de 1995 de l'OMS évalue que l'obésité correspond à un taux de masse grasse supérieur à 25% pour les hommes et 35% pour les femmes (14). En revanche, ce rapport ne propose aucune recommandation en ce qui concerne les critères de pourcentage de graisse corporelle pour le diagnostic de l'obésité car il se base sur une étude suédoise qui évalue une moyenne de pourcentage de masse grasse chez l'homme et chez la

femme âgés de 45 à 49 ans (51). Néanmoins, cette évaluation de l’OMS est devenue une norme pour plusieurs auteurs (52-57). En 2004, Basdevant et al. ont défini, pour une population caucasienne, l’obésité médicale comme une accumulation excessive de masse grasse dépassant 25% du poids corporel pour les hommes et 30% pour les femmes (35).

Bio-impédancemétrie pour contourner certaines limites de l’IMC

L’utilisation de l’IMC est simple et pratique à mettre en place d’un point de vue populationnel. De plus, l’utilisation de l’IMC joue un rôle majeur dans la surveillance mondiale de l’épidémie d’obésité croissante. Malgré tout, cet indice fait l’objet de nombreuses critiques. En effet, l’IMC fournit des informations essentiellement sur la corpulence corporelle en fonction du poids et de la taille de l’individu. Il ne renseigne pas sur les raisons de l’excès de poids (dû à une augmentation de la masse grasse ou à la perte de masse maigre) tandis que le lien critique avec une maladie métabolique est reconnu comme étant l’excès de masse grasse. L’IMC ne tient pas compte de la composition corporelle contrairement aux indicateurs de mesure de la masse grasse tels que l’IMG ou l’IMM, qui permettent donc de contourner certaines limites de l’IMC (37, 50). En effet, l’IMC sous-estime significativement la prévalence de l’obésité. La sous-estimation de l’obésité est plus prononcée chez les femmes et devient plus fréquente avec l’augmentation en âge (58, 59). D’un autre côté, l’IMC peut tendre à classer les individus avec un haut niveau de masse musculaire comme étant en surpoids ou obèse (60).

I.3. Déterminants de l'obésité

Les causes du déséquilibre entre l'apport et la dépense énergétique sont complexes et multiples (61). Au-delà de l'aspect individuel, de nombreux facteurs environnementaux sont impliqués dans l'occurrence et le développement de l'obésité (2). Depuis une quarantaine d'années de recherches sur le sujet dans des domaines variés tels que la médecine, l'économie, la psychologie, la géographie de la santé, la sociologie, etc., de nombreux déterminants ont été identifiés comme affectant directement ou indirectement l'obésité. Par exemple, Ritenbaugh et Kumanyika ont développé en 1999 une « toile causale de l'obésité » (figure 2) afin de mieux comprendre les interactions complexes entre les différents facteurs causaux de l'obésité par niveaux d'organisation sociale (62).

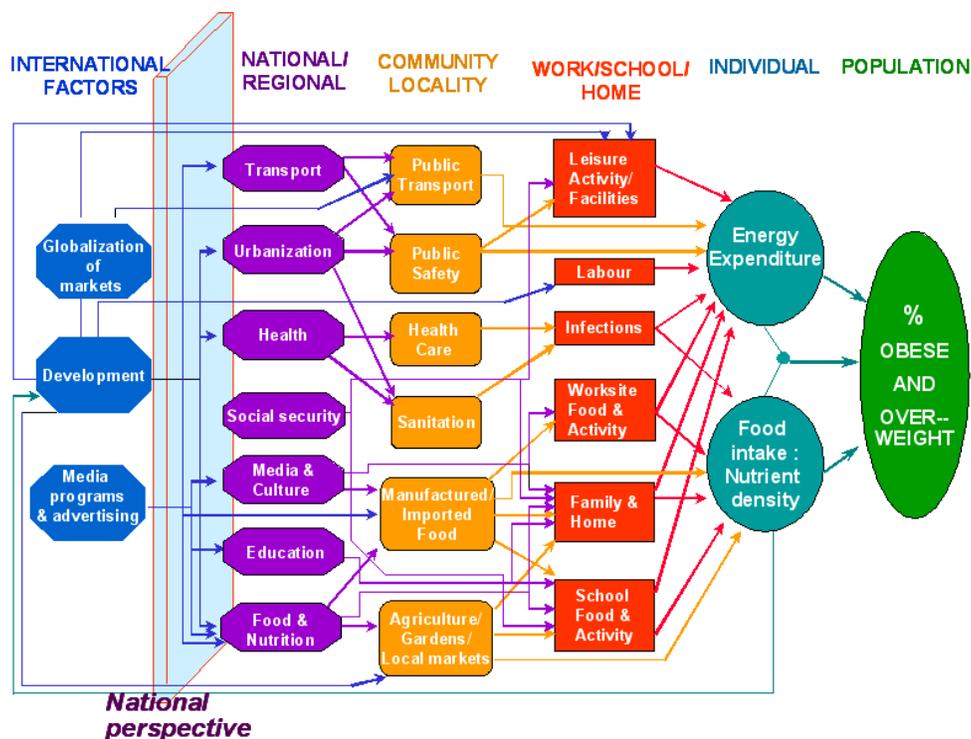


Figure 2 : Toile causale de l'obésité (“Web of causal obesity”) (62)

De même, lors d'une recherche prospective conduite en 2007, le gouvernement britannique a identifié plus de 100 variables influençant l'obésité (annexe 1) (63). Les principaux déterminants de l'obésité, qui s'agencent les uns aux autres, constituent un ensemble d'interactions d'effets sur l'obésité à un temps donné. Ils se situent à un niveau individuel (mutation génétique, variation sociodémographiques et comportementales) et à un niveau environnemental (milieux de vie). De plus, comme le montre la figure 2, le travail (en temps qu'emploi ou « labour » et organisation sociale ou « work ») est associé à l'obésité. Cependant, peu d'études ont exploré le lien entre les secteurs d'activité économique et l'obésité (64-66).

I.3.1 - Niveau individuel

Génétique

De nombreuses études ont montré que la génétique avait une responsabilité importante dans la survenue et le développement de l'obésité (67). Dans un premier temps, la prise en compte de la dimension génétique a été favorisée par la reconnaissance scientifique de l'agrégation familiale de l'obésité : 7 personnes obèses sur 10 ont un parent obèse (68-70). Cette observation est incomplète car, outre les gènes, une famille partage également des habitudes alimentaires et un contexte socio-économique. Dans un second temps, le rôle de la variabilité génétique dans l'expression de l'obésité a été mis en évidence par, d'une part, la concordance « quasi-parfaite » d'IMC de jumeaux monozygotes ayant vécu dans deux familles différentes (71-73) et, d'autre part, par la découverte, en 2007, du gène FTO (74) puis, en 2014, du gène IRX3, tous deux liés et associés à la prise de poids (75). Ces études montrent que les individus pour lesquels le gène FTO est surexprimé auraient un risque entre 30 et 70% supérieur d'être obèses. Grâce à ces récentes

découvertes, la génétique aide à mieux comprendre l'obésité (76). De plus, le facteur génétique constitue une susceptibilité et agit en lien avec l'interaction gène-environnement (26).

Variation sociodémographiques

De nombreuses études et revues de littérature ont regardé les liens entre les variations sociodémographiques individuelles et l'obésité. Les variations individuelles démographiques (âge, sexe), sociales (niveau d'éducation, statut matrimonial) et économiques (revenu du foyer, salaire) affectent l'obésité en influençant ses facteurs déterminants. Ces études montrent que la prévalence de l'obésité augmente avec l'âge (jusqu'à 60-65 ans), que cette maladie est plus fréquente chez les femmes et que les individus avec un faible niveau d'instruction sont les plus obèses (30, 77, 78). Dans le même sens, dans de nombreuses études, la prévalence de l'obésité est plus élevée chez les individus avec les revenus les plus bas (30, 79-83).

Comportement

L'obésité se développe principalement via deux types de comportement : un excès d'apport alimentaire (et/ou une augmentation de l'alimentation grasse et sucrée) et une insuffisance de la dépense énergétique due à un manque d'activité physique (26). Les excès d'apports alimentaires sont variables d'un individu à l'autre et la notion de bilan énergétique positif est individuelle et non-normative (26). Néanmoins, certaines études montrent qu'une alimentation plus grasse et plus sucrée ou la fréquentation de restaurants de restauration rapide favorisent la prise de poids et l'occurrence de l'obésité (84). De plus, le manque d'activité physique est associée à une augmentation de la prévalence de l'obésité (85). En ce qui concerne la baisse de la dépense énergétique, l'OMS constate l'impact des changements dans les modes de transports (l'utilisation de la voiture au détriment de la marche) et dans les activités de loisirs (jouer aux jeux vidéo ou

regarder la télévision à la place de pratiquer un sport) (86). Cependant, les associations entre une alimentation grasse et l'obésité d'un individu sont bien plus évidentes quand l'activité physique de celui-ci est faible (87). De plus, les études constatent que les comportements individuels associées aux variables socio-économiques tels les niveaux d'éducation, le revenu du foyer, etc. sont liés à un sur-risque de la prévalence de l'obésité. En effet, de nombreuses études montrent que ce sont les classes sociales les plus défavorisées qui utilisent leur temps de loisirs de façon sédentaire (88-91) et qui fréquentent le plus les services de restauration rapide et les supermarchés discount (92, 93).

I.3.2 - Niveau environnemental

Les différents milieux de vie

L'environnement est défini comme étant, d'une part, « le milieu dans lequel l'individu et/ou le groupe évoluent, ce milieu incluant l'air, l'eau, le sol, leurs interfaces, les ressources naturelles, la faune, la flore, les champignons, les microbes et les êtres humains, les écosystèmes et la biosphère ». Il peut également être défini, d'un point de vue plus sociétal, comme le « milieu physique, construit, naturel et humain dans lequel un individu et/ou un groupe (une famille, un quartier, une société, une collectivité, une entreprise, une administration, etc.) fonctionnent ; incluant l'air, l'eau, le sol, le sous-sol, la faune, la flore, les autres organismes vivants, les êtres humains et leurs interrelations » (94). La deuxième définition de cette encyclopédie scientifique sous-entend l'individu au centre de l'environnement physique, construit, naturel et humain. Dans ce sens, on peut parler du « milieu de vie ».

L'étude des expositions liées à l'environnement, en tant que milieu de vie pour les individus qui y vivent ou qui le fréquentent, implique de déterminer les caractéristiques de cet

environnement susceptibles d'influencer ces individus ou ces groupes. Les expositions environnementales peuvent agir, soit par la proximité des individus et leurs interactions entre eux, soit par les opportunités proposées à ces individus d'atteindre leur potentiel de santé maximum. C'est pourquoi, en épidémiologie, tenant compte de la mobilité individuelle, l'étude des caractéristiques environnementales basée sur les lieux les plus fréquentés quotidiennement est importante. Certains chercheurs affirment que trois principaux types d'environnements sont : les environnements de résidence, de travail et de loisirs (95, 96). Analyser l'individu dans son environnement résidentiel ainsi que dans son environnement professionnel permet de le situer dans son environnement physique et social. Il apparaît important de considérer ces environnements séparément car ils semblent potentiellement différents, et en plus, ils permettent aux individus un développement personnel et professionnel où plusieurs activités quotidiennes peuvent être menées. Malgré ce constat, une très large majorité des études qui analysent les liens entre l'environnement et l'obésité s'est focalisée sur les environnements résidentiels (97). Ces études ne considèrent pas les caractéristiques associées au lieu de travail et ne permettent donc pas de prendre en compte, entre autres, l'environnement géographique professionnel d'une part, et les conditions de travail d'autre part. Il s'agit d'une limite que cette thèse tente de surmonter en considérant au-delà des environnements résidentiels, les environnements professionnels.

Les différentes dimensions de l'environnement

Les caractéristiques socio-économiques de l'environnement

Depuis les années 80-90, l'impact des caractéristiques du milieu de vie sur l'obésité et sur la santé en général a été largement étudié (96, 98-100). Les caractéristiques socio-économiques du milieu de vie (revenu, niveau d'instruction, etc.) émergent des caractéristiques individuelles des habitants qui partagent un territoire. Ces caractéristiques environnementales peuvent être

associées à des problèmes de santé indépendamment des caractéristiques individuelles qui les composent (101). Dans le cas de l'obésité, en contrôlant sur les caractéristiques individuelles démographiques, sociales et économiques, il apparaît que la prévalence de l'obésité est plus importante dans les environnements socio-économiquement défavorisés (82, 83, 102). En effet, un environnement socio-économiquement défavorisé peut influencer la prise de poids selon plusieurs mécanismes (103, 104). Par exemple, les populations vivant dans des quartiers défavorisés sont plus exposées aux nuisances environnementales (perception d'insécurité, manque d'infrastructures piétonnières, esthétique du quartier) qui découragent la pratique d'une activité physique régulière (101, 105). Ces populations sont également plus enclines à consommer une alimentation plus grasse (101) et à un mode de vie stressant, moins sain et elles sont plus exposées aux maladies causées par l'obésité (diabète et hypertension) du fait de lacunes dans la connaissance des risques liés à l'obésité (101). Néanmoins, des analyses séparées entre les hommes et les femmes montrent des forces d'associations entre un environnement socio-économiquement défavorisé et l'obésité différentes selon le sexe. Par exemple, la défavorisation du quartier peut être associée à une prévalence de l'obésité plus élevée chez les femmes (106, 107).

Les caractéristiques physiques de l'environnement

L'environnement physique se définit par la nature de ses différentes composantes et par la notion de proximité. Une distinction utile concerne les composantes naturelles et construites de l'environnement physique. Les composantes naturelles ne sont pas directement produites par l'homme, mais peuvent être modifiées par l'activité humaine (108). Dans cette conceptualisation, l'environnement construit fait référence, entre autres, aux caractéristiques du bâti et à la densité et à la diversité des activités. De nombreuses études montrent qu'il existe une influence du milieu

physique sur la prévalence de l'obésité (97, 109). Les caractéristiques du bâti font référence aux parcs, aux espaces publics, aux bâtiments (habitations, écoles, commerces, etc.) et aux infrastructures de transport (rues, pistes cyclables, etc.). Plusieurs études ont montré que les caractéristiques du bâti jouent un rôle dans la pratique d'une activité physique et donc, peuvent avoir un impact sur l'obésité par l'influence qu'elles ont sur l'activité physique (110, 111). La présence de parcs, de trottoirs et de rues bien éclairées sont des facteurs environnementaux protecteurs par rapport à l'obésité car ils jouent sur le comportement et incitent les résidents à marcher plus et à avoir un niveau d'activité physique plus élevé (111-116). D'autres facteurs péjoratifs dans la relation environnement/obésité incluent un faible niveau de « marchabilité » du réseau de rues alentour (117). La densité fait référence à la mesure d'une caractéristique environnementale dans un espace donné, par exemple au niveau du quartier, à un moment donné. Les indicateurs de densité les plus utilisés sont la densité de population et la densité de destinations correspondant à un nombre d'individus ou à un nombre de destinations dans une aire déterminée, tels que le nombre d'habitants ou de destinations par km². Les études montrent qu'une faible densité de population ou une faible densité de destinations est associée à une prévalence élevée de l'obésité (118-120). La diversité des activités correspond à l'aménagement spatial de l'utilisation du sol. L'indicateur le plus utilisé est la mixité de l'utilisation du sol (109). Une plus grande diversité des activités est associée à une augmentation de l'utilisation des modes de transport actifs (vélo, marche, etc.) (100) et à une diminution de la prévalence de l'obésité (115, 121, 122). En revanche, l'absence de transports en commun, un manque d'équipements sportifs à proximité (123), la présence d'une grande proportion de restaurants de restauration rapide (124), l'absence de magasins vendant des fruits et des légumes frais, et un mauvais accès à

des supermarchés vendant des produits conformes aux recommandations nutritionnelles (125) sont des facteurs de risque de l'obésité.

Etude en milieu français

Dans le contexte français, l'étude Obépi montre que la prévalence d'obésité chez les adultes de plus de 18 ans ne cesse d'augmenter (30). Cette étude met à jour des inégalités géographiques de santé, par région ou même par quartier. En effet, malgré une augmentation globale de la prévalence d'obésité sur le territoire français, il y a une certaine hétérogénéité géographique dans cette évolution. Par exemple, de 1997 à 2012, les zones ayant subi les plus fortes hausses sont : la Champagne-Ardenne avec une augmentation de plus de 145%, la Franche Comté avec une augmentation de 144%, l'Alsace avec une augmentation de 108% et la Région Parisienne avec une augmentation de 100,5%. De plus, l'étude Obépi montre que la prévalence d'obésité a augmenté en France de façon plus soutenue chez les inactifs, les ouvriers et les employés que chez les cadres (30). De ce fait, l'observation d'un accroissement de la ségrégation sociale des populations dans l'espace (126) et la progression actuelle de l'épidémie d'obésité en France conduit à des disparités importantes entre les quartiers socialement favorisés et défavorisés.

La modification de l'environnement peut aussi jouer un rôle majeur dans la progression de la prévalence d'obésité (127). En effet, des études françaises ont préalablement mis en évidence des caractéristiques contextuelles relatives à l'environnement socio-économique, à l'environnement physique, à l'environnement de services et à l'environnement d'interaction sociale qui permettent d'expliquer les disparités de prévalence d'obésité entre les quartiers favorisés et défavorisés. Une étude de 2011, réalisée dans la Région Parisienne, sur une population adulte âgée de 30 à 79 ans, montre qu'un faible niveau d'éducation de la population

habitant dans le quartier de résidence était associé à un taux d'IMC ou de tour de taille plus élevé. Ces associations étaient plus fortes chez les femmes (128). De plus, une étude montre que même après ajustement sur les caractéristiques socio-économiques individuelles et du quartier, l'IMC ainsi que le tour de taille étaient négativement associés avec les caractéristiques physiques et de services de l'environnement (tels que la proportion de surface construite, la densité de magasin de fruits et légumes, et les restaurants) (129). Dans un même sens, une étude conduite en Région Parisienne montre que les participants âgés de 30 à 79 ans, faisant leurs courses dans les supermarchés « hard discount » et dans les supermarchés dont la zone de chalandise comprend les quartiers résidentiels avec les plus faibles niveaux d'éducation avaient un IMC ainsi qu'un tour de taille plus élevé (d'autant plus si les individus avaient un faible niveau d'éducation) (93). Ces études sont cohérentes avec les résultats d'une étude concernant les enfants qui montre que la probabilité d'être en surpoids était plus élevée chez les enfants vivant dans un environnement où l'accessibilité spatiale aux installations urbaines d'activité physique et aux magasins d'alimentation générale était plus faible (mais seulement chez les enfants d'ouvriers). Aucune relation n'a été trouvée concernant le surpoids et les infrastructures naturelles d'activité physique ou les restaurants issus de la restauration rapide (130). D'autres facteurs environnementaux, tel que le bâti, favorisent les déséquilibres entre les quartiers favorisés et défavorisés. En effet, une étude française conduite en 2012 montre que les individus résidant dans un environnement avec une forte répartition d'accessibilité spatiale aux espaces verts, aux installations de proximité et avec une forte concentration de piste cyclable étaient plus susceptibles de marcher et de faire du vélo. Par contre, l'IMC des individus n'était pas pour autant plus bas chez les individus résidant dans ces types de quartiers (113).

I.3.3 - Niveau des secteurs d'activité économique

Variabilité des expositions selon les secteurs d'activité économique

L'enquête « Conditions de Travail » de 1978 de la Direction de l'Animation de la Recherche, des Études et des Statistiques (DARES) du ministère du travail marque le début des enquêtes sur les conditions de travail en France. Depuis, en complément de l'enquête « Travail-Emploi » de l'Institut National de la Statistique et des Études Économiques (INSEE), l'enquête « Conditions de Travail » est répétée tous les 7 ans, la dernière édition datant de 2013. Dès 1978, cette enquête s'est basée sur le postulat d'une inégalité des expositions au travail en fonction des catégories socio-professionnelles et des secteurs d'activité économique. L'enjeu de cette enquête est de connaître les différentes expositions des travailleurs sur le plan socio-professionnel et par secteurs d'activité susceptibles de porter atteinte à la santé physique des travailleurs (131, 132). A partir de 1984, les enquêtes ont évolué et se sont ouvertes aux troubles psychiques professionnels avec des indicateurs tels que la latitude dans les décisions, la demande psychologique et charge de travail mentale, le soutien avec les supérieurs et la tension avec les collègues, etc. Cette évolution est liée à la prise en considération en Santé publique des risques psychosociaux liés au travail. En France, en parallèle à ces travaux, d'autres études ont comparé les expositions professionnelles par secteurs d'activité (133, 134). Finalement, en 2013, la DARES a fait une nouvelle enquête spécifiquement sur les expositions liées aux risques professionnels selon les secteurs d'activité économique (135). Selon les contraintes spécifiques qui leur sont liées, les expositions d'ordre physiques, psychosociales et comportementales varient fortement d'un secteur à l'autre (135). Cette étude montre que les contraintes physiques intenses concernent

principalement le secteur de la construction et le secteur de l'agriculture. L'exposition aux produits chimiques est plus fréquente chez les employés des secteurs de l'industrie et de l'agriculture. Les agressions sur le lieu de travail touchent principalement les employés de l'éducation, de la fonction publique et du commerce et des transports. En ce qui concerne les horaires atypiques de travail, les secteurs de la fonction publique hospitalière et du commerce et des transports sont les plus touchés (135). L'exposition au bruit est plus présente dans les secteurs de l'industrie, de l'agriculture et de la construction (131, 133). De plus, le niveau de stress est plus élevée dans le secteur de l'éducation, de la santé et de l'agriculture (135). Des différences comportementales entre les secteurs sont aussi observées. En effet, les études de Lauzeville et al. en 2009 et de Goffette en 2012 montrent des disparités de comportements par secteurs d'activité économique (136, 137). Les proportions de fumeurs sont plus importantes dans les secteurs de l'hôtellerie et de la construction. En ce qui concerne l'alcool et les drogues, les personnes travaillant dans le secteur de l'hôtellerie et de la restauration en sont les plus grands consommateurs. Le secteur des arts et des spectacles présente la plus grande consommation de cannabis et le taux d'ivresse par année le plus élevé. Les profils de consommateurs sont souvent homogènes (138), c'est-à-dire que lorsqu'un secteur présente une surconsommation par rapport à la moyenne pour un type de produit, il y a de fortes chances pour que la consommation soit supérieure à la moyenne pour les autres produits étudiés (et inversement pour la sous-consommation). Pour dresser un profil de consommateur de substances psychoactives par secteurs d'activité économique en fonction des résultats de cette étude, chez les hommes, le secteur de la construction se distingue par une surconsommation de substances psychoactives très nette et, dans une moindre mesure, les secteurs de l'hébergement et de l'art et du spectacle. Le secteur le plus « protégé » est l'administration publique et la défense. Pour les femmes, les

secteurs les plus touchés par la consommation de substances psychoactives sont les secteurs de l'hébergement et de l'administration publique. De ce fait, le comportement du travailleur peut révéler une attitude générale vis-à-vis du risque spécifique par secteurs d'activité économique.

Liens entre expositions professionnelles et risque sanitaire

La littérature épidémiologique sur les effets des facteurs physiques, psychosociaux, comportementaux et des expositions professionnelles liées aux statuts socio-économiques sur l'obésité et les maladies cardiovasculaires est abondante (139-145). En revanche, très peu d'études se sont intéressées aux relations entre les secteurs d'activité économique et l'obésité ainsi que les facteurs de risque cardiovasculaire associés. Pourtant, dans les pays industrialisés où les secteurs d'activité économique sont comparables, de fortes différences de mortalité sont observées (146, 147). Par exemple, la mortalité entre secteurs par cancer serait explicable, pour un tiers, par les expositions professionnelles (148). De plus, il apparaît que les expositions physiques, psychologiques et comportementales liées au secteur d'activité économique peuvent constituer des éléments médiateurs de la relation entre les secteurs d'activité économique et l'obésité ainsi que des maladies cardiovasculaires. Par exemple, de nombreuses études établissent un lien entre la contrainte physique (rythme d'horaire décalé, travail de nuit) (149), psychosociale (stress au travail) (150) et comportementale (alimentaire, consommation d'alcool) (151) et des prévalences élevées de l'obésité et de comorbidités cardiovasculaires. Plus spécifiquement, des études ont montré des associations entre des facteurs psychosociaux liés aux activités professionnelles (stress, manque de soutien social, heure de travail décalée, ...) et l'obésité (152) mais aussi avec les maladies cardiovasculaires telles que l'hypertension, l'athérosclérose, l'angine de poitrine, les maladies coronariennes, les accidents vasculaires cérébraux ainsi que le développement d'un diabète (132). Cependant, dans le contexte français, il n'existe pas de

données permettant d'estimer le véritable impact des secteurs d'activité économique sur l'obésité et les facteurs de risque cardiovasculaire. En effet, les statistiques sur les atteintes de santé dues au travail, compilées par la Caisse Nationale d'Assurance Maladie des Travailleurs Salariés (CNAMTS), des travailleurs salariés (soit le régime général de la Sécurité Sociale) ne décrivent que les accidents du travail et les maladies reconnues et indemnisées (132, 146).

I.4. Analyses transversales et longitudinales des déterminants de l'obésité

I.4.1 – Schémas d'analyses utilisés

De nombreuses études ont analysé les liens entre les déterminants socio-économiques individuels et l'obésité (80-82), mais la plupart de ces études ont utilisé un schéma d'analyse transversal (82). La limite du schéma d'analyse transversal vient du fait qu'il ne permet pas de déterminer la direction des effets causaux impliqués dans les associations. Les études longitudinales issues de grandes cohortes, telles que la *National Health and Nutrition Examination Surveys* (NHAES), la *Framingham Heart Study* (FHS), ou en France l'*Etude Prospective Parisienne* (EPP), sont importantes en épidémiologie car elles permettent de suivre une population dans le temps en prenant des mesures de santé de façon répétées. L'évolution des mesures dans le temps permettra de déterminer de façon plus fiable si l'association est liée à un effet causal.

Dès 1989, les résultats d'une revue de littérature de Sobal and Stunkard basée uniquement sur des études transversales montrent que l'obésité est liée au statut socio-économique. Ils montraient l'existence de relation directe entre le statut socio-économique et l'obésité dans les pays en voie de développement. Les hommes, les femmes et les enfants avec un niveau du statut

socio-économique élevé avaient une plus grande probabilité d'être obèse que ceux avec un faible niveau de statut socio-économique. Par contre, dans les pays développés, une relation inverse était trouvée entre le statut socio-économique et l'obésité pour les femmes (153). Depuis cette publication, considérant l'augmentation rapide de la prévalence de l'obésité mondiale, les relations entre le statut socio-économique et l'obésité ont changé (80). En effet, les études transversales de 1988 à 2004 montrent que la tendance générale des résultats, tant pour les hommes que pour les femmes, allait vers une augmentation des associations positives ainsi qu'une diminution des associations négatives entre le statut socio-économique et l'obésité dans les pays ayant des niveaux élevés de développement socio-économique et dans les pays à moyen et bas niveaux de développement socio-économique, respectivement. (81). De plus, ces études montrent que les associations entre le statut socio-économique et l'obésité variaient en fonction de l'indicateur étudié. Par exemple, un faible niveau de statut socio-économique associé à une corpulence plus élevée chez les femmes des pays développés était plus fréquent avec les indicateurs du niveau d'éducation et de la profession (81). Malgré l'importance de ces résultats, la limite la plus importante de ces études est qu'elles ne représentent une population qu'à un moment donné dans le temps et ne permettent pas d'étudier le changement de poids au fil du temps.

Considérant ces limites, Mujahid et al, étudient les associations entre le statut socio-économique et l'IMC de manière transversale et longitudinale chez 13167 sujets (âgés de 45 à 64 ans) d'ethnie noire et blanche. Les résultats de cette étude montrent que le niveau socio-économique individuel et du quartier sont indépendamment et inversement associés à un niveau élevé d'IMC à l'inclusion, chez les femmes. Chez les hommes blancs l'IMC à l'inclusion était négativement associé avec le revenu, tandis que chez les hommes noirs il était positivement

associé avec l'éducation et le revenu. En ce qui concerne l'évolution de l'IMC au cours du temps, les résultats de cette étude montrent que l'IMC augmentait indépendamment du sexe et de l'ethnie. L'augmentation de l'IMC n'était pas associée au niveau socio-économique chez les hommes blancs tandis que chez les hommes noirs, les augmentations d'IMC étaient plus prononcées chez les personnes avec un statut socio-économique plus élevé (82). Les résultats d'études longitudinales entre le niveau du statut socio-économique et le gain de poids montraient des associations inverses relativement constantes entre la profession et le gain de poids pour les hommes et les femmes dans les pays développés, chez les personnes non-noires. De plus, chez les enfants, l'augmentation de l'adiposité était aussi associée à un niveau de statut socio-économique plus faible (154, 155). Mais les études longitudinales sur les facteurs socio-économiques et le changement d'IMC ne rapportent pas toujours des résultats cohérents. Quand le statut socio-économique était évalué par le niveau d'éducation, les associations étaient moins constantes mais permettaient néanmoins de confirmer la tendance d'une relation inverse entre le statut socio-économique et le gain de poids. Par contre, quand le revenu était utilisé comme indicateur de statut socio-économique, les relations n'étaient pas liées à une augmentation de la corpulence corporelle (80, 154, 156-158).

I.4.2 – Données manquantes

Une limite de ces études longitudinales vient de la temporalité liée au schéma d'étude. En effet, les mesures devant s'effectuer sur le même individu lors de différentes vagues d'étude (parfois plusieurs années entre chaque vague), certains peuvent être perdus de vue. La mesure ne peut alors plus être faite et on se retrouve avec des données manquantes. L'attrition correspond aux situations selon lesquelles lors d'une étude prospective les individus quittent le suivi de l'étude

avant son terme prévu (159) et où, les données ne sont plus disponibles. Le biais dû au phénomène d'attrition correspond à une erreur systématique induite par la différence entre les participants des différentes vagues résultant de la perte au cours du suivi (160). Le risque du biais lié aux données manquantes dépend de la raison pour laquelle la donnée est manquante. En 1987, Little et Rubin, décrivent le mécanisme des données manquantes selon trois types de combinaison possibles (161) : 1) manquantes complètement au hasard (« Missing Completely At Random », MCAR), dans ce cas, les résultats ne seront pas biaisés car lors des analyses, les données manquantes ne sont alors pas différentes des données non-manquantes ; 2) manquantes au hasard (« Missing At Random », MAR), où toutes les différences systématiques entre les valeurs manquantes et les valeurs observées peuvent s'expliquer par des différences dans les données observées. Dans ce cas, le fait de prendre en compte des valeurs qui causent les données manquantes dans les analyses permettra d'obtenir des résultats non-biaisés ; 3) manquantes pas au hasard (« Missing Not At Random », MNAR) où dans ce cas, les données manquantes dépendent d'un événement que le chercheur n'a pas mesuré. De ce fait, même après la prise en compte des données observées, une différence systématique demeure entre les données manquantes et les données observées. Le mécanisme de donnée manquante MNAR est aussi appelé « inaccessible », car la cause des données manquantes ne pourra pas être pris en compte lors des analyses et le biais d'attrition ne pourra pas être contourné (162). De plus, du fait que les données soient manquantes, il n'est pas possible de déterminer avec certitude si le mécanisme est MAR ou MNAR, car il s'agit seulement d'une hypothèse (163).

De plus, malgré l'impact des données manquantes dans les résultats, la grande majorité des études longitudinales ne prennent pas en considération le problème potentiel du biais

d'attrition dans les analyses statistiques (82, 164) alors que les données manquantes sont fréquentes et souvent inévitables dans les études épidémiologiques (165, 166).

OBJECTIFS

L'objectif général de ce travail était, à l'aide d'études transversales et longitudinales, d'évaluer les relations entre les environnements géographiques de vie et l'excès de poids et de graisse corporelle et de risque cardiovasculaire. En parallèle, d'un point de vue méthodologique, nous avons cherché à tenir compte de l'environnement géographique de travail en plus de l'environnement de résidence, et nous avons évalué l'impact des données manquantes dans les analyses longitudinales.

Les objectifs spécifiques de ce travail de thèse étaient les suivants :

- Dans un premier temps, nous avons examiné les associations entre, d'une part, les facteurs individuels sociodémographiques, les caractéristiques des quartiers résidentiels et professionnels, et le secteur d'activité économique des participants et, d'autre part, la masse grasse évaluée par bio-impédancemétrie. Nous avons également examiné si l'IMC permet de saisir pleinement les effets sociaux et environnementaux sur la graisse corporelle.
- Dans un second temps, nous avons estimé et comparé les associations entre les secteurs d'activité économique et un large panel de facteurs de risques cardiovasculaire.
- Enfin, dans un troisième temps, à l'aide d'une étude longitudinale, nous nous sommes intéressés aux déterminants de l'évolution des paramètres anthropométriques chez les participants de la cohorte RECORD. Ensuite, d'un point de vue méthodologique, basé sur ces données observées et à l'aide d'une étude de simulation, nous avons évalué l'impact

des données manquantes dans les analyses longitudinales et comparé différentes méthodes d'analyses qui permettent de prendre en compte ces données manquantes.

CHAPITRE II - METHODES

II.1. Étude RECORD : contexte et objectifs

Dans le contexte français où les inégalités sociales ont un niveau considéré comme intermédiaire, les répercussions qu'elles engendrent sur l'état de santé des populations vont en augmentant³. Depuis les deux dernières décennies, ces disparités, qui traduisent une séparation sociale dans l'espace, se sont également accentuées entre les communes des régions françaises. C'est le cas, tout particulièrement, pour la région Île-de-France⁴, qui sert de cadre aux présents travaux. En effet, bien que cette région soit la plus riche de France, elle est aussi celle où les inégalités socio-environnementales sont les plus marquées (167).

L'Étude RECORD est issue d'une collaboration entre l'UMR-S 1136 et le Centre d'Investigations Préventives et Cliniques (IPC). Elle a été conçue à partir du contexte découlant de diverses études conduites en Amérique du Nord et en Europe du Nord consacrées aux disparités sociales et aux influences des environnements géographiques de vie sur la santé (168, 169) ainsi que du constat du manque d'études sur le sujet en France ajouté à la situation croissante des inégalités sociales franciliennes. Son objectif principal est d'étudier les disparités géographiques de santé qui existent au sein de la région Île-de-France, avec un intérêt particulier pour les différences observées entre quartiers favorisés et quartiers défavorisés (170).

³ Niveau intermédiaire par rapport à des pays comme le Japon, niveau d'inégalité faible et le Brésil, niveau fort.

⁴ L'Île-de-France, ou la région parisienne, représente 2,8% du territoire de la France et compte près de 11,9 millions d'habitants, soit environ 18 % de la population française.

Concernant les problématiques sanitaires, l'Étude RECORD s'intéresse principalement aux maladies cardiovasculaires et à leurs facteurs de risque comportementaux, cliniques et biologiques (142) (tels que l'obésité (128), l'hypertension artérielle (171), l'activité physique (172), etc.). Elle vise à étudier également les comportements de recours aux soins des personnes, en matière de traitement des facteurs de risque cardiovasculaire tels que l'hypertension. Enfin, au-delà de la description des disparités de santé entre quartiers, l'objectif central de l'étude est de comprendre par quels mécanismes les environnements géographiques de vie influencent la santé.

II.2. Territoire d'étude et population

L'Étude RECORD est une étude de cohorte qui comporte deux vagues de collectes de données ; une première vague de 2007 à 2008 et une deuxième vague de 2011 à 2014.

Les 7290 participants de la première vague de l'Étude RECORD ont été recrutés entre mars 2007 et février 2008 dans le cadre des Examens Périodiques de Santé de la Sécurité Sociale dans quatre sites du Centre IPC de la région Île-de-France localisés à Paris, Argenteuil, Trappes et Mantes-la-Jolie (Figure 3).

Le découpage du territoire de l'Étude RECORD pour la sélection des communes participantes a été fait à partir des codes postaux renseignés dans la base de données du Centre IPC. Avant le début de l'étude, en 2006, afin d'assurer de larges disparités socio-économiques entre les communes à retenir dans l'étude, une présélection de codes postaux de résidence a été effectuée à partir des listes des personnes venues au Centre IPC. La sélection a pris en compte, selon le recensement 1999, les codes postaux de plus de 15 000 habitants qui avaient fourni la

majorité des effectifs de patients au Centre IPC en 2006. Les localités avec les revenus médians les plus faibles ont été surreprésentées.

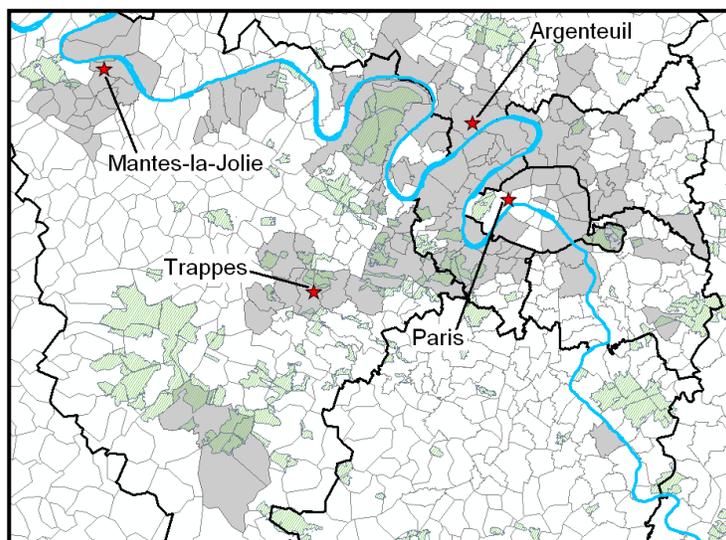


Figure 3. Localisation des 112 communes sélectionnées dans l'Étude RECORD et des 4 sites du Centre IPC participant au recrutement.

L'échantillon des communes sélectionnées est très diversifié. Il comprend aussi bien des communes défavorisées (par exemple, Sarcelle, Mantes-la-Jolie) que des communes favorisées (par exemple, Versailles, Neuilly-sur-Seine), ainsi que des communes fortement urbanisées et des territoires moins urbanisés. Au final, pour garantir une disparité des environnements géographiques de vie, 112 communes comprenant 1915 quartiers correspondant au découpage géographique infra-communal IRIS⁵ défini par l'INSEE et caractérisés par une grande diversité sociale et territoriale, ont été sélectionnées (figure 3). La sélection des participants s'est donc faite selon trois critères spécifiques : 1) l'âge des participants devait se situer entre 30 et 79 ans

⁵ « Ilots Regroupés pour l'Information Statistique » correspond à découpage, infra-communal, qui constitue une partition des communes en "quartiers" dont la population est de l'ordre de 2 000 habitants.

au moment de l'examen ; 2) les participants devaient résider dans l'une des 112 communes sélectionnées *a priori* ; 3) les participants devaient être en mesure de répondre par eux-mêmes (ou avec une aide minimale) au questionnaire de l'étude proposé en langue française. Lorsque que les individus utilisant les services du Centre IPC remplissaient ces trois conditions, il leur était proposé de participer à l'Étude RECORD. 10,9% ont été exclus du fait de leur incapacité à répondre à un questionnaire auto-administré en français. Au final, 83,6% des personnes invitées à participer à l'étude ont complété le protocole de collecte de données. Les enquêteurs de l'étude ont reçu une formation spécifique pour expliquer au mieux le principe de l'étude aux participants sélectionnés. Toutes les personnes qui ont accepté de participer ont signé un formulaire de consentement.

Dans le cadre de la seconde vague de l'Étude RECORD qui a débuté en février 2011, l'ensemble des participants ont été invités à un second examen de santé au centre IPC de février 2011 à décembre 2014. 51% des participants vus en 2007 – 2008 ont été ré-enquêtés dans la seconde vague de l'étude au 31 mars 2013 soit 3746 participants pris en compte dans la présente thèse.

II.3. Les données collectées

II.3.1 - Les données individuelles

Lors des passages des participants au Centre IPC, de nombreuses données biologiques, médicales, socio-administratives, comportementales et psychologiques ont été collectées. Les données cliniques et biologiques ont été recueillies lors des examens de la première et de la

deuxième vagues de l'Étude RECORD. Ces examens comprenaient un bilan biologique⁶, des mesures de tension artérielle, un électrocardiogramme, la spirométrie, un examen dentaire, un visio-test, des mesures anthropométriques (poids, taille, tour de hanche, tour de taille) et des mesures d'impédancemétrie (seulement lors des examens de la première vague de l'étude). D'autres données ont été obtenues à partir de questionnaires, comme le niveau d'activité physique pratiqué, les expositions professionnelles de façon très succincte, la consommation d'alcool et de tabac, les antécédents personnels et familiaux, l'état de santé actuel et les traitements médicaux prescrits, le statut d'emploi, la profession, le niveau d'études, le type de couverture sociale, le statut d'occupation du logement, le nombre d'enfants, la présence de difficultés financières, les problèmes d'accès aux soins, les activités récréatives, le soutien social, etc.

Les participants, par la suite, ont répondu à un questionnaire spécifique à l'Étude RECORD concernant leur statut socio-économique, leurs connaissances sur la santé cardiovasculaire, leurs perceptions et leurs attitudes par rapport à la santé, leur niveau d'activité physique et le vécu au sein de leur quartier (attachement au quartier, sentiment d'appartenance, opinions au sujet du quartier, réputation du quartier, insertion sociale au sein du quartier, activités associatives, confiance-méfiance et hostilité-bienveillance par rapport aux voisins, soutien social de la part du voisinage, impact du quartier sur le bien-être psychologique). L'ensemble du protocole de recueil de données a été approuvé au préalable par la Commission Nationale de l'Informatique et des Libertés (CNIL).

⁶ Bilan réalisé après 12 heures de jeûne et comprenant les mesures suivantes : hémogramme, urée, créatininémie, glycémie, potassium, cholestérolémie, triglycérides, aspartate aminotransférase (ASAT), gamma glutamyl transférases (GT), albuminurie, glycosurie, hématurie, etc.

II.3.2 - Informations sur le lieu de travail des participants RECORD

L'information sur le lieu de travail a été fournie par la Caisse Nationale d'Assurance Vieillesse (CNAV). Les fichiers administratifs reçus de la CNAV indiquaient l'employeur (ou les employeurs, avec un maximum de trois) de chaque participant par année, avec les numéros d'établissements correspondants. Les données ne nous permettaient pas de connaître la date de début ou de fin de contrat durant l'année de recrutement. En conséquence, les données ne nous permettaient pas de confirmer avec certitude que le participant était toujours ou déjà employé au cours de l'année du recrutement. Pour chaque année, nous n'avons retenu que l'employeur principal, qui était celui dont les participants avaient reçu le salaire le plus important. Afin de nous assurer de ne considérer que les lieux de travail où le participant travaillait (ou avait travaillé) au moment du recrutement, nous avons attribué à chaque individu l'établissement principal de travail de l'année précédant son inclusion dans l'Étude RECORD.

Nous avons ensuite géocodé le lieu de travail en utilisant notamment la base de données d'entreprises SIRENE (Système Informatique pour le Répertoire des Entreprises et de leurs Établissements) fournie par l'Institut National de la Statistique et des Études Économiques (INSEE). Cela nous a permis d'apparier les coordonnées spatiales du lieu de travail à une partie de notre population d'étude. Dans la première vague de l'étude, pour un total de 4536 participants travaillants pour lesquels un établissement de travail avait été identifié, nous avons pu faire correspondre les coordonnées du lieu de travail de 3837 participants. Pour 254 autres participants, les lieux de travail ont été géocodés en utilisant Google Maps en fonction des adresses trouvées dans les annuaires d'entreprises disponibles sur Internet. Les lieux de travail de respectivement 238 et 123 participants ont été géocodés au niveau de la zone IRIS et à la commune. Enfin, 84

participants pour lesquels le lieu de travail ne pouvait être trouvé et 124 participants pour lesquels le lieu de travail était situé en dehors de la région Île-de-France ont été exclus. Au final, les données sur le lieu de travail étaient disponibles pour 4331 participants de la vague 1 de la cohorte RECORD.

II.3.3 - Secteurs d'activité économique et catégories légales des établissements

Par le biais des données de la CNAV et des codes d'identification des établissements de travail des participants, les secteurs d'activité économique et les catégories légales des établissements ont été récupérés. Les 17 niveaux d'activité économique de la Nomenclature des Activités Française (NAF) de 2003 ont été utilisés. Nous avons utilisé la nomenclature de 2003 qui constituait la nomenclature des activités au moment de l'inclusion des participants dans l'étude. Parmi ces 17 niveaux d'activité, 11 étaient représentés dans notre échantillon : 1) Industrie manufacturière ; 2) Construction ; 3) Commerce, réparations automobiles et d'articles, domestiques ; 4) Hôtels et restaurants ; 5) Transports et communications ; 6) Activités financières ; 7) Immobilier, location et services aux entreprises ; 8) Administration publique ; 9) Education ; 10) Santé et action sociale et 11) Services collectifs, sociaux et personnels. Les secteurs suivants étaient donc exclus de nos échantillons : Agriculture, chasse, sylviculture ; Pêche, aquaculture, services annexes ; Industries extractives ; Production et distribution d'électricité, de gaz et d'eau ; Activités de ménage ; Activités extraterritoriales. Il est important de souligner que, premièrement, les secteurs d'activité économique ont été assignés au niveau de l'établissement ou du lieu de travail (établissement en tant qu'entité économique située à un endroit déterminé et subordonné à une autorité unique) plutôt qu'au niveau plus large de l'entreprise. Deuxièmement, les sites de travail sont classifiés selon la principale activité

économique de l'établissement. A titre d'exemple, les services de nettoyage ou les services administratifs intégrés dans un établissement du secteur de la construction sont classés comme faisant partie du secteur de la construction. Troisièmement, le secteur d'activité économique est indépendant de la catégorie socio-professionnelle de la personne. De ce fait, chaque secteur d'activité économique comprend des personnes de différentes classes socio-professionnelles (par exemple : des ouvriers, des employés, des professions intermédiaires et des cadres).

Les bases de données de l'INSEE et de la CNAV nous ont aussi permis de distinguer le secteur privé et le secteur public.

II.3.4 - Les données environnementales

Recueil des données environnementales

Les données environnementales ont été obtenues auprès des partenaires de l'Étude RECORD, notamment l'INSEE, l'Institut d'Aménagement et d'Urbanisme (IAU) d'Île-de-France et l'Institut Géographique National (IGN). A partir des bases de données environnementales obtenues auprès de ces organismes, des variables contextuelles rapportées aux adresses résidentielles et professionnelles (pour celles disponibles) des participants ont pu être constituées.

En utilisant les bases de données géographiques de l'Insee, l'ensemble des adresses résidentielles a été géocodé à l'aide de l'application GEOLOC. L'avantage de cette application réside dans le fait qu'elle permet d'attribuer les adresses à des zonages de types « pâtés de maisons » ou « petits quartiers » (îlots et IRIS) de façon beaucoup plus fiable que les procédures classiques de géocodage qui attribuent directement des coordonnées spatiales aux adresses. En

plus de ces codes de zone, l'application GEOLOC attribue aussi aux adresses un code correspondant à la voie et des coordonnées spatiales (x, y).

A partir de cette localisation précise des participants et à l'aide des différentes sources de données environnementales appariées, il est possible d'analyser les associations entre le quartier résidentiel et/ou l'environnement géographique du lieu de travail (voir procédure de géocodage ci-dessus) et l'obésité et les maladies cardiovasculaires.

Construction des variables environnementales

Presque toutes les variables environnementales ont été obtenues à l'aide d'un logiciel SIG (Système d'Information Géographique). Le SIG est un outil qui permet de relier des données géographiques spatialement référencées à une base de données permettant de faire des analyses sur ces données (173, 174). Dans le cadre de cette étude, nous avons utilisé le SIG ArcGIS 10. A l'aide de ce logiciel, nous avons pu définir les quartiers de résidence et de travail des participants de l'Étude RECORD et combiner les données environnementales venant de différentes bases de données afin d'obtenir les mesures d'expositions appropriées pour chaque dimension environnementale traitée.

Différentes zones géographiques à différentes échèles autour des lieux de résidence et des lieux de travail des participants ont été déterminées. En utilisant les coordonnées (x, y) de chaque adresse résidentielle ainsi que de chaque lieu de travail, nous avons pu créer autour de ces adresses une zone tampon (« buffer ») avec un rayon tenant compte du réseau de rues (175), allant de 500m à 2000m. Des variables environnementales ont ensuite été calculées au sein de ces zones tampons. Ces variables nous ont permis, par la suite, de conduire des analyses de

sensibilité pour comparer les associations observées en utilisant des zones tampons de différentes tailles.

L'ensemble de ces mesures environnementales construites dans des zones tampons est basé sur la notion de quartier proche, où sont regroupées les expositions avec lesquelles nous sommes en contact fréquent. De ce fait, dans chaque quartier, les expositions environnementales, parce qu'égo-centrées, peuvent varier selon les individus. C'est une approche qui veut identifier le plus fidèlement possible les expositions potentielles auxquelles les individus sont soumis au sein de leur environnement résidentiel ou professionnel en considérant, au centre, les coordonnées exactes du lieu de résidence ou du lieu de travail et les déplacements dans cette zone. Cette approche par zone tampon diffère de l'utilisation de découpages géographiques prédéterminés, comme les zonages administratifs (par exemple, le découpage en communes, en TRIRIS⁷ ou en IRIS), pour lesquels la mesure de l'exposition environnementale est moins performante car celle-ci reste la même pour deux personnes qui habitent à des endroits différents d'un même TRIRIS (l'un habitant par exemple au centre et l'autre près de la frontière de l'unité de zonage). Celui qui habite à la frontière aura plus de chance d'être également exposé à l'unité de zonage adjacente (176).

Afin de créer une typologie relative au niveau d'urbanité des quartiers résidentiels et professionnels (qui renvoie au niveau de densité de l'espace urbain), nous avons eu recours à des méthodes statistiques d'analyse par grappes (« *cluster analysis* ») basées sur quatre variables de densité (la densité de population⁸, la densité de destinations⁹, la densité d'intersections de rues¹⁰

⁷ Un TRIRIS est une zone géographique intra-communale définie par l'INSEE qui correspond au regroupement de trois IRIS.

⁸ Correspond au nombre d'habitants par km² présents dans la zone tampon basée sur le réseau de rues

et le ratio de l'aire de la zone tampon basée sur le réseau de rues sur l'aire de la zone tampon circulaire¹¹). Ces analyses ont été faites séparément pour le quartier résidentiel et le quartier professionnel. Nous avons utilisé la méthode dite hiérarchique de Ward qui sélectionne d'abord chaque observation multidimensionnelle comme étant une seule grappe (niveau individuel), puis fusionne successivement les deux prochaines grappes les plus proches en terme de distance euclidienne jusqu'à ce qu'il n'y ait plus qu'une seule grappe, qui regroupe toutes les observations (177). Le nombre final de grappe a été choisi afin de sélectionner les quartiers les plus contrasté entre eux ou plus de la moitié des variations des variables ont été prise en compte ($r > 0,5$).

II.4. Analyses statistiques

La principale question de recherche de cette thèse est de comprendre certains mécanismes qui expliquent l'augmentation de la prévalence de l'obésité et de facteurs de risque cardiovasculaire chez certains groupes d'une population donnée. Plus spécifiquement, il s'agit de quantifier et d'interpréter les disparités socio-spatiales d'excès de graisse corporelle en tenant compte à la fois des caractéristiques socio-économiques individuelles, de l'environnement de résidence et du lieu de travail, et du secteur d'activité économique de travail. Ensuite, nous appréhendons et comparons les relations entre ces secteurs d'activité économique et différents facteurs de risque cardiovasculaire. Finalement, nous évaluons les déterminants de l'évolution des paramètres anthropométriques au cours du temps et l'importance des valeurs manquantes dans les

⁹ Correspond au nombre de lieux d'activités auxquels un individu peut se rendre (lieux administratifs, magasins public/privé, lieux de divertissement, etc.) présents dans la zone tampon basée sur le réseau de rues

¹⁰ Correspond au nombre d'intersections (d'au moins 3 route) dans la zone tampon basée sur le réseau de rues

¹¹ Correspond au ratio de l'aire de la zone tampon basée sur le réseau de rues sur l'aire de la zone tampon circulaire.

analyses. Du fait de l'identification multicritère de la population étudiée au cours de cette thèse, nous avons choisi d'utiliser des modèles de régression linéaire, de type multiniveaux car ceux-ci permettent de prendre en considération la structure hiérarchique de nos données (participants résidant au sein des mêmes zones). Les modèles multiniveaux fournissent des outils utiles à la description et à l'estimation des effets de l'environnement sur la santé des individus. Les estimations des forces d'associations et leurs écarts-types au moyen de ces modèles permettent de prendre en compte la corrélation intra-zone de la variable dépendante étudiée. De plus, la prise en compte de la corrélation intra-zone de la variable dépendante peut être utile pour l'interprétation des associations entre les facteurs environnementaux et les variables dépendantes (178, 179). Dans le cadre de nos analyses, nous avons estimé des modèles multiniveaux dans lesquels les individus étaient groupés par TRIRIS. Nos modèles ont été estimés à partir du maximum de vraisemblance.

Toutes nos variables dépendantes, à savoir, dans un premier temps, le pourcentage de masse grasse (%MG), l'indice de masse grasse (IMG), dans un second temps, l'indice de masse corporelle (IMC), le tour de taille (TT), la pression systolique (PAS) et diastolique (PAD), la pression pulsée (PP), le taux de cholestérol total, de glycémie, et le rythme cardiaque au repos et dans un troisième temps, le changement d'IMC entre la vague 1 et 2, ont été analysées sous forme de variables continues. Plusieurs raisons ont motivé ce choix. En effet, d'un point de vue méthodologique, choisir des variables dépendantes continues permet de prendre en compte toutes les variations existant dans les données et pas seulement les variations autour des seuils généraux établis par l'OMS. De plus, d'un point de vue statistique, l'estimation des modèles multiniveaux avec des variables réponses continues permet une meilleure appréciation des variances inter et intra-zone.

Lors de nos analyses, nos modèles ont été ajustés sur plusieurs variables sociodémographiques individuelles ainsi que sur plusieurs variables démographiques et socio-économiques des quartiers. Nous détaillons ces variables dans les prochains chapitres consacrés aux articles empiriques. Afin de voir l'impact des choix des variables sur nos analyses, notre stratégie comportait à chaque fois plusieurs étapes : 1) La variance inter-quartier des variables de santé a été déterminée en ajustant nos modèles seulement sur l'âge et le sexe. 2) Ensuite, nous avons estimé les modèles en ajoutant des caractéristiques individuelles. Ces variables individuelles ont été forcées dans tous les modèles pour permettre une comparaison entre les modèles estimés pour les différentes variables dépendantes. 3) Finalement, pour compléter les modèles d'analyse, dans notre premier article, nous avons testé nos différentes variables contextuelles. Nous avons d'abord introduit ces variables une à une dans nos modèles de base, puis avons testé simultanément celles que nous avons retenues afin de converger vers un modèle ne retenant que les variables contextuelles indépendamment associées à chaque variable réponse (modèle parcimonieux). Dans notre deuxième article, afin de faciliter la comparaison entre les différents facteurs de risque cardiovasculaire, nous avons utilisé un modèle d'analyse unique, préalablement sélectionné, répété pour chaque variable de santé. Finalement, dans notre troisième article, des modèles linéaires simples, ajustés sur l'âge et le sexe, ont été utilisés afin d'estimer le lien entre le niveau d'éducation individuel et l'évolution de l'IMC dans le temps entre la vague 1 et 2. Par la suite, afin de gérer les données manquantes dans nos analyses et de comparer différentes approches, nous avons utilisé les trois méthodes de gestion des données manquantes les plus communément mobilisées (180, 181). Premièrement, la méthode des cas complet (Complete Case Analyses, CCA) qui consiste à simplement exclure les individus avec des données manquantes a été utilisée. Deuxièmement, la pondération par la probabilité inverse

(Inverse Probability Weighting, IPW) qui pondère les cas complets par la probabilité inverse d'être un cas complet a été testée. Troisièmement, nous avons eu recours à la méthode par imputation multiple (Multiple Imputation, MI) qui utilise les données sur tous les sujets (y compris ceux avec des données manquantes) afin de créer un certain nombre de jeux de données imputées en générant plusieurs valeurs imputées pour chaque donnée manquante. Chaque jeu de données imputées est analysé séparément, et leurs estimations sont combinées en utilisant les règles de Rubin de 1987 (182).

Toutes nos analyses statistiques ont été réalisées à l'aide du logiciel SAS 9.3 et du logiciel R. Le critère d'information d'Akaike (AIC) a permis de comparer les modèles estimés au maximum de vraisemblance sous SAS. Plus l'AIC est petit, plus l'adéquation du modèle aux données pénalisée par le nombre de paramètres est satisfaisante.

CHAPITRE III - Relations entre les caractéristiques environnementales résidentielles et professionnelles, les secteurs d'activité économique et la masse grasse

III.1. Présentation de l'article

Dans ce chapitre, nous présentons un travail visant à mieux comprendre les relations qui existent, au-delà des caractéristiques socio-économiques individuelles, entre les différents aspects de l'environnement de résidence, de l'environnement professionnel, et les secteurs d'activité économique et l'obésité appréhendée par des indicateurs de masse grasse. De nombreuses études ont analysé les relations entre les caractéristiques environnementales résidentielles et l'obésité mesurée par l'IMC. En revanche, très peu d'études ont tenu compte de l'environnement autour du lieu de travail dans ces relations et quasiment aucune n'a étudié simultanément les associations entre ces deux environnements géographiques (résidence et travail) et l'obésité. De plus l'IMC, qui est l'indicateur de l'obésité systématiquement utilisé dans ces études, comporte de nombreuses limites.

L'objectif de notre article était d'examiner les associations entre, d'une part, les facteurs individuels sociodémographiques, les caractéristiques du quartier résidentiel et du quartier autour du lieu travail (y compris la distance domicile-travail), et le secteur d'activité économique des

participants et, d'autre part, la masse grasse (MG) évaluée par l'Indice de Masse Grasse (IMG) et le pourcentage de masse grasse (%MG).

L'intérêt du travail était d'appréhender des caractéristiques environnementales différentes qui influencent la situation pondérale et qui renvoient, au-delà du quartier de résidence, à des circonstances de vie associées au lieu de travail, c'est-à-dire à l'environnement géographique du lieu de travail, mais aussi aux conditions de travail. Dans le cadre de cette étude, dans un premier temps, l'établissement de travail des participants a été identifié et géolocalisé à partir de la base de données de la Caisse Nationale d'Assurance Vieillesse (CNAV). Des facteurs environnementaux ont ensuite été pris en compte dans des zones de 1km de rayon tenant compte du réseau de rues centrées sur le domicile et sur le lieu de travail. Dans un second temps, les données de la CNAV nous ont permis (sans connaître le détail de l'activité quotidienne) d'identifier le secteur d'activité économique [référéncé par la Nomenclature des Activités Française (NAF)] de l'établissement de travail pour chaque participant.

Afin d'étudier ces phénomènes et de démêler les effets des différentes dimensions de l'environnement, du fait des corrélations élevées observées entre les variables, nous avons construit, avec la méthode d'analyses par grappes, une typologie sur le degré d'urbanité du quartier de résidence et de travail. Une subdivision en quatre grappes a été faite, représentant les types de quartiers les plus contrastés entre eux et qui peuvent prendre en compte plus de la moitié des variations des variables de quartiers sélectionnées. Par la suite, lors de nos analyses statistiques, nous avons estimé des modèles incluant les variables sociodémographiques individuelles qui étaient associées avec l'IMG et le %MG (i). L'âge, le niveau d'éducation individuel, le niveau d'éducation parental, le stress financier, le statut matrimonial et l'indice de

développement humain (IDH) ont servi de base pour tester les caractéristiques environnementales (ii). Les secteurs d'activité économique ont ensuite été introduits dans les modèles (iii). Des interactions d'effets ont été testées entre le niveau d'éducation individuel et le secteur d'activité économique (iv). Nous avons également testé les associations entre la distance résidence-travail et l'IMG et le %MG (v). Finalement, les modèles ont été ajustés sur l'IMC pour voir si la masse grasse était différemment associée avec les facteurs sociaux et environnementaux que l'IMC (vi). Du fait de la morphologie différente entre hommes et femmes, toutes les analyses ont été stratifiées sur le sexe.

Les résultats obtenus dans le cadre de ces analyses ont montré que, au-delà des caractéristiques sociodémographiques individuelles, le quartier résidentiel, les secteurs d'activité économique et la distance résidence – travail étaient indépendamment associés avec l'IMG et le %MG. Chez les hommes, le niveau d'éducation et la densité de population dans le quartier résidentiel étaient indépendamment associés à l'IMG et au %MG. L'IMG et le %MG augmentaient avec la baisse du niveau d'éducation et de densité de population dans le quartier résidentiel. La densité de population résidentielle était plus fortement associée à l'IMG et au %MG que les autres variables de densité. Chez les hommes, aucune autre variable résidentielle n'était indépendamment associée à la masse grasse. En ce qui concerne les secteurs d'activité économique, après ajustement sur les variables individuelles et résidentielles, l'IMG et/ou le %MG étaient particulièrement élevés chez les hommes qui travaillent dans la construction et dans les secteurs du transport et de la communication et, dans une moindre mesure, parmi ceux qui travaillent dans les secteurs de l'industrie manufacturière ; de l'immobilier, de la location et des services aux entreprises ; de la réparation et du commerce de véhicules automobiles et de moto ;

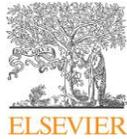
et des activités financières par rapport à ceux qui travaillent dans le secteur de l'éducation. Aucune interaction d'effet n'a été observée entre le niveau d'éducation et le secteur d'activité économique. La distance domicile-travail n'était associée ni avec l'IMG, ni avec le %MG. Après ajustement sur l'IMC, il y avait persistance des associations entre les secteurs d'activité économique et l'IMG et le %MG. Aucune variable environnementale professionnelle n'était associée à l'IMG ni au %MG.

Les femmes vivant dans des quartiers avec un faible niveau d'éducation avaient un %MG plus élevé chez les femmes, le niveau d'éducation du quartier résidentiel était associé avec l'IMG avant ajustement sur la distance domicile-travail, mais pas après. Aucune autre variable résidentielle n'était liée à la masse grasse. Les secteurs d'activité économique n'étaient pas été associés au %MG chez les femmes. Cependant, comparativement aux femmes travaillant dans le secteur des transports et des communications, celles qui travaillent dans le secteur du travail social et de la santé avaient un IMG plus élevé. Celles qui travaillent dans l'administration publique tendaient aussi à avoir un IMG plus élevé. Il n'y avait pas d'interaction d'effets entre les niveaux d'éducation individuels et les secteurs d'activité économique chez les femmes. Le degré d'urbanisation dans le quartier résidentiel ou dans le quartier du lieu de travail n'était pas associé à l'IMG ni au %MG. Une relation dose-réponse a été observée entre la distance domicile-travail et l'IMG chez les femmes: plus les femmes vivent près de leur lieu de travail, moins l'IMG était élevé.

En conclusion, par les associations mise en évidence, cette étude empirique montre l'intérêt de prendre en considération l'environnement résidentiel ainsi que les secteurs d'activité économique en plus des caractéristiques sociodémographiques individuelles afin d'appréhender

les disparités de poids et d'excès de graisses corporelles, et de rendre possible une analyse multicritère des populations à risque. De plus, cette étude suggère que la composition corporelle évaluée par l'IMG et le %MG, en plus de l'IMC, permet de saisir de façon plus fine les effets des secteurs d'activité économique sur la présence de graisses corporelles, puisque des effets des secteurs d'activité sur la masse grasse ont été identifiés à corpulence donnée.

La suite de ce chapitre présente l'intégralité de l'article publié dans la revue *Annals of Epidemiology*



Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org



Original article

Residential neighborhood, geographic work environment, and work economic sector: associations with body fat measured by bioelectrical impedance in the RECORD Study

Antoine Lewin MSc^{a,b,*}, Bruno Pannier MD, PhD^c, Julie Méline PhD^{a,b}, Noëlla Karusisi PhD^{a,b},
Frédérique Thomas PhD^c, Basile Chaix PhD^{a,b}

^aInserm, U707, Research Unit in Epidemiology, Information Systems, and Modeling, Paris, France

^bUniversité Pierre et Marie Curie - Paris6, Faculty of Medicine, UMR-S 707, Paris, France

^cCentre d'Investigations Préventives et Cliniques, Paris, France

ARTICLE INFO

Article history:

Received 16 April 2013

Accepted 23 December 2013

Available online 30 December 2013

Keywords:

Epidemiology

Body composition

Residence neighborhood

Workplace neighborhood

Work economic sector

ABSTRACT

Purpose: Studies of associations between geographic environment and obesity have mostly examined body mass index and focused on residential neighborhoods. We investigated associations between residential neighborhoods, geographic work environments, and work economic sectors and the fat mass index (FMI) and percentage of fat mass (%FM).

Methods: Data on 4331 participants from the French RECORD Study geolocated at their residence and workplace were analyzed. Body composition was assessed by bioelectrical impedance analyzers. Multilevel linear regression was used to investigate the determinants of FMI and %FM.

Results: After adjustment, among men, the FMI and %FM increased independently with decreasing density of population and educational level in the residential neighborhood. Among women, the residential educational level was related to the FMI and %FM. Among men, a higher FMI and %FM were observed among participants working in the construction and transportation/communication sectors than in the education sector. For women, the FMI was higher among participants working in the public administration and health/social work sectors than in the transport/communication sector. A long home-work distance was associated with a higher FMI among women. There was evidence that body mass index cannot fully capture work economic sector effects on fat mass.

Conclusions: Public health interventions to reduce social/territorial disparities in obesity should also consider the different contexts to which the participants belong, such as residential environments and work economic sectors.

© 2014 Elsevier Inc. All rights reserved.

Introduction

Numerous studies have investigated relationships between environmental resources or exposures and obesity [1–6]. A literature review reported that 90% of 131 studies on the relationships between geographic environments and metabolic risk factors only considered residential neighborhoods to define environmental exposure, that 6% exclusively examined nonresidential exposures, and that only 4% accounted for both residential and nonresidential environments (in most cases, school environments) [6]. Among adults, two studies related to the determinants of obesity assessed environmental factors around both the residence and the

workplace [7,8]. Accounting for the environmental correlates of obesity related to both the residential and workplace environment is important to identify the environmental barriers to healthy behaviors that may be present in the different geographic life contexts of individuals [6,9].

Body mass index (BMI) is the most widely used and accepted index to characterize obesity in individuals [10,11]. However, BMI has important limitations, as it is not a direct measure of fat mass (FM). On the one hand, it could lead to classify individuals with a high muscle mass as overweight or obese. On the other hand, subjects with a high percentage of fat could have a BMI within the normal range [12–14]. It is recognized that obesity levels are more accurately assessed with bioelectrical impedance analysis [15]. For example, in a study of 433 women (18–73 years) and 264 men (18–83 years) (i.e., with an age range that encompasses the one considered in our study), 64% of the women and 24% of the men

* Corresponding author. UMR-S 707, Faculté de Médecine Saint-Antoine, 27 rue Chaligny, 75012 Paris, France. Tel.: +33 1 44 73 89 54; fax: +33 1 44 73 84 62.
E-mail address: lewin@u707.jussieu.fr (A. Lewin).

1047-2797/\$ – see front matter © 2014 Elsevier Inc. All rights reserved.
<http://dx.doi.org/10.1016/j.annepidem.2013.12.001>

with a BMI below 30 were found to be obese according to bioelectrical impedance analysis [16].

The aim of the present study was to investigate the associations between, on the one hand, individual sociodemographic factors, residential neighborhood, and workplace neighborhood characteristics (including the home-work distance), and the work economic sector of participants and, on the other hand, FM assessed as the fat mass index (FMI) and the percentage of fat mass (%FM). We also assessed whether BMI allows researchers to fully capture environmental effects on body fat.

Methods and procedures

Study population

The RECORD Study

Data from the first wave of the RECORD (Residential Environment and Coronary Heart Disease) Cohort Study (www.record-study.org) were used for cross-sectional analyses. As described elsewhere [17–24], 7290 participants aged 30–79 years were recruited without *a priori* sampling in 2007–2008 during free preventive medical checkups conducted by the Centre d'Investigations Préventives et Cliniques in the Paris metropolitan area (convenience sample without *a priori* randomization). As an *a priori* eligibility criterion, only participants residing in one of 10 (20) administrative districts of Paris or in one of 111 other municipalities in the metropolitan area were selected for the study. Of the persons selected for participation, 83.6% agreed to participate and completed the data collection protocol. The study protocol was approved by the French Data Protection Authority.

Selection of participants

Administrative files from CNAV (National Old Age Insurance System) and INSEE (French National Institute of Statistics and Economic Studies) were used to assess and geocode participants' workplaces (see Appendix 1 for the detailed procedure). To be sure to only consider workplaces where the participant was already working or had worked at the time of recruitment in the study (and thus avoid reverse causality problems), we assigned to each individual the main work establishment of the year preceding his/her inclusion in the study. A workplace was geocoded for 4331 participants residing and working in the Ile-de-France region. After excluding participants with missing information on FM, 4078 participants were included in the analyses.

In the RECORD Study, the participants can be attributed to a variety of administrative area units, including census block groups or IRIS neighborhoods and census tracts or TRIRIS neighborhoods. IRIS neighborhoods (known as IRIS2000) were determined by INSEE from the 1999 French Census for dividing the country into units of approximately equal population size. In French, IRIS is an acronym of "aggregated units for statistical information," and "2000" refers to the target size of 2000 residents per unit [25]. TRIRIS neighborhoods (referred to below as census tracts) were defined by INSEE by grouping three contiguous [26] IRIS neighborhoods from the same municipality (sometimes more or less) [25]. The average area size of census tracts in our study was 1.2 km² (standard deviation = 2.5 km²).

The 4078 participants were living in 645 TRIRIS census tract neighborhoods. Thus, these census tracts comprised an average of 6.3 participants. Among those 645 census tracts, 414 (64%) had at least five participants. It was not the primary aim of this study to assess between-neighborhood variations in the FMI and %FM.

Measures

Outcome

Data related to height and weight were obtained from medical examinations. Height (using a wall-mounted stadiometer) and weight (using calibrated scales) were recorded by a nurse [27]. FM was measured using a Tanita foot-to-foot bioelectrical impedance analyzer (TBF-300 Tanita Corporation of America, Inc., Arlington Heights, IL). This analyzer measured body weight and FM in kilograms. Age, sex, and height were entered manually into the keypad interface by a nurse. Participants were asked to leave their shoes and socks and stand barefoot on the bioelectrical balance to allow performing the measurement. For the analyses, we used the FMI [28] (FM/height² in kg/m²) and the %FM [29] (FM/body weight × 100) as continuous variables.

Individual sociodemographic variables

The age of the participants was divided into three classes: 30–44, 45–59, and 60–79 years. Marital status was coded in two classes (living alone or in couple). Individual education was coded into four classes: no education, primary education or lower secondary education, higher secondary education or lower tertiary education, and upper tertiary education. Parental education as the average of the education level of the father and of the mother was divided into three classes. Regarding occupation, four categories were distinguished: blue-collar workers; low white-collar workers; intermediate occupations; and high white-collar workers. Household income adjusted for household size was divided into four categories based on the quartiles. A binary variable was determined for financial strain (reporting financial difficulties). The 2004 Human Development Index (HDI) of each participant's country of birth was taken into account, as a proxy for the country's social development level [30]. The HDI was divided into three classes: low-development countries (HDI < 0.5), medium-development countries (HDI = 0.5–0.8), and high-development countries (HDI > 0.8) [31].

Residential and workplace neighborhood variables

The socioeconomic status of the neighborhood was assessed with the educational level (2006 Population Census) and with the household income (General Directory of Taxation) of the residents. Four variables were taken into account to assess density: density of population, density of destinations (all types of public and private services), density of three-way street intersections, and the ratio of the area size of a 1 km street network buffer to the area size of a 1 km radius circular buffer (as a measure of the extent to which the street network allows one to reach the surrounding space). Three specific exposure variables were additionally considered: the surface of green spaces, the number of supermarkets/hypermarkets, and proportion of fast-food restaurants among all restaurants. Details on the definition of the variables are provided in Table 1.

All the residential and workplace environmental variables were computed within street network buffers with a radius of 1 km centered on participants' residences and participants' workplaces (these buffers comprise the area that is accessible within 1 km along the street network) [32]. Street network buffers may allow researchers to account for certain physical barriers in the definition of neighborhoods or reflect that certain portions of the street network (enclaves) are not significantly connected to the other parts of the city. It was possible to determine the socioeconomic variables within street network buffers based on census or tax registry data geocoded at the building address level. We assumed that environmental conditions may be associated with health when factors are measured in "walkable" areas from the residence and the workplace, and 1 km along the street network was conceptualized as an acceptable distance to reach services in urban areas. Moreover, a

Table 1

Description of neighborhood socioeconomic indicators, neighborhood density indicators, and of other neighborhood exposures (all assessed in 1 km radius street network buffers around participants' residences and workplaces)

Neighborhood characteristics	Data	Measure used
Socioeconomic variables		
Neighborhood level of education	Population Census of 2006 geocoded at the building address by INSEE	Aggregation of individual data within street network buffers: proportion of residents aged ≥ 15 y with university education
Neighborhood income level	Population Census of 2006 geocoded at the building address by INSEE	Median household income per consumption unit in the street network buffers, Tax Registry of General Directorate of Taxation (2006)
Neighborhood density		
Density of population	Population Census of 2006 geocoded at the residential address by INSEE	Aggregation of population data within street network buffers: numbers of inhabitants by kilometers
Density of destinations	Geocoded destinations from the 2008 Permanent Database of Facilities of INSEE	GIS processing: count of destinations (administrations, public/private shops, entertainment facilities, etc.) within street network buffers
Density of street intersections	Data from IGN on the street and road network, 2008	GIS count the number of intersection (at least three way) in a 1 km street network buffer
Ratio of areas of street network to circular buffers	Data from IGN on the street and road network, 2008	Ratio of the area of a 1 km street network buffer to the area of a 1 km circular buffer
Other specific exposures		
Surface of green spaces	Linear and polygonal data from IAU-IdF on public parks and green spaces in 2008	GIS processing: surface within street network buffers
Number of supermarkets/hypermarkets	Geocoded supermarkets/hypermarkets in 2008 from the Permanent Database of Facilities of INSEE	GIS processing: count of supermarkets/hypermarkets within street network buffers
Proportion of fast-food restaurants (compared with the total number of restaurants)	Geocoded restaurants in 2007 from the SIRENE database of INSEE	GIS processing: ratio between the count of fast-food restaurants and the total count of restaurants within street network buffers

GIS = Geographic Information System; IAU-IdF = Institute of Urban Planning of region Ile-de-France; IGN = National Geographic Institute; INSEE = National Institute of Statistics and Economic Studies.

recent sensitivity analysis based on the RECORD Study showed that the accessibility to sport facilities was optimally assessed when considering a 1 km street network distance from the residence [33].

ArcInfo 10 software (ESRI, Redlands, CA, USA) and its Network analyst applied to street network data from the National Geographic Institute were used to derive such buffers. This software also allowed us to compute the street network distance between home and work.

Work economic sector and legal category variables

As noted above, CNAV data from 2006–2007 allowed us to identify the work economic sector of each participant. We used the 17 levels of the 2003 French classification of activities (NAF) to classify work economic sectors. Of those 17 economic sectors, only 11 were represented in our sample. The models used the work economic sector with the lowest FMI and %FM as the reference group to emphasize the hierarchy of work economic sectors according to FM. The CNAV data also enabled us to distinguish between the private and the public sector.

Statistical analysis

Definition of a typology on the urbanicity degree of the neighborhood

Cluster analytic methods were applied to assess the urbanicity degree of the residential and workplace neighborhoods based on the four density variables (separate analysis for the residential and workplace). Ward's methods starts with each multidimensional (individual-level) observation as a single cluster and then repeatedly merges the next two closest clusters in terms of Euclidian distances between observations until a single, all-encompassing cluster remains [34]. Application of this method resulted in a neighborhood typology wherein neighborhoods that are substantively comparable on the four selected density characteristics were grouped together although they were not necessarily geographically adjacent [35,36]. This approach considers the characteristics of places but not their spatial location and, therefore, establishes

groups of noncontiguous neighborhoods [36]. This nongeographic clustering technique yielded a model-derived measure (i.e., urbanicity degree) that was used as a contextual variable to predict FM in our multilevel models [37]. Based on an assessment of the corresponding dendrograms, we retained a subdivision into four clusters, as representing the most contrasted neighborhood types with over half of the variations in the selected neighborhood variables being accounted for (R^2 was equal to 0.63 and 0.66, respectively, for the residential and workplace neighborhood cluster analyses). Additional information on this typology, henceforth referred to as the urbanicity degree, is provided in Appendix 2 and Appendix 3.

Multilevel regression analysis of fat mass

Multilevel linear regression models with a random effect at the TRIRIS census tract level (level 2) were estimated to account for within-neighborhood correlation in the FMI and %FM. In the present study, TRIRIS neighborhoods were used only to define the hierarchical structure of multilevel models (not for the measurement of environmental variables). All the analyses were stratified by sex. The statistical analyses involved seven steps. First, we estimated models including the individual sociodemographic variables that were associated with the FMI and %FM. Second, the residential neighborhood variables were tested. Third, the different work economic sectors were introduced into the models. Fourth, we tested interactions of effects between, on the one hand, the individual education or occupational group, and on the other hand, the work economic sector. Fifth, we tested the different variables related to the geographic work environment. Sixth, we additionally tested the associations between the street network distance between home and work and the FMI and %FM. Seventh and finally, the models for the FMI and %FM were adjusted for BMI to assess whether FM was differentially associated with social/environmental factors than BMI.

At each step, we only retained the variables that were associated with the FMI or %FM. All the analyses were conducted with SAS 9.3 (SAS Institute, Cary, NC).

Results

In our sample, the median FMI was 5.65 kg/m² (interquartile range: 4.27–7.59 kg/m²). The median %FM was 22.8% (interquartile range: 18.3%–28.5%). Variations in the FMI and %FM according to the selected individual and contextual characteristics are shown in [Appendix 4](#).

Associations with the FMI and %FM among men

The multilevel linear regression models for the FMI and %FM are presented in [Table 2](#) and [Table 3](#), respectively. After adjustment for individual socioeconomic characteristics, the education level and density of population in the residential neighborhood were independently associated with the FMI and %FM. The FMI and %FM increased with decreasing neighborhood education and decreasing density of population. The residential density of population was more strongly associated with the FMI and %FM than the other density variables and the overall urbanicity degree. No other residential variables were independently associated with FM. Regarding work economic sectors, after adjustment for individual

Table 2
Beta coefficients and 95% CIs for associations with the FMI from multiple multilevel linear regression models, RECORD Cohort Study^a

Variables	Men, β (95% CI)	Women, β (95% CI)
Age (vs. 30–44), y		
45–59	0.65 (0.46 to 0.82)	0.91 (0.49 to 1.33)
60–79	1.14 (0.82 to 1.47)	1.62 (0.91 to 2.32)
Individual education (vs. high education)		
Medium-high education	0.33 (0.11 to 0.55)	0.04 (–0.45 to 0.54)
Medium-low education	0.39 (0.13 to 0.65)	0.79 (0.17 to 1.40)
Low education	0.22 (–0.16 to 0.61)	1.51 (0.65 to 2.38)
Parental education (vs. high education)		
Intermediate education	0.32 (0.11 to 0.53)	0.39 (–0.11 to 0.88)
Low education	0.28 (0.03 to 0.53)	0.66 (0.09 to 1.23)
Perceived financial strain	0.28 (0.02 to 0.53)	0.78 (0.28 to 1.28)
Living alone (vs. as a couple)	–0.47 (–0.67 to –0.26)	—
HDI of country of birth (vs. high development)		
Middle	—	–0.03 (–0.61 to 0.54)
Low	—	2.23 (1.36 to 3.10)
Residential education level (vs. high)		
Mid-high	–0.01 (–0.25 to 0.22)	—
Mid-low	0.02 (–0.23 to 0.28)	—
Low	0.54 (0.24 to 0.83)	—
Residential density of population (vs. high)		
Mid-high	0.28 (0.42 to 0.52)	—
Mid-low	0.24 (–0.02 to 0.51)	—
Low	0.36 (0.09 to 0.63)	—
Work economic sector		
Manufacturing industry	0.88 (0.04 to 1.73)	0.31 (–0.78 to 1.40)
Construction	1.15 (0.48 to 1.82)	0.18 (–2.44 to 2.81)
Commercial repair of motor vehicles and motorcycles	0.60 (0.00 to 1.20)	0.43 (–0.60 to 1.46)
Hotels and restaurants	0.30 (–0.35 to 0.95)	1.08 (–0.21 to 2.360)
Transport and communications	0.93 (0.29 to 1.57)	Ref.
Financial activities	0.57 (–0.03 to 1.18)	0.55 (–0.53 to 1.62)
Real estate, renting, and business services	0.66 (0.10 to 1.23)	0.39 (–0.54 to 1.33)
Public administration	0.08 (–0.76 to 0.93)	1.40 (–0.11 to 2.68)
Education	Ref.	1.30 (0.030 to 2.58)
Health and social work	0.10 (–0.69 to 0.89)	1.38 (0.26 to 2.51)
Collective, social, and personal services	0.23 (–0.37 to 0.83)	0.35 (–0.68 to 1.40)
Home-work distance (vs. low)		
Medium-low	—	0.17 (–0.38 to 0.72)
Medium-high	—	0.27 (–0.28 to 0.82)
High	—	0.84 (0.32 to 1.43)

β = beta coefficient; 95% CI = 95% confidence interval.

^a Models were adjusted for all the individual and contextual variables that were independently associated with the outcome.

Table 3
Beta coefficients and 95% CIs for associations with the %FM from multiple multilevel linear regression models, RECORD Cohort Study^a

Variables	Men, β (95% CI)	Women, β (95% CI)
Age (vs. 30–44), y		
45–59	2.09 (1.63 to 2.56)	2.85 (1.88 to 3.83)
60–79	3.62 (2.78 to 4.46)	5.04 (3.41 to 6.67)
Individual education (vs. high education)		
Medium-high education	0.95 (0.39 to 1.51)	0.37 (–0.75 to 1.50)
Medium-low education	1.15 (0.48 to 1.82)	1.92 (0.55 to 3.30)
Low education	0.66 (–0.32 to 1.66)	4.07 (2.22 to 5.92)
Living alone (vs. as a couple)	–1.21 (–1.74 to –0.68)	—
Parental education (vs. high education)		
Intermediate education	0.78 (0.23 to 1.33)	—
Low education	0.80 (0.16 to 1.45)	—
Perceived financial strain	—	1.66 (0.49 to 2.82)
HDI of country of birth (vs. high development)		
Middle	—	–0.09 (–1.43 to 1.24)
Low	—	8.28 (3.28 to 7.27)
Residential education level (vs. high)		
Mid-high	–0.06 (–0.67 to 0.53)	0.22 (–1.05 to 1.51)
Mid-low	–0.01 (–0.67 to 0.65)	0.54 (–0.79 to 1.88)
Low	1.27 (0.51 to 2.02)	2.51 (1.09 to 3.93)
Residential density of population (vs. high)		
Mid-high	0.75 (0.13 to 1.38)	—
Mid-low	0.75 (0.06 to 1.44)	—
Low	1.00 (0.31 to 1.70)	—
Work economic sector		
Manufacturing industry	1.26 (–0.28 to 2.80)	—
Construction	2.96 (1.23 to 4.69)	—
Commercial repair of motor vehicles and motorcycles	1.81 (0.25 to 3.36)	—
Hotels and restaurants	1.02 (–0.67 to 2.72)	—
Transport and communications	2.35 (0.69 to 4.00)	—
Financial activities	1.55 (–0.01 to 3.12)	—
Real estate, renting, and business services	1.83 (0.36 to 3.30)	—
Public administration	0.23 (–1.95 to 2.43)	—
Education	Ref.	—
Health and social work	0.30 (–1.74 to 2.35)	—
Collective, social, and personal services	0.87 (–0.69 to 2.43)	—

β = beta coefficient; 95% CI = 95% confidence interval.

^a Models were adjusted for all the individual and contextual variables that were independently associated with the outcome.

and residential variables, the FMI and/or %FM were particularly higher among men working in the Construction and in the Transport/communication sectors and, to a lesser extent, among those working in the Manufacturing industry, in the Real estate, renting, and business services, in the Commercial repair of motor vehicle and motorcycle, and in the Financial activities sectors than among those working in the Education sector. No interaction of effects was found between the education level or the occupational group and the work economic sector among men. Among men, the home-work distance was associated neither with the FMI nor with the %FM.

After additional adjustment for BMI, there were persisting associations between work economic sectors and the FMI and %FM ([Appendix 5](#) and [Appendix 6](#), respectively), suggesting a lower FM in the Education sector and Public administrations compared with all other sectors (the correlations between the FMI or %FM and BMI were 0.92 and 0.84, respectively). No workplace neighborhood variable was associated with the FMI and %FM.

Associations with the FMI and %FM among women

As shown in [Table 2](#) and [Table 3](#), after adjusting for individual socioeconomic variables, residential neighborhood education level was associated with the %FM. Women living in neighborhoods with

a low-education level had a higher %FM. The residential neighborhood education level was associated with the FMI before adjustment for the home-work distance but not after. No other residential variables were related to FM. The work economic sector was not associated with the %FM among women. However, compared with the women working in the Transport and communications sector, those working in the Health and social work sector had a higher FMI and those working in Public administrations tended as well to have a higher FMI. There was no interaction of effects between individual socioeconomic groups and work economic sectors among women. The degree of urbanicity in the residential or in the workplace neighborhood was not associated with the FMI or the %FM among women. A dose-response association was observed between distance from home to work and the FMI among women: the closer women lived from their workplace, the lower their FMI and/or %FM was.

Differently than among men, controlling for BMI removed all the associations documented between, on the one hand, the residential neighborhood, work economic sectors, or home-work distance and on the other hand, the FMI and %FM (the correlations between the FMI or %FM and BMI were 0.95 and 0.84, respectively).

Discussion

The present study suggests that the residential neighborhood, work economic sector, and home-work distance were independently associated with the FMI and %FM after adjusting for individual socioeconomic variables.

Strengths and limitations

Regarding study strengths, first, unlike most epidemiologic studies that investigated environmental effects on obesity with BMI [6], our study was performed with body composition factors. Although rare studies, including ours [6,17], have analyzed waist circumference, our study is the very first to examine neighborhood determinants of FM. Second, to our knowledge, this is the first study that simultaneously uses variables for the residential neighborhood, geographic work environment, and work economic sector in relation to body weight or fat. This is important because neighborhood or contextual factors related to the obesity risk are likely not limited to the neighborhood of residence [38–40]. Accounting for work economic sectors in studies of contextual effects on obesity is particularly innovative and useful for health promotion initiatives [41]. Third, cluster analysis allowed us to construct a variable on the degree of urbanicity of the residential and workplace neighborhoods to examine the combined exposure to multiple components of density that are highly correlated [36,42].

Regarding study limitations, first, the cross-sectional design of the study makes it impossible to determine the direction of the causal effects involved in the associations. Second, our sample recruited in health care centers is not representative of the Paris metropolitan area [19]. However, a large panel of municipalities from the region was *a priori* selected to ensure the presence in the sample of people from all socioeconomic backgrounds. Moreover, the present analysis controlled for the individual and neighborhood factors that were found to influence participation in the study [43]. Third, the present study considered only participants' current residential and workplace neighborhoods and disregarded their past residences and workplaces. Fourth, only the participants who were working and who could be geolocated at their workplace were included in the present study. Fifth, the workplace was retrieved through the linkage of administrative information and we could not be sure that all the participants were still employed at that workplace at the recruitment in the study. Sixth, the cluster analysis that

was used to create the urbanicity degree variable did not take into consideration the within-neighborhood correlation or spatial autocorrelation in the environmental variables considered.

Main findings

Associations between residential neighborhood variables and the FMI or %FM

Previous studies have reported that BMI or waist circumference increased with decreasing neighborhood socioeconomic level [44]. In the four models estimated (FMI and %FM for men and women), body fat was higher in low-education neighborhoods. However, in one of these models (FMI among women), the relationship between neighborhood education and FM disappeared after adjustment for the home-work distance, suggesting that the residential neighborhood education effect in this group was in fact attributable to the short distance to work for women of high education neighborhoods.

Associations between residential neighborhood education and the FMI or %FM did not follow a dose-response pattern. According to a nonlinear pattern of increase, only men and women living in low-education neighborhoods had a higher FM than those living in high-education neighborhoods. A previous study also reported an elevated body weight only in the most disadvantaged group of neighborhoods [45].

After adjustment for neighborhood education level, a lower density of population in the residential neighborhood was associated with a higher FMI and %FM but only among men. A review of literature showed that the urbanicity or density degree was the most consistent predictor of BMI in affluent countries [6]. It is not clear at this stage why density effects were better captured by population density than by the other components of density. It is important to note that, after controlling for generic environmental dimensions (socioeconomic status, density), the spatial accessibility to green spaces and fast-food restaurants was not associated with FM.

Associations between work economic sectors and the FMI and %FM

To our knowledge, although a number of studies have described differences in overweight and obesity prevalence between occupational status groups (i.e., blue-collar vs. white-collar workers or skilled workers vs. unskilled workers) [46,47], extremely few studies were able to correlate work economic sectors with BMI (and no study reported on differences in FM between economic sectors). One study of Dutch workers found that the prevalence of overweight and obesity differed between occupational groups and economic sectors, even after adjustment for age, gender, and education [41]. People working in trade, industries, or transportation as well as the legislators and senior managers had a higher BMI and a higher prevalence of overweight and obesity. Another study from the same authors suggests that differences in work-related physical activity between work economic sectors may contribute to explain disparities in BMI [48].

In our study, associations with work economic sectors were documented after adjusting for individual socioeconomic variables (including individual and parental education, as neither occupation nor income was related to body fat) and residential neighborhood variables. We found that the FMI and/or %FM were higher among men working in a number of sectors (Construction, Transport, and communications, Manufacturing industry, Commercial/repair of motor vehicle and motorcycle, and Real estate, renting, and business services) than among men working in the Education sector. Differently, among women, those working in the Health and social work sector and possibly in the Public administration sector had a higher FMI than those working in the Transportation and

communications sector. It should be noted, however, that no association was documented between work economic sectors and the %FM among women, which might be attributable, among other explanations, to the fact that the FMI assesses obesity relatively accurately, but that the %FM, as the BMI but to a lower extent, underestimates obesity [15]. Especially among men, it seems rather unlikely that physical activity and energy expenditure at work intervene as the major contributors to the observed differences between economic sectors. An additional justification for rejecting this hypothesis is that the differences between economic sectors were not found to be different among low and high occupational/educational groups.

Of particular interest, there were associations persisting between work economic sectors and the FMI or %FM among men after adjustment for BMI (our aim with such adjustment was to assess whether FM was differentially associated with social/environmental variables than BMI). Although the public/private sector variable was not associated with body fat, these analyses show that people working in the Education and Public administration sectors, that is, two sectors typically managed by the Public State, had a lower FM than men working in all other (typically private) sectors. Our findings suggest that body composition assessed with the FMI and %FM allowed us to capture information on body fat that could not be provided by BMI and that BMI does not allow to fully assess social/contextual effects on body fat. Those results are in line with Rosemond et al. [49] who found that obesity (BMI) and central obesity (waist-to-hip ratio) were differently associated with occupational and social factors.

An absence of association between workplace neighborhood variables and fat mass

To our knowledge, only two studies analyzed both the residential and the workplace neighborhoods in relation to obesity [7,8]. One of these studies examined whether living or working near fast-food restaurants was associated with body weight [7]. No association was documented between the proximity of fast-food restaurants to home or work and BMI. Our study did not find any association between the proportion of fast-food restaurants among all restaurants available and FM (our initial hypothesis was that the density of fast-food restaurants may be more influential when assessed around the workplace than around the residence). The other workplace neighborhood variables, including those related to the socioeconomic status or the urbanicity degree, were not associated as well with body fat. A previous study documented associations in the expected direction between the workplace geographic environment and BMI but the associations often had large confidence intervals [8]. Additional work will be needed to examine whether only focusing on residential environments implies or not a major misclassification of environmental exposures related to the obesity risk.

Regarding the home-work distance, previous studies have reported that there is an inverse association between walking to work or cycling to work and overweight or obesity and that people working at home have a higher prevalence of overweight or obesity [50]. In our study, the mean street network distance to work was 10.9 km among men and 9.3 km among women. A long home-work distance was found to be associated with a higher FM, but only among women and only for the FMI (which again may better capture true differences in FM). Even if we had no information on the transportation mode that was typically used to go to work, this finding is coherent with the use of motorized modes for long home-work distances. A potential explanation for the observed gender difference may be that men more heavily rely on motorized modes for short distance trips. Our findings may indicate that walking or cycling to work is an opportunity to

perform physical activity at a level that may be beneficial for health [51].

As a conclusion, public health interventions programs attempting to reduce the obesity prevalence and social/territorial inequalities in obesity should consider work economic sectors in addition to the residential neighborhood. Future etiologic work on the social and environmental determinants of excess weight and fat should simultaneously investigate the influence of various life contexts.

Acknowledgments

Funding Sources: This work was supported by a doctoral grant of Région Île-de-France attributed to A.L. The RECORD Study is funded by the Institute for Public Health Research (IReSP, Institut de Recherche en Santé Publique); the National Institute for Prevention and Health Education (INPES, Institut National de Prévention et d'Éducation pour la Santé) (Prevention Program 2007; 2010–2011 financial support; 2011–2013 financial support; 2012–2014 financial support); the National Institute of Public Health Surveillance (InVS, Institut de Veille Sanitaire) (Territory and Health Program); the French Ministries of Research and Health (Epidemiologic Cohorts Grant 2008); the National Health Insurance Office for Salaried Workers (CNAM-TS, Caisse Nationale d'Assurance Maladies des Travailleurs Saliés); the Ile-de-France Regional Health Agency (ARS, Agence Régionale de Santé); the Ile-de-France Regional Council (Conseil Régional d'Île-de-France, DIM SenT and CODDIM); the National Research Agency (ANR, Agence Nationale de la Recherche) (Health–Environment Program 2005); the City of Paris (Ville de Paris); and the Ile-de-France Youth, Sports, and Social Cohesion Regional Direction (DRJSCS, Direction Régionale de la Jeunesse, des Sports et de la Cohésion Sociale).

The authors are grateful to the French National Institute for Prevention and Health Education (and Pierre Arwidson) for its continued support since the beginning of the study. We also thank Méliani Alberto from the ARS of Ile-de-France. The authors are grateful to INSEE, the French National Institute of Statistics and Economic Studies, which provided support for the geocoding of the RECORD participants and allowed us to access to relevant geographical data (with special thanks to Pascale Breuil). The authors thank Geoconcept for allowing us to access to the Universal Geocoder software. The authors also thank CNAM-TS and the Caisse Primaire d'Assurance Maladie de Paris (CPAM-P, France) for helping make this study possible. The authors thank Isabelle Bridenne and Vincent Poubelle who allowed us to access to the data from the National Old Age Insurance System, (Caisse Nationale d'Assurance Vieillesse, CNAV). Regarding the data used the present study, the authors also thank the National Geographic Institute (IGN) and the Ile-de-France Institute of Land Use Planning and Urbanism.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.annepidem.2013.12.001>.

References

- [1] 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(6):531–8.
- [2] Coogan PF, Cozier YC, Krishnan S, Wise LA, Adams-Campbell LL, Rosenberg L, et al. Neighborhood socioeconomic status in relation to 10-year weight gain in the Black Women's Health Study. *Obesity (Silver Spring)* 2010;18(10):2064–5.
- [3] Grow HM, Cook AJ, Arterburn DE, Saelens BE, Drewnowski A, Lozano P. Child obesity associated with social disadvantage of children's neighborhoods. *Soc Sci Med* 2010;71(3):584–91.

- [4] Oliver LN, Hayes MV. Effects of neighbourhood income on reported body mass index: an eight year longitudinal study of Canadian children. *BMC Public Health* 2008;8:16.
- [5] Stafford M, Brunner EJ, Head J, Ross NA. Deprivation and the development of obesity: a multilevel, longitudinal study in England. *Am J Prev Med* 2010;39(2):130–9.
- [6] Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obes Rev* 2011;12(3):217–30.
- [7] Jeffery RW, Baxter J, McGuire M, Linde J. Are fast food restaurants an environmental risk factor for obesity? *Int J Behav Nutr Phys Act* 2006;3:2.
- [8] Moore K, Diez Roux AV, Auchincloss A, Evenson KR, Kaufman J, Mujahid M, Williams K. Home and work neighbourhood environments in relation to body mass index: the Multi-Ethnic Study of Atherosclerosis (MESA). *J Epidemiol Community Health* 2013;67(10):846–53.
- [9] Brownson RC, Haire-Joshu D, Luke DA. Shaping the context of health: a review of environmental and policy approaches in the prevention of chronic diseases. *Annu Rev Public Health* 2006;27:341–70.
- [10] Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. Indices of relative weight and obesity. *J Chronic Dis* 1972;25(6):329–43.
- [11] Bouchard C. BMI, fat mass, abdominal adiposity and visceral fat: where is the 'beef'? *Int J Obes (Lond)* 2007;31(10):1552–3.
- [12] Garrido-Chamorro RP, Sirvent-Belando JE, Gonzalez-Lorenzo M, Martin-Carratala ML, Roche E. Correlation between body mass index and body composition in elite athletes. *J Sports Med Phys Fitness* 2009;49(3):278–84.
- [13] Camhi SM, Bray GA, Bouchard C, Greenway FL, Johnson WD, Newton RL, et al. The relationship of waist circumference and BMI to visceral, subcutaneous, and total body fat: sex and race differences. *Obesity (Silver Spring)* 2011;19(2):402–8.
- [14] Jackson AS, Stanforth PR, Gagnon J, Rankinen T, Leon AS, Rao DC, et al. The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study. *Int J Obes Relat Metab Disord* 2002;26(6):789–96.
- [15] Peltz G, Aguirre MT, Sanderson M, Fadden MK. The role of fat mass index in determining obesity. *Am J Hum Biol* 2010;22(5):639–47.
- [16] Carrasco F, Reyes E, Rimler O, Rios F. [Predictive accuracy of body mass index in estimating body fatness measured by bioelectrical impedance]. *Arch Latinoam Nutr* 2004;54(3):280–6.
- [17] Leal C, Bean K, Thomas F, Chaix B. Are associations between neighborhood socioeconomic characteristics and body mass index or waist circumference based on model extrapolations? *Epidemiology* 2011;22(5):694–703.
- [18] Chaix B, Bean K, Leal C, Thomas F, Havad S, Evans D, et al. Individual/neighborhood social factors and blood pressure in the RECORD Cohort Study: which risk factors explain the associations? *Hypertension* 2010;55(3):769–75.
- [19] Chaix B, Billaudeau N, Thomas F, Havad S, Evans D, Kestens Y, et al. Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology* 2011;22(1):18–26.
- [20] Chaix B, Jouven X, Thomas F, Leal C, Billaudeau N, Bean K, et al. Why socially deprived populations have a faster resting heart rate: impact of behaviour, life course anthropometry, and biology—the RECORD Cohort Study. *Soc Sci Med* 2011;73(10):1543–50.
- [21] Chaix B, Kestens Y, Bean K, Leal C, Karusisi N, Meghief K, et al. Cohort profile: residential and non-residential environments, individual activity spaces and cardiovascular risk factors and diseases—the RECORD Cohort Study. *Int J Epidemiol* 2012;41(5):1283–92.
- [22] Havad S, Reich BJ, Bean K, Chaix B. Social inequalities in residential exposure to road traffic noise: an environmental justice analysis based on the RECORD Cohort Study. *Occup Environ Med* 2011;68(5):366–74.
- [23] Leal C, Bean K, Thomas F, Chaix B. Multicollinearity in associations between multiple environmental features and body weight and abdominal fat: using matching techniques to assess whether the associations are separable. *Am J Epidemiol* 2012;175(11):1152–62.
- [24] Karusisi N, Bean K, Oppert JM, Pannier B, Chaix B. Multiple dimensions of residential environments, neighborhood experiences, and jogging behavior in the RECORD Study. *Prev Med* 2012;55(1):50–5.
- [25] INSEE. IRIS—Definition. Paris, France: National Institute of Statistics and Economic Studies; 2011.
- [26] Siordia C, Fox A. Public Use Microdata Area fragmentation: research and policy implications of polygon discontinuity. *Spatial Demography* 2013;1(1):42–56.
- [27] Thomas F, Bean K, Pannier B, Oppert JM, Guize L, Benetos A. Cardiovascular mortality in overweight subjects: the key role of associated risk factors. *Hypertension* 2005;46(4):654–9.
- [28] Schutz Y, Kyle UU, Pichard C. Fat-free mass index and fat mass index percentiles in Caucasians aged 18–98 y. *Int J Obes Relat Metab Disord* 2002;26(7):953–60.
- [29] Lazzar S, Boirie Y, Meyer M, Vermorel M. Evaluation of two foot-to-foot bioelectrical impedance analysers to assess body composition in overweight and obese adolescents. *Br J Nutr* 2003;90(5):987–92.
- [30] Beckman A, Hakansson A, Rastam L, Lithman T, Merlo J. The role country of birth plays in receiving disability pensions in relation to patterns of health care utilisation and socioeconomic differences: a multilevel analysis of Malmö, Sweden. *BMC Public Health* 2006;6:71.
- [31] United-Nations-Development-Program. Human Development Report 2007/2008. New York, NY: United Nations; 2008.
- [32] Siordia C, Saenz J. What is a "Neighborhood"? Definition in studies about depressive symptoms in older persons. *The Journal of Frailty and Aging* 2013.
- [33] Karusisi N, Thomas F, Meline J, Chaix B. Spatial accessibility to specific sport facilities and corresponding sport practice: the RECORD Study. *Int J Behav Nutr Phys Act* 2013;10:48.
- [34] Kumar V. Cluster analysis: basic concepts and algorithms. In: Tan P-N, Steinbach M, Kumar V, editors. *Introduction to Data Mining*. Boston: Addison Wesley; 2005. pp. 487–568.
- [35] Li YS, Chuang YC. Neighborhood effects on an individual's health using neighborhood measurements developed by factor analysis and cluster analysis. *J Urban Health* 2009;86(1):5–18.
- [36] Van Hulst A, Thomas F, Barnett TA, Kestens Y, Gauvin L, Pannier B, et al. A typology of neighborhoods and blood pressure in the RECORD Cohort Study. *J Hypertens* 2012;30(7):1336–46.
- [37] Siordia C. Sociospatial inequality: combining multilevel and spatial analysis. *Journal of Maps* 2013;9(1):50–4.
- [38] 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.
- [39] Chaix B, Kestens Y, Perchoux C, Karusisi N, Merlo J, Labadi K. An interactive mapping tool to assess individual mobility patterns in neighborhood studies. *Am J Prev Med* 2012;43(4):440–50.
- [40] Chaix B, Meline J, Duncan S, Merrien C, Karusisi N, Perchoux C, et al. GPS tracking in neighborhood and health studies: a step forward for environmental exposure assessment, a step backward for causal inference? *Health Place* 2013;21C:46–51.
- [41] Proper KI, Hildebrandt VH. Overweight and obesity among Dutch workers: differences between occupational groups and sectors. *Int Arch Occup Environ Health* 2010;83(1):61–8.
- [42] Gershoff ET, Pedersen S, Lawrence Aber J. Creating neighborhood typologies of GIS-based data in the absence of neighborhood-based sampling: a factor and cluster analytic strategy. *J Prev Interv Community* 2009;37(1):35–47.
- [43] Chaix B, Evans D, Merlo J, Suzuki E. Commentary: weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. *Epidemiology* 2012;23(1):129–31. discussion 32–7.
- [44] Kershaw KN, Albrecht SS, Carnethon MR. Racial and ethnic residential segregation, the neighborhood socioeconomic environment, and obesity among Blacks and Mexican Americans. *Am J Epidemiol* 2013.
- [45] Coogan PF, Cozier YC, Krishnan S, Wise LA, Adams-Campbell LL, Rosenberg L, Palmer JR. Neighborhood socioeconomic status in relation to 10-year weight gain in the Black Women's Health Study. *Obesity (Silver Spring)* 2010;18(10):2064–5.
- [46] Nakamura K, Shimai S, Kikuchi S, Tominaga K, Takahashi H, Tanaka M, et al. Shift work and risk factors for coronary heart disease in Japanese blue-collar workers: serum lipids and anthropometric characteristics. *Occup Med (Lond)* 1997;47(3):142–6.
- [47] Nakamura S, Nakamura K, Tanaka M. Increased risk of coronary heart disease in Japanese blue-collar workers. *Occup Med (Lond)* 2000;50(1):11–7.
- [48] Proper KI, Hildebrandt VH. Physical activity among Dutch workers—differences between occupations. *Prev Med* 2006;43(1):42–5.
- [49] Rosmond R, Lapidus L, Bjorntorp P. The influence of occupational and social factors on obesity and body fat distribution in middle-aged men. *Int J Obes Relat Metab Disord* 1996;20(7):599–607.
- [50] Wen LM, Rissel C. Inverse associations between cycling to work, public transport, and overweight and obesity: findings from a population based study in Australia. *Prev Med* 2008;46(1):29–32.
- [51] Oja P, Vuori I, Paronen O. Daily walking and cycling to work: their utility as health-enhancing physical activity. *Patient Educ Couns* 1998;33(1 Suppl):S87–94.

Appendix 1. *Working participants with information on their workplace*

Administrative files from CNAV (National Old Age Insurance System) and Insee (French National Institute of Statistics and Economic Studies) were used to assess and geocode participants' workplaces. First, we identified the establishment of work based on the CNAV database of occupational careers. The file received from CNAV indicated the employer (or employers, with a maximum of 3) of each participant for each year, with the corresponding establishment identification codes. The file did not provide information on the dates of beginning and end of the contracts during the year. The data therefore did not allow us to confirm for sure that the participant was employed, nor with which employer he/she was employed (if several employer were reported), at the exact date of enrollment in the study. We retained for every year only the main employer, which was the one from which the participant received the most important salary. To be sure to only consider workplaces where the participant was already working (or had worked) at the time of recruitment in the study (and thus avoid reverse causality problems), we assigned to each individual the main work establishment of the year preceding his/her inclusion in the study.

We then used databases of facilities or companies from Insee (Permanent Database of Facilities, SIRENE register) or Trade Dimension to geocode the workplaces. These databases allowed us to retrieve the spatial coordinates of the workplace of the participants. For the 4536 participants for whom a work establishment was identified, we were able to retrieve in these databases the coordinates of the workplace of 3837 participants (geocoding at the workplace address). For 254 other participants, the workplace was geocoded using Google Maps based on addresses found in company directories available on the internet. The workplace of respectively

238 and 123 participants was geocoded at the centroid of the corresponding census block group neighborhood or at the centroid of the corresponding municipality. Eighty-four participants for whom the workplace could not be identified or located and 124 participants for whom the workplace was located outside the Paris Ile-de-France region were excluded from the analyses.

Overall, a workplace was geocoded for 4331 participants residing and working in the Ile-de-France region. After excluding participants with missing information on fat mass, 4078 participants were included in the analyses. They were living in 645 census tract neighborhoods (defined by Insee by grouping three census block group neighborhoods) (183). These census tracts comprised an average of 6.3 participants. The study protocol was approved by the French Data Protection Authority.

Appendix 2. Description of the residential neighborhood urbanicity degree variable determined from cluster analysis

	Low density (n=691)		Middle-low density (n=853)		Middle-high density (n=1444)		High density (n=1085)	
	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
Density of population per km ²	5101	1	10376	2	13707	3	32917	4
Density of destinations	105.6	1	258.2	2	442.1	3	1882.5	4
Density of 3-way street intersections per km ²	113.3	1	178.0	3	159.2	2	27.0	4
Ratio of areas of street network to circular buffers	0.47	1	0.48	2	0.59	3	0.04	4

Appendix 3. Description of the workplace neighborhood urbanicity degree variable determined from cluster analysis

	Low density (n=723)		Middle-low density (n=1671)		Middle-high density (n=1019)		High density (n=665)	
	Mean	Rank	Mean	Rank	Mean	Rank	Mean	Rank
Density of population per km ²	3074	1	12850	2	29147	4	15198	3
Density of destinations	56.9	1	411.5	2	1871.7	3	3660.6	4
Density of 3-way street intersections per km ²	95.1	1	179.3	2	181.9	3	202.8	4
Ratio of areas of street network to circular buffers	0.38	1	0.52	2	0.63	3	0.67	4

Appendix 4. Average FMI and %FM among men and women according to selected individual and contextual characteristics in selected participants from the RECORD Cohort Study

	Men			Women		
	n	FMI ¹	%FM ²	n	FMI ¹	%FM ²
Age						
30–44	126	5.24 (2.43)	20.03 (6.30)	514	7.03 (3.40)	28.48 (8.25)
45–59	141	5.95 (2.53)	22.33 (6.05)	525	8.13 (3.74)	31.42 (8.16)
60–79	240	6.32 (2.16)	23.57 (5.79)	115	8.61 (3.61)	32.84 (8.16)
Individual education						
High	133	5.31 (2.09)	20.49 (5.62)	390	6.72 (2.94)	28.16 (7.63)
Mid-high	782	5.84 (2.51)	21.86 (5.62)	404	7.38 (3.30)	29.69 (8.16)
Mid-low	591	6.12 (2.71)	22.64 (7.10)	253	8.74 (4.12)	32.50 (8.77)
Low	191	6.04 (2.64)	22.34 (6.61)	96	9.81 (4.30)	34.53 (8.25)
Residential education level						
High	839	5.38 (2.20)	20.74 (5.84)	318	7.14 (3.56)	28.98 (8.41)
Mid-high	786	5.46 (2.17)	20.93 (5.87)	302	7.04 (2.97)	29.02 (7.99)
Mid-low	637	5.61 (2.39)	21.24 (6.38)	278	7.80 (3.70)	30.43 (8.22)
Low	661	6.35 (2.77)	23.09 (6.83)	275	8.99 (3.98)	33.09 (8.34)
Residential density of population						
High	757	5.17 (2.13)	20.10 (5.88)	315	7.09 (3.51)	28.80 (8.47)
Mid-high	795	5.68 (2.40)	21.46 (6.23)	316	7.59 (3.81)	29.74 (8.96)
Mid-low	729	5.88 (2.56)	21.99 (6.57)	270	8.03 (3.54)	31.23 (8.00)
Low	642	6.02 (2.43)	22.35 (6.14)	254	8.18 (3.53)	31.66 (7.61)
Work economic sector						
Manufacturing industry	364	5.59 (2.22)	21.38 (5.84)	100	7.19 (3.34)	29.15 (7.81)
Construction	154	6.63 (2.82)	23.78 (6.59)	7	6.91 (3.11)	29.08 (8.61)
Commercial, repair of motor vehicles and motorcycles	335	5.80 (2.43)	21.87 (6.45)	139	7.58 (3.41)	30.24 (8.57)
Hotels and restaurants	165	5.56 (2.24)	21.16 (6.16)	52	9.12 (4.12)	33.25 (9.20)
Transport and communications	187	6.04 (6.04)	22.17 (7.24)	59	6.69 (2.79)	28.23 (7.77)
Financial activities	293	5.71 (2.49)	21.18 (5.88)	106	7.61 (3.31)	30.34 (7.57)
Real estate, renting, and business services	873	5.71 (2.49)	21.46 (6.35)	333	7.25 (3.48)	29.18 (8.39)
Public administration	49	5.32 (2.19)	20.26 (5.97)	48	8.81 (3.76)	32.68 (6.86)
Education	70	5.07 (1.83)	19.75 (5.42)	52	8.27 (4.07)	31.60 (8.48)
Health and social work	63	5.38 (2.56)	20.50 (6.54)	91	9.41 (4.41)	33.79 (8.70)

Collective, social, and personal services	320	5.59 (2.22)	20.36 (5.56)	137	7.32 (3.48)	29.45 (8.52)
Legal status of the company						
Private	2664	5.71 (2.42)	21.54 (6.28)	914	7.46 (3.49)	29.74 (8.34)
Public	259	5.28 (2.25)	20.30 (6.06)	241	8.56 (3.98)	32.18 (8.29)

¹ Mean (SD) in kg/m².

² Mean (SD) in %.

Appendix 5. Beta coefficients and 95% confidence intervals for the associations with the FMI after additional adjustment for BMI from multiple multilevel regression models, RECORD Cohort Study¹

Variables	Men β (95% IC) ²	Women β (95% IC) ²
Age (vs. 30–44)		
45–59	0.24 (0.17, 0.31)	0.25 (0.12, 0.38)
60–79	0.40 (0.27, 0.53)	0.42 (0.20, 0.63)
Individual education (vs. high education)		
Medium-high education	0.13 (0.05, 0.22)	0.01 (-0.14, 0.16)
Medium-low education	0.22 (0.12, 0.32)	0.02 (-0.17, 0.20)
Low education	0.10 (-0.05, 0.25)	-0.01 (-0.27, 0.25)
Parental education (vs. high education)		
Intermediate education	0.06 (-0.02, 0.14)	0.04 (-0.11, 0.19)
Low education	0.09 (-0.00, 0.19)	-0.09 (-0.27, 0.08)
Perceived financial strain	0.06 (-0.04, 0.16)	0.04 (-0.11, 0.19)
Living alone (vs. as a couple)	0.01 (-0.07, 0.09)	–
Human Development Index of country of birth (vs. high development)		
Middle	–	-0.07 (-0.24, 0.10)
Low	–	0.41 (0.15, 0.68)
Residential education level (vs. high)		
Mid-high	-0.06 (-0.15, 0.02)	–
Mid-low	-0.08 (-0.18, 0.01)	–
Low	0.00 (-0.11, 0.11)	–
Residential density of population (vs. high)		
Mid-high	0.03 (-0.05, 0.13)	–
Mid-low	0.08 (0.00, 0.18)	–
Low	0.04 (-0.06, 0.14)	–
Work economic sector		
Manufacturing industry	0.38 (0.14, 0.61)	-0.15 (-0.48, 0.18)
Construction	0.40 (0.14, 0.67)	0.17 (-0.62, 0.95)
Commercial, repair of motor vehicles and motorcycles	0.37 (0.13, 0.61)	0.03 (-0.28, 0.34)
Hotels and restaurants	0.28 (0.02, 0.53)	-0.06 (-0.45, 0.33)
Transport and communications	0.37 (0.12, 0.62)	REF
Financial activities	0.32 (0.08, 0.55)	-0.07 (-0.39, 0.26)
Real estate, renting, and business services	0.38 (0.15, 0.60)	-0.12 (-0.40, 0.16)
Public administration	0.02 (-0.31, 0.36)	-0.32 (-0.71, 0.07)

Education	REF	-0.08 (-0.47, 0.31)
Health and social work	0.32 (0.01, 0.63)	-0.09 (-0.44, 0.25)
Collective, social, and personal services	0.37 (0.13, 0.61)	-0.02 (-0.33, 0.29)
Home – work distance (vs. low)		
Medium-low		-0.03 (-0.13; 0.20)
Medium-high		0.09 (-0.08; 0.26)
High		0.11 (-0.13; 0.20)
BMI	0.61 (0.60, 0.62)	0.71 (0.70, 0.73)

¹Models were adjusted for all the individual and contextual variables that were independently associated with the outcome.

² β : beta coefficients; 95% CI: 95% confidence interval.

Appendix 6. Beta coefficients and 95% confidence intervals for the associations with the %FM after additional adjustment for BMI from multiple multilevel models, RECORD Cohort Study¹

Variables	Men β (95% IC) ²	Women β (95% IC) ²
Age (vs. 30–44)		
45–59	1.15 (0.90, 1.41)	1.27 (0.71, 1.83)
60–79	1.91 (1.45, 2.37)	2.05 (1.10, 2.99)
Individual education (vs. high education)		
Medium-high education	0.43 (0.12, 0.73)	0.14 (-0.50, 0.79)
Medium-low education	0.68 (0.32, 1.05)	0.02 (-0.76, 0.81)
Low education	0.26 (-0.28, 0.80)	-0.07 (-1.14, 1.00)
Living alone (vs. as a couple)	-0.11 (-0.40, 0.17)	–
Parental education (vs. high education)		
Intermediate education	0.18 (-0.11, 0.48)	–
Low education	0.36 (0.00, 0.71)	–
Perceived financial strain	–	0.11 (-0.55, 0.78)
Human Development Index of country of birth (vs. high development)		
Middle	–	-0.13 (-0.89, 0.63)
Low	–	1.63 (0.48, 2.79)
Residential education level (vs. high)		
Mid-high	-0.18 (-0.51, 0.15)	0.15 (-0.56, 0.86)
Mid-low	-0.29 (0.65, 0.07)	-0.17 (-0.92, 0.57)
Low	-0.05 (-0.47, 0.36)	0.27 (-0.53, 1.07)
Residential density of population (vs. high)		
Mid-high	0.18 (-0.15, 0.52)	–
Mid-low	0.37 (-0.00, 0.75)	–
Low	0.25 (-0.13, 0.63)	–
Work economic sector		–
Manufacturing industry	1.29 (0.45, 2.14)	–
Construction	1.29 (0.34, 2.23)	–
Commercial, repair of motor vehicles and motorcycles	1.33 (0.48, 2.18)	–
Hotels and restaurants	0.99 (0.06, 1.92)	–
Transport and communications	1.10 (0.19, 2.00)	–
Financial activities	1.04 (0.18, 1.90)	–
Real estate, renting, and business services	1.19 (0.39, 2.00)	–
Public administration	0.01 (-1.19, 1.21)	–
Education	REF	–
Health and social work	0.82 (-0.30, 1.94)	–
Collective, social, and personal services	1.26 (0.40, 2.11)	–

BMI

1.44 (1.41, 1.47)

1.42 (1.37, 1.48)

¹Models were adjusted for all the individual and contextual variables that were independently associated with the outcome.

² β , beta coefficient; 95% CI, 95% confidence interval.

Appendix 7 Association between distance from home and work and FMI and %FM for men and women

Distance (vs high)	Men ¹		Women ¹	
	FMI β (95% IC) ²	%FM β (95% IC) ²	FMI β (95% IC) ²	%FM β (95% IC) ²
Medium-high	-0.29 (-0.53, -0.05)	-0.85 (-1.46 -0.24)	-0.56 (-1.16, 0.03)	-1.34 (-2.73, 0.05)
Medium-low	-0.45 (-0.69, -0.21)	-1.10 (-1.72, -0.48)	-0.67 (-1.25, -0.09)	-1.75 (-3.12, -0.38)
low	-0.32 (-0.57, -0.07)	-0.73 (-1.37, -0.09)	-0.85 (-1.43, -0.28)	-1.96 (-3.30, -0.62)

¹Models were adjust for all individual variables

² β , beta coefficient; 95% CI, 95% confidence interval

CHAPITRE IV - Relations entre les secteurs d'activité économique et les facteurs de risque cardiovasculaire

IV.1. Présentation de l'article

Nous avons, dans ce chapitre, réalisé une étude afin de mieux comprendre les déterminants des facteurs de risque cardiovasculaire. Cet article fait suite à l'article précédent qui se focalisait exclusivement sur la mesure des liens entre l'obésité et les différents secteurs d'activité économique des participants. En ce qui concerne les disparités de santé en milieu professionnel, très peu d'études se sont intéressées aux secteurs d'activité économique. Pourtant, étudier les secteurs d'activité économique, même à une échelle large, nous apparaît d'une grande utilité, car il est prouvé que la prévalence d'expositions (facteurs psychosociaux, pénibilité du travail, bruit, rythme, demande physique, consommation de tabac et d'alcool, etc.) varie d'un secteur à l'autre. Ce travail, présenté dans l'article qui suit, est fondé sur l'idée, qu'en plus d'une littérature florissante sur les relations entre les expositions professionnelles spécifiques et la santé, il est pertinent d'examiner les disparités de santé par secteurs d'activités économique..

En épidémiologie sociale, il est fréquent d'examiner des causes qui sont distantes des phénomènes de santé étudiés, avec une distinction entre les déterminants dits en amont (par exemple, la classe sociale, les secteurs d'activité économique, etc.) et les déterminants de médiation, en aval, qui sont plus proches des phénomènes de santé. Considérer les secteurs d'activité économique, en plus des catégories socio-professionnelles ainsi que d'autres grandes

caractéristiques populationnelles, peut être utile pour identifier des populations en mauvaise santé sur lesquelles cibler les interventions et les efforts de prévention. De plus, l'intérêt opérationnel de l'utilisation des secteurs d'activité économique pour cibler les interventions est également lié au fait que cette information est souvent rapportée et disponible dans de nombreuses bases de données. Enfin, les secteurs d'activité économique sont également liés à des capacités d'interventions concrètes, à travers les services de la médecine du travail, représentée dans chaque entreprise. Dans l'ensemble, la documentation des relations entre les secteurs d'activité économique et les facteurs de risque cardiovasculaire apparaît comme potentiellement importante dans le but de cibler des interventions de Santé publique. L'objectif de cette étude a donc été d'estimer et de comparer, chez les hommes et chez les femmes, au-delà des caractéristiques sociodémographiques individuelles et des caractéristiques de l'environnement résidentiel, les associations entre les secteurs d'activité économique et un large panel de facteurs de risque cardiovasculaire.

Pour répondre à cet objectif, les relations entre 11 secteurs d'activité économique et 10 facteurs de risque cardiovasculaire ont été estimées. Afin d'établir un profil des facteurs de risque cardiovasculaire, 10 variables réponses ont été examinées dans cette étude soit : des indicateurs anthropométriques (indice de masse corporelle [IMC] et tour de taille), des indicateurs de pression sanguine (pression artérielle systolique [PAS] et diastolique [PAD] et pression pulsée), des indicateurs lipidiques (taux de cholestérol total, taux de lipoprotéine de haute densité [HDL] et taux de lipoprotéine de faible densité [LDL]), le taux de glycémie et la fréquence cardiaque au repos. A partir de la base de données de la CNAV et de la base SIRENE de l'INSEE nous avons été en mesure d'identifier les secteurs d'activité économique de chaque participant. 11 des 17

niveaux de la Nomenclature des Activités Française (NAF) de 2003 étaient représentés dans notre échantillon : Santé et travail social ; Industrie manufacturière ; Construction ; Commerce, réparations automobile ; Hôtels et restaurants ; Transports et communications ; Activités financières ; Immobilier, location et services aux entreprises ; Administration publique ; Education ; et Services collectifs, sociaux et personnels.

Notre stratégie analytique a été d'estimer des modèles de régression linéaire multiniveau avec un effet aléatoire au niveau des zones infra-communales TRIRIS pour tenir compte de la corrélation intra-quartier dans les facteurs de risque cardiovasculaire examinés. Compte tenu des différences dans la profession et dans les profils cardiovasculaires entre les femmes et les hommes, toutes nos analyses ont été stratifiées par sexe. Afin d'estimer et de pouvoir comparer les relations entre les secteurs d'activité économique et les facteurs de risque cardiovasculaire, toutes nos analyses ont été ajustées sur les mêmes variables sociodémographiques individuelles (âge, niveau d'éducation individuel, profession, revenu du foyer, statut matrimonial, stress financier) ainsi que sur le niveau d'éducation du quartier. Les modèles pour les indicateurs de pression sanguine ont également été ajustés sur l'utilisation de médicaments antihypertenseurs.

Dans le cadre de ces analyses, les résultats obtenus ont montré qu'outre certaines caractéristiques sociodémographiques individuelles et résidentielles, des associations étaient présentes entre les secteurs d'activité économique et certains facteurs de risque cardiovasculaire (mais pas tous).

En ce qui concerne les indicateurs anthropométriques, les hommes travaillant dans le secteur de la construction et dans le secteur des transports et des communications avaient un IMC

plus élevé et avaient tendance à avoir un plus grand tour de taille. A l'inverse, les hommes travaillant dans le secteur de la santé et du travail social (groupe de référence) et dans le secteur des services collectifs, sociaux et personnels, ont un IMC et un tour de taille plus faibles. Pour les femmes, les associations entre les variables anthropométriques et les secteurs d'activité économique étaient étonnamment différentes. Les femmes qui travaillent dans le secteur de la Santé et du travail social avaient le plus grand (et non le plus faible) IMC et tour de taille. Contrairement aux hommes, les femmes travaillant dans le secteur du transport et de la communication avaient parmi les plus bas niveaux d'IMC et de tour de taille. Concernant les indicateurs de pression sanguine, les hommes et les femmes travaillant dans le secteur de la santé et du travail social avaient les niveaux de PAS les plus hauts. A l'inverse, les hommes et les femmes travaillant dans le secteur des services collectifs, sociaux et personnels avaient des niveaux de PAS relativement faibles. Les hommes travaillant dans le secteur des hôtels et des restaurants avaient également un taux relativement faible de PAS. Aucune association n'a été documentée entre les secteurs d'activité économique et la PAD chez les hommes. Chez les femmes, un certain nombre de secteurs d'activité économique (notamment le commerce, la réparation de véhicules ; les services collectifs, sociaux et personnels ; les transports et communications ; et l'immobilier, la location et les services aux entreprises) étaient associés à un niveau de PAD inférieur à celui des femmes travaillant dans le secteur de la santé et du travail social.

Pour les indicateurs lipidiques, les secteurs d'activité économique n'étaient pas associés au cholestérol total ou au cholestérol LDL chez les hommes et les femmes. En revanche, les hommes travaillant dans le secteur de la Santé et du travail social avaient le taux de cholestérol

HDL le plus élevé, tandis que les plus faibles taux de cholestérol HDL étaient observés chez les hommes travaillant dans le secteur des hôtels et des restaurants et dans le secteur de l'éducation. Aucune association n'a été trouvée avec le taux de glycémie chez les femmes. Chez les hommes, ceux qui travaillent dans le secteur des transports et des communications et dans le secteur des hôtels et des restaurants avaient les niveaux les plus élevés de glycémie, tandis que ceux qui travaillent dans le secteur de la santé et du travail social avaient le niveau le plus bas. En ce qui concerne la fréquence cardiaque au repos, aucune association n'a été documentée avec les secteurs d'activité économique chez les hommes. Chez les femmes, seulement celles travaillant dans le secteur des hôtels et des restaurants et dans le secteur des services collectifs, sociaux et personnels avaient une fréquence cardiaque au repos plus faible.

En conclusion, les résultats de cette étude suggèrent que les secteurs économiques de travail contribuent à façonner les paramètres métaboliques et cardiovasculaires, même après ajustement sur les caractéristiques sociodémographiques individuelles et résidentielles. Toutefois, les associations variaient selon le facteur de risque examiné ainsi qu'entre les hommes et les femmes. Dans l'ensemble, même si notre étude n'a pas tenu compte des caractéristiques spécifiques des environnements de travail, l'examen des relations entre les principaux secteurs économiques de travail et un large panel de facteurs de risque cardiovasculaire demeure utile, en plus de la prise en compte des caractéristiques individuelles et résidentielles, pour déterminer les profils des populations à cibler pour les interventions de Santé publique et les efforts de prévention.

La suite de ce chapitre présente l'intégralité de l'article publié dans la revue *BMC Public Health*.

RESEARCH ARTICLE

Open Access

Work economic sectors and cardiovascular risk factors: cross-sectional analysis based on the RECORD Study

Antoine Lewin^{1,2*}, Frédérique Thomas³, Bruno Pannier³ and Basile Chaix^{1,2}

Abstract

Background: Little is known on the comparative effect of work economic sectors on multiple cardiovascular risk factors. Such information may be useful to target Public health interventions, e.g., through the occupational medicine. We investigated whether and how a large panel of cardiovascular risk factors varied between 11 work economic sectors.

Methods: Data on 4360 participants from the French RECORD Study geolocated at their residence were analyzed. Ten outcomes were assessed: body mass index (BMI), waist circumference, systolic and diastolic blood pressure (BP), pulse pressure, total cholesterol, glycaemia, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and resting heart rate. Multilevel linear regression models stratified by sex and adjusted for individual and neighborhood sociodemographic characteristics were estimated.

Results: Among men, the Health and social work sector was found to be the most protective sector for BMI, waist circumference, and glycaemia (while the Construction sector and the Transport and communications sector tended to be unfavorable for these outcomes). The Health and social work sector was also associated with higher HDL cholesterol among men. However, men working in the Health and social work sector showed the highest systolic BP and pulse pressure. Women working in the Health and social work sector had the highest BMI, the largest waist circumference, and the most elevated systolic and diastolic BP. The Commercial and repair of vehicles sector, the Transport and communication sector, and the Collective, social, and personal services sector were associated with a more favorable profile for these risk factors among women.

Conclusion: Work economic sectors contribute to shape metabolic and cardiovascular parameters after adjustment for individual/neighborhood sociodemographic characteristics. However, patterns of associations varied strikingly according to the risk factor examined and between men and women. Such findings may be useful to target interventions for reducing cardiovascular risk, e.g., through the occupational medicine.

Keywords: Cardiovascular risk factors, Socioeconomic factors, Work economic sectors

Background

In Europe and North America, low socioeconomic status groups have a higher prevalence of chronic disease risk factors such as smoking, physical inactivity, obesity, hypertension, hypercholesterolemia, and diabetes [1-3], and a higher cardiovascular risk score [2,4].

Associations are well established between various categorizations of occupation and related social class and cardiovascular risk factors [5-9]. Based on occupational classifications that distinguish blue-collar from white-collar workers or professional, technical, and manual workers [5,9], there is substantial evidence for an inverse relationship between the occupational social level and cardiovascular disease risk factors [10] such as high blood pressure [5,9,11,12], smoking [13-16], cholesterol [17], hemostatic factors [18], and obesity [19].

* Correspondence: lewin@u707.jussieu.fr

¹Sorbonne Universités, UPMC Univ Paris 06, UMR_S 113, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris 75012, France

²Inserm, UMR_S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris 75012, France

Full list of author information is available at the end of the article



© 2014 Lewin et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Compared to occupational social class, very few studies have focused on work economic sectors. Work economic sectors, even considered on a broad scale, may be of importance, because numerous exposures may vary in prevalence from one sector to the other. Studies have suggested that occupational exposures such as psychosocial factors [20,21], work hardness, high physical demand, noise at work [21,22], working rhythm [23,24], and the prevalence of health behavior such as tobacco [25], alcohol, and drug consumption [26] show a different prevalence in the various work economic sectors [27].

The present work is grounded on the idea that, in addition to this flourishing literature investigating the relationship between specific occupational exposures and health; it is relevant to consider the broad economic sectors where people work. In Social epidemiology, it is very common to examine causes that are distant from the health outcomes investigated, with a distinction between the so-called upstream determinants (e.g., social class, work economic sectors) and downstream determinants of diseases that mediate the effects of the former. Considering work economic sectors in addition to social class and other broad characteristics of populations may be useful to identify an additional marker of populations in poor health, in order to target interventions to populations who critically need them. The operational interest of work economic sectors for targeting interventions is also related to the fact that this information is commonly reported and available in a number of databases. Finally, work economic sectors are also connected to a potential approach to develop interventions, through the occupational medicine departments represented in each company. Overall, documenting relationships between work economic sectors and cardiovascular risk factors is potentially important to develop Public health interventions.

Regarding cardiovascular health as the focus of the present article, one study investigated work economic sectors in relation to hypertension. Takashima et al. found that the age-adjusted prevalence of hypertension among transportation and communication workers was four times higher than that of service workers, who had the lowest prevalence of hypertension [11]. In another study, occupation-based social class and work economic sectors were combined into one classification, not allowing the authors to disentangle their independent effect on hypertension [7].

One study examined work economic sectors in relation to obesity in a Dutch working population [28]. The study found that there were three sectors with a relatively low body mass index (BMI): the catering industry, the healthcare sector, and the culture, sport, and recreation sector. On the opposite, workers of the transportation sector showed a high BMI and a high prevalence of overweight and obesity.

Only one study conducted among women examined work economic sectors in relation to multiple cardiovascular risk factors including cholesterol, blood pressure, heart rate, and anthropometric variables aggregated in order to reflect allostatic load [29]. The study found that working in the health care sector was associated with a higher allostatic load than working in the information technology and media sector. However, this study only compared two work economic sectors.

To our knowledge, no study investigated the association between a panel of different work economic sectors and multiple cardiovascular risks factors. Moreover, no study adjusted the relationships between work economic sectors and health for individual and neighborhood socioeconomic characteristics or compared the associations between men and women. Overall, our aim was to investigate whether and how a large panel of cardiovascular risk factors including weight status and fat, blood pressure (BP), cholesterol, glycaemia, and resting heart rate varied between 11 work economic sectors after adjustment for individual and neighborhood sociodemographic characteristics in a large working population.

Methods

Study population

The RECORD Study

Data from the first wave of the RECORD (Residential Environment and CORonary heart Disease) Cohort Study (www.record-study.org) were used for cross-sectional analyses. As described elsewhere [30-39], 7,290 participants aged 30–79 years were recruited without *a priori* sampling in 2007–2008 during free preventive medical checkups conducted by the Centre d'Investigations Préventives et Cliniques in the Paris metropolitan area. As an *a priori* eligibility criterion, only participants residing in 10 (out of 20) administrative districts of Paris or in 111 other municipalities in the region were selected for the study. Of the persons selected for participation, 83.6% agreed to participate and completed the data collection protocol.

Working participants with information on their workplace

Administrative files from CNAV (National Old Age Insurance System) were used to identify the establishment of work. We relied on a database of occupational careers routinely used for the computation of retirement pensions. The file received from CNAV indicated the employer (or employers, with a maximum of 3) of each participant for each year, with the corresponding establishment identification code and work economic sector. The file did not provide information on the dates of beginning and end of the contracts during the year. The data therefore did not allow us to confirm for sure that the participant was employed, nor with which employer he/she was employed (if several employers were reported), at the exact date of

enrollment in the study. We retained for every year only the main employer, which was the one from which the participant received the most important salary. To be sure to only consider a workplace where the participant was already working (or had worked) at the time of recruitment in the study (and thus avoid reverse causality problems), we assigned to each individual the main work establishment of the year preceding his/her inclusion in the study.

After excluding participants with missing information on the work economic sector ($n = 176$), the final sample used for the analyses comprised 4,360 participants residing in the Ile-de-France region. Descriptive information on study participants is presented in Additional file 1. They were living in 648 census tract neighborhoods (defined by Insee by grouping three census block group neighborhoods) [40]. These census tracts comprised an average of 6.8 participants, and an average of 8205 residents. The study protocol was approved by the French Data Protection Authority.

Measures

Cardiovascular risk factors

In order to assess a complete panel of the basic cardiovascular risk factors, ten outcomes were examined in this study [41]: body mass index (BMI), waist circumference, systolic and diastolic blood pressure (SBP and DBP), pulse pressure, total cholesterol, glycaemia, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and resting heart rate. Height (using a wall-mounted stadiometer) and weight (using calibrated scales) recorded by a nurse during the medical examination were used to calculate BMI (kg/m^2) [42,43]. Waist circumference was measured in cm using an inelastic tape placed midway between the lower ribs and iliac crest on the midaxillary line [44]. Supine brachial BP was measured by trained nurses three times in the right arm after a 10 minute rest period, using a manual mercury sphygmomanometer [42]. SBP and DBP were defined as the first and fifth Korotkoff phases, respectively, using the mean of the last two BP measurements [44]. Pulse pressure was defined as the difference between systolic and diastolic BP. Total cholesterol, HDL cholesterol and glycaemia (enzymatic method, automat Hitachi 917, Hitachi, Tokyo, Japan) were measured under fasting conditions. LDL cholesterol was calculated from the Friedewald equation [45]. Trained nurses measured resting heart rate in bpm by electrocardiogram, using a Cardionics CardioPlug device. The measurement was made in a quiet room after a 5- to 7-min rest period in the supine position [46].

Individual sociodemographic variables

Several sociodemographic characteristics were considered. The age of the participants was divided into 3 classes:

30–44, 45–59, and 60–79 years. Personal education was divided into four classes: no education, primary education and lower secondary education, higher secondary education and lower tertiary education, and upper tertiary education. Regarding occupation, four categories were distinguished: blue-collar workers, low-white collar workers, intermediate occupations, and high white-collar workers. Household income adjusted for household size was divided into four categories based on the quartiles. Marital status was coded in two classes (living alone or in couple). Self-reported financial strain (reporting financial difficulties) was coded as a binary variable.

Antihypertensive medication use

Antihypertensive medication use was determined by merging the administrative SNIIR-AM national health insurance database on all healthcare reimbursements in 2006–2009 to the RECORD Study database at the individual level. A binary variable was created indicating whether or not individuals had been reimbursed for any antihypertensive medication over the previous year.

Neighborhood socioeconomic variable

We hypothesized that neighborhood socioeconomic status may confound the relationship between work economic sectors and cardiovascular risk factors (if the place where people live influence individuals' job opportunities). An alternative hypothesis may be that neighborhood socioeconomic status intervenes as a mediator in the relationship between work economic sectors and cardiovascular risk (if work economic sectors, through their influence on the salary, determine where people can afford to live). Under the latter hypothesis, neighborhood socioeconomic status may be seen as an "indirect biasing pathway" and would also have to be adjusted for, because the effect of work economic sectors we are interested to isolate operates independently of the socioeconomic status of the neighborhood [47,48].

The socioeconomic status of the neighborhood was assessed with the educational level of the residents. The variable was computed within buffers with a radius of 1000 m centered on participants' residences. These buffers took into account the street network, i.e., comprise the area that is accessible within 1000 m along the street network. ArcInfo10 and its Network analyst applied to street network data from the National Geographic Institute were used to derive such buffers.

Work economic sectors and legal category of the establishment

Based on the business identification code of each workplace retrieved from the CNAV data, we then used databases of facilities or companies from Insee (Permanent Database of Facilities, SIRENE register) or Trade Dimension to identify

the work economic sector of each participant. We used the 17 levels of the 2003 French classification of activities (NAF) to classify work economic sectors. Of those 17 levels of activities, only 11 were represented in our sample: Health and social work; Manufacturing industry; Construction; Commercial, repair of motor vehicles and motorcycles; Hotels and restaurants; Transport and communications; Financial activity; Real estate, renting, and business services; Public administration; Education; and Collective, social and personal services. The Health and social work sector was taken as the reference category because preliminary descriptive analyses showed that this sector showed both the lowest level of risk for certain cardiovascular risk factors and the highest level of risk for other risk factors, thus that it was relevant to report the findings as contrasts with this particular sector.

It is important to note, first, that the work economic sector was assigned at the level of the establishment or worksite (the establishment is an economic entity located in a definite location and subordinated to a unique authority) rather than at the broader level of the company. Second, worksites were classified according to the main economic activity of the establishment. Accordingly, cleaning services or administrative services integrated in a company from the Construction sector would be classified as part of the Construction sector. Third, the work economic sector is independent of the personal social group of the person, and each work economic sector comprises people of different classes (e.g., managers and blue-collar workers, low and high white collar workers).

Additional analyses were conducted with the work economic sectors combined into a smaller number of groups: the primary sector (extraction and production of raw materials), the secondary sector (transformation of raw or intermediate materials into goods) and the tertiary sector (supply of services to consumers and businesses). The primary sector was excluded from the analyses because of a too small sample size ($n = 14$ for men and $n = 14$ for women).

The Insee databases also enabled us to distinguish between the private and the public sector.

Statistical analysis

Multilevel linear regression models with a random effect at the census tract level were estimated to account for within-neighborhood correlation in the cardiovascular risk factors examined. Given differences in occupation and cardiovascular profiles between females and males, all the analyses were stratified by sex. Models of the relationships between work economic sectors and cardiovascular risk factors were adjusted for individual sociodemographic variables and neighborhood education level. Models for blood pressure were further adjusted for antihypertensive medication use. All the analyses were conducted with SAS 9.3 (SAS Institute, Cary, North Carolina).

Results

The associations between individual or neighborhood socio-demographic variables and cardiovascular risk factors are shown in Additional files 2, 3, 4 and 5. The most consistent associations were documented with age, individual education, and neighborhood education. All risk factor variables increased with age, and HDL cholesterol also increased with age. In men and/or women, a low personal education was associated with a higher BMI, waist circumference, SBP, DPB, and pulse pressure, and with a lower HDL cholesterol. After mutual adjustment, in men and/or women, living in a low education neighborhood was related to a higher BMI, waist circumference, SBP, DBP, and resting heart rate, and to a lower HDL cholesterol. Additionally, men living alone had a lower BMI, waist circumference, total cholesterol, and glycaemia than men living in couple.

Anthropometric risk factors

Men working in the Construction sector and in the Transport and communications sector had a higher BMI and tended to have a larger waist circumference (Table 1). On the opposite, men working in the Health and social work sector (the reference group) and in the Collective, social, and personal services sector had the lowest BMI and waist circumference.

For women, the associations between work economic sectors and anthropometric variables were strikingly different. Women working in the Health and social work sector had the highest, not the lowest, BMI and waist circumference. Contrary to men, women working in the Transport and communications sector had among the lowest BMI and smallest waist circumference. Other work economic sectors associated with a lower BMI and waist circumference among women included the following sectors: Collective, social and personal services; Commercial and repair of vehicles; and Real estate, renting and business service.

Blood pressure

Both men and women working in the Health and social work sector had the highest SBP (Table 2). On the opposite, both men and women working in the Collective, social and personal service sector had a relatively low SBP. Men working in the Hotels and restaurants sector also had a relatively low SBP.

No associations were documented between work economic sectors and DBP among men. Among women, a number of work economic sectors (especially Commercial and repair of vehicle; Collective, social, and personal services; Transport and communications; and Real estate, renting and business services) were associated with a lower DBP than in the Health and social work sector.

Table 1 Associations with BMI and waist circumference estimated from multilevel regression models among men and women*

	BMI		Waist circumference	
	Men β (95% CI)	Women β (95% CI)	Men β (95% CI)	Women β (95% CI)
Work economic sector				
Ref.: Health and social work				
Manufacturing industry	0.32 -0.65 – 1.30	-1.43 -2.82 – -0.03	-0.17 -3.01 – 2.66	-2.59 -5.87 – 0.69
Construction	1.50 0.43 – 2.57	-2.12 -5.49 – 1.24	3.62 0.52 – 6.73	-3.85 -11.76 – 4.05
Commercial, repair of motor vehicles and motorcycles	0.64 -0.34 – 1.62	-1.55 -2.85 – -0.26	1.39 -1.46 – 4.24	-3.47 -6.53 – -0.42
Hotels and restaurants	0.31 -0.73 – 1.34	-0.67 -2.26 – 0.92	0.23 -2.77 – 3.24	-0.02 -3.76 – 3.71
Transport and communications	1.15 0.13 – 2.18	-1.89 -3.46 – -0.33	2.44 -0.53 – 5.42	-4.45 -8.14 – -0.77
Financial activities	0.50 -0.49 – 1.50	-0.98 -2.37 – 0.41	0.20 -2.70 – 3.10	-3.08 -6.36 – 0.19
Real estate, renting and business services	0.58 -0.35 – 1.51	-1.17 -2.32 – -0.01	1.05 -1.65 – 3.75	-2.36 -5.10 – 0.36
Public administration	0.61 -0.58 – 1.81	-0.12 -1.55 – 1.31	1.45 -2.02 – 4.93	-0.96 -4.34 – 2.42
Education	0.59 -0.53 – 1.72	-0.95 -2.42 – 0.52	0.85 -2.40 – 4.11	-1.64 -5.13 – 1.83
Collective, social, and personal services	-0.13 -1.07 – 0.79	-1.99 -3.17 – -0.82	-1.04 -3.75 – 1.67	-4.35 -7.12 – -1.58
Private (vs. public)	0.09 -0.46 – 0.66	-0.68 -1.57 – 0.22	0.88 -0.74 – 2.52	-0.94 -3.04 – 1.16

*Models were adjusted for age, education, income, perceived financial strain, marital status, occupational status, and neighborhood level of education.
 Note. BMI, body mass index; CI, confidence interval.

Table 2 Associations with SBP, DBP, and pulse pressure estimated from multilevel regression models among men and women*

	SBP		DBP		Pulse pressure	
	Men β (95% CI)	Women β (95% CI)	Men β (95% CI)	Women β (95% CI)	Men β (95% CI)	Women β (95% CI)
Work economic sector						
Ref.: Health and social work						
Manufacturing industry	-1.97 -6.02 – 2.08	-0.71 -5.89 – 4.46	1.37 -1.38 – 4.12	-3.09 -6.20 – 0.01	-3.33 -5.97 – -0.69	2.38 -1.11 – 5.88
Construction	-2.52 -6.96 – 1.91	-8.92 -21.33 – 3.48	1.39 -1.61 – 4.41	-5.20 -12.65 – 2.24	-3.91 -6.80 – -1.02	-3.71 -12.09 – 4.65
Commercial, repair of motor vehicles and motorcycles	-2.20 -6.27 – 1.87	-4.14 -8.96 – 0.67	0.87 -1.89 – 3.64	-5.16 -8.06 – -2.27	-2.98 -5.63 – -0.33	1.02 -2.23 – 4.27
Hotels and restaurants	-4.85 -9.15 – -0.56	-2.48 -8.36 – 3.40	0.29 -2.62 – 3.21	-2.08 -5.61 – 1.44	-5.15 -7.94 – -2.39	-0.39 -4.36 – 3.57
Transport and communications	-1.62 -5.87 – 2.61	-1.60 -7.42 – 4.21	1.43 -1.45 – 4.31	-4.74 -8.24 – -1.26	-3.05 -5.82 – -0.29	3.14 -0.78 – 7.07
Financial activities	-2.89 -7.03 – 1.25	-1.90 -7.05 – 3.24	0.96 -1.85 – 3.77	-2.02 -5.11 – 1.06	-3.86 -6.56 – -1.16	0.11 -3.35 – 3.58
Real estate, renting and business services	-3.24 -7.09 – 0.62	-2.01 -6.30 – 2.28	0.40 -2.21 – 3.02	-3.53 -6.10 – -0.95	-3.61 -6.13 – -1.10	1.52 -1.37 – 4.42
Public administration	-1.25 -6.22 – 3.72	-5.07 -10.37 – 0.23	0.31 -3.06 – 3.69	-2.79 -5.97 – 0.39	-1.62 -4.87 – 1.61	-2.27 -5.85 – 1.30
Education	-2.92 -7.57 – 1.72	-1.90 -7.39 – 3.59	0.34 -2.81 – 3.49	-1.87 -5.17 – 1.42	-3.29 -6.31 – -0.26	-0.02 -3.73 – 3.68
Collective, social and personal services	-3.88 -7.75 – -0.01	-6.17 -10.56 – -1.79	-0.12 -2.75 – 2.51	-4.90 -7.53 – -2.27	-3.77 -6.29 – -1.25	-1.27 -4.23 – 1.68
Private (vs. public)	0.60 -1.71 – 2.92	-0.92 -4.23 – 2.40	0.08 -1.49 – 1.65	1.32 -0.67 – 3.31	0.48 -1.02 – 1.99	-2.24 -4.48 – -0.01

*Models were adjusted for age, education, income, perceived financial strain, marital status, occupational status, antihypertensive medication use and neighborhood level of education.
 Note. SBP, systolic blood pressure; DBP, diastolic blood pressure; CI, confidence interval.

As a result of the different patterns of associations with SBP and DBP among men and women, no association was documented between work economic sectors and pulse pressure among women, while all economic sectors (especially Hotels and restaurants) were related to a lower pulse pressure than in the Health and social work sector among men. Women working in the private sector had a lower pulse pressure than those working in the public sector (private – public sector differences were documented for no other risk factor).

Cholesterol

Work economic sectors were not associated with total cholesterol or LDL cholesterol among men and women (Table 3). Associations were documented between work economic sectors and HDL cholesterol, but only among men. Men working in the Health and social work sector had the highest HDL cholesterol, while the lowest HDL cholesterol levels were observed in the Hotels and restaurants sector and in the Education sector.

Glycaemia

No association was documented with glycaemia among women (Table 4). Among men, those working in the Transportation and communications sector and in the Hotels and restaurants sector had the highest glycaemia

levels while those working in the Health and social work sector had the lowest level.

Resting heart rate

No association was documented between work economic sectors and resting heart rate among men. As the only associations that were observed among women, those working in Hotels or restaurants and in the Collective, social, and personal services sector had the lowest resting heart rate (Table 4).

When work economic sectors were combined to distinguish between the secondary sector and the tertiary sector, no association was found with any of the cardiovascular risk factors among men and women (Additional file 6).

Discussion

Because, as social class, work economic sectors may be seen as an upstream cause of diseases, and because knowledge on the relationships between work economic sectors and health may be useful to target Public health interventions, our aim was to investigate disparities between work economic sectors and most of the basic anthropometric and biological risk factors of cardiovascular diseases. In order to assess whether work economic sectors truly contributed to the variations in cardiovascular risk, our analyses were carefully adjusted for individual and neighborhood socioeconomic factors.

Table 3 Associations with total cholesterol, LDL cholesterol, and HDL cholesterol estimated from multilevel regression models among men and women*

	Total cholesterol		LDL cholesterol		HDL cholesterol	
	Men	Women	Men	Women	Men	Women
	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
Work economic sector						
Ref: Health and social work						
Manufacturing industry	2.27 -9.03 – 13.58	-1.12 -12.28 – 10.03	3.71 -6.16 – 13.58	-3.08 -13.05 – 6.89	-1.74 -5.12 – 1.65	0.22 -3.82 – 4.26
Construction	8.05 -4.35 – 20.46	12.04 -14.48 – 38.55	9.00 -1.84 – 19.84	12.87 -10.75 – 36.50	-1.86 -5.58 – 1.85	0.89 -8.69 – 10.48
Commercial, repair of motor vehicles and motorcycles	1.26 -10.09 – 12.61	4.41 -5.95 – 14.79	3.03 -6.88 – 12.94	1.28 -8.00 – 10.57	-2.28 -5.68 – 1.12	2.53 -1.23 – 6.30
Hotels and restaurants	0.69 -11.33 – 12.72	12.84 -0.26 – 25.93	3.98 -6.52 – 14.49	8.83 -2.90 – 20.57	-4.98 -8.58 – -1.37	3.04 -1.72 – 7.80
Transport and communications	-4.53 -16.40 – 7.34	-0.45 -13.06 – 12.17	-1.64 -12.03 – 8.75	-3.25 -14.57 – 8.07	-3.52 -7.08 – 0.04	2.59 -1.99 – 7.18
Financial activities	4.41 -7.13 – 15.95	1.56 -9.67 – 12.81	6.87 -3.20 – 16.95	1.74 -8.31 – 11.80	-2.30 -5.76 – 1.15	-0.50 -4.58 – 3.58
Real estate, renting and business services	0.70 -10.06 – 11.46	4.84 -4.43 – 14.13	3.06 -6.34 – 12.45	2.45 -5.87 – 10.78	-2.61 -5.84 – 0.61	1.83 -1.54 – 5.21
Public administration	1.92 -12.09 – 15.92	-5.20 -16.69 – 6.29	1.38 -10.85 – 13.61	-2.01 -12.28 – 8.26	-2.55 -6.75 – 1.64	-2.46 -6.63 – 1.70
Education	-3.54 -16.43 – 9.33	4.04 -7.87 – 15.95	1.52 -9.72 – 12.77	3.70 -6.94 – 14.34	-3.87 -7.74 – -0.01	-0.77 -5.09 – 3.54
Collective, social and personal services	5.20 -5.57 – 15.97	-1.55 -11.01 – 7.89	7.30 -2.11 – 16.70	-2.51 -10.97 – 5.94	-2.89 -6.11 – 0.33	1.28 -2.14 – 4.72
Private (vs. public)	3.17 -3.29 – 9.64	-1.89 -9.11 – 5.32	2.16 -3.49 – 7.81	0.34 -6.10 – 6.77	-0.41 -2.35 – 1.52	-0.79 -3.40 – 1.82

*Models were adjusted for age, education, income, perceived financial strain, marital status, occupational status and neighborhood level of education.
 Note. LDL, low-density lipoprotein; HDL, high-density lipoprotein; CI, confidence interval.

Table 4 Associations with glycaemia and resting heart rate estimated from multilevel regression models among men and women*

	Glycaemia		Resting heart rate	
	Men β (95% CI)	Women β (95% CI)	Men β (95% CI)	Women β (95% CI)
Work economic sector				
Ref.: Health and social work				
Manufacturing industry	2.55 -1.32 – 6.43	0.65 -3.58 – 4.89	-1.36 -4.16 – 1.44	-2.22 -5.15 – 0.70
Construction	4.08 -0.17 – 8.33	1.28 -8.80 – 11.37	-1.84 -4.90 – 1.22	5.47 -1.59 – 12.53
Commercial, repair of motor vehicles and motorcycle	3.13 -0.75 – 7.02	1.56 -2.37 – 5.50	-0.83 -3.64 – 1.98	-2.14 -4.88 – 0.60
Hotels and restaurants	4.98 0.85 – 9.11	-0.66 -5.64 – 4.33	-1.39 -4.36 – 1.58	-3.68 -7.02 – -0.34
Transport and communications	6.30 2.23 – 10.37	-0.47 -5.24 – 4.28	-1.13 -4.07 – 1.80	-0.85 -4.14 – 2.43
Financial activities	3.34 -0.61 – 7.30	-0.25 -4.53 – 4.02	-1.08 -3.95 – 1.77	0.09 -2.83 – 3.02
Real estate, renting and business services	2.92 -0.76 – 6.61	0.26 -3.26 – 3.78	-0.95 -3.61 – 1.71	-0.22 -2.66 – 2.22
Public administration	0.87 -3.90 – 5.64	1.81 -2.55 – 6.18	0.32 -3.12 – 3.76	0.40 -2.62 – 3.42
Education	0.56 -3.88 – 4.99	-2.05 -6.58 – 2.47	-1.51 -4.73 – 1.70	0.72 -2.35 – 3.79
Collective, social and personal services	3.11 -0.57 – 6.80	-0.57 -4.16 – 3.01	-1.71 -4.38 – 0.96	-2.49 -4.96 – -0.02
Private (vs. public)	-0.39 -2.62 – 1.84	-1.31 -4.04 – 1.41	1.19 -0.42 – 2.80	0.36 -1.54 – 2.25

*Models were adjusted for age, education, income, perceived financial strain, marital status, occupational status, and neighborhood level of education.
 Note. CI, confidence interval.

Main findings

We found disparities in certain (but not all) cardiovascular risk factors between work economics sectors, and the patterns of such disparities were different in men and women and according to the risk factor.

Regarding similarities observed from one risk factor to the other, the Health and social work sector was found to be the most protective sector for BMI and waist circumference but also for glycaemia among men, while unfavorable profiles for these three risk factors among men were relatively consistently observed in the Construction sector and in the Transport and communications sector. Among men, the HDL cholesterol was also found to be the highest for those who were working in the Health and social work sector. However, this pattern did not apply to the other risk factors since no clear relationships were documented among men for total cholesterol, LDL cholesterol, and resting heart rate, and since, on the opposite, men working in the Health and social work sector showed the highest systolic BP and pulse pressure (rather than the lowest as for BMI, waist circumference, and glycaemia).

There were also discrepancies in the patterns of disparities between men and women. While the Health and social work sector was the most protective among men for different risk factors, women working in the Health and social work sector had the highest BMI, the largest waist circumference, and the most elevated systolic but also diastolic BP. On the opposite, the Commercial and repair of vehicles sector, the Transport and communication sector, and the Collective, social, and personal services sector were fairly consistently associated with a

more favorable profile for these risk factors among women. Overall, it is not clear whether the higher SBP (and higher DBP among women) documented in the Health and social work sector may be due to the specific occupational risks of the sector, including strong constraints in the schedule and a high physical demand and job strain [22,31].

Some of the disparities between work economic sectors may be due to differences in total energy intake or energy expenditure between work economic sectors. Regarding energy intake, a Japanese study [49] based on food records indicated that the daily average energy intake was higher in Craftsmen, in the Production process and construction personnel, and in Laborers (2432 kcal), in the service personnel (2444 kcal), in the professional and technical personnel (2174 kcal), and in the transport and communication personnel (2200 kcal). These data derived from selected individuals of a different country with markedly different dietary habits likely not represent the actual energy intake of the French subjects in the present study. If these findings on dietary habits are unable to explain, e.g., the higher BMI and larger waist circumference of men working in the Construction sector and Transportation and communications sector documented in our study, at least they show that energy intake can vary to a large extent between work economic sectors, as a potential explanation for the observed disparities. Similarly, our study does not empirically confirm that the trend of relationship between working in hotels and restaurants and higher total cholesterol and the relationship between working in this sector and lower HDL cholesterol documented among

women is attributable to the specific dietary habits of these women.

Because work can be a major contributor to the total daily physical activity [50], an hypothesis may be that differences in physical activity levels between work economic sectors contribute to the reported economic disparities in cardiovascular risk factors. Based on this hypothesis, one would expect work economic sectors typically related with heavy manual work to be associated with a lower prevalence of cardiovascular risk factors, especially with a lower BMI and thinner waist circumference. However, we did not validate empirically this hypothesis, since for example men working in the Construction sector were found to have the highest BMI and largest waist circumference. This is in line with a previous study that indicated that, unlike recreational physical activity, occupational physical activity was not beneficial for cardiovascular risk factors, including those related to excess weight [51].

It should be noted that in addition to the social and physical characteristics of work conditions, the characteristics of the geographic environment around the typical locations of the worksites of each work economic sector may contribute to the reported associations.

Finally, also as an indication of the relevance of our examination of the different work economic sectors, we found that associations were lost when contrasting the secondary sector to the tertiary sector, which distinction was therefore not meaningful for cardiovascular risk factors.

Strengths and limitations

Regarding study strengths, first, unlike previous studies on work economic sectors and cardiovascular risk factors, we examined in a comparative way a large panel of cardiovascular risk factors including BMI, waist circumference, SBP and DBP, pulse pressure, total cholesterol, HDL cholesterol, LDL cholesterol, glycaemia, and resting heart rate. While a large number of studies [5-9] have focused on differences in cardiovascular risk factors between occupational groups (as a *social* category), much less studies [7,11,28] have analyzed cardiovascular risk according to the work economic sector (as an *economic* variable). Our study is the very first to examine the relationships between work economic sectors and a large set of cardiovascular risk factors after controlling for individual and residential neighborhood socioeconomic variables. Another strength of our study is that information on the work economic sector was obtained through the linkage of administrative data to a population sample, leading to the presence of a very large number of workplaces ($n = 3553$ work establishments) and companies ($n = 3005$) in the sample, thus improving the generalizability of the findings.

Regarding study limitations, a first shortcoming of the work is related to the notion of work economic sector and

operationalization of it that were used. It may be difficult to build a universal classification of work economic sectors because of the large difference in socioeconomic background between the countries [49]. In addition, it is sometimes problematic to classify a subject into only one work economic sector, since a number of companies and work establishments may be related to different economic sectors. Moreover, as the work life of workers is complex to assess and is the combination of a number of activities and circumstances, even within the same work economic sector, there may be notable variations in the lifestyles, in the magnitude of psychological stress, and in the intensity of physical activity (for a given occupation) between different territories, especially between different countries [11].

As a second limitation, the cross-sectional design of the study made it impossible to determine the direction of the causal effects involved in the associations reported (a poor cardiovascular health may differentially encourage to withdraw from professional activity according to the work economic sector, or may lead to seek for a job in specific sectors). Third, our sample recruited in preventive healthcare centers was not representative of the Paris Ile-de-France region [32]. However, a large panel of municipalities from the region was *a priori* selected to ensure the presence in the sample of people from all socioeconomic backgrounds. Moreover, the present analysis controlled for the individual and neighborhood factors that were found to influence participation in the study [52]. Fourth, the workplace was retrieved through the linkage of administrative information and we could not be formally sure that all the participants were still employed in the company and in the related work economic sector at the recruitment in the study. It may be expected, however, that participants changing from one company to another would tend to remain in the same work economic sector. Moreover, even if the participants were working in another work economic sector or were no longer working when they were enrolled in the study, they had been exposed to the work economic sector taken into account in the study the year prior to their recruitment (exposure to work economic sectors was measured homogeneously in the whole sample). Fifth and finally, the duration of employment in the work economic sector was not taken into account in the present study.

Conclusion

In conclusion, the results of this study suggest that work economic sectors contribute to shape metabolic and cardiovascular parameters, even after adjustment for individual and neighborhood sociodemographic characteristics. The patterns of associations were found to vary, however, according to the risk factor examined and between men and women.

Overall, even if our study did not consider specific characteristics of work environments, the examination of the relationships between broad work economic sectors and a panel of cardiovascular risk factors is useful, in addition to the consideration of individual and residential characteristics, to determine profiles of populations to target Public health interventions and prevention efforts. As an example, while our previous work showed that having a low education, residing in neighborhoods with a low education and a low urbanicity degree, and shopping in hard discount supermarkets were all associated with a higher BMI and larger waist circumference [30,36,38], the present study further suggests that working in the Construction and in the Transport and communication sectors among males was also related to a larger body weight and fat, thus refining our knowledge on the profile of people at risk.

Additional files

Additional file 1: Table S1. Tabulation of socioeconomic status by work economic sectors: % of participants in each category after excluding participants with missing information for the variable.

Additional file 2: Associations between individual and neighborhood sociodemographic variables and BMI and waist circumference among men and women.

Additional file 3: Associations between individual and neighborhood sociodemographic variables and SBP, DBP, and pulse pressure among men and women.

Additional file 4: Associations between individual and neighborhood sociodemographic variables and total cholesterol, LDL cholesterol, and HDL cholesterol among men and women.

Additional file 5: Associations between individual and neighborhood sociodemographic variables and glycaemia and resting heart rate among men and women.

Additional file 6: Associations between working in the secondary rather than in the tertiary sector and cardiovascular risk factors estimated from multilevel regression models among men and women.

Abbreviations

BMI: Body mass index; SBP: Systolic blood pressure; DBP: Diastolic blood pressure; HDL: High-density lipoprotein; LDL: Low-density lipoprotein; CI: Confidence interval; CNAV: National Old Age Insurance System; Insee: French National Institute of Statistics and Economic Studies.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

BC designed the overall research project; BC and AL designed the analytical strategy of the present study; AL analyzed the data; AL wrote the paper, and BC revised the first draft of the manuscript; all other authors, including BP and FT, critically revised the manuscript for important intellectual content. All authors have read and approved the final manuscript. BC and AL have primary responsibility for the final content of the article.

Acknowledgements

We are grateful to INPES (and Pierre Arwidson) for its continued support since the beginning of the study. We also thank Mélanie Alberto from the ARS of Ile-de-France. We are grateful to Insee, the French National Institute of Statistics and Economic Studies, which provided support for the geocoding of the RECORD participants and allowed us to access to relevant geographical data (with special thanks to Pascale Breuil). We thank Geoconcept for allowing

us to access to the Universal Geocoder software. We also thank CNAM-TS and the Caisse Primaire d'Assurance Maladie de Paris (CPAM-P, France) for helping make this study possible. We thank Isabelle Bridenne and Vincent Poubelle who allowed us to access to the data from the National Old Age Insurance System (Caisse Nationale d'Assurance Vieillesse, CNAV). Regarding the data used the present study, we also thank the National Geographic Institute (IGN). This work was supported by a doctoral grant of Région Ile-de-France (CORDDIM) attributed to Antoine Lewin. The RECORD study is funded by the Institute for Public Health Research (IReSP, Institut de Recherche en Santé Publique); the National Institute for Prevention and Health Education (INPES, Institut National de Prévention et d'Education pour la Santé); the National Institute of Public Health Surveillance (InVS, Institut de Veille Sanitaire); the French Ministries of Research and Health (Epidemiologic Cohorts Grant 2008); the National Health Insurance Office for Salaried Workers (CNAM-TS, Caisse Nationale d'Assurance Maladie des Travailleurs Saliés); the Ile-de-France Regional Health Agency (ARS, Agence Régionale de Santé); the National Research Agency (ANR, Agence Nationale de la Recherche); the City of Paris (Ville de Paris); and the Ile-de-France Youth, Sports, and Social Cohesion Regional Direction (DRJSCS, Direction Régionale de la Jeunesse, des Sports et de la Cohésion Sociale).

Author details

¹Sorbonne Universités, UPMC Univ Paris 06, UMR_S 113, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris 75012, France. ²Inserm, UMR_S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris 75012, France. ³Centre d'Investigations Préventives et Cliniques, 6 rue La Pérouse, Paris 75116, France.

Received: 9 April 2014 Accepted: 2 July 2014

Published: 24 July 2014

References

1. Yusuf S, Islam S, Chow CK, Rangarajan S, Dagenais G, Diaz R, Gupta R, Kelishadi R, Iqbal R, Avezum A, Kruger A, Kuttly R, Lanas F, Lisheng L, Wei L, Lopez-Jaramillo P, Oguz A, Rahman O, Swidan H, Yusuf K, Zatoński W, Rosengren A, Teo KK, Prospective Urban Rural Epidemiology (PURE) Study Investigators: Use of secondary prevention drugs for cardiovascular disease in the community in high-income, middle-income, and low-income countries (the PURE Study): a prospective epidemiological survey. *Lancet* 2011, **378**(9798):1231–1243.
2. Gupta R, Kaul V, Agrawal A, Guptha S, Gupta VP: Cardiovascular risk according to educational status in India. *Prev Med* 2010, **51**(5):408–411.
3. Reddy KS, Prabhakaran D, Jeemon P, Thankappan KR, Joshi P, Chaturvedi V, Ramakrishnan L, Ahmed F: Educational status and cardiovascular risk profile in Indians. *Proc Natl Acad Sci U S A* 2007, **104**(41):16263–16268.
4. Jeemon P, Prabhakaran D, Huffman MD, Ramakrishnan L, Goenka S, Thankappan KR, Mohan V, Joshi PP, Mohan BV, Ahmed F, Ramanathan M, Ahuja R, Chaturvedi V, Lloyd-Jones DM, Reddy KS, Sentinel Surveillance in Industrial Populations Study Group: Distribution of 10-year and lifetime predicted risk for cardiovascular disease in the Indian Sentinel Surveillance Study population (cross-sectional survey results). *BMJ Open* 2011, **1**(1):e000068.
5. Frommer MS, Edye BV, Mandryk JA, Grammeno GL, Berry G, Ferguson DA: Systolic blood pressure in relation to occupation and perceived work stress. *Scand J Work Environ Health* 1986, **12**(5):476–485.
6. Hu G, Jousilahti P, Antikainen R, Tuomilehto J: Occupational, commuting, and leisure-time physical activity in relation to cardiovascular mortality among Finnish subjects with hypertension. *Am J Hypertens* 2007, **20**(12):1242–1250.
7. Leigh JP, Du J: Hypertension and occupation among seniors. *J Occup Environ Med* 2009, **51**(6):661–671.
8. Schnall PL, Landsbergis PA, Baker D: Job strain and cardiovascular disease. *Annu Rev Public Health* 1994, **15**:381–411.
9. Tsutsumi A, Kayaba K, Tsutsumi K, Igarashi M: Association between job strain and prevalence of hypertension: a cross sectional analysis in a Japanese working population with a wide range of occupations: the Jichi Medical School cohort study. *Occup Environ Med* 2001, **58**(6):367–373.
10. Kaplan GA, Keil JE: Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation* 1993, **88**(4 Pt 1):1973–1998.
11. Takahima Y, Yoshida M, Kokaze A, Orido Y, Tsugane S, Ishikawa M, Takeuchi Y, Takagi Y, Tanaka N, Watanabe S, Akamatsu T: Relationship of occupation

- to blood pressure among middle-aged Japanese men—the significance of the differences in body mass index and alcohol consumption. *J Epidemiol* 1998, **8**(4):216–226.
12. Haglund BJ: Geographical and socioeconomic distribution of high blood pressure and borderline high blood pressure in a Swedish rural county. *Scand J Soc Med* 1985, **13**(2):53–66.
 13. Covey LS, Wynder EL: Smoking habits and occupational status. *J Occup Med* 1981, **23**(8):537–542.
 14. Dobson AJ, Gibberd RW, Leeder SR, O'Connell DL: Occupational differences in ischemic heart disease mortality and risk factors in Australia. *Am J Epidemiol* 1985, **122**(2):283–290.
 15. Simons LA, Simons J, Magnus P, Bennett SA: Education level and coronary risk factors in Australians. *Med J Aust* 1986, **145**(9):448–450.
 16. Helmert U, Shea S, Herman B, Greiser E: Relationship of social class characteristics and risk factors for coronary heart disease in West Germany. *Public Health* 1990, **104**(6):399–416.
 17. Kritz-Silverstein D, Wingard DL, Barrett-Connor E: Employment status and heart disease risk factors in middle-aged women: the Rancho Bernardo Study. *Am J Public Health* 1992, **82**(2):215–219.
 18. Wilson TW, Kaplan GA, Kauhanen J, Cohen RD, Wu M, Salonen R, Salonen JT: Association between plasma fibrinogen concentration and five socioeconomic indices in the Kuopio Ischemic Heart Disease Risk Factor Study. *Am J Epidemiol* 1993, **137**(3):292–300.
 19. Medalie JH, Papier C, Herman JB, Goldbourt U, Tamir S, Neufeld HN, Riss E: Diabetes mellitus among 10,000 adult men. I Five-year incidence and associated variables *Isr J Med Sci* 1974, **10**(7):681–697.
 20. Mlczarek M, Schneider E, González ER: *Stress at Work. Volume 9. Edited by Report ERO.* Luxembourg: European Agency for Safety and Health at Work; 2009:132.
 21. Holman D, McClelland C: *Job Quality in Growing and Declining Economic Sectors of the EU.* In *Work and Life Quality in new & Growing Jobs.* Manchester: 2011:720.
 22. DARES: *Conditions de travail et santé.* Edited by Direction de l'animation de la recherche des études et des statistiques. Paris: Ministère du Travail, de l'Emploi, de la Formation Professionnelle et du Dialogue Social; 2009.
 23. Klein T, Long K: *Working conditions, work organization and use of ICT as business.* In *Center for Strategic Analysis, Department Labour Employment.* Paris: 2013.
 24. Sandor E: *European Company Survey 2009. Part-time work in Europe.* In *European Foundation for the Improvement of Living and Working Conditions.* Dublin: Eurofound; 2009.
 25. Lauzeville D, Marchand JL, Ferrand M: *Consommation de tabac par catégorie socioprofessionnelle et secteur d'activité: outil méthodologique pour l'épidémiologie.* 2009.
 26. Beck F: *Résultats du Baromètre Santé 2010 - Liens entre usages de substances psychoactives (SPA) et milieu professionnel.* In *Enquêtes et analyses statistiques.* Paris: INPES; 2012.
 27. Goffette C: *Le contexte social du tabagisme: Le rôle de l'environnement familial et professionnel.* Science po Paris, Sociology; 2012.
 28. Proper KI, Hildebrandt VH: Overweight and obesity among Dutch workers: differences between occupational groups and sectors. *Int Arch Occup Environ Health* 2010, **83**(1):61–68.
 29. Hasson D, Von Thiele SU, Lindfors P: Self-rated health and allostatic load in women working in two occupational sectors. *J Health Psychol* 2009, **14**(4):568–577.
 30. Leal C, Bean K, Thomas F, Chaix B: Are associations between neighborhood socioeconomic characteristics and body mass index or waist circumference based on model extrapolations? *Epidemiology* 2011, **22**(5):694–703.
 31. Chaix B, Bean K, Leal C, Thomas F, Havard S, Evans D, Jago B, Pannier B: Individual/neighborhood social factors and blood pressure in the RECORD Cohort Study: which risk factors explain the associations? *Hypertension* 2010, **55**(3):769–775.
 32. Chaix B, Billaudeau N, Thomas F, Havard S, Evans D, Kestens Y, Bean K: Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology* 2011, **22**(1):18–26.
 33. Chaix B, Jouven X, Thomas F, Leal C, Billaudeau N, Bean K, Kestens Y, Jago B, Pannier B, Danchin N: Why socially deprived populations have a faster resting heart rate: impact of behaviour, life course anthropometry, and biology—the RECORD Cohort Study. *Soc Sci Med* 2011, **73**(10):1543–1550.
 34. Chaix B, Kestens Y, Bean K, Leal C, Karusisi N, Meghrief K, Burban J, Fon Sing M, Perchoux C, Thomas F, Merlo J, Pannier B: Cohort profile: residential and non-residential environments, individual activity spaces and cardiovascular risk factors and diseases—the RECORD Cohort Study. *Int J Epidemiol* 2012, **41**(5):1283–1292.
 35. Havard S, Reich BJ, Bean K, Chaix B: Social inequalities in residential exposure to road traffic noise: an environmental justice analysis based on the RECORD Cohort Study. *Occup Environ Med* 2011, **68**(5):366–374.
 36. Chaix B, Bean K, Daniel M, Zenk SN, Kestens Y, Charreire H, Leal C, Thomas F, Karusisi N, Weber C, Oppert JM, Simon C, Merlo J, Pannier B: Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). *PLoS One* 2012, **7**(4):e32908.
 37. Karusisi N, Bean K, Oppert JM, Pannier B, Chaix B: Multiple dimensions of residential environments, neighborhood experiences, and jogging behavior in the RECORD Study. *Prev Med* 2012, **55**(1):50–55.
 38. Leal C, Bean K, Thomas F, Chaix B: Multicollinearity in associations between multiple environmental features and body weight and abdominal fat: using matching techniques to assess whether the associations are separable. *Am J Epidemiol* 2012, **175**(11):1152–1162.
 39. Karusisi N, Thomas F, Meline J, Chaix B: Spatial accessibility to specific sport facilities and corresponding sport practice: the RECORD Study. *Int J Behav Nutr Phys Act* 2013, **10**(1):48.
 40. INSEE: *IRIS—Definition.* Paris, France: National Institute of Statistics and Economic Studies; 2011.
 41. *Cardiovascular disease.* [http://www.who.int/mediacentre/factsheets/fs317/en/]
 42. Thomas F, Bean K, Pannier B, Oppert JM, Guize L, Benetos A: Cardiovascular mortality in overweight subjects: the key role of associated risk factors. *Hypertension* 2005, **46**(4):654–659.
 43. WHO: *Obesity: preventing and managing the global epidemic. Report of a WHO consultation.* *World Health Organ Tech Rep Ser* 2000, **894**:1–253. i-xi.
 44. Benetos A, Thomas F, Pannier B, Bean K, Jago B, Guize L: All-cause and cardiovascular mortality using the different definitions of metabolic syndrome. *Am J Cardiol* 2008, **102**(2):188–191.
 45. Pannier B, Thomas F, Bean K, Jago B, Benetos A, Guize L: The metabolic syndrome: similar deleterious impact on all-cause mortality in hypertensive and normotensive subjects. *J Hypertens* 2008, **26**(6):1223–1228.
 46. Benetos A, Rudnichi A, Thomas F, Safar M, Guize L: Influence of heart rate on mortality in a French population: role of age, gender, and blood pressure. *Hypertension* 1999, **33**(1):44–52.
 47. Chaix B, Simon C, Charreire H, Thomas F, Kestens Y, Karusisi N, Vallee J, Oppert JM, Weber C, Pannier B: The environmental correlates of overall and neighborhood based recreational walking (a cross-sectional analysis of the RECORD Study). *Int J Behav Nutr Phys Act* 2014, **11**(1):20.
 48. Chaix B, Leal C, Evans D: Neighborhood-level confounding in epidemiologic studies: unavoidable challenges, uncertain solutions. *Epidemiology* 2010, **21**(1):124–127.
 49. Takashima Y, Iwase Y, Yoshida M, Kokaze A, Takagi Y, Taubono Y, Tsugane S, Takahashi T, Itoi Y, Akabane M, Watanabe S, Akamatsu T: Relationship of food intake and dietary patterns with blood pressure levels among middle-aged Japanese men. *J Epidemiol* 1998, **8**(2):106–115.
 50. Proper KI, Hildebrandt VH: Physical activity among Dutch workers—differences between occupations. *Prev Med* 2006, **43**(1):42–45.
 51. Oppert JM, Thomas F, Charles MA, Benetos A, Basdevant A, Simon C: Leisure-time and occupational physical activity in relation to cardiovascular risk factors and eating habits in French adults. *Public Health Nutr* 2006, **9**(6):746–754.
 52. Chaix B, Evans D, Merlo J, Suzuki E: Commentary: Weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. *Epidemiology* 2012, **23**(1):129–131. discussion 132–127.

doi:10.1186/1471-2458-14-750
Cite this article as: Lewin et al.: Work economic sectors and cardiovascular risk factors: cross-sectional analysis based on the RECORD Study. *BMC Public Health* 2014 **14**:750.

Additional Table 1. Tabulation of socioeconomic status by work economic sectors: % of participants in each category after excluding participants with missing information for the variable

Variables	Health and social work (n =180)	Manufacturing industry (n =496)	Construction (n =173)	Commercial, repair of motor vehicles and motorcycles (n =508)	Hotels and restaurants (n =239)	Transport and communications (n =256)	Financial activities (n =408)	Real estate, renting and business services (n =1332)	Public administration (n =128)	Education (n =142)	Collective, social, and personal services (n =498)
Age											
30-44	39	37	47	46	55	49	34	49	39	42	46
45-59	51	52	47	45	37	44	58	43	48	37	46
60-79	10	11	6	9	8	7	8	8	13	21	8
Individual education											
Low	12	4	18	8	19	6	2	8	8	4	7
Medium-Low	33	24	43	25	37	18	17	15	32	14	20
Medium-High	28	27	25	34	31	31	27	28	33	28	33
High	27	45	14	33	13	45	54	49	27	54	40
Occupation											
Blue-collar	16	13	54	16	38	15	2	13	15	6	11
Low white-collar	56	19	13	32	47	22	26	30	55	51	47
Intermediate	13	10	11	9	4	9	4	6	6	6	9
High white-collar	15	58	22	43	11	54	68	51	24	37	33
Household income											
Low	51	17	39	26	52	23	9	28	47	39	21

Medium-low	23	24	29	29	25	20	19	21	27	25	28
Medium-high	17	25	19	21	13	25	32	22	16	15	27
High	9	34	13	24	10	32	40	29	10	21	24
Living alone	41	11	21	26	33	26	7	30	49	37	32
Perceived financial strain	32	10	16	18	31	13	28	18	36	26	16
Antihypertensive medication use	10	6	8	5	5	6	7	5	14	6	4
Educational level of local residents											
Low	41	23	53	29	33	23	12	25	35	18	18
Medium-low	27	23	21	20	26	23	21	21	26	22	22
Medium-high	17	26	16	27	21	25	28	26	24	24	31
High	15	28	10	24	20	29	39	28	15	36	29

Additional file 2 Associations between individual and neighborhood sociodemographic variables and BMI and waist circumference among men and women.

	BMI				Waist circumference			
	Men		Women		Men		Women	
	β	(95% CI)	β	(95% CI)	β	(95% CI)	β	(95% CI)
Age (vs. 30 to 44)								
45 to 59	0.77	0.51 – 1.03	0.91	0.36 – 1.46	3.77	3.01 – 4.53	2.62	1.32 – 3.92
60 to 79	1.21	0.74 – 1.68	1.97	1.04 – 2.90	5.87	4.51 – 7.24	5.66	3.47 – 7.85
Individual education (vs. high education)								
Medium-high education	0.50	0.17 – 0.83	0.18	-0.51 – 0.87	1.22	0.26 – 2.18	0.36	-1.26 – 1.99
Medium-low education	0.61	0.20 – 1.02	1.36	0.53 – 2.20	1.98	0.79 – 3.17	3.09	1.12 – 5.06
Low education	0.56	-0.02 – 1.15	2.68	1.53 – 3.83	0.92	-0.79 – 2.64	5.05	2.35 – 7.75
Household income (vs. high income)								
Medium-high income	-0.13	-0.48 – 0.22	0.17	-0.63 – 0.97	-0.18	-1.21 – 0.84	0.17	-1.72 – 2.05
Medium-low income	-0.32	-0.69 – 0.05	0.96	0.15 – 1.78	-0.58	-1.67 – 0.51	1.49	-0.42 – 3.40
Low income	-0.66	-1.11 – -0.22	0.62	-0.26 – 1.50	-0.98	-2.27 – 0.31	1.31	-0.76 – 3.38
Perceived financial strain	0.51	0.13 – 0.90	1.11	0.43 – 1.78	1.03	-0.09 – 2.15	2.59	0.99 – 4.18
Living alone (vs. as a couple)	-0.92	-1.22 – -0.62	-0.20	-0.73 – 0.33	-2.13	-3.01 – -1.25	-0.92	-2.17 – 0.32

Occupational status (vs. high white-collar)

Intermediate	-0.20 -0.74 – 0.33	-0.05 -1.16 – 1.06	-0.98 -2.53 – 0.58	-0.77 -3.39 – 1.85
Low-white collar	0.09 -0.28 – 0.46	-0.32 -1.03 – 0.38	-0.08 -1.16 – 1.01	-0.22 -1.87 – 1.43
Blue-collar	-0.45 -0.93 – 0.02	-0.14 -1.41 – 1.13	-2.08 -3.46 – -0.71	-0.53 -3.52 – 2.46

Residential education level (vs. high)

Medium-high	0.13 -0.20 – 0.47	-0.15 -0.87 – 0.57	-0.31 -1.29 – 0.68	-0.54 -2.24 – 1.16
Medium-low	0.26 -0.10 – 0.63	0.34 -0.41 – 1.09	0.28 -0.78 – 1.35	0.14 -1.64 – 1.91
Low	1.01 0.62 – 1.39	1.78 0.98 – 2.58	1.55 0.43 – 2.67	3.59 1.70 – 5.48

Note. CI, confidence interval; BMI, Body Mass Index.

Additional file 3 Associations between individual and neighborhood sociodemographic variables and SBP, DBP, and pulse pressure among men and women.

	SBP		DBP		Pulse pressure	
	Men β (95% IC)	Women β (95% IC)	Men β (95% IC)	Women β (95% IC)	Men β (95% IC)	Women β (95% IC)
Age (vs. 30 to 44)						
45 to 59	5.26 4.16 – 6.35	7.86 5.80 – 9.93	5.26 4.51 – 6.00	2.89 1.66 – 4.14	0.03 -0.68 – 0.74	4.96 3.57 – 6.36
60 to 79	10.02 8.04 – 11.99	14.06 10.56 – 17.57	5.67 4.32 – 7.01	4.86 2.76 – 6.97	4.34 3.05 – 5.63	9.20 6.83 – 11.56
Individual education (vs. high education)						
Medium-high education	1.34 -0.03 – 2.71	-0.20 -2.78 – 2.37	0.87 -0.05 – 1.81	-0.40 -1.95 – 1.14	0.48 -0.41 – 1.37	0.19 -1.54 – 1.93
Medium-low education	3.02 1.32 – 4.73	2.26 -0.86 – 5.39	1.33 0.18 – 2.49	0.86 -1.01 – 2.74	1.73 0.62 – 2.84	1.40 -0.70 – 3.51
Low education	2.83 0.364 – 5.29	3.76 -0.50 – 8.01	0.60 -1.07 – 2.27	2.64 0.09 – 5.20	2.21 0.60 – 3.82	1.11 -1.76 – 3.98
Household income (vs. high income)						
Medium-high income	-1.74 -3.20 – -0.27	-1.41 -4.40 – 1.58	-0.41 -1.41 – 0.57	-0.22 -2.02 – 1.57	-1.37 -2.32 – -0.41	-1.18 -3.20 – 0.83
Medium-low income	0.08 -1.48 – 1.64	0.06 -2.96 – 3.07	0.16 -0.90 – 1.22	1.06 -0.74 – 2.87	-0.15 -1.17 – 0.87	-1.00 -3.04 – 1.02
Low income	0.38 -1.47 – 2.23	-0.41 -3.69 – 2.88	0.69 -0.57 – 1.94	0.24 -1.72 – 2.21	-0.39 -1.60 – 0.81	-0.65 -2.87 – 1.56
Perceived financial	-0.84 -2.45 – 0.76	0.77 -1.75 – 3.29	-0.02 -1.11 – 1.07	-0.50 -2.01 – 1.01	-0.82 -1.86 – 0.23	1.27 -0.43 – 2.97

strain												
Living alone (vs. as a couple)	-0.26	-1.51 – 0.99	1.12	-0.85 – 3.08	-0.11	-0.97 – 0.74	0.97	-0.21 – 2.14	-0.17	-0.99 – 0.64	0.15	-1.17 – 1.47
Occupational status (vs. white-collar)												
Intermediate	0.21	-2.02 – 2.43	1.71	-2.41 – 5.83	0.02	-1.48 – 1.53	1.33	-1.13 – 3.81	0.19	-1.25 – 1.64	0.37	-2.40 – 3.15
Low-white collar	-0.03	-1.58 – 1.52	1.62	-1.00 – 4.24	-0.94	-1.99 – 0.11	1.02	-0.55 – 2.59	0.94	-0.07 – 1.96	0.60	-1.16 – 2.37
Blue-collar	-0.00	-1.97 – 1.96	6.54	1.82 – 11.27	-0.45	-1.79 – 0.88	3.67	0.83 – 6.51	0.56	-0.72 – 1.85	2.87	-0.31 – 6.06
Antihypertensive medication	10.47	8.04 – 12.89	12.86	9.12 – 16.60	5.22	3.58 – 6.87	7.25	5.01 – 9.49	5.22	3.64 – 6.79	5.61	3.09 – 8.13
Residential education level (vs. high)												
Medium-high	1.73	0.31 – 3.15	-1.84	-4.52 – 0.83	1.07	0.12 – 2.02	-1.21	-2.82 – 0.39	0.69	-0.21 – 1.61	-0.63	-2.43 – 1.17
Medium -low	1.83	0.30 – 3.35	-0.93	-3.73 – 1.86	0.60	-0.42 – 1.63	-0.68	-2.36 – 0.99	1.28	0.30 – 2.27	-0.25	-2.13 – 1.63
Low	2.41	0.80 – 4.02	0.64	-2.34 – 3.61	1.46	0.37 – 2.54	1.49	-0.29 – 3.27	0.93	-0.10 – 1.97	-0.85	-2.86 – 1.15

Note. SBP, diastolic blood pressure; DBP, systolic blood pressure; CI, confidence interval.

Additional file 4 Associations between individual and neighborhood sociodemographic variables and total cholesterol, LDL cholesterol, and HDL cholesterol among men and women.

	Total cholesterol		LDL cholesterol		HDL cholesterol	
	Men	Women	Men	Women	Men	Women
	β (95% IC)	β (95% IC)	β (95% IC)	β (95% IC)	β (95% IC)	β (95% IC)
Age (vs. 30 to 44)						
45 to 59	11.37 8.37 – 14.37	25.15 20.72 – 29.59	8.90 6.27 – 11.52	19.63 15.67 – 23.58	0.94 0.04 – 1.84	3.88 2.28 – 5.49
60 to 79	9.78 4.26 – 15.30	37.44 29.98 – 44.91	5.94 1.11 – 10.77	29.57 22.90 – 36.24	3.02 1.36 – 4.68	5.10 2.39 – 7.80
Individual education (vs. high education)						
Medium-high education	5.17 1.38 – 8.97	5.83 0.30 – 11.37	4.15 0.83 – 7.47	5.66 0.72 – 10.61	-1.00 -2.14 – 0.13	-1.48 -3.48 – 0.52
Medium-low education	6.97 2.28 – 11.66	5.63 -1.08 – 12.34	3.97 -0.13 – 8.07	7.05 1.06 – 13.05	0.27 -1.13 – 1.68	-3.50 -5.93 – -1.07
Low education	2.13 -4.59 – 8.87	1.74 -7.47 – 10.96	1.26 -4.67 – 7.19	4.98 -3.24 – 13.21	-0.80 -2.83 – 1.23	-5.60 -8.94 – -2.26
Household income (vs. high income)						
Medium-high income	-0.10 -4.15 – 3.95	-2.62 -9.05 – 3.80	0.17 -3.37 – 3.71	-0.40 -6.12 – 5.32	0.61 -0.60 – 1.83	-1.69 -4.01 – 0.63
Medium-low income	-0.47 -4.81 – 3.86	-2.91 -9.43 – 3.59	-0.57 -4.36 – 3.22	-0.40 -6.21 – 5.40	-0.03 -1.34 – 1.26	-1.73 -4.09 – 0.62

Low income	-2.62	-7.74 – 2.49	-4.38	-11.47 – 2.71	2.97	-7.45 – 1.50	-0.84	-7.16 – 5.49	-0.39	-1.92 – 1.14	-2.86	-5.42 – -0.29
Perceived financial strain	-2.90	-7.34 – 1.52	1.87	-3.64 – 7.38	-2.05	-5.94 – 1.83	2.87	-2.04 – 7.79	-2.29	-3.62 – -0.96	-2.10	-4.09 – -0.10
Living alone (vs. as a couple)	-4.10	-7.56 – -0.63	-1.44	-5.69 – 2.81	-3.99	-7.02 – -0.96	-2.48	-6.28 – 1.31	0.60	-0.44 – 1.64	0.71	-0.83 – 2.25
Occupational status (vs. white-collar)												
Intermediate	-2.29	-8.42 – 3.83	-1.49	-10.34 – 7.36	-0.31	-5.67 – 5.03	-3.61	-11.49 – 4.26	0.32	-1.51 – 2.16	1.29	-1.89 – 4.49
Low-white collar	-1.26	-5.56 – 3.03	-3.66	-9.32 – 1.99	-1.33	-5.09 – 2.43	-4.10	-9.14 – 0.94	0.23	-1.06 – 1.52	0.34	-1.70 – 2.39
Blue-collar	-2.71	-8.13 – 2.71	2.18	-8.03 – 12.40	-2.69	-7.44 – 2.04	-0.13	-9.26 – 9.00	1.80	0.17 – 3.43	2.63	-1.07 – 6.34
Residential education level (vs. high)												
Medium-high	1.95	-1.93 – 5.85	-1.08	-6.92 – 4.75	1.94	-1.46 – 5.35	-0.12	-5.25 – 5.01	-0.24	-1.43 – 0.94	-0.69	-2.77 – 1.39
Medium -low	-0.83	-5.04 – 3.37	-2.81	-8.94 – 3.31	0.22	-3.47 – 3.91	-0.62	-6.03 – 4.77	-0.73	-2.01 – 0.55	-1.37	-3.57 – 0.81
Low	-3.72	-8.17 – 0.71	-2.86	-9.33 – 3.60	-1.18	-5.07 – 2.70	-0.13	-5.83 – 5.58	-1.63	-2.98 – -0.28	-2.54	-4.86 – -0.23

Note. LDL, low-density lipoprotein; HDL, high-density lipoprotein; CI, confidence interval.

Additional file 5 Associations between individual and neighborhood sociodemographic variables and glycaemia and resting heart rate among men and women.

	Glycaemia				Resting heart rate			
	Men		Women		Men		Women	
	β	(95% CI)	β	(95% CI)	β	(95% CI)	β	(95% CI)
Age (vs. 30 to 44)								
45 to 59	5.00	3.97 – 6.04	4.75	3.07 – 6.44	1.15	0.40 – 1.91	-0.37	-1.54 – 0.78
60 to 79	7.24	5.33 – 9.14	7.83	5.00 – 10.66	1.73	0.38 – 3.09	-0.05	-2.01 – 1.89
Individual education (vs. high education)								
Medium-high education	0.31	-0.99 – 1.62	-0.44	-2.55 – 1.65	1.14	0.19 – 2.09	-0.27	-1.73 – 1.19
Medium-low education	1.29	-0.32 – 2.91	-0.23	-2.78 – 2.31	1.64	0.46 – 2.81	-0.97	-2.74 – 0.79
Low education	0.75	-1.57 – 3.08	-0.63	-4.13 – 2.86	0.19	-1.50 – 1.90	1.54	-0.89 – 3.98
Household income (vs. high income)								
Medium-high income	-1.27	-2.67 – 0.12	-2.62	-5.05 – -0.19	-0.12	-1.13 – 0.89	-1.36	-3.04 – 0.33
Medium-low income	-1.07	-2.57 – 0.42	-1.95	-4.43 – 0.53	0.80	-0.27 – 1.89	-0.80	-2.51 – 0.90
Low income	-1.68	-3.45 – 0.08	-2.26	-4.96 – 0.43	2.59	1.31 – 3.87	0.68	-1.17 – 2.53
Perceived financial strain	0.21	-1.31 – 1.74	-0.63	-4.13 – 2.86	0.04	-1.07 – 1.15	-0.86	-2.29 – 0.57
Living alone (vs. as a couple)	-1.36	-2.55 – -0.16	0.24	-1.37 – 1.86	0.57	-0.29 – 1.44	-0.27	-1.39 – 0.84

Occupational status (vs. high white-collar)

Intermediate	-0.16 -2.27 – 1.94	2.49 -0.88 – 5.86	0.29 -1.25 – 1.83	1.13 -1.19 – 3.46
Low-white collar	-1.76 -3.24 – -0.28	3.05 0.90 – 5.19	-0.82 -1.89 – 0.25	1.70 0.21 – 3.18
Blue-collar	-0.10 -1.96 – 1.77	4.69 0.81 – 8.58	0.45 -0.91 – 1.81	4.70 2.01 – 7.40

Residential education level (vs. high)

Medium-high	0.59 -0.80 – 1.98	-0.88 -3.07 – 1.30	0.14 -0.85 – 1.14	0.19 -1.32 – 1.70
Medium-low	0.81 -0.69 – 2.31	-2.91 -5.22 – -0.61	0.89 -0.18 – 1.97	1.14 -0.44 – 2.73
Low	0.02 -1.55 – 1.60	-1.73 -4.16 – 0.69	1.13 0.01 – 2.26	2.28 0.59 – 3.97

Note. CI, confidence interval.

Additional file 6 Associations between working in the secondary rather than in the tertiary sector and cardiovascular risk factors estimated from multilevel regression models among men and women*

Cardiovascular risk factors	β (95% CI)	
	Men	Women
BMI	0.14 -0.18 – 0.47	-0.50 -1.38 – 0.37
Waist circumference	0.23 -0.72 – 1.18	-0.26 -2.30 – 1.79
SBP	0.66 -0.68 – 2.01	0.84 -2.40 – 4.08
DBP	0.77 -0.14 – 1.68	0.14 -1.80 – 2.10
Pulse pressure	-0.12 -0.99 – 0.76	0.69 -1.49 – 2.88
Total cholesterol	3.65 -0.08 – 7.39	-1.48 -8.38 – 5.42
Glycemia	-0.29 -1.58 – 1.00	0.36 -2.26 – 2.99
HDL cholesterol	0.84 -0.27 – 1.95	-1.56 -7.70 – 4.58
LDL cholesterol	2.63 -0.63 – 5.91	-1.76 -7.91 – 4.39
Resting heart rate	-0.20 -1.13 – 0.73	-0.64 -2.47 – 1.18

*Model were adjusted for age, education, income, perceived financial strain, occupational status, and neighborhood level of education. Models for SBP, DBP, and pulse pressure were further adjusted for antihypertensive medication use.

Note. CI, confidence interval; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

CHAPITRE V - Les biais d'attrition dans la relation entre le niveau d'éducation et le changement d'IMC : étude longitudinale et étude de simulation

V.1. Présentation de l'article

A l'aide d'une étude longitudinale, dans ce chapitre, nous examinons, dans un premier temps, dans quelle mesure le niveau d'éducation individuel influe sur l'évolution des paramètres anthropométriques et la présence de biais d'attrition dans les données observées. Dans un second temps, par une étude de simulation permettant d'amplifier le phénomène d'attrition, nous évaluons le biais potentiel sur les résultats selon différents mécanismes d'attrition et différentes méthodes statistiques prenant en compte ces données manquantes.

En effet, de nombreuses études ont montré une association entre un faible statut socio-économique individuel et un poids corporel élevé. Ces résultats sont basés essentiellement sur des études transversales. De plus, une large majorité des études longitudinales n'a pas abordé la question du biais potentiel dû à l'attrition sélective. Dans les études épidémiologiques, les données manquantes sont fréquentes, inévitables peuvent compromettre la validité des résultats. Le biais introduit par les données manquantes dépend directement de la cause de ces données manquantes. Les données manquantes sont catégorisées selon 3 types de mécanisme : 1) « Missing Completely At Random » (MCAR) : qui signifie que les données manquantes ne sont pas différentes des données non manquantes. 2) « Missing At Random » (MAR) : dans ce cas, les

différences systématiques entre les valeurs manquantes et les valeurs observées peuvent être expliquées par des différences dans les données observées. 3) « Missing Not At Random » (MNAR) : les données manquantes dépendent d'un événement que le chercheur n'a pas mesuré. Dans cette catégorisation, le mécanisme MNAR de données manquantes est aussi appelé « inaccessible » tandis que MCAR et MAR sont appelés « accessibles » car les causes des données manquantes peuvent être prises en compte.

Après l'analyse de l'effet du niveau d'éducation sur le changement d'IMC dans les données observées, notre objectif était, à l'aide d'une étude de simulation basée sur les données empiriques, d'étudier l'efficacité de trois approches statistiques communément utilisées pour traiter les données manquantes à savoir l'analyse des cas complet (Complete Case Analyses, CCA), la pondération par l'inverse de probabilité (Inverse Probability Weighting, IPW) et l'imputation multiple (Multiple Imputation, MI). L'étude de simulation avait pour but d'augmenter le biais d'attrition selon sept scénarios d'attrition d'intensité différente afin d'évaluer l'importance du biais introduit dans l'association entre le niveau d'éducation individuel et le changement d'IMC, en suivant les mécanismes de données manquantes MAR et MNAR.

Pour répondre à cet objectif, les relations entre le niveau d'éducation individuel et le changement d'IMC entre la vague 1 et la vague 2 ont été estimées à partir de notre échantillon sans donnée manquante.

Dans notre stratégie analytique, une étude de simulation excluant les valeurs de la variable réponse (changement d'IMC) pour certaines observations a été réalisée, basée sur les données complètes observées de 3693 individus avec une mesure d'IMC à la vague 1 et à la vague 2 de

l'Étude RECORD. Des bases de données incomplètes ont ainsi été générées. Chaque base de données incomplètes a été aléatoirement dérivée en simulant la valeur de la variable dépendante comme étant manquante suivant deux modèles MAR et un modèle MNAR (voir le Graphe Acyclique Orienté en appendice de l'article 3) en augmentant le niveau d'attrition d'un scénario à l'autre. Pour chaque modèle, 7 scénarios de niveau d'attrition ont été évalués en augmentant à chaque fois la relation entre 2 variables préalablement choisies et le risque d'être manquant au suivi (Odd Ratio augmenté de 1 à 4 par intervalle de 0.5). Afin d'éviter les effets de la variabilité d'échantillonnage, 200 bases de données incomplètes ont été générées pour chaque scénario d'attrition. Les covariables des modèles d'exclusion ont été sélectionnées à partir d'études précédentes qui ont analysé l'impact du statut socio-économique individuel et résidentiel sur le poids corporel. Sur la base de ces observations, deux types de mécanisme MAR ont été créés : le fait que la valeur du changement d'IMC soit manquante dépend 1) à la fois du niveau d'éducation du quartier et du niveau d'éducation individuel et 2) de l'IMC à l'inclusion et du niveau d'éducation individuel. Le troisième modèle est basé sur un mécanisme d'attrition MNAR. Dans ce cas, le fait que le changement d'IMC soit manquant dépend du niveau d'éducation individuel et du changement d'IMC lui-même.

Des régressions linéaires ont été utilisées pour estimer les associations entre le niveau d'éducation individuel et le changement d'IMC. Toutes les analyses des différents modèles ont été ajustées sur l'âge et le sexe. Tout d'abord, CCA, IPW et MI ont été appliqués à l'échantillon complet (N=7172 pour MI, N=3693 pour IPW et N=3693 pour CCA). Ensuite ces méthodes ont été appliquées aux bases de données simulées par ajout des données manquantes. La médiane des 200 simulations a été utilisée pour définir l'estimation finale des associations et les 2.5^{ème} et

97.5^{ème} percentiles ont été utilisés pour représenter l'incertitude dans ces estimations. Nous ne pouvions pas évaluer exactement la performance des trois approches en relation avec l'attrition observée car la vraie estimation des associations entre le niveau d'éducation individuel et le changement d'IMC n'était pas connu. Néanmoins, en ce qui concerne nos simulations, l'estimation obtenue à partir de la base de données complète (N=3693) a été considérée comme étant la vraie estimation dans le but d'évaluer les méthodes.

En ce qui concerne les résultats pour le premier mécanisme d'attrition, résultant de l'effet du niveau d'éducation individuel et de l'éducation du quartier sur la participation à la deuxième vague (modèle MAR), l'association inverse entre le niveau d'éducation individuel et le changement d'IMC a été retrouvée pour toutes les méthodes et tous les scénarios d'attrition. L'augmentation de l'attrition n'entraînait pas une augmentation du biais. L'incertitude dans les estimations a augmenté pour toutes les méthodes d'analyse avec l'augmentation de l'attrition. Le second mécanisme de simulation d'attrition résulte des effets du niveau d'éducation individuel et de l'IMC au recrutement sur la participation à la deuxième vague (modèle MAR). Avec l'augmentation des niveaux d'attrition, le biais introduit par les données manquantes avait tendance à augmenter et ne pouvait pas être corrigé par la méthode d'analyse des cas complets (CCA) ni par l'imputation multiple (MI). La méthode de la pondération par l'inverse de probabilité (IPW) semblait être la seule des trois méthodes d'analyse à pouvoir gérer correctement le biais introduit par les données manquantes MAR. En raison de la perte de données, l'incertitude des estimations augmentait avec le niveau d'attrition pour les trois méthodes d'analyse. Enfin, pour le troisième mécanisme d'attrition, le fait que la valeur du changement de l'IMC soit manquante dépendaient du niveau d'éducation individuel et du

changement d'IMC lui-même (modèle MNAR). On observe que le biais introduit par les valeurs manquantes était plus grand que pour les deux mécanismes d'attrition précédents et augmentait rapidement avec l'augmentation du niveau d'attrition pour chacune des trois méthodes d'analyse (CCA, MI, IPW). Ni IPW ni MI ne permettaient de contrôler le biais d'attrition.

En conclusion, les résultats de cette étude suggèrent que, lorsque les données sont manquantes seulement sur la variable dépendante, MI n'est pas en mesure de corriger le biais introduit par un modèle d'attrition MAR. Ainsi, bien que MI soit la méthode la plus couramment préconisée pour gérer les données manquantes dans les études de cohorte, cette approche ne devrait pas être utilisée dans ces circonstances particulières (variable dépendante manquante sous un modèle MAR) car elle n'est pas plus performante que CCA et peut même être pire. Lorsque les données sont manquantes suivant le mécanisme MNAR, ce qui n'est pas improbable, aucune des méthodes testées dans cet article ne peut corriger le biais induit par l'attrition. Ces résultats sont particulièrement préoccupants, car nous avons constaté que l'ampleur du biais d'attrition augmente beaucoup plus rapidement avec le niveau d'attrition suivant un mécanisme MNAR. Dans l'ensemble, ces résultats soulignent donc la nécessité de consacrer plus d'efforts pendant la conception des études à prévoir une collecte d'informations sur les causes d'abandon afin d'essayer au mieux de connaître le mécanisme d'attrition impliqué.

La suite de ce chapitre présente l'intégralité de l'article qui suit a été soumis à une revue.

V.2. Article intégral

Accounting for attrition bias in the association between education and change in body mass index: a longitudinal simulation study based on the RECORD Cohort Study

Antoine Lewin, MSc^{1,2,*}

*Corresponding author

Email: antoine.lewin@upmc.fr

Ruben Brondeel, MSc^{1,2,3}

Email: ruben.brondeel@upmc.fr

Frédérique Thomas, PhD⁴

Email: thomas@ipc.asso.fr

Basile Chaix, PhD^{1,2}

Email: basile.chaix@upmc.fr

¹*Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1136, Pierre Louis Institute of Epidemiology and Public Health, 75012, Paris, France*

²*Inserm, UMR_S 1136, Pierre Louis Institute of Epidemiology and Public Health, 75012, Paris, France*

³*EHESP School of Public Health, 35000, Rennes, France*

⁴*Centre d'Investigations Préventives et Cliniques, 6 rue La Pérouse, 75116 Paris, France*

Text Word Count: 3992

Abstract Word Count: 376

Tables and Figures: 2 Tables and 3 Figures

Abstract

Introduction: Longitudinal studies have shown that low individual and neighborhood socioeconomic status (SES) are associated with high body weight. However, most of the few longitudinal studies available did not address the issue of the potential bias due to selective attrition. Using empirical data, this article examined the performance of the 3 common approaches to handle missing data by increasingly introducing attrition biases, through a manipulation of the data, in the association between individual education and the change in body mass index (BMI).

Methods: We used longitudinal data from the first and second waves of the French RECORD Cohort Study (N = 3,693 participants, retention rate = 51%). After analyzing attrition bias in the observed data, different scenarios of attrition were simulated by generating additional missing data in BMI at the follow-up under Missing At Random (MAR) and Missing Not At Random (MNAR) mechanisms (with, in this application, the probability of having an observed outcome respectively independent of or dependent on the outcome). Complete Case Analyses (CCA), Inverse Probability Weighting (IPW), and Multiple Imputation (MI) analyses were assessed in their ability to account for attrition bias. In the simulation part, the final estimates were derived from 200 replications of our scenarios of attrition.

Results: When no missing observations were added to those already present in the data, a negative association was found between individual education and the change in BMI, in CCA as well as with IPW and MI (the correction had little impact). When generating additional

observations with unobserved follow-up outcomes under a MAR pattern, the bias increased with the attrition level when the analyses were based on CCA or MI. IPW correctly handled the attrition bias. MI and IPW were not useful to overcome the bias introduced by missing outcome data under MNAR.

Conclusion: MI is the widely advocated method to handle missing data in cohort studies, but when data are missing only for the outcome (loss of follow-up), MI does not do better than CCA, even under MAR. When the data are missing under MNAR, none of the methods examined can fix the attrition bias. These results are consistent with the recommendation that more effort is needed during the design phase to understand the attrition mechanisms by collecting information on the reasons for dropout.

Keywords: Attrition bias; Individual education; Body Mass Index; Inverse Probability Weighting; Multiple Imputation

INTRODUCTION

Many studies have shown that low individual and neighborhood socioeconomic status (SES) are associated with high body weight and fat [1-7]. Those findings are mostly based on cross-sectional studies [3]. Moreover, the large majority of studies with a longitudinal design did not address potential biases due to selective attrition [3, 8]. The potential of missing data to compromise the validity of research results has often been overlooked [9, 10]. The risk of bias due to missing data depends on the cause of missing data. Little and Rubin [9, 11] describe three categories of missing data mechanisms: “Missing Completely At Random” (MCAR), i.e., missing cases are not different than non-missing cases; “Missing At Random” (MAR), i.e., any systematic differences between the missing values and the observed values can be explained by differences in observed data; “Missing Not At Random” (MNAR), i.e., the probability of missingness depends on an event that the researcher has not measured, for example on the true value of the missing data. Unfortunately, because the distribution of missing data is only an assumption, it is impossible to determine with certainty whether the data are MAR or MNAR [12-14].

Three common statistical approaches for analyzing datasets with missing data are: Complete Case Analysis (CCA), Inverse Probability Weighting (IPW), and Multiple Imputation (MI) [15, 16]. The most common approach is CCA [16]. In a review of 262 studies published in 2010 in three leading journals, Eekhout et al. found that 81% of these studies used CCA [8]. The method simply excludes individuals with missing data. CCA is usually the default method in statistical

packages [15]. Inferences based on estimations obtained from CCA are valid when data are MCAR because complete cases are a representative subsample of the sample [9, 17]. However, estimates of such analyses may be biased if the excluded individuals are systematically different from those included (MAR or MNAR) [9, 13, 14, 16].

IPW [18] is one of the methods used to mitigate certain attrition biases. In this method complete cases are weighted by the inverse probability of being a complete case [16]. This method is typically implemented when a MAR pattern is assumed. IPW does not model the distribution of the partially observed variables, but models the determinants of missingness [15, 16]. However, in general, it has been suggested that IPW is less efficient than MI [15].

MI [17, 19] uses data on all subjects (including those with missing data). This method creates a number of imputed datasets by generating multiple imputed values for each missing data. Each imputed dataset is analyzed separately, and their estimates are combined using Rubin's rules (1987) [15, 19]. MI yields correct estimators if the imputation model is correctly specified and the data are MAR [20]. Generally, in the case of missing explanatory covariates, MI is more efficient than CCA and IPW since MI uses a model for the joint distribution of the data [20] and uses the observed data from incomplete cases and potentially also from auxiliary variables. This has led to the wide use of MI [9, 15, 20, 21].

Using empirical data, this article examines the performance of 3 common approaches to handle missing data, first by investigating the observed attrition in our data and second by artificially

introducing increasing selection biases in the association between individual education level and change in BMI following MAR and MNAR missing data mechanisms.

METHODS

Study population

Data from the first and second waves of the RECORD Cohort Study (www.record-study.org) were used for longitudinal analyses. During the first wave, as described elsewhere [22-24], 7,290 participants aged 30-79 years at their inclusion were recruited without *a priori* sampling in 2007-2008 during free preventive medical checkups conducted by the Centre d'Investigation Préventive et Clinique in the Paris metropolitan area. As an *a priori* eligibility criterion, only participants residing in 10 (out of 20) administrative districts of Paris or in 111 other municipalities in the region were selected.

In the first wave, 83% of the eligible participants at the health centers agreed to participate, completed the data collection, and were geocoded using their residential address. During the second wave, 3,746 participants were reexamined between 2011 and 2013. The overall revisit rate at wave two was 51%. The French Data Protection Authority approved the study protocol. After excluding participants with missing values for Body Mass Index (BMI) in the first wave, the sample comprised 7,172 participants (this sample was used to investigate attrition in the observed sample). After excluding participants with missing BMI in both the first and second

waves, the sample size was of 3,693 participants (this sample was used to simulate attrition in our data).

Longitudinal changes in BMI

BMI (kg/m²) [25] was calculated at each wave, using height (measured with a wall mounted stadiometer) and weight (measured with calibrated scales) recorded by a nurse [26]. This allowed us to compute the longitudinal change in BMI from baseline to follow-up.

Individual sociodemographic variables

The following baseline sociodemographic characteristics were included in the analyses: age (in years), sex, and personal education coded into two classes (higher secondary school and above, corresponding to 68% of the sample, or not).

Neighborhood socioeconomic variable

The socioeconomic status of the neighborhood was assessed with the educational level of residents (percentage of residents with >2 University years). The variable was computed within a street network buffer with a radius of 1000m centered on the participants' residences. ArcInfo 10 and its Network analyst were used to derive such buffers based on street network data from the National Geographic Institute. Neighborhood education was used as a continuous variable.

Statistical Analysis

Definition of simulated datasets to investigate attrition bias

After investigating attrition bias in the observed data, we defined additional datasets based on the observed complete cases in the first and second wave of the cohort (n=3,693). These additional datasets were incomplete datasets that were generated from the complete dataset by randomly simulating the outcome value (BMI in the second wave, thus change in BMI) to be missing based on two MAR and one MNAR missing mechanisms (see Directed Acyclic Graph in Appendix 1), and for each mechanism of attrition under seven scenarios of attrition level. The attrition rate was defined on the basis of the risk of having missing data at follow up. For each mechanism of attrition, seven scenarios of attrition level were examined, each time increasing the relation between 2 chosen variables (depending on the mechanism of attrition, see below) and the risk of having missing data in wave 2. The 7 scenarios were defined on the basis of odds ratio [OR] ranging from 1.0 to 4.0 (increasing by 0.5) for these associations with missingness (Table 1, see the footnote for a description of the units of variables). To circumvent random sampling variability, 200 such incomplete datasets were generated for each mechanism and for each scenario of attrition level.

We selected two determinants of attrition for each of the three mechanisms of attrition. Previous studies have shown that low individual education and low neighborhood education were associated with a higher BMI or a higher change in BMI and that a higher BMI at baseline was associated with a higher change in BMI [1-4, 7]. Moreover, as shown with a regression model reported in Appendix 2, these three baseline characteristics were independently associated with the fact of having a follow-up BMI measurement in the RECORD Study: the odds of

participating in wave 2 were higher at high individual education and at high neighborhood education levels and among participants who were not obese at baseline (Appendix 2). Based on these observations, two types of MAR mechanisms were generated: (1) missingness in the outcome was set to be conditional on both individual education level and neighborhood education and (2) missingness was set to be conditional on both individual education and baseline BMI. The probability for the change in BMI to be set to missing (1) among low individual education level and low neighborhood education level participants and (2) among high baseline BMI and low individual education participants increased gradually from one scenario of attrition level to the other.

For the third mechanism of attrition (MNAR), missingness in the outcome depended both on the individual education level and on the outcome itself (change in BMI), with the probability of missing among low individual education and high BMI change participants increasing gradually from one scenario of attrition level to the other.

Analytic methods to handle missing data

The three methods below were first applied to the observed dataset and second to the simulated datasets.

In CCA, only subjects with an observed value for the outcome were included in the analysis.

With IPW, the contribution of the complete cases to the estimation was weighted by the inverse of their probability of being a complete case. We modeled the probability of being followed in

the second wave as a logistic function of the following predictors: age, sex, individual education, and residential neighborhood education for the first mechanism of attrition; age, sex, individual education, and baseline BMI for the second mechanism of attrition; and age, sex, and individual education for the third mechanism of attrition (considering that the change of BMI was inaccessible to the researchers when missing).

With MI, a small number of imputed datasets (3 to 5) is sufficient [11, 13, 27]. We created five imputed datasets based on linear regression with change in BMI as the outcome. The covariates used in these imputation models for the three mechanisms of attrition were the same than those listed above for the models for the probability of follow-up in IPW. Multiple imputation (by chained equations) was carried out in R using the mi package [28].

Analyses

For these three statistical approaches, a linear model was used to estimate the association between individual education and the change in BMI. All the analyses were adjusted for age and gender. First, CCA, IPW, and MI were applied to the observed dataset (N = 7,172 for MI and N = 3,693 for CCA and IPW) (Table 2). Second, these methods were applied to the simulated datasets generated from the database of complete cases (N = 3,693 for MI and N < 3,693 for CCA and IPW). The median of the coefficient over the 200 simulated datasets was used as the final estimate. The uncertainty in this estimate was assessed with the 2.5th and 97.5th percentiles over the simulated datasets.

We could not exactly assess the performance of the three approaches in relation to the observed attrition, because the true estimate of the association between individual education and change in BMI was unknown. Regarding the simulation, the estimate in the sample of 3,693 participants was considered to be the true estimate for the purpose of evaluating the methods. To determine the performance of the three methods, we compared the estimate yielded by each analytical approach with each mechanism of attrition under each scenario of attrition level with the estimate based on the original dataset (N = 3,693). All analyses were conducted using R, version 3.1.1 (<http://www.R-project.org>).

RESULTS

In our sample of 7,172 participants, median BMI at baseline was 25.0 (interdecile range: 20.7, 30.7). In the sample of 3,693 complete cases, the median change in BMI was 0.1 (interdecile range: -1.4, 1.9).

In the CCA (N = 3,693), after adjustment for age and gender, the change in BMI was lower for participants with a high compared to a low educational level [-0.26, 95% confidence interval (CI): -0.36, -0.15]. As shown in Table 2, IPW and MI also yielded a comparable negative association between individual education and change in BMI. The final estimate was relatively similar whatever the correction method (IPW or MI) and whatever the covariates used in correction models (to define inverse probability weights and to perform imputation).

The results of the analyses in the simulated datasets based on three mechanisms of attrition under seven different scenarios of attrition level (OR = 1.0 to 4) are shown in Figures 1 to 3. The three graphs represent, for each mechanism of attrition respectively, the median of the estimated coefficient as well as the 2.5th and 97.5th percentiles as measures of uncertainty (over 200 realizations of each scenario) with each of the three analytic methods. The horizontal thick line represent the coefficient from the analysis without missing data (N = 3,693), assumed to represent the true coefficient of reference for the simulation work.

First attrition mechanism

The first attrition mechanism was grounded on the effect of individual education and neighborhood education on participation in the second wave. Figure 1 shows an inverse association between the individual education level and the change in BMI after adjustment for age and sex, with all methods and under all scenarios of attrition level. Increasing attrition did not result in an increase in bias, as the gap between the true coefficient and the coefficient from the CCA did not widen with increasing attrition. The uncertainty in the estimates increased for all analytic methods with increasing attrition bias.

Second attrition mechanism

The second simulated mechanism of attrition resulted from the effects of individual education and baseline BMI on attrition (MAR pattern). As shown in Figure 2, the gap between the true coefficient and the CCA coefficient increased weakly but regularly (from -0.26 to -0.25) as a

result of the simulated bias. MI could not correct the bias, and was on the opposite associated with a stronger bias than the CCA itself.

Analyses with IPW handled correctly the small bias introduced by the increasing level of attrition. The IPW estimates for the seven scenarios of attrition level were close from the horizontal reference line (coefficient equal to -0.26 whatever the degree of attrition). However, due to the loss of data, the uncertainty in the estimates increased with the level of attrition for all three analytic methods, especially with IPW.

Third attrition mechanism

With the third attrition mechanism, information of BMI change was missing following a MNAR mechanism based on individual education and the change in BMI itself. As shown in Figure 3, the bias with this MNAR mechanism of attrition was greater than for the two previous mechanisms of attrition. For example, the coefficient for the association between individual education and the change in BMI changed from -0.25 to -0.07 under the seven scenarios of attrition level. Neither IPW nor MI could correct this attrition bias. The increase in the uncertainty of the estimates with increasing attrition was smaller than that observed with the first and second mechanisms of attrition.

DISCUSSION

Previous studies have reported that BMI or body fat increased with decreasing individual and neighborhood socioeconomic levels [1, 2, 7, 29]. The majority of those studies were cross-sectional [3]. To our knowledge, the present study is the first to rely on a simulation approach based on longitudinal observed data in order to test the ability of statistical methods to correct an attrition bias due to missing information in the outcome at the follow-up. Different missing data mechanisms (including MAR and MNAR patterns) and various scenarios of attrition level were considered. Based on simulations grounded on observed data, we assessed the performance of three statistical approaches to handle this bias.

Main findings

Observed dataset

Our findings indicate that in the overall sample participants with a high education level had a weaker increase in BMI over the follow-up. Our analyses based on IPW and MI provide indication that MAR mechanisms of attrition based on effects of individual education, neighborhood education, and baseline BMI on participation in the study are unlikely to affect the direction and even the strength of the association. This is likely because there were only weak effects of individual education, neighborhood education, and baseline BMI on participation in the second wave were weak, and weak effects of neighborhood education and baseline BMI on the change in BMI. Unfortunately, we could not determine whether a MNAR pattern of attrition, e.g.,

related to an influence of the magnitude of change in BMI on participation, was present in the data.

Simulation work: first and second attrition mechanisms (MAR)

There was evidence of a modest attrition bias for the coefficient of interest especially for one of the two MAR mechanisms of attrition examined (the one specifying an effect baseline BMI on participation). A likely explanation for this is that the effect of baseline BMI on the change in BMI was stronger than the effect of neighborhood education on the change in BMI involved in the other MAR attrition mechanism examined (the correlation between neighborhood education and change in BMI was 0.06 while the correlation between baseline BMI and change in BMI was 0.09). The comparison of Figure 1 with Figure 2 shows that the attrition bias was in the opposite direction in these two cases, which is attributable to the fact that the covariate that was used as a determinant of the attrition rate was associated in the opposite direction with the change in BMI (see Directed Acyclic Graph in Appendix 1).

It should be noted that the bias was very small even when a strong selective attrition was simulated in the data. These results emphasize that, even in a presence of a very strong selective attrition, the bias for the association of interest will remain of small magnitude if the correlation between the covariates influencing study dropout and the outcome is weak.

In this example with missing data only in the dependent variable, only IPW was able to correct the weak attrition bias. On the opposite, there was evidence that MI could not correct and even

amplified the observed bias, which is consistent with comparable studies [30, 31]. In practice, MI is the most commonly advocated and employed method to handle missing data [9, 15]. Our results show that MI should be used carefully. When data are missing in the dependent variable only and that a MAR pattern applies, MI cannot correct the attrition bias even if the imputation model is correct. Moreover, we show that, in specific case, MI can be worse than a simple CCA. This result is consistent with previous methodological articles. Allison et al. and Von Hippel et al. showed that if there is no missing data in the independent variables and if there are no strongly correlated auxiliary predictors, then there is no additional information in the imputed values. In that case, MI would tend to add needless noise to the estimates [32, 33]. Von Hippel et al. and Kullback et al. showed that observations with an imputed outcome contain no information about the regression of Y on X and so the log-likelihood for these observations is exactly zero [32, 34]. In this case, using MI is equivalent to performing a CCA with listwise deletion of observations with a missing outcome, while the imputation of missing outcome values adds noise to these estimates.

Like our results showed, IPW does not have the same problem. Our simulation study suggests that a potential MAR attrition bias would have been corrected by IPW. IPW is based on a participation model and does not impute values for variables with missing data. Differently, IPW adjusts the contribution of subjects followed until study completion in order to represent subjects who were lost to follow-up [13, 16]. IPW weights complete cases on the basis of a model for the

probability of being a complete case. Therefore, IPW relies only on the information from complete cases in the regression analysis itself [16, 18, 20].

Overall, with the MAR pattern involved, MI was unsuccessful to correct the bias because it relied on the poorly informative knowledge available to predict the level of the outcome based on covariates. On the opposite, IPW was more successful in addressing the bias in the data by relying on the available knowledge on the determinants of missingness to reweight the available observations in the analyses.

Simulation work: third attrition mechanism (MNAR)

With this MNAR missingness mechanism, the magnitude of bias increased rapidly with increasing attrition level. With this MNAR pattern, statistical methods such as IPW and MI were not useful to overcome the bias. The uncertainty in the coefficient of interest increased only slightly with increasing attrition level. Comparable to other studies [31], the increase in confidence intervals was smaller for this MNAR pattern than for the MAR simulated samples. These results are important for the epidemiological practice because MNAR represents a realistic assumption in many cases [35]. Our results emphasize that relying on IPW or MI with an incorrect assumption of MAR mechanism (when MNAR in fact actually applies) results in biased estimates that cannot be corrected with these statistical methods.

The magnitude of bias depended heavily on the missingness mechanism involved. Therefore, since it is not possible to determine whether a MAR or MNAR pattern applies on the basis of the

available data, it is important that researchers devote efforts and resources to describe and investigate the potential causes of the study dropout [36]. An option is to collect data that can be used to explore the possible attrition mechanisms. As several studies have suggested, a safe option is to test the assumption in a sensitivity analysis that the observed attrition arises from a MNAR mechanism [15, 31, 35, 37]. In our study, the hypothesized MNAR mechanism was found to bias the association of interest towards the null; thus, under the hypothesis of a MNAR pattern in our observed data, the true association would be stronger (more negative) than the one that was reported.

Strengths and limitations

Regarding study strengths, first, unlike previous longitudinal studies on individual education and change in BMI, we attempted to address attrition biases, comparing three statistical methods to handle missing data. Second, our study was based on observed longitudinal data from the RECORD Cohort Study. We used the available knowledge on the determinants of attrition in this cohort to create scenarios of attrition corresponding to realistic situations occurring in practice, allowing us to compare statistical methods to handle missing data in “real life scenarios”. Third, in this simulation study, three different (including two MAR and one MNAR) missingness mechanisms of attrition bias were explored. Four, to reduce the impact of sampling variability on the examination of these scenarios, 200 samples were drawn and analyzed for each attrition scenario.

Regarding limitations, a first shortcoming of the work is related to the recruitment of the study sample. Our convenience sample, recruited in preventive healthcare centers, was not representative of the Paris Ile-de-France region [23]. However, a large panel of municipalities from the region was *a priori* selected to ensure the presence in the sample of people from all socioeconomic backgrounds. Second, for the sake of simplicity and pedagogy, a small number of covariates was introduced in our statistical models, leading to a simplified examination of the relationships between the study variables. While introducing other covariates might have improved the model for the weight in IPW and the imputation model in MI in our analysis of the observed data, the reduced number of covariates was not an issue in the simulations since the attrition mechanisms were controlled.

Conclusion

In conclusion, the results of this study suggest that, when data are missing only in the outcome, MI may not be able to correct the bias introduced by a MAR pattern of attrition. Thus, although MI is the most commonly advocated method to handle missing data in cohort studies, it should not be used in these particular circumstances (outcome missingness under a MAR pattern) where it does not do better and can even do worse than CCA. When the data are missing under MNAR, which is probably extremely common, none of the methods tested in our article can fix the attrition bias. This is of particular concern because we found that the magnitude of bias increased

much faster with attrition level under MNAR than under MAR. Overall, these results therefore emphasize the need to devote more effort during the design phase to plan the collection of relevant information on the cause of study dropout, in order to evaluate the attrition mechanism involved.

Acknowledgments

We are grateful to INPES (and Pierre Arwidson) for its continued support since the beginning of the study. We are grateful to Insee, the French National Institute of Statistics and Economic Studies, which provided support for the geocoding of the RECORD participants and allowed us to access to relevant geographical data (with special thanks to Pascale Breuil). We thank Geoconcept for allowing us to access to the Universal Geocoder software. We also thank CNAM-TS and the Caisse Primaire d'Assurance Maladie de Paris (CPAM-P, France) for helping make this study possible.

This work was supported by a doctoral grant of Région Île-de-France (CORDDIM) attributed to Antoine Lewin. The RECORD study is funded by the Institute for Public Health Research (IReSP, Institut de Recherche en Santé Publique); the National Institute for Prevention and Health Education (INPES, Institut National de Prévention et d'Education pour la Santé); the National Institute of Public Health Surveillance (InVS, Institut de Veille Sanitaire); the French Ministries of Research and Health (Epidemiologic Cohorts Grant 2008); the National Health Insurance Office for Salaried Workers (CNAM-TS, Caisse Nationale d'Assurance Maladie des Travailleurs Salariés); the Ile-de-France Regional Health Agency (ARS, Agence Régionale de Santé); the National Research Agency (ANR, Agence Nationale de la Recherche); the City of Paris (Ville de Paris); and the Ile-de-France Youth, Sports, and Social Cohesion Regional Direction (DRJSCS, Direction Régionale de la Jeunesse, des Sports et de la Cohésion Sociale).

REFERENCES

1. Ball K, Crawford D: **Socioeconomic status and weight change in adults: a review**. *Soc Sci Med* 2005, **60**(9):1987-2010.
2. McLaren L: **Socioeconomic status and obesity**. *Epidemiol Rev* 2007, **29**:29-48.
3. Mujahid MS, Diez Roux AV, Borrell LN, Nieto FJ: **Cross-sectional and longitudinal associations of BMI with socioeconomic characteristics**. *Obes Res* 2005, **13**(8):1412-1421.
4. Leal C, Bean K, Thomas F, Chaix B: **Are associations between neighborhood socioeconomic characteristics and body mass index or waist circumference based on model extrapolations?** *Epidemiology* 2011, **22**(5):694-703.
5. Leal C, Chaix B: **The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda**. *Obes Rev* 2011, **12**(3):217-230.
6. Chaix B, Bean K, Daniel M, Zenk SN, Kestens Y, Charreire H, Leal C, Thomas F, Karusisi N, Weber C *et al*: **Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study)**. *PLoS One* 2012, **7**(4):e32908.
7. Lewin A, Pannier B, Meline J, Karusisi N, Thomas F, Chaix B: **Residential neighborhood, geographic work environment, and work economic sector: associations with body fat measured by bioelectrical impedance in the RECORD Study**. *Ann Epidemiol* 2014, **24**(3):180-186.

8. Eekhout I, de Boer RM, Twisk JW, de Vet HC, Heymans MW: **Missing data: a systematic review of how they are reported and handled.** *Epidemiology* 2012, **23**(5):729-732.
9. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR: **Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.** *BMJ* 2009, **338**:b2393.
10. Wood AM, White IR, Thompson SG: **Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals.** *Clin Trials* 2004, **1**(4):368-376.
11. Little RJA, DB R: *Statistical Analysis with Missing Data.* New Jersey: John Wiley & Sons; 1987.
12. Rubin DB: **Inference and Missing Data.** *Biometrika* 1976, **63**(3):581-592.
13. Schafer JL, Graham JW: **Missing data: our view of the state of the art.** *Psychol Methods* 2002, **7**(2):147-177.
14. van der Heijden GJ, Donders AR, Stijnen T, Moons KG: **Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example.** *J Clin Epidemiol* 2006, **59**(10):1102-1109.
15. Bartlett JW, Carpenter JR, Tilling K, Vansteelandt S: **Improving upon the efficiency of complete case analysis when covariates are MNAR.** *Biostatistics* 2014, **15**(4):719-730.
16. Seaman SR, White IR: **Review of inverse probability weighting for dealing with missing data.** *Stat Methods Med Res* 2013, **22**(3):278-295.

17. Little RJA, Rubin DB: *Statistical analysis with missing data. Volume 2nd edition.* Chichester: Wiley; 2002.
18. Hofler M, Pfister H, Lieb R, Wittchen HU: **The use of weights to account for non-response and drop-out.** *Soc Psychiatry Psychiatr Epidemiol* 2005, **40**(4):291-299.
19. Rubin D.B.: *Multiple imputation for non response in surveys.* New York: Wiley; 1987.
20. Seaman SR, White IR, Copas AJ, Li L: **Combining multiple imputation and inverse-probability weighting.** *Biometrics* 2012, **68**(1):129-137.
21. White IR, Carlin JB: **Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values.** *Stat Med* 2010, **29**(28):2920-2931.
22. Chaix B, Bean K, Leal C, Thomas F, Havard S, Evans D, Jégo B, Pannier B: **Individual/neighborhood social factors and blood pressure in the RECORD Cohort Study: which risk factors explain the associations?** *Hypertension* 2010, **55**(3):769-775.
23. Chaix B, Billaudeau N, Thomas F, Havard S, Evans D, Kestens Y, Bean K: **Neighborhood effects on health: correcting bias from neighborhood effects on participation.** *Epidemiology* 2011, **22**(1):18-26.
24. Havard S, Reich BJ, Bean K, Chaix B: **Social inequalities in residential exposure to road traffic noise: An environmental justice analysis based on the RECORD Cohort Study.** *Occup Environ Med* 2011, **68**(5):366-374.
25. WHO: **Obesity: preventing and managing the global epidemic.** Edited by consultation W. Geneva: World Health Organisation; 2000.

26. Thomas F, Bean K, Pannier B, Oppert JM, Guize L, Benetos A: **Cardiovascular mortality in overweight subjects: the key role of associated risk factors.** *Hypertension* 2005, **46**(4):654-659.
27. Schafer JL: **Multiple imputation: a primer.** *Stat Methods Med Res* 1999, **8**(1):3-15.
28. Su Y.S., Gelman A., Hill J., Yajima M.: **Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box.** *Journal of Statistical Software* 2011, **45**(2):6-15.
29. Kershaw KN, Albrecht SS, Carnethon MR: **Racial and ethnic residential segregation, the neighborhood socioeconomic environment, and obesity among Blacks and Mexican Americans.** *Am J Epidemiol* 2013, **177**(4):299-309.
30. Twisk J, de Vente W: **Attrition in longitudinal studies. How to deal with missing data.** *J Clin Epidemiol* 2002, **55**(4):329-337.
31. Kristman VL, Manno M, Cote P: **Methods to account for attrition in longitudinal data: do they work? A simulation study.** *Eur J Epidemiol* 2005, **20**(8):657-662.
32. Hipple PT: **Regression with Missing Y's: An Improved Strategy for Analysing Multiply Imputed Data.** *Sociological Methodology* 2007, **37**:83-117.
33. Allison PD: *Missing data.* Thousand Oaks, CA; 2001.
34. Kullback S: *Information Theory and Statistics.* New York; 1959.
35. Osborne JW: *Best Practices in Data Cleaning. Dealing with Missing or Incomplete Data.* Louisville, KY: Sage Publication, Inc; 2013.
36. Little RJA: **Modeling the drop-out mechanism in repeated-measures studies.** *J Am Stat Assoc* 1995, **90**:1112-1121.

37. Kendra KS, Baojiang C: **Missing Data: A Non-ignorable Issue in Modern Biostatistics.**
Journal of Applied Bioinformatics & Computation Biology 2012, **1**:1.

Table 1. Number of observations with a missing outcome at the follow-up for each mechanism of attrition under each attrition level

Attrition level	First attrition mechanism (MAR, individual education, neighborhood education)	Second attrition mechanism (MAR, individual education, baseline BMI)	Third attrition mechanism (MNAR, individual education, change in BMI)
OR = 1.0*	67	67	67
OR = 1.5	141	141	140
OR = 2.0	248	246	244
OR = 2.5	385	374	370
OR = 3.0	536	520	509
OR = 3.5	686	661	654
OR = 4.0	831	803	795

* The OR refers to a multiplicative coefficient for the odds of having a missing outcome in wave 2 for having a low rather than a high individual education level (binary variable), for a 2-standard deviation decrease in the proportion of high educated residents in the neighborhood, and for a 2-standard deviation increase in baseline BMI (a 2-standard deviation was chosen to mimic the effect of a binary variable with approximately half of the population in each group). The distribution of the standardized baseline BMI and standardized change in BMI were shifted towards the positive, so as to have a minimum value equal to 0, yielding the minimal (and non-negative) probability of attrition when these recoded variables were equal to 0. The distribution of standardized neighborhood education was shifted towards the negative, so as to have a maximum value equal to 0, yielding (due to a negative coefficient in our model) the minimal probability of attrition when this recoded variable was equal to 0 (i.e., when neighborhood education was the highest) and an increase in the odds of attrition with more negative. The model to define the probability of attrition for each participant was a logistic model with an intercept equal to -4.

Table 2. Associations between individual education (high vs. low) and change in BMI (adjusted for age and gender) corrected from the observed attrition bias through IPW and MI, with different covariates entered in the model to define inverse probability weights or to perform the multiple imputation

	Covariates in the weight model or imputation model		
	Age and sex β (95% CI)	Age, sex, and neighborhood education β (95% CI)	Age, sex, and, baseline BMI β (95% CI)
IPW (N = 3,693)	-0.25 (-0.36 ; -0.14)	-0.25 (-0.35 ; -0.14)	-0.25 (-0.36 ; -0.14)
MI (N = 7,172)	-0.24 (-0.33 ; -0.14)	-0.26 (-0.43 ; -0.09)	-0.28 (-0.39 ; -0.15)

β , beta coefficient; CI, confidence interval.

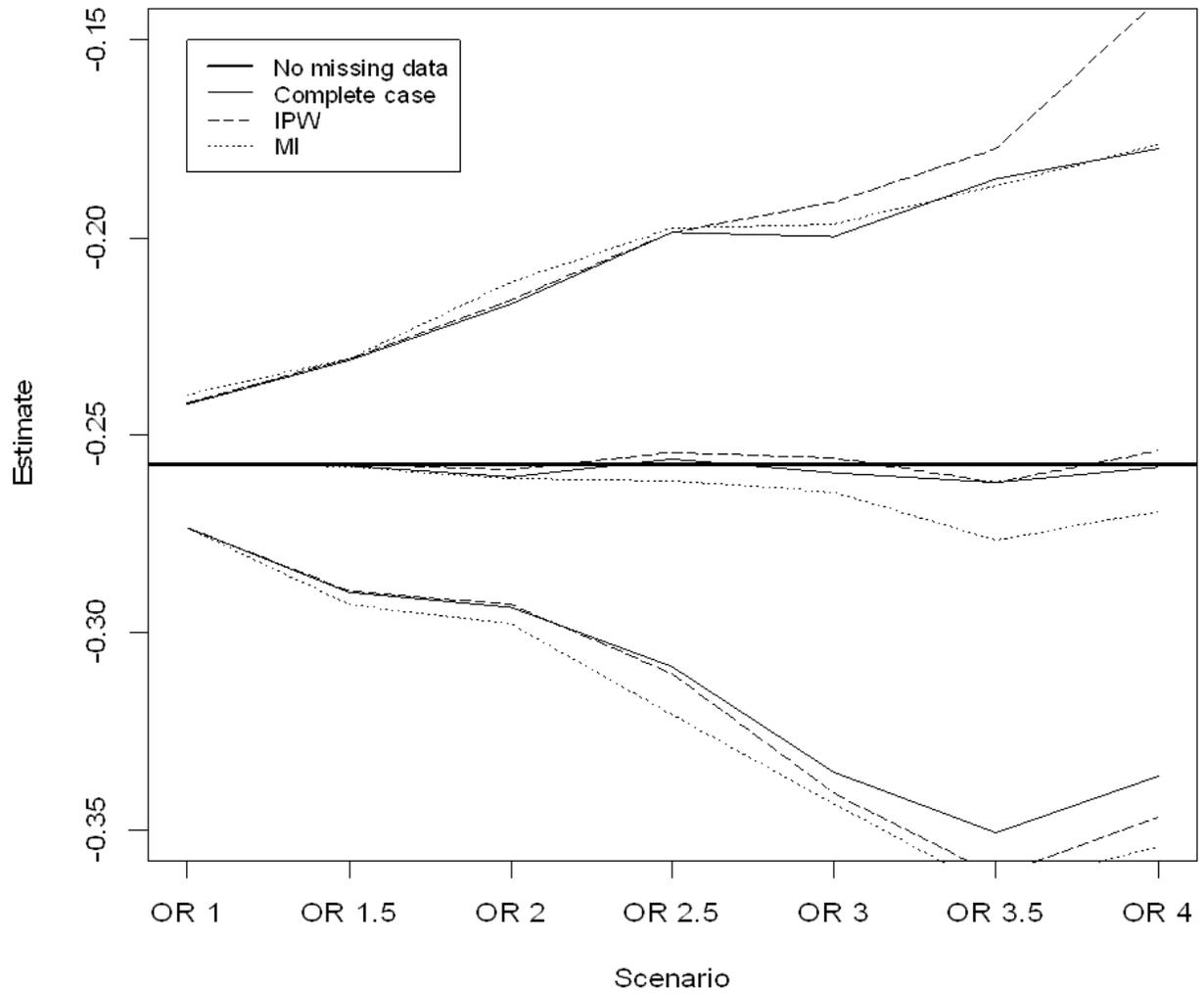


FIGURE 1. Associations between individual education and change in BMI in simulated datasets based on the first attrition mechanism (MAR, individual education, neighborhood education) under seven scenarios of attrition level

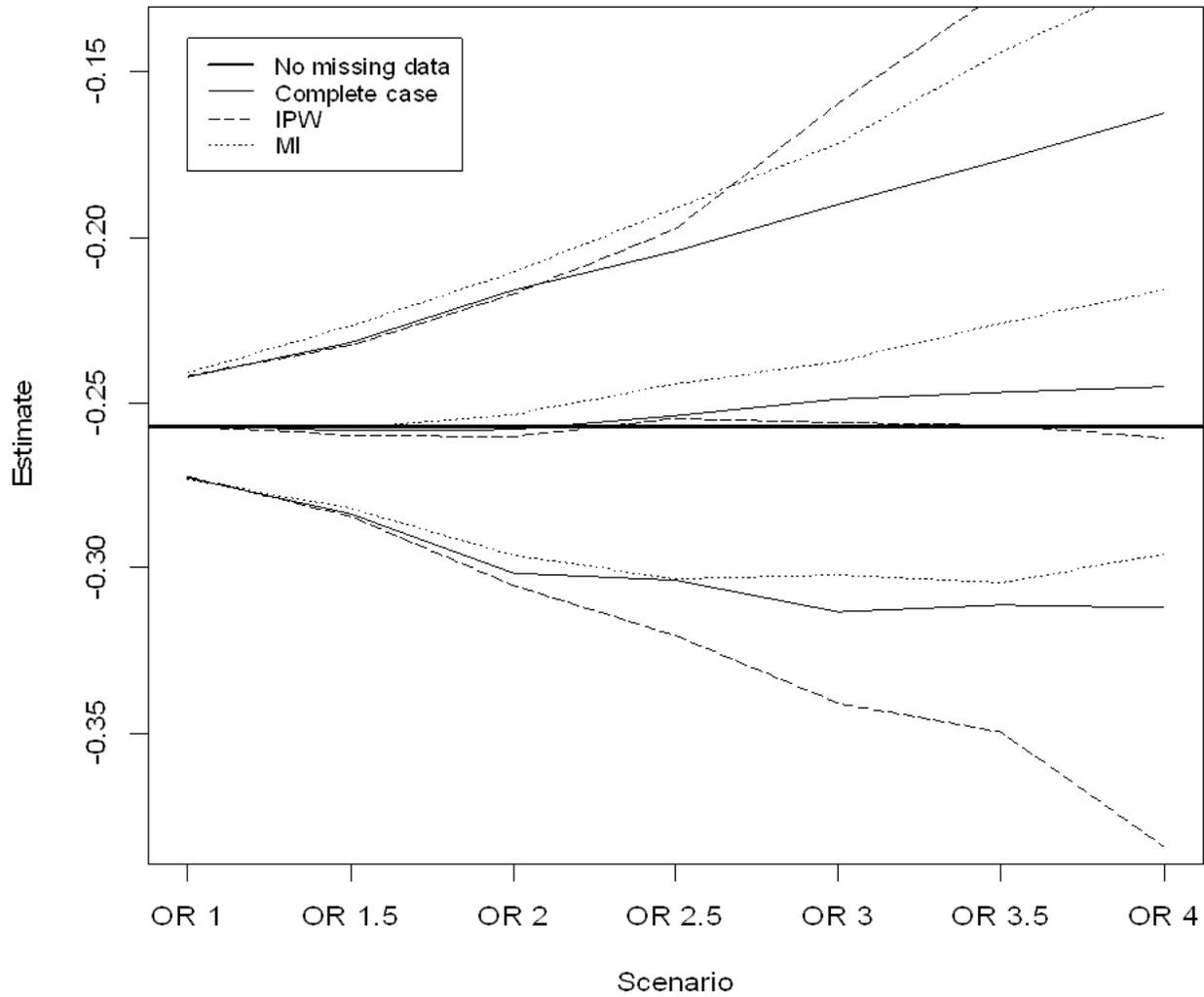


FIGURE 2. Associations between individual education and change in BMI in simulated datasets based on the second attrition mechanism (MAR, individual education, baseline BMI) under seven scenarios of attrition level

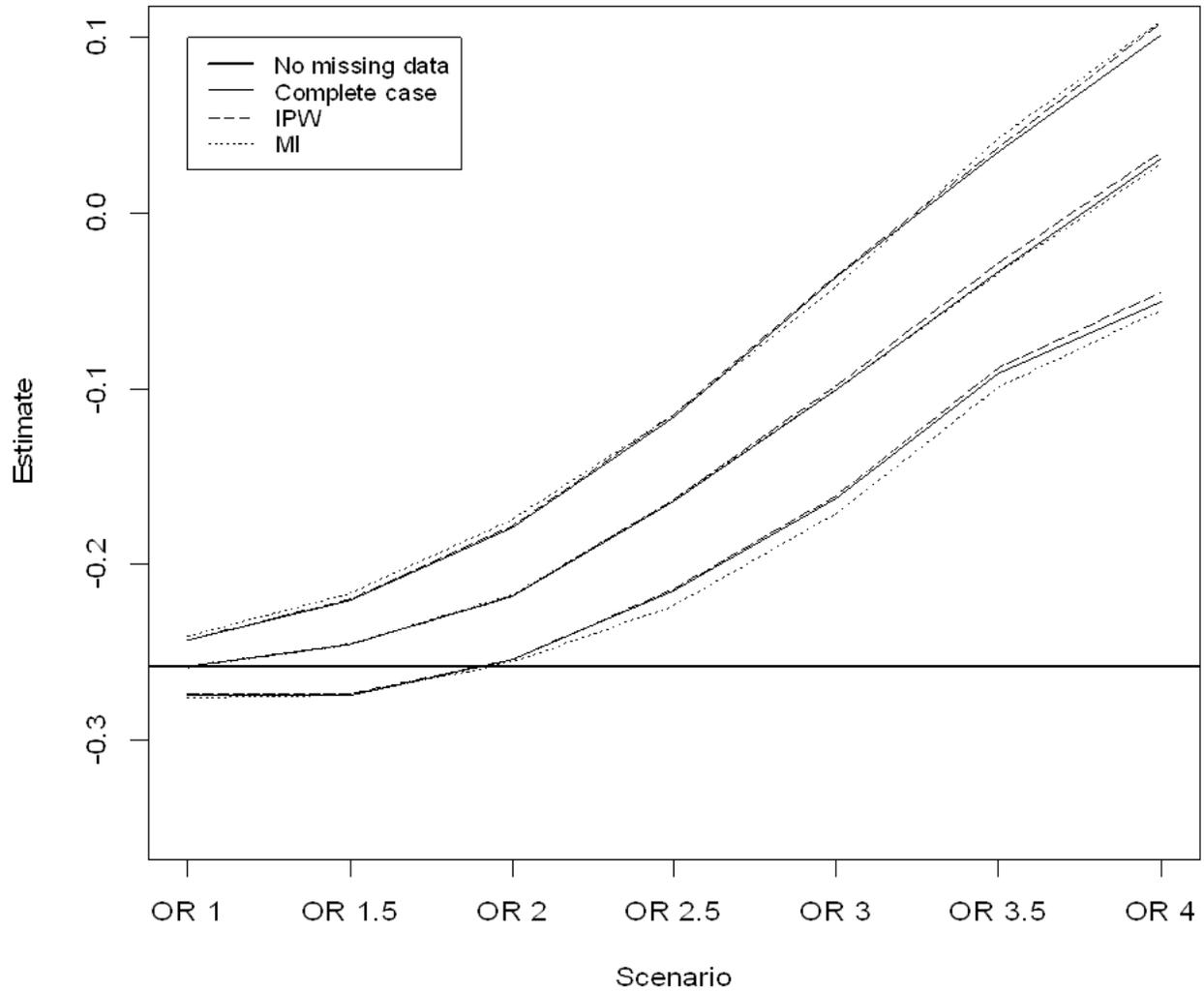


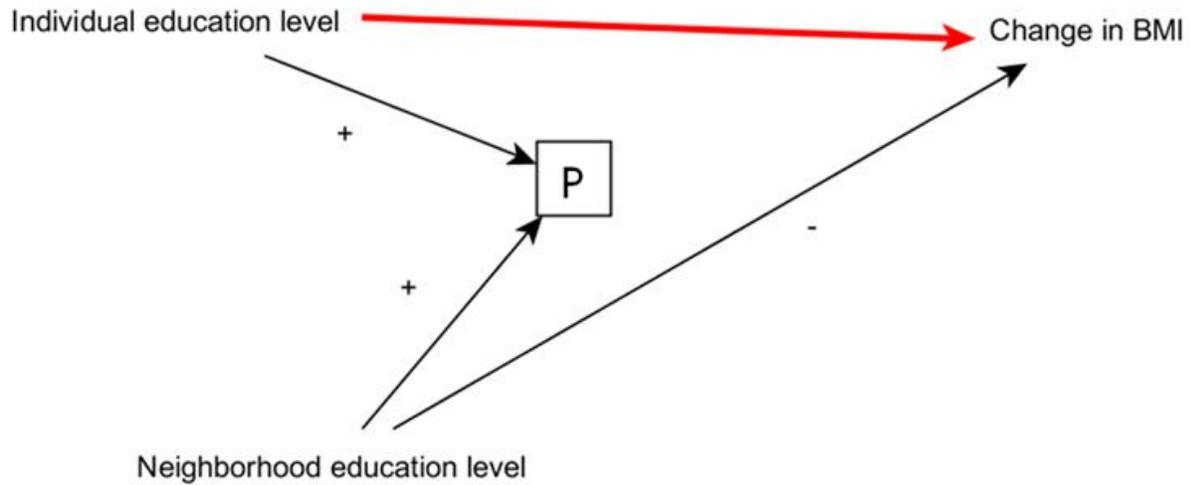
FIGURE 3. Associations between individual education and change in BMI in simulated datasets based on the third attrition mechanism (MNAR, individual education, change in BMI) under seven scenarios of attrition level

Appendix 1

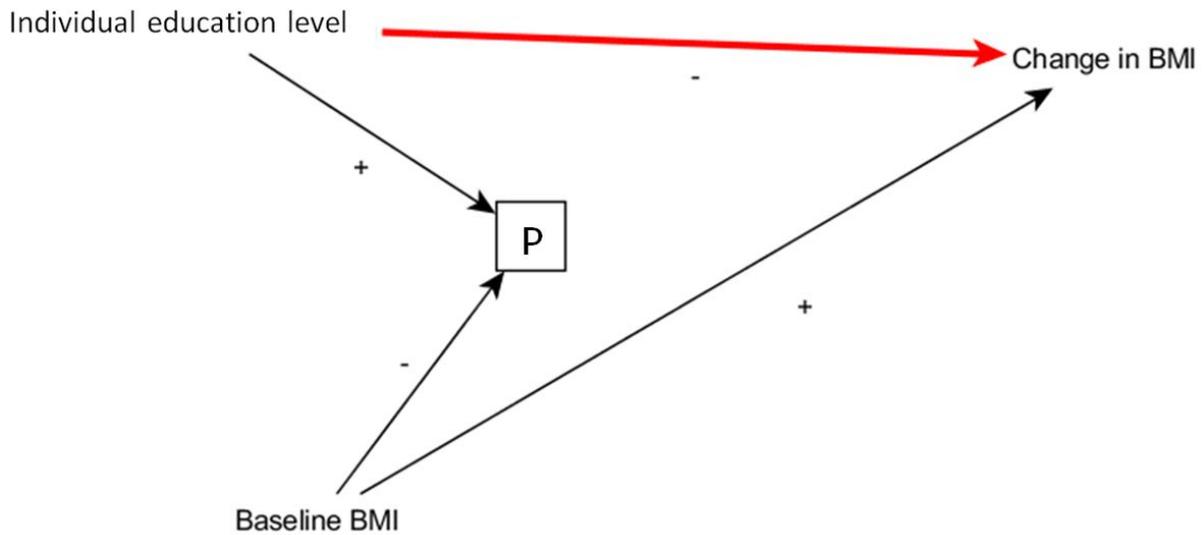
Appendix Figures 1A to 1C represent the different attrition biases that were introduced with the three mechanisms of attrition examined.

Appendix Figures 1A and 1B represent an attrition bias following a Missing At Random (MAR) mechanism on the effect of individual education on change in BMI. If participants drop out from the study, the change in BMI cannot be measured (censored) [Participation (P) = 0]. Participants with a high level of neighborhood education or a low baseline BMI (respectively in the first and second MAR patterns) and participants with high individual level of education are more likely to participate in the second wave. The attrition bias in the estimate of interest is introduced by conditioning on participation (P = 1), which is a common effect of the variable of interest (individual education) and a covariate that also influences the change in BMI [neighborhood education (Appendix Figure 1A) or baseline BMI (Appendix Figure 1B)].

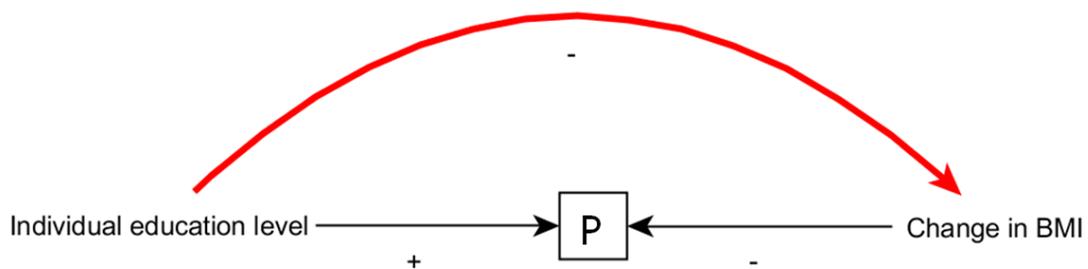
Figure 1C represents the attrition bias following a Missing Not At Random (MNAR) mechanism on the effect of individual education on change of BMI. Individual education and change in BMI are causes of non-participation in the second wave (P = 0). Participants with a high individual education level and a lower increase in BMI are more likely to participate. The MNAR attrition bias in the estimate of interest is introduced by conditioning on participation (P = 1), which is caused by both the explanatory variable of interest and the outcome.



Appendix Figure 1A. First mechanism of attrition examined, related to a MAR pattern based on individual education and neighborhood education effects on participation



Appendix Figure 1B. Second mechanism of attrition examined, related to a MAR pattern based on individual education and baseline BMI effects on participation



Appendix Figure 1C. Third mechanism of attrition examined, related to a MNAR pattern based on individual education and change in BMI effects on participation

Appendix 2

Appendix Table 1. Logistic model for the determinants of participation in wave 2 (N = 7,172)

	Participation in wave 2 β (SD)
Low individual education (vs. high)	-0.069 (0.013)***
Body Mass Index in wave 1	
Obese (vs. normal weight)	-0.067 (0.018)**
Neighborhood proportion of high educated*	0.098 (0.043)*

*While the two other explanatory variables were coded as categorical variables, neighborhood education was expressed as a continuous variable comprised between 0 and 1.

CHAPITRE VI - DISCUSSION

Dans ce chapitre, nous discutons du travail accompli, en passant en revue les principaux résultats obtenus et les principales limites rencontrées dans le cadre de cette thèse. De façon plus large, nous discutons ces résultats dans une perspective globale de Santé publique.

VI.1. Les principaux résultats

Au vu des principaux résultats de cette thèse, nous avons pu constater que dans l'Étude de cohorte RECORD, il existe des associations entre l'environnement du quartier résidentiel, les secteurs d'activité économique, la distance domicile-travail et la masse grasse et/ou les facteurs de risque cardiovasculaire. Nous avons aussi pu constater qu'il existait une association entre le niveau d'éducation et l'évolution de l'IMC.

Les résultats de notre première étude empirique montrent que le quartier résidentiel, les secteurs d'activité économique et la distance entre le domicile et le travail étaient indépendamment associés à l'Indice de masse grasse (IMG) et au pourcentage de masse grasse (%MG) après ajustement sur les variables socio-économiques individuelles. Ces résultats sont cohérent avec les nombreuses études antérieures qui ont rapporté : 1) une augmentation de l'IMC ou du tour de taille avec la diminution de niveau socio-économique quartier de résidence (184) ;

2) des prévalences du surpoids et de l'obésité évalués par l'IMC plus élevées chez les individus travaillant dans le commerce, les industries, ou le transport ainsi que chez les juristes et les cadres supérieurs même après ajustement sur l'âge, le sexe et l'éducation (185, 186) ; 3) des associations inverses entre le fait de marcher ou de faire du vélo pour se rendre au travail et le surpoids ou l'obésité (121).

En effet, au travers de notre première étude, nous avons pu observer que les hommes ainsi que les femmes vivant dans les quartiers à faible niveau d'éducation avaient des niveaux de graisse corporelle (IMG et %MG) plus élevés. De plus, après ajustement sur le niveau d'éducation du quartier de résidence, une faible densité de population dans les quartiers résidentiels était associée à un IMG et un %MG plus élevés chez les hommes ceci peut être expliqué par le fait qu'une faible densité est un facteur contextuel qui influence la marche à pied par le développement d'environnement plus propice à la marche. En effet, des indicateurs de densité tels que la densité de population ou la densité de destination sont positivement associés à la marche (187, 188). En ce qui concerne les associations entre les secteurs d'activité économique et la masse grasse, après ajustement sur les variables socio-économiques individuelles et les variables résidentielles, l'IMG et/ou le %MG étaient plus élevés chez les hommes travaillant dans le secteur de la construction ; des transports et des communications ; de l'industrie manufacturière ; du commerce et de la réparation de véhicules ; de l'immobilier et de la location et des services aux entreprises que chez les hommes travaillant dans le secteur de l'éducation. Différemment, chez les femmes, celles travaillant dans le secteur du travail social et de la santé avaient un IMG plus élevé que celles du secteur des transports et des communications. Malgré

qu'aucune étude publiée n'ait évalué les différences de masse grasse entre les secteurs d'activité économiques, des pistes d'explications peuvent être avancées. Proper et al, 2006, propose qu'une des causes des disparités d'IMC entre les secteurs d'activité économique pourrait se trouver dans les différences d'activité physique liée au travail (189). Aucune association n'a été trouvée entre les secteurs économiques de travail et le %MG chez les femmes, ce qui pourrait être attribuable, entre autres, au fait que l'IMG évalue l'obésité de façon relativement précise, mais que le %MG, tel l'IMC (mais moins que lui) sous-estime l'obésité (190). Par ailleurs, les associations persistaient entre les secteurs d'activité économique et l'IMG ou le %MG après ajustement sur l'IMC chez les hommes. Nos résultats suggèrent que la composition corporelle évaluée avec l'IMG et le %MG nous permet de recueillir des informations sur les déterminants de la graisse corporelle qui ne pouvait pas être captées par l'IMC car cet indicateur ne permet pas d'évaluer pleinement les effets sociaux et contextuels sur la graisse corporelle. Ces résultats confirment ceux de l'étude de Rosemond et al. (191) qui a constaté que l'obésité (évalué par IMC) et l'obésité centrale (évaluée par le rapport taille-hanche) étaient différemment liées aux facteurs professionnels et sociaux. Malgré tout, comme le montre une étude de 1997, qui évalue le lien entre la masse grasse mais aussi la masse maigre et la mortalité, le fait que nos associations persistent après ajustement sur l'IMC pourrait également être attribuable à la masse maigre (192). Finalement, en ce qui concerne la distance domicile-travail, les résultats de notre étude montrent qu'une longue distance domicile-travail était associée à un taux de masse grasse plus élevé, mais seulement chez les femmes et uniquement avec l'IMG (qui peut à nouveau mieux saisir les

différences véritables en matière de masse grasse). Cette constatation est cohérente avec l'utilisation des modes de transport motorisés pour de longues distances domicile-travail.

Après avoir montré que les inégalités liées aux caractéristiques environnementales et aux secteurs d'activité économique peuvent être un déterminant dans la prévalence d'obésité, notre objectif suivant a été d'étudier les relations entre les secteurs d'activité économique et plupart des facteurs de risque cardiovasculaire. Afin d'évaluer la contribution de ces secteurs professionnels dans le cumul de facteurs de risque cardiovasculaire, nos analyses ont été ajustées sur les caractéristiques socio-économiques individuelles et résidentielles. Nous avons trouvé des associations entre les secteurs d'activité économique et certains facteurs de risque cardiovasculaire. Les motifs de ces disparités étaient différents chez les hommes et les femmes suivant le facteur de risque étudié. Certaines des disparités entre les secteurs économiques de travail peuvent être dues à des différences dans l'apport énergétique total ou la dépense d'énergie. Les résultats d'une étude passée montrent que l'apport énergétique, comme une explication possible aux disparités observées, peut varier dans une large mesure entre les secteurs d'activité économique (193). Par ailleurs, du fait que le travail est un contributeur majeur à l'activité physique quotidienne totale (189), une autre hypothèse peut être que les différences dans les niveaux d'activité physique entre les secteurs d'activité économique contribuent aux disparités observées dans les facteurs de risque cardiovasculaire. Cependant, nous n'avons pas validé empiriquement ces hypothèses, puisque par exemple, les hommes travaillant dans le secteur de la construction avaient l'IMC et le tour de taille les plus élevés. Ces résultats confirment ceux d'une étude précédente qui a trouvé que, contrairement à l'activité physique de loisir, l'activité physique

au travail n'avait pas d'effet protecteur en ce qui concerne les facteurs de risque cardiovasculaire, y compris ceux liés à l'excès de poids (194).

Finalement, afin d'étudier les déterminants de l'évolution des paramètres anthropométriques chez les participants de l'Étude RECORD et de progresser dans l'identification d'effets causaux, nous avons analysé les données de manière longitudinale. En effet, de nombreuses études ont rapporté que l'IMC ou la graisse corporelle augmentaient avec la baisse du niveau socio-économique individuel et du quartier (80, 81, 184). Cependant, la grande majorité de ces études sont transversales (82). De plus, au vu des données manquantes présentes dans notre méthode, nous avons testé la capacité de méthodes statistiques à corriger le biais d'attrition. Nous avons simulé divers mécanismes de données manquantes (deux mécanismes MAR et un mécanisme MNAR) et différents scénarios de niveau d'attrition. Les résultats concernant les données observées (c'est à dire sans simulation de données manquantes) montrent que l'IMC augmentait plus chez les participants avec un niveau d'éducation individuel faible de façon quasi équivalente pour les 3 méthodes analytiques. En effet, depuis de nombreuses années l'effet de l'éducation sur la prise de poids est bien connu (153). Un faible niveau d'éducation peut être en relation avec une faible connaissance en matière de santé ce qui aura pour effet une plus grande difficulté à acquérir des aliments plus couteux et moins énergétiques tels que les fruits, les légumes, les céréales, etc. De plus, ces personnes auront tendance à avoir moins de temps de loisir qu'ils peuvent consacrer à un exercice récréatif (15, 153). Enfin, les personnes avec un niveau d'éducation plus faibles auront aussi tendance à moins pouvoir résister aux

environnements obésogènes car ils ont moins de flexibilité dans leurs choix alimentaires et dans leurs activités de loisir (195).

En ce qui concerne le mécanisme MAR (modèles 1 et 2), les résultats de notre étude de simulation ont montré peu de différence entre les estimations à partir des données complètes et celles avec de faibles niveaux d'attrition. Avec l'augmentation des niveaux d'attrition, le biais introduit par les données manquantes avait tendance à augmenter et ne pouvait pas être corrigé par la méthode d'analyse des cas complets (CCA) et l'imputation multiple (MI). La méthode de la pondération par l'inverse de probabilité (IPW) était la seule des trois méthodes statistiques à pouvoir gérer le biais introduit par les données manquantes correctement. Comme dans d'autres études, (196, 197) ces résultats suggèrent que, lorsque les données manquantes sont seulement sur la variable dépendante, ni MI ni CCA ne peuvent être utiles pour corriger le biais d'attrition. En ce qui concerne le biais introduit avec le mécanisme MNAR, on observe que ce biais augmentait rapidement avec le niveau d'attrition pour chacune des trois méthodes statistiques (CCA, MI, IPW). Ces résultats sont importants en termes de méthodes d'analyse épidémiologique car la littérature montre que l'hypothèse la plus raisonnable et réaliste dans les mécanismes des données manquantes est MNAR (198) et nos résultats montrent qu'aucune de ces 3 méthodes ne peut contourner le biais introduit par MNAR et que ces approches donnent des estimations biaisées des associations étudiées. C'est pourquoi il est important de décrire les causes potentielles d'abandon (199) en amont des analyses afin de décrire le mécanisme de données manquantes. Nos résultats ont montré que MI doit être utilisé avec précaution. Lorsque les données sont manquantes uniquement sur la variable dépendante, alors MI ne peut pas

corriger le biais d'attrition même si le modèle d'imputation est correctement spécifié et si le mécanisme de données manquantes est MAR. Ce résultat est cohérent avec des articles méthodologiques précédents. En effet, Alison et al. et Hippel et al. ont montré que si les données manquantes sont essentiellement sur la variable dépendante et qu'il n'y a pas de prédicteurs auxiliaires fortement corrélés à celle-ci dans le modèle d'imputation, alors aucune information supplémentaire n'est apportée dans les valeurs imputées sous MI (200, 201). En revanche, IPW n'a pas ce problème car IPW est basé sur un modèle de participation et n'impute pas de nouvelles valeurs pour les données manquantes (dépendantes ou indépendantes). En effet, IPW ajuste la contribution des sujets présents dans l'étude pour représenter les sujets perdus de vue (181, 202). IPW pondère les cas complets en s'appuyant sur un modèle de la probabilité d'être un cas complet basé sur des valeurs observées, donc IPW n'utilise que des informations sur les cas complets (181, 203, 204).

En résumé, ces résultats enrichissent nos connaissances sur les déterminants individuels et contextuels des problèmes d'excès de poids et de graisse corporelle dans la région Ile de France.

VI.2. Limites

Malgré les précautions méthodologiques que nous avons prises et l'atout considérable que représentent les données de la cohorte RECORD, certaines limites demeurent aux différentes étapes de ce travail.

La première limite concerne la représentativité de l'échantillon sélectionné dans la base de données de l'Étude RECORD. Cet échantillon ne représente pas directement la population d'Ile-de-France. Cette non représentativité est explicable par la méthode de recrutement des participants au Centre IPC sans randomisation préalable en population générale ; par les exclusions résultant de l'incapacité de certaines personnes à répondre aux questionnaires ; et par le refus de participation à l'étude (205). Cependant, afin de garantir un échantillon représentant un large panel de profils socio-économiques, un grand nombre de municipalités de la région Ile-de-France ont été sélectionnées *a priori*.

Concernant les analyses, des biais de mesures associés aux mesures d'exposition environnementale dues à la mobilité des individus au cours du temps peuvent exister. En effet, les analyses considèrent essentiellement l'environnement dans lequel l'individu résidait ou travaillait au moment de l'étude, sans s'intéresser aux caractéristiques des environnements précédemment fréquentés. De ce fait, les caractéristiques environnementales prise en compte peuvent renvoyer à des périodes d'exposition plus ou moins longue selon les individus. De plus, la notion de quartier n'a pas été définie géographiquement par l'individu lui-même.

Une autre limite vient du fait le lieu de travail et le secteur d'activité ont été récupérés par appariement d'informations administratives. Nous ne pouvons pas être certains que ces individus étaient toujours employés à ce lieu de travail ou dans ce secteur d'activité économique au moment du recrutement. En ce qui concerne les secteurs d'activité économique, cette limite est probablement mineure car il peut être attendu qu'un participant qui change de travail restera dans

le même secteur d'activité économique. De plus, même si un individu avait changé de secteur d'activité économique ou s'était arrêté de travailler au moment de l'étude, il aurait tout de même été exposé au secteur d'activité économique pris en compte l'année précédant son recrutement dans l'étude.

Finalement, une limite non négligeable qui concerne non seulement ce travail de thèse mais aussi la plupart des études épidémiologiques qui évaluent le lien entre l'environnement et l'obésité porte sur le phénomène de migration sélective (97). Ce phénomène se définit par le fait que des personnes avec des caractéristiques individuelles spécifiques emménagent (volontairement ou non) dans des quartiers qui possèdent aussi des caractéristiques spécifiques (95). Du fait du manque de données sur les déménagements des participants, nous ne pouvons pas contourner cette limite, si ce n'est par un bon ajustement des modèles au niveau individuel.

VI. 3. Intérêts des travaux réalisés

La gravité de l'épidémie mondiale d'obésité ainsi que ses coûts économiques, sociaux et sanitaires sont communément établis. En effet, en 2008, 1,4 milliards d'adultes étaient en surpoids et plus de 500 millions obèses. La prévalence d'obésité a doublé de 1980 à 2008 et l'OMS estime à 3,3 milliards le nombre de personnes en surpoids d'ici 2030. L'obésité touche également les individus dès le plus jeune âge : en 2013, 42 millions d'enfants de moins de 5 ans dans le monde étaient en surpoids (86). Au niveau mondial, toutes causes confondues, chaque année 3.4 millions d'individus meurent chaque année. Parmi ces décès, 2,8 millions au moins

sont attribuables aux conséquences du surpoids ou de l'obésité (diabète, cardiopathies ischémiques, cancers), ce qui classe l'obésité comme premier facteur de mortalité mondiale (86). Mais l'obésité n'est pas qu'un problème de Santé publique : en 2014, le coût économique lié à l'épidémie d'obésité dans le monde était évalué à plus de 2000 milliards de dollars par an, ce qui la place largement devant l'alcoolisme ou les accidents de la route (206). D'où l'intérêt d'étudier, à l'aide de nouveaux indicateurs, les relations entre les facteurs individuels, professionnels, environnementaux, etc. et l'obésité et les facteurs de risque cardiovasculaire afin de définir les profils multiniveaux de l'excès de poids et de risque cardiovasculaire.

En lien avec les perspectives de l'OMS qui sont une approche d'étude multisectorielle, multidisciplinaire et populationnelle de l'obésité (86), les résultats de ce travail de thèse, pourraient contribuer à l'élaboration de nouvelles stratégies de politique publique. Par exemple, décrire les disparités de prévalence de facteurs de risque cardiovasculaire ou de l'obésité en fonction du secteur d'activité économique, de la profession et des caractéristiques environnementales du lieu de travail peut donner lieu à des stratégies d'intervention potentiellement efficaces notamment via la médecine du travail. En effet, la médecine du travail constitue un pivot d'intervention qui peut permettre des approches spécifiques par secteur d'activité. Il semble alors envisageable de cibler certaines catégories socio-professionnelles de certains secteurs d'activité via la médecine du travail. D'autre part, nos études tenant compte des environnements géographiques résidentiels et professionnels permettront d'affiner le ciblage territorial des interventions visant à lutter contre l'épidémie d'obésité. De plus, d'un point de vue méthodologique, parmi les différents outils utilisés cette thèse montre les limites des méthodes

statistiques qui permettent de prendre en compte les données manquantes lors d'études longitudinales. Les analyses conduites durant cette thèse pourront aider les futures études longitudinales à prendre en compte les risques de biais systématiques inhérents aux données manquantes lors des analyses.

Au total, nos résultats mettent en évidence (à l'aide d'indicateurs de composition corporelle et anthropométriques) des associations entre les caractéristiques socio-économiques individuelles, les environnements géographiques résidentiels, et les secteurs d'activité économique et l'obésité ainsi que les facteurs de risque cardiovasculaire. D'autre part, ils soulignent l'importance de prendre en considération les biais d'attrition pouvant intervenir dans les études longitudinales. Ces réflexions sont des outils nouveaux pour les débats et les actions de Santé publiques actuels.

CONCLUSION ET PERSPECTIVES

Ce travail de thèse visait à apporter de nouveaux éléments dans la compréhension des voies par lesquelles les environnements géographiques de vie et le milieu professionnel pourraient influencer la prise de poids et le risque cardiovasculaire. Ce travail poursuivait également différents objectifs méthodologiques, notamment d'évaluer le lien entre les valeurs manquantes et les biais d'attrition dans les associations longitudinales.

Afin de répondre à notre question de recherche, nous avons tout d'abord cherché à évaluer les associations entre d'une part les facteurs individuels sociodémographiques, les caractéristiques des quartiers résidentiels et professionnels, et les secteurs d'activité économique des participants et d'autre part, la masse grasse à l'aide d'indicateurs d'adiposité. Ce travail nous a permis de mettre en évidence comme attendu que l'environnement socio-économique ainsi que le secteur d'activité économique influençaient la prise de poids. De plus, ce travail a permis de démontrer que l'IMC, à la différence d'indicateurs d'adiposité, ne permettait pas de saisir pleinement les effets des caractéristiques environnementales et des secteurs d'activité économique sur l'accumulation de graisse corporelle. Une hypothèse à tester pour l'avenir pourrait être d'examiner si l'environnement géographique autour du lieu de travail influe sur divers comportements de santé (par exemple alimentation) au-delà du l'environnement de résidence. Ces résultats montrent également l'intérêt qu'il pourrait y avoir à réaliser à partir de

données française une étude des déterminants contextuels de l'évolution de la masse grasse avec une période de suivi plus longue.

Ensuite, au vu des inégalités observées en matière d'accumulation de graisse corporelle selon les secteurs d'activité économique, notre objectif a été de comparer les disparités entre secteur d'activité pour un large panel de facteurs de risque cardiovasculaire. Notre objectif était de tester l'hypothèse selon laquelle un secteur d'activité économique pourrait cumuler un sur-risque pour différents facteurs de risque cardiovasculaire. Ce travail révèle des disparités face aux différents facteurs de risque cardiovasculaire selon ces secteurs d'activité chez les hommes et chez les femmes. Suite à ce travail, il sera intéressant d'étudier les inégalités professionnelles face au risque cardiovasculaire en testant des modèles qui tiennent compte à la fois des secteurs d'activité économique et d'un ensemble de variables reflétant les conditions de travail.

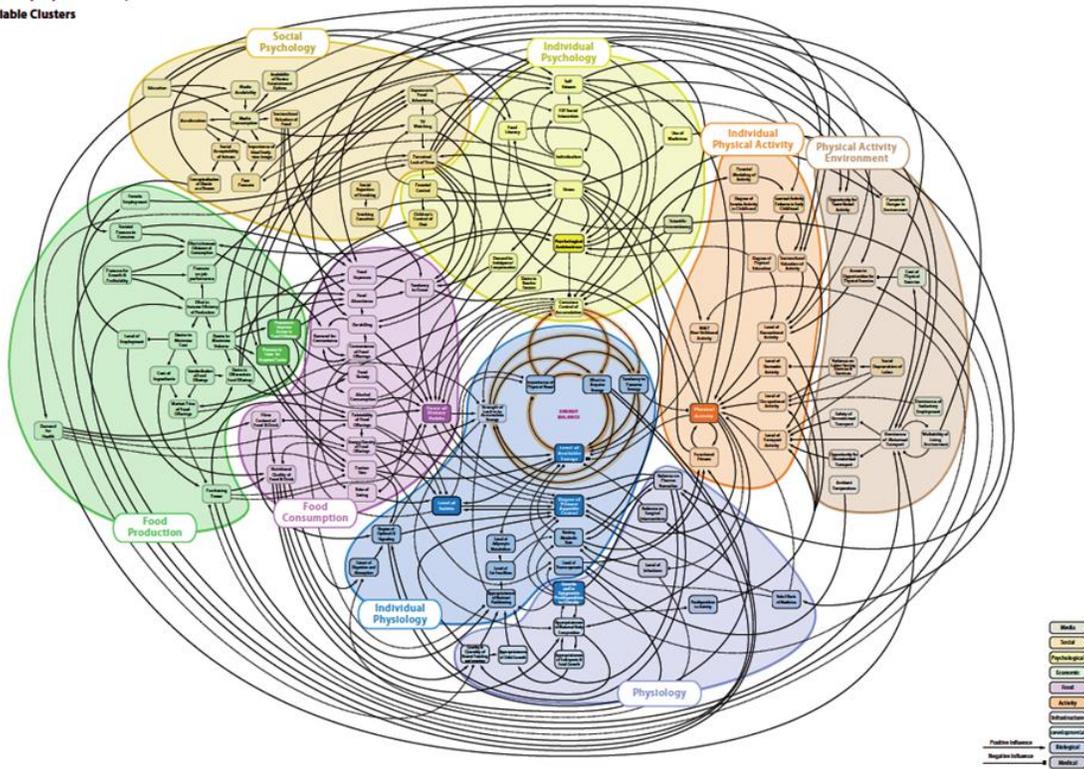
Enfin, nous avons cherché à tester l'influence du niveau d'instruction individuel sur l'évolution des graisses corporelles à l'aide d'une étude longitudinale. D'un point de vue méthodologique, notre objectif a été de montrer l'impact du biais d'attrition et de tester les méthodes statistiques utilisées pour prendre en compte ce biais dans les études longitudinales. Ce travail a montré que le niveau d'éducation individuel était négativement associé au changement d'IMC. Par ailleurs, ce travail a permis de révéler l'influence de certains motifs de données manquantes sur les associations et l'utilité de certaines méthodes statistiques sous certains scénarios d'attritions pour contourner ce biais.

L'originalité et les forces de ce travail sont dues à plusieurs facteurs : la taille de l'échantillon, les indicateurs de composition corporelle utilisés pour évaluer l'obésité, les mesures des facteurs environnementaux dans différents lieux de vies (résidentiel et professionnel), la construction d'indicateurs environnementaux au moyen de système d'information géographique ou de méthodes statistiques (analyses de grappes), la prise en compte au niveau populationnel des secteurs d'activité économique et la comparaison de méthodes statistiques adaptées aux données longitudinales observées. A notre connaissance, aucune étude n'a regardé à l'aide de l'ensemble de ces méthodologies l'excès de poids corporel et les problèmes cardiovasculaires.

Ce travail a permis de conclure d'une part, à l'existence d'effets non négligeables des caractéristiques individuelles, de l'environnement résidentiel et des secteurs d'activité économique sur l'accumulation de masse grasse, le changement de corpulence ainsi que sur les facteurs de risque cardiovasculaire ; et d'autre part de prendre en compte l'impact des données manquantes dans les analyses longitudinales.

ANNEXE

Obesity System Map
Variable Clusters



Annexe 1 : Obesity system map (63).

RÉFÉRENCES BIBLIOGRAPHIQUES

Cités dans le manuscrit

1. Organisation Mondiale de la Santé, (OMS). Aide-mémoire n°311 de l'OMS sur l' obésité et surpoids 2014.
2. Basdevant A, Ricquier D, editors. Pour une approche scientifique de l'obésité. Paris: Editions scientifiques et médicales Elsevier SAS; 2003.
3. LeGendre. Nouveau Traité De Médecine, Fascicule 7, Avitaminoses, Maladies Par Agents Physiques, Troubles De La Nutrition. Broché, editor. Paris 1924.
4. Antipatis VJ, Gill TP. Obesity as a Global Problem, in International Textbook of Obesity Björntorp P, editor. Chichester, UK: John Wiley & Sons, Ltd; 2001.
5. Lew EA, Garfinkel L. Variations in mortality by weight among 750,000 men and women. J Chronic Dis. 1979;32(8):563-76.
6. Basdevant A, Guy-Grand B. Chapitre 15 : Analyse de la composition corporelle. Médecine-Sciences Flammarion ed. l'obésité Md, editor. Paris 2004.
7. Eckel RH. Obesity and heart disease: a statement for healthcare professionals from the Nutrition Committee, American Heart Association. Circulation. 1997 Nov 4;96(9):3248-50.
8. Poirier P, Giles TD, Bray GA, Hong Y, Stern JS, Pi-Sunyer FX, et al. Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. Circulation. 2006 Feb 14;113(6):898-918.
9. Vigarello C. Les métamorphoses du gras. Histoire de l'obésité. Paris: Points; 2013.
10. Eknayan G. Adolphe Quetelet (1796-1874)--the average man and indices of obesity. Nephrol Dial Transplant. 2008 Jan;23(1):47-51.
11. Ducimetiere P, Richard J, Cambien F. The pattern of subcutaneous fat distribution in middle-aged men and the risk of coronary heart disease: the Paris Prospective Study. Int J Obes. 1986;10(3):229-40.

12. Luft FC, Walker PD, Hamburger RJ, Kleit SA. Thrombosis of the renal veins and vena cava. Occurrence in morbid obesity. JAMA. 1975 Dec 15;234(11):1158-60.
13. Food and Drug Administration F. Report of the Endocrinologic and Metabolic Drug Advisory Committee. WASHINGTON1985.
14. Organisation Mondiale de la Santé, (OMS). Physical status: The use and interpretation of anthropometry. Washington1995.
15. Organisation Mondiale de la Santé, (OMS). Obesity: preventing and managing the global epidemic. Report of a WHO Consultation (WHO Technical Report Series 894)2000.
16. Caballero B. The Global Epidemic of Obesity: An Overview. epidemiologic review. 2007;29(1):5.
17. Seidell JC. Time trends in obesity: an epidemiological perspective. Horm Metab Res. 1997 Apr;29(4):155-8.
18. Seidell JC. Obesity: a growing problem. Acta Paediatr Suppl. 1999 Feb;88(428):46-50.
19. Seidell JC, Flegal KM. Assessing obesity: classification and epidemiology. Br Med Bull. 1997;53(2):238-52.
20. Ziegler O. L'histoire naturelle de l'obésité est en train de changer. Obésité. 2008;3(1):6.
21. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. JAMA. 2014 Feb 26;311(8):806-14.
22. Center for Disease Control and Prevention C. Overweight and Obesity - Adult Obesity Facts. Atlanta, GA2014.
23. Treppoz S. [RePOP, a network for prevention and management of obesity in children in Lyon area]. Arch Pediatr. 2006 Jun;13(6):641-3.
24. Rigby NJ, Kumanyika S, James WP. Confronting the epidemic: the need for global solutions. J Public Health Policy. 2004;25(3-4):418-34.
25. Organisation de coopération et de développement économiques O. Obesity update. Paris2014.
26. Basdevant A. [Obesity epidemic: origins and consequences]. C R Biol. 2006 Aug;329(8):562-9; discussion 653-5.

27. Organisation for Economic Co-operation and Development, (OECD). Health policies and data - OECD health statistics. 2014.
28. Eurostat. Overweight and obesity - BMI statistics2011: Available from: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Overweight_and_obesity_-_BMI_statistics.
29. l'Europe T. Santé publique : comparatif de l'obésité en Europe. Paris [cited 2014]; Available from: <http://www.touteurope.eu>.
30. ObEpi-Roche. Enquête épidémiologique nationale sur le surpoids et l'obésité. Paris: INSERM / TNS HEALTHCARE(KANTARHEALTH)2012.
31. Rolland-Cachera MF, Castetbon K, Arnault N, Bellisle F, Romano MC, Lehingue Y, et al. Body mass index in 7-9-y-old French children: frequency of obesity, overweight and thinness. *Int J Obes Relat Metab Disord*. 2002 Dec;26(12):1610-6.
32. Lopez AA, Cespedes ML, Vicente T, Tomas M, Bannasar-Veny M, Tauler P, et al. Body adiposity index utilization in a Spanish Mediterranean population: comparison with the body mass index. *PLoS One*. 2012;7(4):e35281.
33. Malnick SD, Knobler H. The medical complications of obesity. *QJM*. 2006 Sep;99(9):565-79.
34. Duggleby SL, Jackson AA, Godfrey KM, Robinson SM, Inskip HM. Cut-off points for anthropometric indices of adiposity: differential classification in a large population of young women. *Br J Nutr*. 2009 Feb;101(3):424-30.
35. Basdevant A, Guy-Grand B. Chapitre 1 : Définition et classification des obésités. *Médecine de l'obésité*. Paris: Médecine-Sciences Flammarion; 2004.
36. Basdevant A, Laville M, Lerebours E. Chapitre 42 : Les obésités. *Traité de nutrition clinique de l'adulte*. Paris: Médecine-Sciences Flammarion; 2001.
37. Shah NR, Braverman ER. Measuring adiposity in patients: the utility of body mass index (BMI), percent body fat, and leptin. *PLoS One*. 2012;7(4):e33308.
38. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *JAMA*. 2013 Jan 2;309(1):71-82.

39. Romero-Corral A, Montori VM, Somers VK, Korinek J, Thomas RJ, Allison TG, et al. Association of bodyweight with total mortality and with cardiovascular events in coronary artery disease: a systematic review of cohort studies. *Lancet*. 2006 Aug 19;368(9536):666-78.
40. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*. 2008 Feb 16;371(9612):569-78.
41. Weyer C, Funahashi T, Tanaka S, Hotta K, Matsuzawa Y, Pratley RE, et al. Hypoadiponectinemia in obesity and type 2 diabetes: close association with insulin resistance and hyperinsulinemia. *J Clin Endocrinol Metab*. 2001 May;86(5):1930-5.
42. de Koning L, Merchant AT, Pogue J, Anand SS. Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: meta-regression analysis of prospective studies. *Eur Heart J*. 2007 Apr;28(7):850-6.
43. Litwin SE. Which measures of obesity best predict cardiovascular risk? *J Am Coll Cardiol*. 2008 Aug 19;52(8):616-9.
44. Goldstein DJ. Beneficial health effects of modest weight loss. *Int J Obes Relat Metab Disord*. 1992 Jun;16(6):397-415.
45. Organisation Mondiale de la Santé, (OMS). Obésité : prévention et prise en charge de l'épidémie mondiale. Série de Rapports techniques n°894. Série de Rapports techniques. Washington 2003.
46. Moreno M-V. Etude de la composition corporelle par impedencemetrie sur des adultes et des enfants sains et pathologiques. Compiègne: UTC Compiègne). 2007.
47. Hill GL, Jonathan E, Rhoads Lecture. Body composition research: implications for the practice of clinical nutrition. *JPEN J Parenter Enteral Nutr*. 1992 May-Jun;16(3):197-218.
48. Thomasset MA. [Bioelectric properties of tissue. Impedance measurement in clinical medicine. Significance of curves obtained]. *Lyon Med*. 1962 Jul 15;94:107-18.
49. Meijer JH, de Vries PM, Goovaerts HG, Oe PL, Donker AJ, Schneider H. Measurement of transcellular fluid shift during haemodialysis. Part 1. Method. *Med Biol Eng Comput*. 1989 Mar;27(2):147-51.
50. Lazzar S, Boirie Y, Meyer M, Vermorel M. Evaluation of two foot-to-foot bioelectrical impedance analysers to assess body composition in overweight and obese adolescents. *Br J Nutr*. 2003 Nov;90(5):987-92.

51. Bjorntorp P, Evans W. The effect of exercise on body composition. In: Watkins J, Roubenoff R, Rosenberg IH, . Aging BCTMaMoCW, editor. Boston, MA: Foundation for Nutrition Advancement; 1992.
52. Deurenberg P, Yap M, van Staveren WA. Body mass index and percent body fat: a meta analysis among different ethnic groups. *Int J Obes Relat Metab Disord.* 1998 Dec;22(12):1164-71.
53. Ko GT, Tang J, Chan JC, Sung R, Wu MM, Wai HP, et al. Lower BMI cut-off value to define obesity in Hong Kong Chinese: an analysis based on body fat assessment by bioelectrical impedance. *Br J Nutr.* 2001 Feb;85(2):239-42.
54. Chang CJ, Wu CH, Chang CS, Yao WJ, Yang YC, Wu JS, et al. Low body mass index but high percent body fat in Taiwanese subjects: implications of obesity cutoffs. *Int J Obes Relat Metab Disord.* 2003 Feb;27(2):253-9.
55. Deurenberg P. Universal cut-off BMI points for obesity are not appropriate. *Br J Nutr.* 2001 Feb;85(2):135-6.
56. He M, Tan KC, Li ET, Kung AW. Body fat determination by dual energy X-ray absorptiometry and its relation to body mass index and waist circumference in Hong Kong Chinese. *Int J Obes Relat Metab Disord.* 2001 May;25(5):748-52.
57. Romero-Corral A, Somers VK, Sierra-Johnson J, Thomas RJ, Collazo-Clavell ML, Korinek J, et al. Accuracy of body mass index in diagnosing obesity in the adult general population. *Int J Obes (Lond).* 2008 Jun;32(6):959-66.
58. Camhi SM, Bray GA, Bouchard C, Greenway FL, Johnson WD, Newton RL, et al. The relationship of waist circumference and BMI to visceral, subcutaneous, and total body fat: sex and race differences. *Obesity (Silver Spring).* 2011 Feb;19(2):402-8.
59. Jackson AS, Stanforth PR, Gagnon J, Rankinen T, Leon AS, Rao DC, et al. The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study. *Int J Obes Relat Metab Disord.* 2002 Jun;26(6):789-96.
60. Garrido-Chamorro RP, Sirvent-Belando JE, Gonzalez-Lorenzo M, Martin-Carratala ML, Roche E. Correlation between body mass index and body composition in elite athletes. *J Sports Med Phys Fitness.* 2009 Sep;49(3):278-84.
61. Czernichow S, Ciangura C, Oppert JM. [Obesity in the adult]. *Rev Prat.* 2006 Dec 31;56(20):2275-81.
62. Ritenbaugh C, Kumanyika S. Caught in the causal web : a new perspective on social factors affecting obesity. *Healthy Weight Journal.* 1999;13(88):89.

63. Butland B, Jebb S, Kopelman P, McPherson K, Thomas S, Mardell J, et al. Foresight. Tackling obesity: future choices. London: UK Government Office for Science 2007.
64. Takashima Y, Yoshida M, Kokaze A, Orido Y, Tsugane S, Ishikawa M, et al. Relationship of occupation to blood pressure among middle-aged Japanese men--the significance of the differences in body mass index and alcohol consumption. *J Epidemiol.* 1998 Oct;8(4):216-26.
65. Proper KI, Hildebrandt VH. Overweight and obesity among Dutch workers: differences between occupational groups and sectors. *Int Arch Occup Environ Health.* 2010 Jan;83(1):61-8.
66. Hasson D, Von Thiele Schwarz U, Lindfors P. Self-rated health and allostatic load in women working in two occupational sectors. *J Health Psychol.* 2009 May;14(4):568-77.
67. Zhang Y, Liu J, Yao J, Ji G, Qian L, Wang J, et al. Obesity: Pathophysiology and Intervention. *Nutrients.* 2014;6(11):5153-83.
68. Belsky DW, Moffitt TE, Sugden K, Williams B, Houts R, McCarthy J, et al. Development and evaluation of a genetic risk score for obesity. *Biodemography Soc Biol.* 2013;59(1):85-100.
69. Scharoun-Lee M, Adair LS, Kaufman JS, Gordon-Larsen P. Obesity, race/ethnicity and the multiple dimensions of socioeconomic status during the transition to adulthood: a factor analysis approach. *Soc Sci Med.* 2009 Feb;68(4):708-16.
70. Bouchard C, Pérusse L. Genetic of obesity. *Annual Reviews of Nutrition.* 1993;13(337):54.
71. Bouchard C, Tremblay A, Despres JP, Nadeau A, Lupien PJ, Theriault G, et al. The response to long-term overfeeding in identical twins. *N Engl J Med.* 1990 May 24;322(21):1477-82.
72. Stunkard AJ, Harris JR, Pedersen NL, McClearn GE. The body-mass index of twins who have been reared apart. *N Engl J Med.* 1990 May 24;322(21):1483-7.
73. Hainer V, Stunkard A, Kunesova M, Parizkova J, Stich V, Allison DB. A twin study of weight loss and metabolic efficiency. *Int J Obes Relat Metab Disord.* 2001 Apr;25(4):533-7.
74. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* 2007 May 11;316(5826):889-94.
75. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature.* 2014 Mar 20;507(7492):371-5.

76. Bout B. Rapport de Mme Brigitte BOUT, fait au nom de l'Office parlementaire d'évaluation des choix scientifiques et technologiques. L'organisation de la recherche et ses perspectives en matière de prévention et de traitement de l'obésité. 2010 Contract No.: 158.
77. Shimokata H, Tobin JD, Muller DC, Elahi D, Coon PJ, Andres R. Studies in the distribution of body fat: I. Effects of age, sex, and obesity. *J Gerontol.* 1989 Mar;44(2):M66-73.
78. Wang Y, Beydoun MA. The obesity epidemic in the United States--gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. *Epidemiol Rev.* 2007;29:6-28.
79. Zhang Q, Wang Y. Socioeconomic inequality of obesity in the United States: do gender, age, and ethnicity matter? *Soc Sci Med.* 2004 Mar;58(6):1171-80.
80. Ball K, Crawford D. Socioeconomic status and weight change in adults: a review. *Soc Sci Med.* 2005 May;60(9):1987-2010.
81. McLaren L. Socioeconomic status and obesity. *Epidemiol Rev.* 2007;29:29-48.
82. Mujahid MS, Diez Roux AV, Borrell LN, Nieto FJ. Cross-sectional and longitudinal associations of BMI with socioeconomic characteristics. *Obes Res.* 2005 Aug;13(8):1412-21.
83. Drewnowski A, Moudon AV, Jiao J, Aggarwal A, Charreire H, Chaix B. Food environment and socioeconomic status influence obesity rates in Seattle and in Paris. *Int J Obes (Lond).* 2014 Feb;38(2):306-14.
84. Jeffery RW, French SA. Epidemic obesity in the United States: are fast foods and television viewing contributing? *Am J Public Health.* 1998 Feb;88(2):277-80.
85. Monasta L, Batty GD, Cattaneo A, Lutje V, Ronfani L, Van Lenthe FJ, et al. Early-life determinants of overweight and obesity: a review of systematic reviews. *Obes Rev.* 2010 Oct;11(10):695-708.
86. Organisation Mondiale de la Santé, (OMS). Aide-mémoire N°311 de l'OMS sur l'obésité et surpoids. 2014.
87. Bray GA, Paeratakul S, Popkin BM. Dietary fat and obesity: a review of animal, clinical and epidemiological studies. *Physiol Behav.* 2004 Dec 30;83(4):549-55.
88. Owen N, Leslie E, Salmon J, Fotheringham MJ. Environmental determinants of physical activity and sedentary behavior. *Exerc Sport Sci Rev.* 2000 Oct;28(4):153-8.

89. Giles-Corti B, Donovan RJ. Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. *Prev Med.* 2002 Dec;35(6):601-11.
90. La Torre G, Masala D, De Vito E, Langiano E, Capelli G, Ricciardi W. Extra-curricular physical activity and socioeconomic status in Italian adolescents. *BMC Public Health.* 2006;6:22.
91. Cerin E, Leslie E. How socio-economic status contributes to participation in leisure-time physical activity. *Soc Sci Med.* 2008 Jun;66(12):2596-609.
92. Kestens Y, Lebel A, Chaix B, Clary C, Daniel M, Pampalon R, et al. Association between activity space exposure to food establishments and individual risk of overweight. *PLoS One.* 2012;7(8):e41418.
93. Chaix B, Bean K, Daniel M, Zenk SN, Kestens Y, Charreire H, et al. Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). *PLoS One.* 2012;7(4):e32908.
94. Encyclopédie scientifique. Définition d'environnement. Available from: <http://www.techno-science.net/>.
95. 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.
96. Leal C. Les environnements résidentiels et l'obésité: l'application de technique d'appariement pour estimer les relations. Paris: Université Pierre et Marie Curie - Paris 6; 2011.
97. Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obes Rev.* 2011 Mar;12(3):217-30.
98. Basdevant A. Obésité , santé et société. Recherches et convergences. Paris 2011.
99. Karusisi N. Effets comparés des environnements géographiques de vie sur différentes formes d'activités physique : vers une meilleur prise en compte des environnements de vie. Paris: Université Pierre et Marie Curie - Paris 6; 2013.
100. Lebel A. La géographie d'excès de poids au Québec. Laval: Université de Laval; 2010.
101. Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci.* 2010 Feb;1186:125-45.

102. Wang MC, Kim S, Gonzalez AA, MacLeod KE, Winkleby MA. Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *J Epidemiol Community Health*. 2007 Jun;61(6):491-8.
103. Chaix B, Kestens Y, Bean K, Leal C, Karusisi N, Meghiref K, et al. Cohort profile: residential and non-residential environments, individual activity spaces and cardiovascular risk factors and diseases--the RECORD Cohort Study. *Int J Epidemiol*. 2012 Oct;41(5):1283-92.
104. Chaix B, Kestens Y, Perchoux C, Karusisi N, Merlo J, Labadi K. An interactive mapping tool to assess individual mobility patterns in neighborhood studies. *Am J Prev Med*. 2012 Oct;43(4):440-50.
105. Romero AJ. Low-income neighborhood barriers and resources for adolescents' physical activity. *J Adolesc Health*. 2005 Mar;36(3):253-9.
106. Sellstrom E, Arnoldsson G, Alricsson M, Hjern A. Obesity prevalence in a cohort of women in early pregnancy from a neighbourhood perspective. *BMC Pregnancy Childbirth*. 2009;9:37.
107. Robert SA, Reither EN. A multilevel analysis of race, community disadvantage, and body mass index among adults in the US. *Soc Sci Med*. 2004 Dec;59(12):2421-34.
108. Bernard P, Charafeddine R, Frohlich KL, Daniel M, Kestens Y, Potvin L. Health inequalities and place: a theoretical conception of neighbourhood. *Soc Sci Med*. 2007 Nov;65(9):1839-52.
109. Feng J, Glass TA, Curriero FC, Stewart WF, Schwartz BS. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place*. 2010 Mar;16(2):175-90.
110. Bensimhon DR, Kraus WE, Donahue MP. Obesity and physical activity: a review. *Am Heart J*. 2006 Mar;151(3):598-603.
111. Oppert JM, Charreire H. The importance of the food and physical activity environments. *Nestle Nutr Inst Workshop Ser*. 2012;73:113-21.
112. Sallis JF, Floyd MF, Rodriguez DA, Saelens BE. Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation*. 2012 Feb 7;125(5):729-37.
113. Charreire H, Weber C, Chaix B, Salze P, Casey R, Banos A, et al. Identifying built environmental patterns using cluster analysis and GIS: relationships with walking, cycling and body mass index in French adults. *Int J Behav Nutr Phys Act*. 2012;9:59.

114. Townshend T, Lake AA. Obesogenic urban form: theory, policy and practice. *Health Place*. 2009 Dec;15(4):909-16.
115. Frank LD, Andresen MA, Schmid TL. Obesity relationships with community design, physical activity, and time spent in cars. *Am J Prev Med*. 2004 Aug;27(2):87-96.
116. Kaczynski AT, Potwarka LR, Saelens BE. Association of park size, distance, and features with physical activity in neighborhood parks. *Am J Public Health*. 2008 Aug;98(8):1451-6.
117. Frank L, Engelke P. The built environment and human activity patterns: exploring the impacts of urban form on public health. *Journal of Planning Literature*. 2001;16:202-18.
118. Liu GC, Wilson JS, Qi R, Ying J. Green neighborhoods, food retail and childhood overweight: differences by population density. *Am J Health Promot*. 2007 Mar-Apr;21(4 Suppl):317-25.
119. Zhao Z. Neighborhood effects on obesity. Chicago: University of Illinois at Chicago; 2008.
120. Huang R, Moudon AV, Cook AJ, Drewnowski A. The spatial clustering of obesity: does the built environment matter? *J Hum Nutr Diet*. 2014 Oct 3.
121. Wen LM, Rissel C. Inverse associations between cycling to work, public transport, and overweight and obesity: findings from a population based study in Australia. *Prev Med*. 2008 Jan;46(1):29-32.
122. Oliver M, Badland H, Mavoa S, Witten K, Kearns R, Ellaway A, et al. Environmental and socio-demographic associates of children's active transport to school: a cross-sectional investigation from the URBAN Study. *Int J Behav Nutr Phys Act*. 2014;11:70.
123. Diez Roux AV, Evenson KR, McGinn AP, Brown DG, Moore L, Brines S, et al. Availability of recreational resources and physical activity in adults. *Am J Public Health*. 2007 Mar;97(3):493-9.
124. Glanz K, Sallis JF, Saelens BE, Frank LD. Healthy nutrition environments: concepts and measures. *Am J Health Promot*. 2005 May-Jun;19(5):330-3, ii.
125. Morland K, Diez Roux AV, Wing S. Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study. *Am J Prev Med*. 2006 Apr;30(4):333-9.
126. Etat des lieux de la pauvreté et de l'exclusion en Île-de-France- 2005. Paris: Mission d'Information sur la Pauvreté et l'Exclusion Sociale en Île-de-France,. 2006.

127. Jeffery RW, Baxter J, McGuire M, Linde J. Are fast food restaurants an environmental risk factor for obesity? *Int J Behav Nutr Phys Act.* 2006;3:2.
128. Leal C, Bean K, Thomas F, Chaix B. Are associations between neighborhood socioeconomic characteristics and body mass index or waist circumference based on model extrapolations? *Epidemiology.* 2011 Sep;22(5):694-703.
129. Leal C, Bean K, Thomas F, Chaix B. Multicollinearity in associations between multiple environmental features and body weight and abdominal fat: using matching techniques to assess whether the associations are separable. *Am J Epidemiol.* 2012 Jun 1;175(11):1152-62.
130. Casey R, Chaix B, Weber C, Schweitzer B, Charreire H, Salze P, et al. Spatial accessibility to physical activity facilities and to food outlets and overweight in French youth. *Int J Obes (Lond).* 2012 Jul;36(7):914-9.
131. DARES. Conditions de travail et santé. Paris: Ministère du Travail, de l'Emploi, de la Formation Professionnelle et du Dialogue Social.2009.
132. Geoffroy-Perez B. Analyse de la mortalité et des causes de décès par secteur d'activité de 1968 à 1999 à partir de l'Echantillon démographique permanent - Etude pour la mise en place du programme Cosmop : Cohorte pour la surveillance de la mortalité par profession Paris: INVS2006.
133. Arnaudo B, Mafaud-Camus I, Sandret N. Enquêtes Surveillance médicale des expositions aux risques professionnels (Sumer) Paris: DARES2003.
134. Bué J, Coutrot T, Hamon-Cholet S, Vinck L. Conditions de Travail : une pause dans l'intensification. Paris2007.
135. DARES. Les risques professionnels en 2010 : de fortes différences d'exposition selon les secteurs. Paris: Ministère du Travail, de l'Emploi, de la Formation Professionnelle et du Dialogue Social2013.
136. Lauzeville D, Marchand JL, Ferrand M. Consommation de tabac par catégorie socioprofessionnelle et secteur d'activité : outil méthodologique pour l'épidémiologie. 2009.
137. Goffette C. Le contexte social du tabagisme. Le rôle de l'environnement familial et professionnel. Paris: Science po paris; 2012.
138. Beck F. Résultats du Baromètre Santé 2010 - Liens entre usages de substances psychoactives (SPA) et milieu professionnel. Paris: INPES2012.

139. Carrasco F, Reyes E, Rimler O, Rios F. [Predictive accuracy of body mass index in estimating body fatness measured by bioelectrical impedance]. *Arch Latinoam Nutr.* 2004 Sep;54(3):280-6.
140. Frommer MS, Edye BV, Mandryk JA, Grammeno GL, Berry G, Ferguson DA. Systolic blood pressure in relation to occupation and perceived work stress. *Scand J Work Environ Health.* 1986 Oct;12(5):476-85.
141. Hu G, Jousilahti P, Antikainen R, Tuomilehto J. Occupational, commuting, and leisure-time physical activity in relation to cardiovascular mortality among finnish subjects with hypertension. *Am J Hypertens.* 2007 Dec;20(12):1242-50.
142. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation.* 1993 Oct;88(4 Pt 1):1973-98.
143. Leigh JP, Du J. Hypertension and occupation among seniors. *J Occup Environ Med.* 2009 Jun;51(6):661-71.
144. Schnall PL, Landsbergis PA, Baker D. Job strain and cardiovascular disease. *Annu Rev Public Health.* 1994;15:381-411.
145. Tsutsumi A, Kayaba K, Tsutsumi K, Igarashi M. Association between job strain and prevalence of hypertension: a cross sectional analysis in a Japanese working population with a wide range of occupations: the Jichi Medical School cohort study. *Occup Environ Med.* 2001 Jun;58(6):367-73.
146. Leclerc A, Fassin D, Grandjean H, Kaminski M, Lang T. Les inégalités sociales de santé. *Recherches C*, editor. Paris: Inserm; 2000.
147. Kogevinas M, Pearce N, Susser M, Boffetta P. Social inequalities and cancer. Lyon: IARC1997 Contract No.: 138.
148. Imbernon E. Estimation du nombre de cas de certains cancers attribuables à des facteurs professionnels Paris: INVS2003.
149. Buxton OM, Cain SW, O'Connor SP, Porter JH, Duffy JF, Wang W, et al. Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. *Sci Transl Med.* 2012 Apr 11;4(129):129ra43.
150. Brunner EJ, Chandola T, Marmot MG. Prospective effect of job strain on general and central obesity in the Whitehall II Study. *Am J Epidemiol.* 2007 Apr 1;165(7):828-37.
151. Wanjek C. L'alimentation au travail: des solutions contre la malnutrition, l'obésité et les maladies chroniques. Genève: Organisation Internationale du Travail2005.

152. Belkic K, Nedic O. Workplace stressors and lifestyle-related cancer risk factors among female physicians: assessment using the Occupational Stress Index. *J Occup Health*. 2007 Jan;49(1):61-71.
153. Sobal J, Stunkard AJ. Socioeconomic status and obesity: a review of the literature. *Psychol Bull*. 1989 Mar;105(2):260-75.
154. Burke GL, Bild DE, Hilner JE, Folsom AR, Wagenknecht LE, Sidney S. Differences in weight gain in relation to race, gender, age and education in young adults: the CARDIA Study. *Coronary Artery Risk Development in Young Adults*. *Ethn Health*. 1996 Dec;1(4):327-35.
155. Moore DB, Howell PB, Treiber FA. Changes in overweight in youth over a period of 7 years: impact of ethnicity, gender and socioeconomic status. *Ethn Dis*. 2002 Winter;12(1):S1-83-6.
156. Kahn HS, Williamson DF. The contributions of income, education and changing marital status to weight change among US men. *Int J Obes*. 1990 Dec;14(12):1057-68.
157. van Lenthe FJ, Droomers M, Schrijvers CT, Mackenbach JP. Socio-demographic variables and 6 year change in body mass index: longitudinal results from the GLOBE study. *Int J Obes Relat Metab Disord*. 2000 Aug;24(8):1077-84.
158. Sundquist J, Johansson SE. The influence of socioeconomic status, ethnicity and lifestyle on body mass index in a longitudinal study. *Int J Epidemiol*. 1998 Feb;27(1):57-63.
159. Begaud B. *Dictionnaire de Pharmaco-épidémiologie*. Troisième édition ed. Bordeaux: ARME - Pharmacovigilance; 1998.
160. The Cochrane collaboration. *Cochrane Handbook for Systematic Reviews of Interventions - Part 2: General methods for Cochrane reviews 2011*. Available from: http://handbook.cochrane.org/front_page.htm.
161. Little RJA, DB R. *Statistical Analysis with Missing Data*. New Jersey: John Wiley & Sons; 1987.
162. Graham JW, Donaldson SI. Evaluating interventions with differential attrition: the importance of nonresponse mechanisms and use of follow-up data. *J Appl Psychol*. 1993 Feb;78(1):119-28.
163. van der Heijden GJ, Donders AR, Stijnen T, Moons KG. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*. 2006 Oct;59(10):1102-9.

164. Eekhout I, de Boer RM, Twisk JW, de Vet HC, Heymans MW. Missing data: a systematic review of how they are reported and handled. *Epidemiology*. 2012 Sep;23(5):729-32.
165. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
166. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials*. 2004;1(4):368-76.
167. Ile-de-France) ARdSI-d-FA. Plan Strategique Régional de Santé : Eléments de contexte et principes d'action de l'ARS Ile-de-France. 2014; Available from: <http://www.ars.iledefrance.sante.fr/>.
168. Riva M, Gauvin L, Barnett TA. Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. *J Epidemiol Community Health*. 2007 Oct;61(10):853-61.
169. Diez Roux AV. Residential environments and cardiovascular risk. *J Urban Health*. 2003 Dec;80(4):569-89.
170. L'Etude RECORD. Une étude sur les disparité sociales de santé. 2014; Available from: <http://www.record-study.org>.
171. Chaix B, Bean K, Leal C, Thomas F, Havard S, Evans D, et al. Individual/neighborhood social factors and blood pressure in the RECORD Cohort Study: which risk factors explain the associations? *Hypertension*. 2010 Mar;55(3):769-75.
172. Billaudeau N, Oppert JM, Simon C, Charreire H, Casey R, Salze P, et al. Investigating disparities in spatial accessibility to and characteristics of sport facilities: direction, strength, and spatial scale of associations with area income. *Health Place*. 2011 Jan;17(1):114-21.
173. Apparicio P, Abdelmajid M, Riva M, Shearmur R. Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *Int J Health Geogr*. 2008;7:7.
174. Leslie E, Coffee N, Frank L, Owen N, Bauman A, Hugo G. Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. *Health Place*. 2007 Mar;13(1):111-22.
175. Oliver LN, Schuurman N, Hall AW. Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *Int J Health Geogr*. 2007;6:41.

176. Riva M, Apparicio P, Gauvin L, Brodeur JM. Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: an exemplar for designing optimal zones. *Int J Health Geogr*. 2008;7:43.
177. Tan P-N, Steinbach M, Kumar V, editors. *Cluster analysis: basic concepts and algorithms*. Upper Saddle River, NJ: Addison-Wesley; 2005.
178. 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 Dec;63(12):1043-8.
179. Chaix B. *Modélisation des effets du contexte sur la santé et le recours aux soins*. Paris: Université Paris 6; 2004.
180. Bartlett JW, Carpenter JR, Tilling K, Vansteelandt S. Improving upon the efficiency of complete case analysis when covariates are MNAR. *Biostatistics*. 2014 Oct;15(4):719-30.
181. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013 Jun;22(3):278-95.
182. Rubin D.B. *Multiple imputation for non response in surveys*. New York: Wiley; 1987.
183. INSEE. *IRIS—Definition*. Paris, France: National Institute of Statistics and Economic Studies. 2011.
184. Kershaw KN, Albrecht SS, Carnethon MR. Racial and ethnic residential segregation, the neighborhood socioeconomic environment, and obesity among Blacks and Mexican Americans. *Am J Epidemiol*. 2013 Feb 15;177(4):299-309.
185. Nakamura K, Shimai S, Kikuchi S, Tominaga K, Takahashi H, Tanaka M, et al. Shift work and risk factors for coronary heart disease in Japanese blue-collar workers: serum lipids and anthropometric characteristics. *Occup Med (Lond)*. 1997 Apr;47(3):142-6.
186. Nakamura S, Nakamura K, Tanaka M. Increased risk of coronary heart disease in Japanese blue-collar workers. *Occup Med (Lond)*. 2000 Jan;50(1):11-7.
187. Rundle A, Diez Roux AV, Free LM, Miller D, Neckerman KM, Weiss CC. The urban built environment and obesity in New York City: a multilevel analysis. *Am J Health Promot*. 2007 Mar-Apr;21(4 Suppl):326-34.
188. Karusisi N, Thomas F, Meline J, Brondeel R, Chaix B. Environmental conditions around itineraries to destinations as correlates of walking for transportation among adults: the RECORD cohort study. *PLoS One*. 2014;9(5):e88929.

189. Proper KI, Hildebrandt VH. Physical activity among Dutch workers--differences between occupations. *Prev Med.* 2006 Jul;43(1):42-5.
190. Peltz G, Aguirre MT, Sanderson M, Fadden MK. The role of fat mass index in determining obesity. *Am J Hum Biol.* 2010 Sep-Oct;22(5):639-47.
191. Rosmond R, Lapidus L, Bjorntorp P. The influence of occupational and social factors on obesity and body fat distribution in middle-aged men. *Int J Obes Relat Metab Disord.* 1996 Jul;20(7):599-607.
192. Allison DB, Faith MS, Heo M, Kotler DP. Hypothesis concerning the U-shaped relation between body mass index and mortality. *Am J Epidemiol.* 1997 Aug 15;146(4):339-49.
193. Takashima Y, Iwase Y, Yoshida M, Kokaze A, Takagi Y, Taubono Y, et al. Relationship of food intake and dietary patterns with blood pressure levels among middle-aged Japanese men. *J Epidemiol.* 1998 Jun;8(2):106-15.
194. Oppert JM, Thomas F, Charles MA, Benetos A, Basdevant A, Simon C. Leisure-time and occupational physical activity in relation to cardiovascular risk factors and eating habits in French adults. *Public Health Nutr.* 2006 Sep;9(6):746-54.
195. Organisation mondiale de la santé, (OMS). Diet, nutrition and the prevention of chronic diseases Report of the joint WHO/FAO expert consultation. Geneva2003.
196. Kristman VL, Manno M, Cote P. Methods to account for attrition in longitudinal data: do they work? A simulation study. *Eur J Epidemiol.* 2005;20(8):657-62.
197. Twisk J, de Vente W. Attrition in longitudinal studies. How to deal with missing data. *J Clin Epidemiol.* 2002 Apr;55(4):329-37.
198. Osborne JW. *Best Practices in Data Cleaning. Dealing with Missing or Incomplete Data.* Louisville Uo, editor. Louisville, KY: Sage Publication, Inc; 2013.
199. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *J Am Stat Assoc.* 1995;90:1112-21.
200. Allison PD. *Missing data.* Sciences SUPoQAitS, editor. Thousand Oakc, CA2001.
201. Hipple PT. Regression with Missing Y's: An Improved Strategy for Analysing Multiply Imputed Data. *Sociological Methodology.* 2007;37:83-117.
202. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods.* 2002 Jun;7(2):147-77.

203. Hofler M, Pfister H, Lieb R, Wittchen HU. The use of weights to account for non-response and drop-out. *Soc Psychiatry Psychiatr Epidemiol*. 2005 Apr;40(4):291-9.
204. Seaman SR, White IR, Copas AJ, Li L. Combining multiple imputation and inverse-probability weighting. *Biometrics*. 2012 Mar;68(1):129-37.
205. Chaix B, Billaudeau N, Thomas F, Havard S, Evans D, Kestens Y, et al. Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology*. 2011 Jan;22(1):18-26.
206. McKinsey Global Institute. *Overcoming obesity: An initial economic analysis*. New York, NY 2014.

Cités dans les articles

1. 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(6):531-8.
2. Coogan PF, Cozier YC, Krishnan S, Wise LA, Adams-Campbell LL, Rosenberg L, et al. Neighborhood socioeconomic status in relation to 10-year weight gain in the Black Women's Health Study. *Obesity (Silver Spring)*. 2010 Oct;18(10):2064-5.
3. Grow HM, Cook AJ, Arterburn DE, Saelens BE, Drewnowski A, Lozano P. Child obesity associated with social disadvantage of children's neighborhoods. *Soc Sci Med*. 2010 Aug;71(3):584-91.
4. Oliver LN, Hayes MV. Effects of neighbourhood income on reported body mass index: an eight year longitudinal study of Canadian children. *BMC Public Health*. 2008;8:16.
5. Stafford M, Brunner EJ, Head J, Ross NA. Deprivation and the development of obesity a multilevel, longitudinal study in England. *Am J Prev Med*. 2010 Aug;39(2):130-9.
6. Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obes Rev*. 2011 Mar;12(3):217-30.

7. Jeffery RW, Baxter J, McGuire M, Linde J. Are fast food restaurants an environmental risk factor for obesity? *Int J Behav Nutr Phys Act.* 2006;3:2.
8. Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. Indices of relative weight and obesity. *J Chronic Dis.* 1972 Jul 1;25(6):329-43.
9. Bouchard C. BMI, fat mass, abdominal adiposity and visceral fat: where is the 'beef'? *Int J Obes (Lond).* 2007 Oct;31(10):1552-3.
10. Garrido-Chamorro RP, Sirvent-Belando JE, Gonzalez-Lorenzo M, Martin-Carratala ML, Roche E. Correlation between body mass index and body composition in elite athletes. *J Sports Med Phys Fitness.* 2009 Sep;49(3):278-84.
11. Camhi SM, Bray GA, Bouchard C, Greenway FL, Johnson WD, Newton RL, et al. The relationship of waist circumference and BMI to visceral, subcutaneous, and total body fat: sex and race differences. *Obesity (Silver Spring).* 2011 Feb;19(2):402-8.
12. Jackson AS, Stanforth PR, Gagnon J, Rankinen T, Leon AS, Rao DC, et al. The effect of sex, age and race on estimating percentage body fat from body mass index: The Heritage Family Study. *Int J Obes Relat Metab Disord.* 2002 Jun;26(6):789-96.
13. Leal C, Bean K, Thomas F, Chaix B. Are associations between neighborhood socioeconomic characteristics and body mass index or waist circumference based on model extrapolations? *Epidemiology.* 2011 Sep;22(5):694-703.
14. Chaix B, Bean K, Leal C, Thomas F, Havard S, Evans D, et al. Individual/neighborhood social factors and blood pressure in the RECORD Cohort Study: which risk factors explain the associations? *Hypertension.* 2010 Mar;55(3):769-75.
15. Chaix B, Billaudeau N, Thomas F, Havard S, Evans D, Kestens Y, et al. Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology.* 2011 Jan;22(1):18-26.
16. Chaix B, Jouven X, Thomas F, Leal C, Billaudeau N, Bean K, et al. Why socially deprived populations have a faster resting heart rate: impact of behaviour, life course anthropometry, and biology--the RECORD Cohort Study. *Soc Sci Med.* 2011 Nov;73(10):1543-50.
17. Chaix B, Kestens Y, Bean K, Leal C, Karusisi N, Meghief K, et al. Cohort profile: residential and non-residential environments, individual activity spaces and cardiovascular risk factors and diseases--the RECORD Cohort Study. *Int J Epidemiol.* 2012 Oct;41(5):1283-92.

18. Havard S, Reich BJ, Bean K, Chaix B. Social inequalities in residential exposure to road traffic noise: an environmental justice analysis based on the RECORD Cohort Study. *Occup Environ Med.* 2011 May;68(5):366-74.
19. Chaix B, Bean K, Daniel M, Zenk SN, Kestens Y, Charreire H, et al. Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). *PLoS One.* 2012;7(4):e32908.
20. Karusisi N, Bean K, Oppert JM, Pannier B, Chaix B. Multiple dimensions of residential environments, neighborhood experiences, and jogging behavior in the RECORD Study. *Prev Med.* 2012 Jul;55(1):50-5.
21. INSEE. IRIS—Definition. Paris, France: National Institute of Statistics and Economic Studies. 2011.
22. Thomas F, Bean K, Pannier B, Oppert JM, Guize L, Benetos A. Cardiovascular mortality in overweight subjects: the key role of associated risk factors. *Hypertension.* 2005 Oct;46(4):654-9.
23. Schutz Y, Kyle UU, Pichard C. Fat-free mass index and fat mass index percentiles in Caucasians aged 18-98 y. *Int J Obes Relat Metab Disord.* 2002 Jul;26(7):953-60.
24. Lazzer S, Boirie Y, Meyer M, Vermorel M. Evaluation of two foot-to-foot bioelectrical impedance analysers to assess body composition in overweight and obese adolescents. *Br J Nutr.* 2003 Nov;90(5):987-92.
25. Beckman A, Hakansson A, Rastam L, Lithman T, Merlo J. The role country of birth plays in receiving disability pensions in relation to patterns of health care utilisation and socioeconomic differences: a multilevel analysis of Malmö, Sweden. *BMC Public Health.* 2006;6:71.
26. United-Nations-Development-Program. Human Development Report 2007/2008. 2008(New York, NY: United Nations).
27. Kumar V. Cluster analysis: basic concepts and algorithms. In: Tan P-N, Steinbach M, Kumar V, editors. *Introduction to Data Mining.* 2005;Boston: Addison Wesley:pp.487-568.
28. Li YS, Chuang YC. Neighborhood effects on an individual's health using neighborhood measurements developed by factor analysis and cluster analysis. *J Urban Health.* 2009 Jan;86(1):5-18.
29. Van Hulst A, Thomas F, Barnett TA, Kestens Y, Gauvin L, Pannier B, et al. A typology of neighborhoods and blood pressure in the RECORD Cohort Study. *J Hypertens.* 2012 Jul;30(7):1336-46.

30. 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.
31. Chaix B, Kestens Y, Perchoux C, Karusisi N, Merlo J, Labadi K. An interactive mapping tool to assess individual mobility patterns in neighborhood studies. *Am J Prev Med*. 2012 Oct;43(4):440-50.
32. Chaix B, Meline J, Duncan S, Merrien C, Karusisi N, Perchoux C, et al. GPS tracking in neighborhood and health studies: A step forward for environmental exposure assessment, a step backward for causal inference? *Health Place*. 2013 Jan 17;21C:46-51.
33. Proper KI, Hildebrandt VH. Overweight and obesity among Dutch workers: differences between occupational groups and sectors. *Int Arch Occup Environ Health*. 2010 Jan;83(1):61-8.
34. Gershoff ET, Pedersen S, Lawrence Aber J. Creating neighborhood typologies of GIS-based data in the absence of neighborhood-based sampling: a factor and cluster analytic strategy. *J Prev Interv Community*. 2009;37(1):35-47.
35. Chaix B, Evans D, Merlo J, Suzuki E. Commentary: Weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. *Epidemiology*. 2012 Jan;23(1):129-31; discussion 32-7.
36. Kershaw KN, Albrecht SS, Carnethon MR. Racial and Ethnic Residential Segregation, the Neighborhood Socioeconomic Environment, and Obesity Among Blacks and Mexican Americans. *Am J Epidemiol*. 2013 Jan 20.
37. Rehm CD, Moudon AV, Hurvitz PM, Drewnowski A. Residential property values are associated with obesity among women in King County, WA, USA. *Soc Sci Med*. 2012 Aug;75(3):491-5.
38. Nakamura K, Shimai S, Kikuchi S, Tominaga K, Takahashi H, Tanaka M, et al. Shift work and risk factors for coronary heart disease in Japanese blue-collar workers: serum lipids and anthropometric characteristics. *Occup Med (Lond)*. 1997 Apr;47(3):142-6.
39. Nakamura S, Nakamura K, Tanaka M. Increased risk of coronary heart disease in Japanese blue-collar workers. *Occup Med (Lond)*. 2000 Jan;50(1):11-7.
40. Proper KI, Hildebrandt VH. Physical activity among Dutch workers--differences between occupations. *Prev Med*. 2006 Jul;43(1):42-5.
41. Yusuf S, Islam S, Chow CK, Rangarajan S, Dagenais G, Diaz R, et al. Use of secondary prevention drugs for cardiovascular disease in the community in high-income, middle-income,

and low-income countries (the PURE Study): a prospective epidemiological survey. *Lancet*. 2011 Oct 1;378(9798):1231-43.

42. Gupta R, Kaul V, Agrawal A, Guptha S, Gupta VP. Cardiovascular risk according to educational status in India. *Prev Med*. 2010 Nov;51(5):408-11.

43. Reddy KS, Prabhakaran D, Jeemon P, Thankappan KR, Joshi P, Chaturvedi V, et al. Educational status and cardiovascular risk profile in Indians. *Proc Natl Acad Sci U S A*. 2007 Oct 9;104(41):16263-8.

44. Jeemon P, Prabhakaran D, Huffman MD, Ramakrishnan L, Goenka S, Thankappan KR, et al. Distribution of 10-year and lifetime predicted risk for cardiovascular disease in the Indian Sentinel Surveillance Study population (cross-sectional survey results). *BMJ Open*. 2011 Jan 1;1(1):e000068.

55. Frommer MS, Edey BV, Mandryk JA, Grammeno GL, Berry G, Ferguson DA. Systolic blood pressure in relation to occupation and perceived work stress. *Scand J Work Environ Health*. 1986 Oct;12(5):476-85.

56. Hu G, Jousilahti P, Antikainen R, Tuomilehto J. Occupational, commuting, and leisure-time physical activity in relation to cardiovascular mortality among Finnish subjects with hypertension. *Am J Hypertens*. 2007 Dec;20(12):1242-50.

57. Leigh JP, Du J. Hypertension and occupation among seniors. *J Occup Environ Med*. 2009 Jun;51(6):661-71.

58. Schnall PL, Landsbergis PA, Baker D. Job strain and cardiovascular disease. *Annu Rev Public Health*. 1994;15:381-411.

59. Tsutsumi A, Kayaba K, Tsutsumi K, Igarashi M. Association between job strain and prevalence of hypertension: a cross sectional analysis in a Japanese working population with a wide range of occupations: the Jichi Medical School cohort study. *Occup Environ Med*. 2001 Jun;58(6):367-73.

60. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation*. 1993 Oct;88(4 Pt 1):1973-98.

61. Takashima Y, Yoshida M, Kokaze A, Orido Y, Tsugane S, Ishikawa M, et al. Relationship of occupation to blood pressure among middle-aged Japanese men--the significance of the differences in body mass index and alcohol consumption. *J Epidemiol*. 1998 Oct;8(4):216-26.

62. Haglund BJ. Geographical and socioeconomic distribution of high blood pressure and borderline high blood pressure in a Swedish rural county. *Scand J Soc Med*. 1985;13(2):53-66.

63. Covey LS, Wynder EL. Smoking habits and occupational status. *J Occup Med*. 1981 Aug;23(8):537-42.
64. Dobson AJ, Gibberd RW, Leeder SR, O'Connell DL. Occupational differences in ischemic heart disease mortality and risk factors in Australia. *Am J Epidemiol*. 1985 Aug;122(2):283-90.
65. Simons LA, Simons J, Magnus P, Bennett SA. Education level and coronary risk factors in Australians. *Med J Aust*. 1986 Nov 3;145(9):446, 8-50.
66. Helmert U, Shea S, Herman B, Greiser E. Relationship of social class characteristics and risk factors for coronary heart disease in West Germany. *Public Health*. 1990 Nov;104(6):399-416.
67. Kritz-Silverstein D, Wingard DL, Barrett-Connor E. Employment status and heart disease risk factors in middle-aged women: the Rancho Bernardo Study. *Am J Public Health*. 1992 Feb;82(2):215-9.
68. Wilson TW, Kaplan GA, Kauhanen J, Cohen RD, Wu M, Salonen R, et al. Association between plasma fibrinogen concentration and five socioeconomic indices in the Kuopio Ischemic Heart Disease Risk Factor Study. *Am J Epidemiol*. 1993 Feb 1;137(3):292-300.
69. Medalie JH, Papier C, Herman JB, Goldbourt U, Tamir S, Neufeld HN, et al. Diabetes mellitus among 10,000 adult men. I. Five-year incidence and associated variables. *Isr J Med Sci*. 1974 Jul;10(7):681-97.
70. Milczarek M, Schneider E, González ER. stress at work. Luxembourg: European Agency for Safety and Health at Work2009.
71. Holman D, McClelland C. Job Quality in Growing and Declining Economic Sectors of the EU Manchester2011.
72. DARES. Conditions de travail et santé. Paris: Ministère du Travail, de l'Emploi, de la Formation Professionnelle et du Dialogue Social.2009.
73. Klein T, Long K. Conditions de travail organisation du travail et usages des TIC selon les métiers: Centre d'analyse stratégique2013.
74. Sandor E. European Company Survey 2009. Part-time work in Europe. Dublin: Eurofound2009.
75. Lauzeville D, Marchand JL, Ferrand M. Consommation de tabac par catégorie socioprofessionnelle et secteur d'activité : outil méthodologique pour l'épidémiologie. 2009.

76. Beck F. Résultats du Baromètre Santé 2010 - Liens entre usages de substances sychoactives (SPA) et milieu professionnel. Paris: INPES2012.
77. Goffette C. Le contexte social du tabagisme. Le rôle de l'environnement familial et professionnel. Paris: Science po paris; 2012.
78. Proper KI, Hildebrandt VH. Overweight and obesity among Dutch workers: differences between occupational groups and sectors. *Int Arch Occup Environ Health*. 2010 Jan;83(1):61-8.
79. Hasson D, Von Thiele Schwarz U, Lindfors P. Self-rated health and allostatic load in women working in two occupational sectors. *J Health Psychol*. 2009 May;14(4):568-77.
80. Leal C, Bean K, Thomas F, Chaix B. Are associations between neighborhood socioeconomic characteristics and body mass index or waist circumference based on model extrapolations? *Epidemiology*. 2011 Sep;22(5):694-703.
81. Chaix B, Bean K, Leal C, Thomas F, Havard S, Evans D, et al. Individual/neighborhood social factors and blood pressure in the RECORD Cohort Study: which risk factors explain the associations? *Hypertension*. 2010 Mar;55(3):769-75.
82. Chaix B, Billaudeau N, Thomas F, Havard S, Evans D, Kestens Y, et al. Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology*. 2011 Jan;22(1):18-26.
83. Chaix B, Jouven X, Thomas F, Leal C, Billaudeau N, Bean K, et al. Why socially deprived populations have a faster resting heart rate: impact of behaviour, life course anthropometry, and biology--the RECORD Cohort Study. *Soc Sci Med*. 2011 Nov;73(10):1543-50.
84. Chaix B, Kestens Y, Bean K, Leal C, Karusisi N, Meghrief K, et al. Cohort profile: residential and non-residential environments, individual activity spaces and cardiovascular risk factors and diseases--the RECORD Cohort Study. *Int J Epidemiol*. 2012 Oct;41(5):1283-92.
85. Havard S, Reich BJ, Bean K, Chaix B. Social inequalities in residential exposure to road traffic noise: an environmental justice analysis based on the RECORD Cohort Study. *Occup Environ Med*. 2011 May;68(5):366-74.
86. Chaix B, Bean K, Daniel M, Zenk SN, Kestens Y, Charreire H, et al. Associations of supermarket characteristics with weight status and body fat: a multilevel analysis of individuals within supermarkets (RECORD study). *PLoS One*. 2012;7(4):e32908.
87. Karusisi N, Bean K, Oppert JM, Pannier B, Chaix B. Multiple dimensions of residential environments, neighborhood experiences, and jogging behavior in the RECORD Study. *Prev Med*. 2012 Jul;55(1):50-5.

88. Leal C, Bean K, Thomas F, Chaix B. Multicollinearity in associations between multiple environmental features and body weight and abdominal fat: using matching techniques to assess whether the associations are separable. *Am J Epidemiol*. 2012 Jun 1;175(11):1152-62.
89. Karusisi N, Thomas F, Meline J, Chaix B. Spatial accessibility to specific sport facilities and corresponding sport practice: the RECORD Study. *Int J Behav Nutr Phys Act*. 2013 Apr 20;10(1):48.
90. INSEE. IRIS—Definition. Paris, France: National Institute of Statistics and Economic Studies. 2011.
91. WHO. Cardiovascular disease. 2013 [cited fact sheet n°317]; 4 march 2014]. Available from: <http://www.who.int/mediacentre/factsheets/fs317/en/>.
92. Thomas F, Bean K, Pannier B, Oppert JM, Guize L, Benetos A. Cardiovascular mortality in overweight subjects: the key role of associated risk factors. *Hypertension*. 2005 Oct;46(4):654-9.
93. WHO. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. *World Health Organ Tech Rep Ser*. 2000;894:i-xii, 1-253.
94. Benetos A, Thomas F, Pannier B, Bean K, Jengo B, Guize L. All-cause and cardiovascular mortality using the different definitions of metabolic syndrome. *Am J Cardiol*. 2008 Jul 15;102(2):188-91.
95. Pannier B, Thomas F, Bean K, Jengo B, Benetos A, Guize L. The metabolic syndrome: similar deleterious impact on all-cause mortality in hypertensive and normotensive subjects. *J Hypertens*. 2008 Jun;26(6):1223-8.
96. Benetos A, Rudnichi A, Thomas F, Safar M, Guize L. Influence of heart rate on mortality in a French population: role of age, gender, and blood pressure. *Hypertension*. 1999 Jan;33(1):44-52.
97. Chaix B, Simon C, Charreire H, Thomas F, Kestens Y, Karusisi N, et al. The environmental correlates of overall and neighborhood based recreational walking (a cross-sectional analysis of the RECORD Study). *Int J Behav Nutr Phys Act*. 2014;11(1):20.
98. Chaix B, Leal C, Evans D. Neighborhood-level confounding in epidemiologic studies: unavoidable challenges, uncertain solutions. *Epidemiology*. 2010 Jan;21(1):124-7.
99. Takashima Y, Iwase Y, Yoshida M, Kokaze A, Takagi Y, Taubono Y, et al. Relationship of food intake and dietary patterns with blood pressure levels among middle-aged Japanese men. *J Epidemiol*. 1998 Jun;8(2):106-15.

100. Proper KI, Hildebrandt VH. Physical activity among Dutch workers--differences between occupations. *Prev Med.* 2006 Jul;43(1):42-5.
101. Oppert JM, Thomas F, Charles MA, Benetos A, Basdevant A, Simon C. Leisure-time and occupational physical activity in relation to cardiovascular risk factors and eating habits in French adults. *Public Health Nutr.* 2006 Sep;9(6):746-54.
102. Chaix B, Evans D, Merlo J, Suzuki E. Commentary: Weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. *Epidemiology.* 2012 Jan;23(1):129-31; discussion 32-7.